Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Materials Science of Thin Films

Deposition and Structure

Second Edition

Milton Ohring
Department of Materials Science and Engineering
Stevens Institute of Technology
Hoboken, New Jersey

ACADEMIC PRESS

San Diego San Francisco New York Boston London Sydney Tokyo

This book is printed on acid-free paper. ∞

Copyright © 2002 by Academic Press

All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Requests for permission to make copies of any part of the work should be mailed to the following address: Permissions Department, Harcourt, Inc., 6277 Sea Harbor Drive, Orlando, Florida 32887-6777.

Academic Press
A division of Harcourt, Inc.
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
http://www.academicpress.com

Academic Press
Harcourt Place, 32 Jamestown Road, London, NW1 7BY, UK
http://www.academicpress.com

Library of Congress Catalog Card Number: 2001089414
International Standard Book Number: 0-12-524975-6

Printed in the United States of America
01 02 03 04 05 06 ML 9 8 7 6 5 4 3 2 1

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Historical Perspective

Thin-film technology is simultaneously one of the oldest arts and one of the newest sciences. Involvement with thin films dates to the metal ages of antiquity. Consider the ancient craft of gold beating, which has been practiced continuously for at least four millenia. Gold's great malleability enables it to be hammered into leaf of extraordinary thinness while its beauty and resistance to chemical degradation have earmarked its use for durable ornamentation and protection purposes. The Egyptians appear to have been the earliest practitioners of the art of gold beating and gilding. Many magnificent examples of statuary, royal crowns, and coffin cases that have survived intact attest to the level of skill achieved. The process involves initial mechanical rolling followed by many stages of beating and sectioning composite structures consisting of gold sandwiched between layers of vellum, parchment, and assorted animal skins. Leaf samples from Luxor dating to the Eighteenth Dynasty (1567–1320 B.C.) measured 0.3 microns in thickness (Ref. 1). As a frame of reference for the reader, a human hair is about 75 microns in diameter. Such leaf was carefully applied and bonded to smoothed wax or resin-coated wood surfaces in a mechanical (cold) gilding process. From Egypt the art spread as indicated by numerous accounts of the use of gold leaf in antiquity.

Today, gold leaf can be machine-beaten to 0.1 micron and to 0.05 micron when beaten by a skilled craftsman. In this form it is invisible sideways and quite readily absorbed by the skin. It is no wonder then that British gold beaters were called upon to provide the first metal specimens to be observed in the transmission electron microscope. Presently, gold leaf is used to decorate such diverse structures and objects as statues, churches, public

buildings, tombstones, furniture, hand-tooled leather, picture frames, and, of course, illuminated manuscripts.

Thin-film technologies related to gold beating, but probably not as old, are mercury and fire gilding. Used to decorate copper or bronze statuary, the cold mercury process involved carefully smoothing and polishing the metal surface, after which mercury was rubbed into it (Ref. 2). Some copper dissolved in the mercury, forming a very thin amalgam film that left the surface shiny and smooth as a mirror. Gold leaf was then pressed onto the surface cold and bonded to the mercury-rich adhesive. Alternately, gold was directly amalgamated with mercury, applied, and the excess mercury was then driven off by heating, leaving a film of gold behind. Fire gilding was practiced well into the 19th century despite the grave health risk due to mercury vapor. The hazard to workers finally became intolerable and provided the incentive to develop alternative processes, such as electroplating (Ref. 3).

Distinct from these physical gold-beating and gilding processes are chemical recipes for decorating copper-base alloy objects with gold-rich coatings. One such technique known as depletion gilding capitalizes on the fact that copper oxidizes preferentially relative to gold. Starting with a relatively dilute copper–gold alloy (tumbaga), successive thermal oxidations each followed by dissolution of the resultant copper oxide leaves the surface increasingly enriched in gold. Depletion gilding (Ref. 4) of sheet metals was practiced by the Andean metalsmiths of the Peruvian coast for perhaps two millenia prior to the Spanish conquest of the Incas. Much to the surprise of the Conquistadors when they melted Inca treasure, the bullion contained less gold than originally imagined; a gold-rich veneer only 0.5 to 2 microns thick masked the more plentiful, but relatively worthless copper-rich metal underneath.

The history of gold beating and gilding is replete with experimentation and process development in diverse parts of the ancient world. Practitioners were concerned with the purity and cost of the gold, surface preparation, the uniformity of the applied films, adhesion to the substrate, reactions between and among the gold, mercury, copper, bronze (copper–tin), etc., process safety, color, optical appearance, durability of the final coating, and competitive coating technologies. As we shall see in the ensuing pages, modern thin-film technology addresses these same generic issues, albeit with a great compression of time. And although science is now in the ascendancy, there is still much room for art.

References

1. E. D. Nicholson, *Gold Bull*. **12**, 161 (1979).
2. O. Vittori, *Gold Bull*. **12**, 35 (1979).
3. L. B. Hunt, *Gold Bull*. **9**, 24 (1976).
4. H. Lechtman, *Scientific American* **250**, 56 (1984).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Preface

Technological progress and scientific advances often proceed with different time constants. While the former is often shorter than the latter, they nevertheless march forward in a coupled rhythm. This is perhaps nowhere better illustrated than in thin-film science and technology. And that is why even though much of the subject matter of this book has been dramatically updated, its spirit has remained that of the first edition of *Materials Science of Thin Films*, which appeared a decade ago. Documenting and interpreting the remarkable technological progress of the intervening years in terms of the underlying, largely unchanging physical and chemical sciences remains an invariant feature of this revised edition.

Thin-film microelectronics and optoelectronics industries are among the strongest technological drivers of our economy, a fact manifested by the explosive growth in communications, and information processing, storage, and display applications. Fruits of these technologies have fertilized expanding thin-film uses in diverse areas, e.g., coatings of all kinds (optical, decorative, environmental, and wear resistant), biotechnology, and the generation and conservation of energy. Common to this family of related thin-film applications are issues rooted in materials science and engineering, accounting for the book's flavor and focus. Included among its pages is an information and knowledge base intended for the same interdisciplinary and varied audience served by the first edition, namely,

1. Science and engineering students in advanced undergraduate or first-year graduate level courses on thin films

2. Participants in industrial in-house courses or short courses offered by professional societies
3. Mature scientists and engineers switching career directions who require an overview of the field

Readers should be reasonably conversant with introductory college chemistry and physics and possess a passive cultural familiarity with topics commonly treated in undergraduate physical chemistry and modern physics courses. Short of this, a good course in materials science and engineering will do. Such courses traditionally focus on bulk solids, typically utilizing metals, semiconductors, ceramics, and polymers, taken singly or collectively as illustrative vehicles to convey principles. The same spirit is adopted in this book, except that thin solid films are the vehicle. Of the tetrahedron of processing–structure–properties–performance interactions, the multifaceted processing–structure concerns are the ones this book primarily focuses on. Within this context, I have attempted to weave threads of commonality among seemingly different materials, processes, and structures, as well as draw distinctions when they exhibit outwardly similar behavior. In particular, parallels and contrasts between films and bulk materials are themes of recurring discussions.

An optional introductory review chapter on standard topics in materials science establishes a foundation for subsequent chapters. Following a second chapter devoted to vacuum science and technology, the remaining text is broadly organized into three sections. Chapters 3, 4, 5, and 6 deal primarily with the principles and practices of *film deposition* from the vapor phase by physical and chemical methods. The increasing importance of plasmas and ion beams in recent years to deposit, etch, and modify films, is reflected in the content of the middle two of these chapters. *Film structure* is the subject of Chapters 7, 8, and 9. These three chapters track the events that start with the condensation of atomic clusters on a bare substrate, continue with film growth due to additional deposition, and end with fully developed polycrystalline, single-crystal (epitaxial), or amorphous films and coatings. Thin films are structurally and chemically characterized by the assorted electron and scanning probe microscopies as well as surface analytical techniques that are described in Chapter 10. Finally, the last two chapters broadly expose the underlying connections between film deposition and structure by addressing the roles of mass transport and stress. These chapters also consider the stability of film systems under driving forces that promote structural and chemical change. Exercises of varying difficulty are provided in each chapter, and I believe a deeper sense of the subject matter will be gained by considering them. Three elegant problems (exercises 5, 6,

and 7 of Chapter 12) were developed by Professor W. D. Nix, and I thank him for their use.

I have been most gratified by the reception of the first edition of my book. The present version has been thoroughly revised and no former chapter has remained untouched. Obsolete and unsuitable material has been updated and replaced not only in response to advances in the field, but also to make it better conform to pedagogical demands. In this vein, the last 11 chapters may be viewed as core subject matter applicable to all films and coatings, and therefore suitable for introductory courses. Based on my own experience, I have only been able to present a representative portion, but by no means all aspects, of this core material in a one-semester course on thin films. Former readers will note the omission of chapters dealing with electrical, magnetic, and optical film properties and applications, which were included in the first edition. Because these topics tend to be too specialized for the target audience, time limitations have generally meant their exclusion from course syllabi. Readers interested in these subjects will, however, find a rich but broadly dispersed literature on these subjects.

Because of its emphasis on immutable concepts, I hope this book will be spared the specter of rapid obsolescence. However, if this book will in some small measure help spawn new technology that renders it obsolete, it will have served a useful function.

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Contents

Foreword to First Edition xi

Preface xiii

Acknowledgments xvii

A Historical Perspective xix

Chapter 1

A Review of Materials Science 1
1.1. Introduction 1
1.2. Structure 2
1.3. Defects in Solids 10
1.4. Bonds and Bands in Materials 14
1.5. Thermodynamics of Materials 24
1.6. Kinetics 36
1.7. Nucleation 44
1.8. An Introduction to Mechanical Behavior 47
1.9. Conclusion 52
Exercises 52
References 55

Chapter 2

Vacuum Science and Technology 57
2.1. Introduction 57
2.2. Kinetic Theory of Gases 58
2.3. Gas Transport and Pumping 63

2.4. Vacuum Pumps 70
2.5. Vacuum Systems 81
2.6. Conclusion 88
Exercises 90
References 92

Chapter 3

Thin-Film Evaporation Processes 95

3.1. Introduction 95
3.2. The Physics and Chemistry of Evaporation 97
3.3. Film Thickness Uniformity and Purity 106
3.4. Evaporation Hardware 118
3.5. Evaporation Processes and Applications 128
3.6. Conclusion 139
Exercises 140
References 143

Chapter 4

Discharges, Plasmas, and Ion–Surface Interactions 145

4.1. Introduction 145
4.2. Plasmas, Discharges, and Arcs 147
4.3. Fundamentals of Plasma Physics 152
4.4. Reactions in Plasmas 164
4.5. Physics of Sputtering 170
4.6. Ion Bombardment Modification of Growing Films 184
4.7. Conclusion 196
Exercises 198
References 201

Chapter 5

Plasma and Ion Beam Processing of Thin Films 203

5.1. Introduction 203
5.2. DC, AC, and Reactive Sputtering Processes 205
5.3. Magnetron Sputtering 222
5.4. Plasma Etching 233
5.5. Hybrid and Modified PVD Processes 252
5.6. Conclusion 269
Exercises 270
References 273

Chapter 6

Chemical Vapor Deposition 277

6.1. Introduction 277
6.2. Reaction Types 281
6.3. Thermodynamics of CVD 287
6.4. Gas Transport 293
6.5. Film Growth Kinetics 303
6.6. Thermal CVD Processes 312
6.7. Plasma-Enhanced CVD Processes 323
6.8. Some CVD Materials Issues 334
6.9. Safety 347
6.10. Conclusion 349
Exercises 350
References 353

Chapter 7

Substrate Surfaces and Thin-Film Nucleation 357

7.1. Introduction 357
7.2. An Atomic View of Substrate Surfaces 360
7.3. Thermodynamic Aspects of Nucleation 376
7.4. Kinetic Processes in Nucleation and Growth 386
7.5. Experimental Studies of Nucleation and Growth 400
7.6. Conclusion 409
Exercises 410
References 414

Chapter 8

Epitaxy 417

8.1. Introduction 417
8.2. Manifestations of Epitaxy 420
8.3. Lattice Misfit and Defects in Epitaxial Films 429
8.4. Epitaxy of Compound Semiconductors 439
8.5. High-Temperature Methods for Depositing Epitaxial Semiconductor Films 453
8.6. Low-Temperature Methods for Depositing Epitaxial Semiconductor Films 466
8.7. Mechanisms and Characterization of Epitaxial Film Growth . . . 476
8.8. Conclusion 488
Exercises 489
References 492

Chapter 9

Film Structure 495

9.1. Introduction 495
9.2. Structural Morphology of Deposited Films and Coatings 497
9.3. Computational Simulations of Film Structure 510
9.4. Grain Growth, Texture, and Microstructure Control in Thin Films 520
9.5. Constrained Film Structures 533
9.6. Amorphous Thin Films 540
9.7. Conclusion 552
Exercises 553
References 556

Chapter 10

Characterization of Thin Films and Surfaces 559

10.1. Introduction 559
10.2. Film Thickness 562
10.3. Structural Characterization of Films and Surfaces 583
10.4. Chemical Characterization of Surfaces and Films 606
10.5. Conclusion 633
Exercises 635
References 639

Chapter 11

Interdiffusion, Reactions, and Transformations in Thin Films 641

11.1. Introduction 641
11.2. Fundamentals of Diffusion 643
11.3. Interdiffusion in Thin Metal Films 659
11.4. Compound Formation and Phase Transformations in Thin Films 669
11.5. Metal–Semiconductor Reactions 682
11.6. Mass Transport in Thin Films under Large Driving Forces 695
11.7. Conclusion 704
Exercises 704
References 708

Chapter 12

Mechanical Properties of Thin Films 711

12.1. Introduction 711
12.2. Mechanical Testing and Strength of Thin Films 713
12.3. Analysis of Internal Stress 723
12.4. Techniques for Measuring Internal Stress in Films 735
12.5. Internal Stresses in Thin Films and Their Causes 742
12.6. Mechanical Relaxation Effects in Stressed Films 754
12.7. Adhesion 764
12.8. Conclusion 775
Exercises 776
References 779

Index 783

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 1

A Review of Materials Science

1.1 INTRODUCTION

A cursory examination of the vast body of solid substances reveals what outwardly appears to be an endless multitude of external forms and structures possessing a bewildering variety of properties. The branch of study known as Materials Science evolved in part to classify those features that are common among the structure and properties of different materials in a manner somewhat reminiscent of chemical or biological classification schemes. This dramatically reduces the apparent variety. From this perspective, it turns out that solids can be classified as belonging typically to one of only four different categories (metallic, ionic, covalent, van der Waals) depending on the nature of the electronic structure and resulting interatomic bonding forces. Another scheme based on engineering use would again arguably limit materials to four chief classes, namely, metals, semiconductors, polymers, and ceramics.

Similar divisions occur with respect to structure of solids. Solids are either internally crystalline or noncrystalline. Those that are crystalline can be further subdivided according to one of 14 different geometric arrays or lattices depending on the placement of the atoms. When properties are considered, there are similar descriptors and simplifying categorizations. Thus, materials are either good, intermediate, or poor conductors of electricity, they are either mechanically brittle or can easily be stretched without fracture, they are either optically reflective or transparent, etc. It is, of course, easier to recognize that property differences exist than to understand why they exist. Nevertheless, much progress has been made in this

subject as a result of the research of the past century. Basically, the richness in the diversity of materials properties occurs because countless combinations of the admixture of chemical compositions, bonding types, crystal structures, and morphologies either occur naturally or can be synthesized.

This chapter reviews various aspects of the structure, bonding, and properties of solids with the purpose of providing the background to better understand the remainder of the book. Additional topics dealing with thermodynamics and kinetics of atomic motion in materials are also included. These will later have relevance to aspects of the formation, stability, and solid-state reactions in thin films. This review ends with a discussion of mechanical properties, a subject of significance in phenomena ranging from film deposition to adhesion. Although much of this chapter is a condensed adaptation of standard treatments of bulk materials, it is largely applicable to thin films as well. Nevertheless, many distinctions between bulk materials and films exist and they will be stressed in the ensuing discussion. Readers already familiar with concepts of materials science may wish to skip this chapter. However, it is recommended that those who seek deeper and broader coverage of this background material should consult the general overview texts in the list of references.

1.2 STRUCTURE

1.2.1 Crystalline Solids

Many solid materials possess an ordered internal crystal structure despite external appearances which are not what we associate with the term crystalline, i.e., clear, transparent, faceted, etc. Actual crystal structures can be imagined to arise from a three-dimensional array of points geometrically and repetitively distributed in space such that each point has identical surroundings. There are only fourteen ways to arrange points in space having this property and the resulting point arrays are known as Bravais lattices. They are shown in Fig. 1-1 with lines intentionally drawn to emphasize the symmetry of the lattice. It should be noted that only a single cell for each lattice is reproduced here and that the point array actually stretches in an endlessly repetitive fashion in all directions. If an atom or group of two or more atoms is now placed at each Bravais lattice point, a physically real crystal structure emerges. Thus, if individual copper atoms populated every point of a face-centered cubic (FCC) lattice whose cube-edge dimension, or so-called lattice parameter, were 3.54 Å, the material known as metallic copper would be generated; similarly for other types of lattices and atoms.

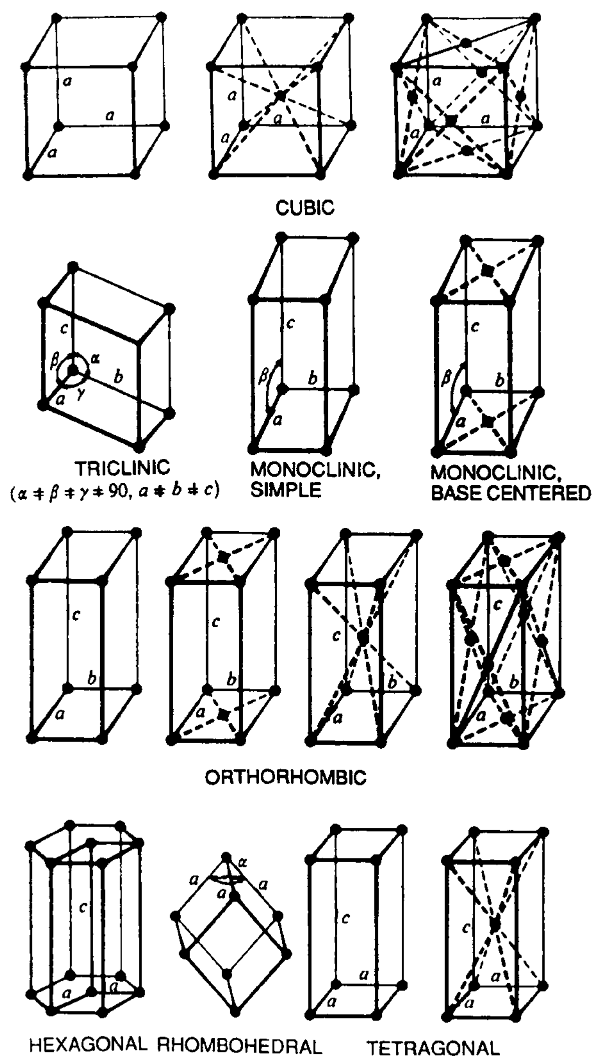


Figure 1.1 The 14 Bravais space lattices.

The reader should, of course, realize that just as there are no lines in actual crystals, there are no spheres. Each sphere in the Cu crystal structure represents the atomic nucleus surrounded by a complement of 28 core electrons [i.e., $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6(3d)^{10}$] and a portion of the free-electron gas contributed by 4s electrons. Furthermore, we must imagine that

these spheres touch in certain crystallographic directions and that their packing is rather dense. In FCC structures the atom spheres touch along the direction of the face diagonals, i.e., [110], but not along the face edge directions, i.e., [100]. This means that the planes containing the three face diagonals shown in Fig. 1-2a, i.e., (111), are close packed. On this plane the atoms touch each other in much the same way that a racked set of billiard balls do on a pool table. All other planes in the FCC structure are less densely packed and thus contain fewer atoms per unit area. Placement of two identical silicon atoms at each FCC point would result in the formation of the diamond-cubic silicon structure (Fig. 1-2c), whereas the rock-salt structure (Fig. 1-2b) is generated if sodium–chlorine groups were substituted for each lattice point. In both cases the positions and orientation of each two-atom motif must be preserved from point to point.

Quantitative identification of atomic positions as well as planes and directions in crystals requires the use of simple concepts of coordinate geometry. First, an orthogonal set of axes is arbitrarily positioned such that each point can now be identified by three coordinates $x = u$, $y = v$, and $z = w$. In a cubic lattice, e.g., FCC, the center of coordinate axes is taken

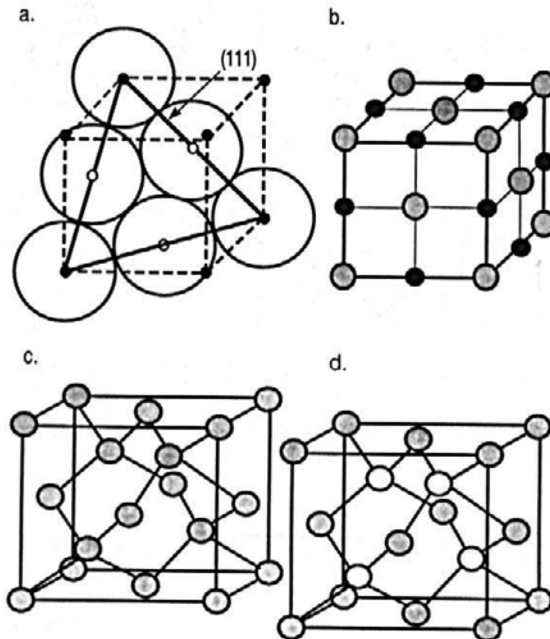


Figure 1-2 (a) (111) plane in FCC lattice; (b) rock-salt structure, e.g., NaCl; Na ●, Cl ⬤; (c) diamond cubic structure, e.g., Si, Ge; (d) zinc blende structure, e.g., GaAs.

as $x = 0$, $y = 0$, $z = 0$, or (0,0,0). Coordinates of other nearest equivalent cube corner points are then (1,0,0)(0,1,0)(1,0,0), etc. In this framework the two Si atoms referred to earlier would occupy the (0,0,0) and (1/4,1/4,1/4) positions. Subsequent translations of this oriented pair of atoms at each FCC lattice point generates the diamond-cubic structure in which each Si atom has four nearest neighbors arranged in a tetrahedral configuration. Similarly, substitution of the (0,0,0)Ga and (1/4,1/4,1/4)As motif for each point of the FCC lattice would result in the zinc-blende GaAs crystal structure (Fig. 1-2d).

Specific crystal planes and directions are frequently noteworthy because phenomena such as crystal growth, chemical reactivity, defect incorporation, deformation, and assorted properties are not isotropic or the same on all planes and in all directions. Therefore, the important need arises to be able to accurately identify and distinguish crystallographic planes and directions. A simple recipe for identifying a given plane in the cubic system is:

1. Determine the intercepts of the plane on the three crystal axes in number of unit cell dimensions.
2. Take reciprocals of those numbers.
3. Reduce these to smallest integers by clearing fractions.

The result is a triad of numbers known as the Miller indices for the plane in question, i.e., (hkl). Several planes with identifying Miller indices are indicated in Fig. 1-3. Note that a negative index is indicated above the integer with a minus sign.

Crystallographic directions shown in Fig. 1-3 are determined by the components of the vector connecting any two lattice points lying along the direction. If the coordinates of these points are (u_1, v_1, w_1) and (u_2, v_2, w_2), then the components of the direction vector are $(u_1 - u_2, v_1 - v_2, w_1 - w_2)$. When reduced to smallest integer numbers and placed within brackets they are known as the Miller indices for the direction, i.e., $[hkl]$. In this notation the direction cosines for the given direction are $h/\sqrt{h^2 + k^2 + l^2}$, $k/\sqrt{h^2 + k^2 + l^2}$, $l/\sqrt{h^2 + k^2 + l^2}$. Thus the angle a between any two directions $[h_1k_1l_1]$ and $[h_2k_2l_2]$ is given by the vector dot product

$$\cos\alpha = \frac{h_1h_2 + k_1k_2 + l_1l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\sqrt{h_2^2 + k_2^2 + l_2^2}}. \tag{1-1}$$

Two other useful relationships in the crystallography of cubic systems are given without proof.

1. The Miller indices of the direction normal to the (hkl) plane is $[hkl]$.
2. The spacing between successive (hkl) planes is $a = a_0/\sqrt{h^2 + k^2 + l^2}$ where a_0 is the lattice parameter.

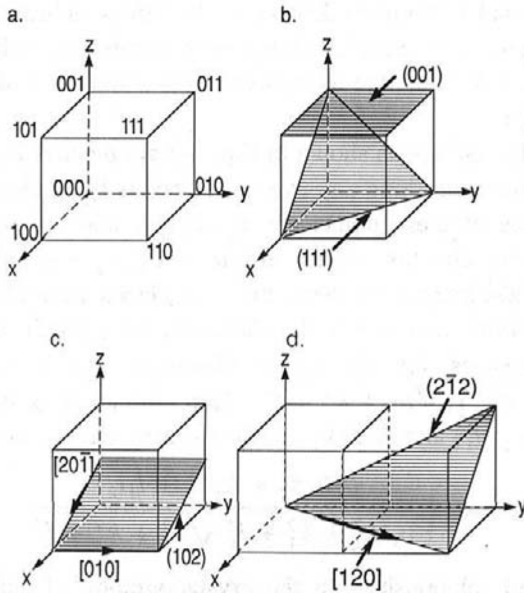


Figure 1-3 (a) Coordinates of lattice sites; (b) Miller indices of planes; (c, d) Miller indices of planes and directions.

As an example, let us calculate the angle between any two neighboring tetrahedral bonds in the diamond cubic lattice. The bonds lie along [111]-type directions which are specifically taken here to be [$\bar{1}\bar{1}$1] and [111]. Therefore, by Eq. 1-1,

$$\cos\alpha = \frac{(1)(-1) + 1(-1) + (1)(1)}{\sqrt{1^2 + 1^2 + 1^2}\sqrt{(-1)^2 + (-1)^2 + 1^2}} = -\frac{1}{3} \quad \text{and} \quad \alpha = 109.5^\circ.$$

These two bond directions lie in a common (110)-type crystal plane. The precise indices of this plane must be ($\bar{1}$10) or (1$\bar{1}$0). This can be seen by noting that the dot product between each bond vector and the vector normal to the plane they lie in must vanish.

1.2.2 X-Ray Diffraction

We close this brief discussion on lattices and crystal geometry with mention of experimental evidence in support of the internal crystalline structure of solids. X-ray diffraction methods have very convincingly dem-

onstrated the crystallinity of solids by exploiting the fact that the spacing between atoms is comparable to the wavelength (λ) of X-rays. This results in easily detected emitted beams of high intensity along certain directions when incident X-rays impinge at critical diffraction angles (θ). Under these conditions the well-known Bragg relation

$$n\lambda = 2a \sin \theta \quad (1\text{-}2)$$

holds, where n is an integer.

In bulk solids, large diffraction effects occur at many values of θ. In thin films, however, very few atoms are present to scatter X-rays into a diffracted beam when θ is large. For this reason the intensities of the diffraction lines or spots will be unacceptably small unless the incident beam strikes the film surface at a near glancing angle. This, in effect, makes the film look thicker. Such X-ray techniques for examination of thin films have evolved and will be discussed in Chapters 10 and 12. Relative to bulk solids, thin films require long counting times to generate enough signal for suitable X-ray diffraction patterns. This thickness limitation in thin films is turned into great advantage, however, in the transmission electron microscope. Here electrons must penetrate through the material under observation and this can occur only in thin films or specially thinned specimens. The short wavelength of the electrons employed enables diffraction effects and high-resolution imaging of the lattice structure to be observed. As an example, consider the electron micrograph of Fig. 1-4 (top) showing apparent perfect crystalline registry between a thin film of cobalt silicide and a silicon substrate. Correspondingly, the atomic positioning in this structure is schematically depicted in Fig. 1-4 (bottom). The phenomenon of a single-crystal film coherently oriented on a single-crystal substrate is known as *epitaxy* and is widely exploited in semiconductor technology. In this example the silicide film/substrate was mechanically and chemically thinned normal to the original film plane to make the cross-section visible. Such evidence should leave no doubt as to the internal crystalline nature of solids.

1.2.3 Amorphous Solids

In some materials the predictable long-range geometric order characteristic of crystalline solids break down. These are the noncrystalline amorphous or glassy solids exemplified by silica glass, inorganic oxide mixtures, and polymers. When such bulk materials are cooled from the melt even at low rates, the more random atomic positions that we associate with a liquid

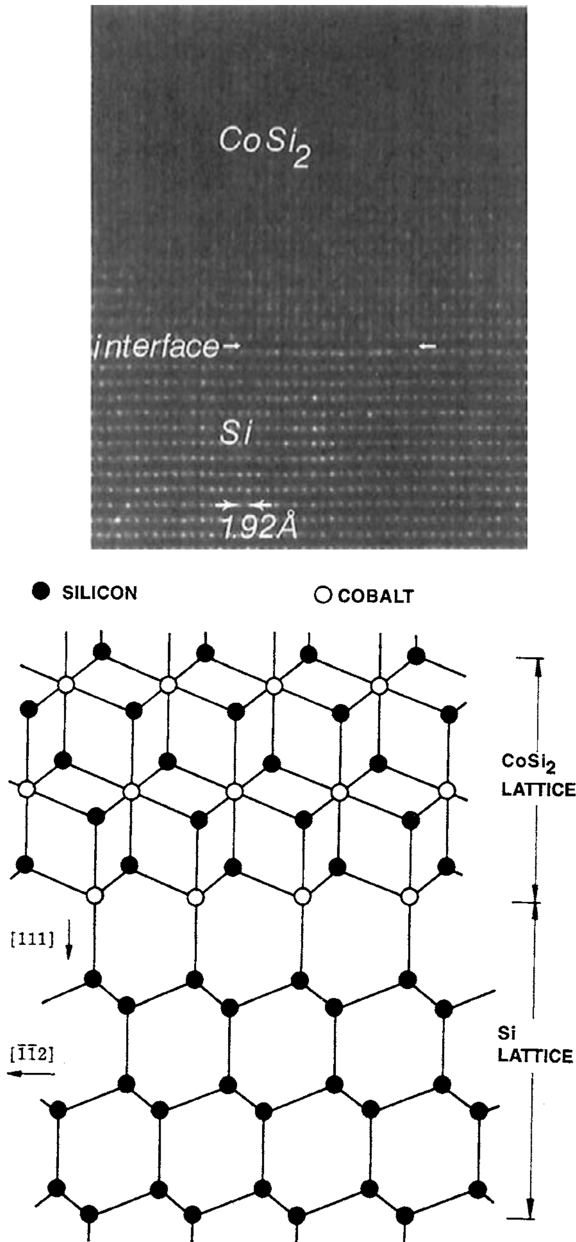


Figure 1-4 *Top:* High-resolution lattice image of epitaxial $CoSi_2$ film on (111) Si (⟨112⟩ projection). Courtesy J. M. Gibson, AT&T Bell Laboratories. *Bottom:* Ball and stick atomic model of the $CoSi_2$–Si heterostructure. After J. M. Gibson, *MRS Bulletin* XVI(3), 27 (1991).

are frozen in place within the solid. On the other hand, certain alloys composed of transition metal and metalloid combinations, e.g., Fe–B, can only be made in glassy form through extremely rapid quenching of melts. The required cooling rates are of the order of 10^{6}°C/s, and therefore heat-transfer considerations limit bulk glassy metals to foil, ribbon, or powder shapes typically ~0.05 mm in thickness or size. In general, amorphous solids can retain their structureless character practically indefinitely at low temperatures even though thermodynamics suggests greater stability for crystalline forms. Crystallization will, however, proceed with release of energy when these materials are heated to appropriate elevated temperatures. The atoms then have the required mobility to seek out equilibrium lattice sites.

Thin films of amorphous metal alloys, semiconductors, oxide and chalcogenide glasses have been readily prepared by common physical vapor deposition (evaporation and sputtering) as well as chemical vapor deposition (CVD) methods. Vapor quenching onto cryogenically cooled glassy substrates has made it possible to make alloys and even pure metals, the most difficult of all materials to amorphize, glassy. In such cases, the surface mobility of depositing atoms is severely restricted and a disordered atomic configuration has a greater probability of being frozen in.

Our present notions of the structure of amorphous inorganic solids are extensions of models first established for silica glass. These depict amorphous SiO_2 to be a random three-dimensional network consisting of tetrahedra that are joined at the corners but share no edges or faces. Each tetrahedron contains a central Si atom bonded to four vertex oxygen atoms, i.e., $(SiO_4)^{4-}$. The oxygens are, in turn, shared by two Si atoms and are thus positioned as the pivotal links between neighboring tetrahedra. In crystalline quartz the tetrahedra cluster in an ordered six-sided ring pattern shown schematically in Fig. 1-5a, which should be contrasted with the completely random network depicted in Fig. 1-5b. On this basis the glassy solid matrix is probably an admixture of these structural extremes possessing considerable short-range order, and microscopic crystalline regions, i.e., less than 100 Å in size (Fig. 1-5c). The loose disordered network structure allows for a considerable number of "holes" or "vacancies" to exist, and it therefore comes as no surprise that the density of glasses will be less than that of their crystalline counterparts. In crystalline quartz, for example, the density is 2.65 g/cm^3, whereas in silica glass it is 2.2 g/cm^3. Amorphous silicon, which has found commercial use in thin-film solar cells, is, like silica, tetrahedrally bonded and believed to possess a similar structure. We shall later return to discuss further structural aspects and properties of amorphous films in various contexts throughout the book, e.g., Section 9.6.

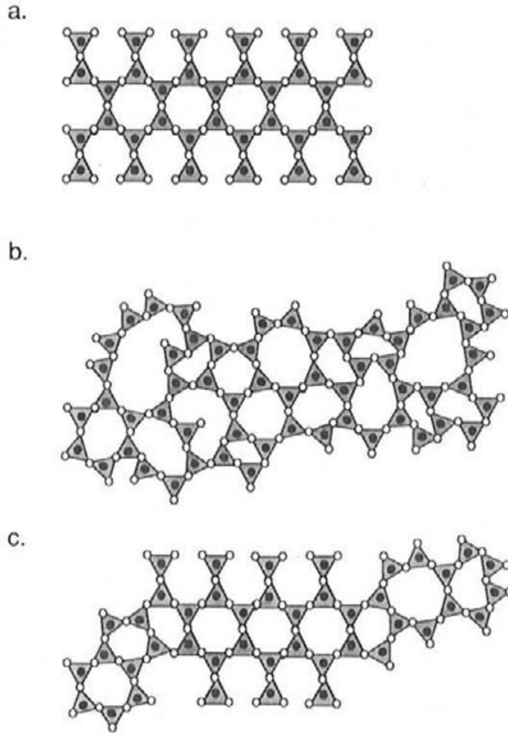


Figure 1-5 Schematic representation of (a) crystalline quartz; (b) random network (amorphous); (c) mixture of crystalline and amorphous regions. (Reprinted with permission from John Wiley & Sons, E. H. Nicollian, and J. R. Brews, *MOS Physics and Technology*, copyright © 1983, John Wiley & Sons.)

1.3 DEFECTS IN SOLIDS

The picture of a perfect crystal structure repeating a particular geometric pattern of atoms without interruption or error is somewhat of an exaggeration. Although there are materials like carefully grown silicon single crystals that have virtually perfect crystallographic structures extending over macroscopic dimensions, this is generally not the case in bulk materials. In thin crystalline films the presence of defects not only serves to disrupt the geometric regularity of the lattice on a microscopic level, but significantly influences many film properties such as chemical reactivity, electrical conduction, and mechanical behavior. The structural defects briefly considered in this section are vacancies, dislocations, and grain boundaries.

1.3.1 VACANCIES

The most elementary of crystalline defects are vacancies, and they arise when lattice sites are unoccupied by atoms. Also known as *point defects*, vacancies form because the energy E_f required to remove atoms from interior sites and place them on the surface is not particularly high. This, coupled with the increase in the statistical entropy of mixing vacancies among lattice sites, gives rise to a thermodynamic probability that an appreciable number of vacancies will exist, at least at elevated temperature. The fraction f of total sites that will be unoccupied as a function of temperature T is predicted to be approximately

$$f = \exp - E_f/k_B T, \tag{1-3}$$

reflecting the statistical thermodynamic nature of vacancy formation. Noting that k_B is the Boltzmann constant and E_f is typically 1 eV per atom gives $f = 10^{-5}$ at 1000 K.

Vacancies play an important role in all processes related to solid-state diffusion, including recrystallization, grain growth, sintering, and phase transformations. In semiconductors, vacancies are electrically neutral as well as charged and can be associated with dopant atoms. This leads to a variety of normal as well as anomalous diffusional-doping effects.

1.3.2 DISLOCATIONS

Next in the hierarchy of defect structures are dislocations. These are *line* defects that bear a definite crystallographic relationship to the lattice. The two fundamental types of dislocations — the edge and the screw — are shown in Fig. 1-6 and are represented by the symbol $\perp$. An edge dislocation can be generated by wedging an extra row of atoms into a perfect crystal lattice, while the screw dislocation requires cutting followed by shearing of the resultant halves with respect to each other. The geometry of a crystal containing a dislocation is such that when attempting a simple closed traverse about its axis in the surrounding lattice, there is a closure failure, i.e., one finally arrives at a lattice site displaced from the starting position by a lattice vector, the so-called Burgers vector **b**. This vector lies perpendicular to the edge dislocation line and parallel to the screw dislocation line. Individual cubic cells representing the original undeformed crystal lattice are now distorted somewhat in the presence of dislocations. Therefore, even without application of external forces on the crystal, a state of internal strain (stress) exists around each dislocation. Furthermore, the strains (stresses)

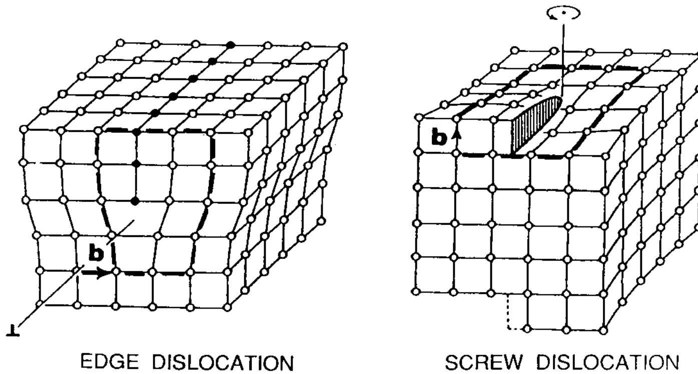


Figure 1-6 (left) Edge dislocation; (right) Screw dislocation. (Reprinted with permission from John Wiley & Sons, H. W. Hayden, W. G. Moffat, and J. Wulff, *The Structure and Properties of Materials*, Vol. III, copyright © 1965, John Wiley & Sons.)

differ around edge and screw dislocations because the lattice distortions differ. Close to the dislocation axis or core the stresses are high but they fall off with distance (r) according to a $1/r$ dependence.

In contrast to vacancies, dislocations are not thermodynamic defects. Because dislocation lines are oriented along specific crystallographic directions, their statistical entropy is low. Coupled with a high formation energy due to the many atoms involved, thermodynamics would predict a dislocation content of less than one per crystal. Thus, while it is possible to create a solid devoid of dislocations, it is impossible to eliminate vacancies.

Dislocations are important because they have provided models to help explain a variety of mechanical phenomena and properties in all classes of crystalline solids. An early application was to the important process of plastic deformation, which occurs after a material is loaded beyond its limit of elastic response. In the plastic range, specific planes shear in characteristic directions relative to each other much as a deck of cards shear from a rectangular prism to a parallelepiped. Rather than have rows of atoms undergo a rigid group displacement to produce the slip offset step at the surface, the same amount of plastic deformation can be achieved with less energy expenditure. This alternate mechanism requires that dislocations undulate through the crystal, making and breaking bonds on the slip plane until a slip step is produced as shown in Fig. 1-7a. Dislocations thus help explain why metals are weak and can be deformed at low stress levels. Paradoxically, dislocations can also explain why metals work-harden or get stronger when they are deformed. These explanations require the presence of dislocations in great profusion. In fact, a density as high as 10^{12}

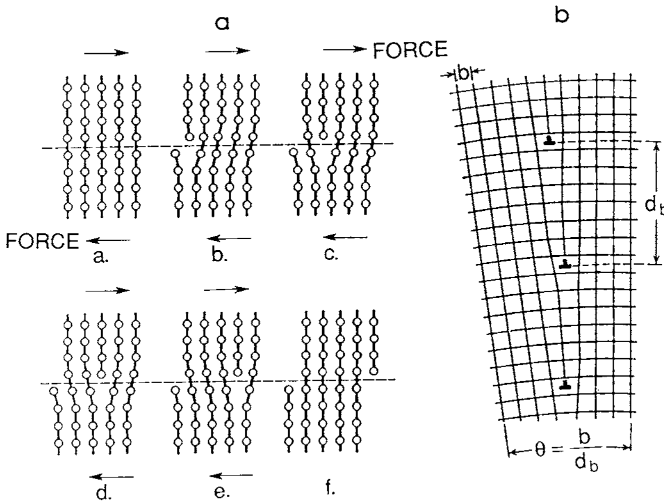


Figure 1-7 (a) Edge dislocation motion through lattice under applied shear stress. Reprinted with permission from J. R. Shackelford, *Introduction to Materials Science for Engineers*, Macmillan, New York, 1985. (b) Dislocation model of a grain boundary. The crystallographic misorientation angle θ between grains is b/d_b.

dislocation lines threading 1 cm^2 of surface area has been observed in highly deformed metals. Many deposited polycrystalline-metal thin films also have high dislocation densities. Some dislocations are stacked vertically giving rise to so-called small-angle grain boundaries (Fig. 1-7b). The superposition of externally applied forces and internal stress fields of individual or groups of dislocations, arrayed in a complex three-dimensional network, makes it generally more difficult for them to move and the lattice to deform easily.

Dislocations play varied roles in thin films. As an example, consider the deposition of atoms onto a single-crystal substrate in order to grow an epitaxial single-crystal film. If the lattice parameter in the film and substrate differ, then some geometric accommodation in bonding may be required at the interface resulting in the formation of interfacial dislocations. The latter are unwelcome defects, particularly if films of high crystalline perfection are required. This is why a good match of lattice parameters is sought for epitaxial growth. Substrate steps and dislocations should also be eliminated where possible prior to film growth. If the substrate has screw dislocations emerging normal to the surface, depositing atoms may perpetuate the extension of the dislocation spiral into the growing film. Like grain boundaries in semiconductors, dislocations can be sites of charge recombination

or generation as a result of uncompensated "dangling bonds." Film stress, thermally induced mechanical relaxation processes, and diffusion of atoms in films are all influenced by dislocations.

1.3.3 Grain Boundaries

Grain boundaries are surface or *area* defects that constitute the interface between two single-crystal grains of different crystallographic orientation. The normal atomic bonding in grains terminates at the grain boundary where more loosely bound atoms prevail. Like atoms on surfaces, they are necessarily more energetic than those within the grain interior. This causes the grain boundary to be a heterogeneous region where various atomic reactions and processes such as solid-state diffusion and phase transformations, precipitation, corrosion, impurity segregation, and mechanical relaxation are favored or accelerated. In addition, electronic transport in metals is impeded through increased scattering at grain boundaries, which also serve as charge recombination centers in semiconductors. Grain sizes in films are typically between 0.01 and 1.0 μm and are smaller, by a factor of more than 100, than common grain sizes in bulk materials. For this reason, thin films tend to be more reactive than their bulk counterparts. The fraction of atoms associated with grain boundaries in spherical grains of diameter l_g is approximately $6a/l_g$, where a is the atomic dimension. For $l_g = 1000$ Å, this corresponds to about 1 in 100.

Controlling grain morphology, orientation, and size are not only important objectives in bulk materials but are quite important in thin-film technology. Indeed a major goal in microelectronic applications is to eliminate grain boundaries altogether through epitaxial growth of single-crystal semiconductor films on oriented single-crystal substrates. Many special deposition methods are employed in this effort, which continues to be a major focus of thin-film semiconductor technology.

1.4 BONDS AND BANDS IN MATERIALS

1.4.1 Bonding at the Atomic Level

The reason that widely spaced isolated atoms condense to form solids is the energy reduction accompanying bond formation. Thus, if N atoms of type A in the gas phase (g) condense to form a solid (s), the binding energy E_b is released according to the equation

$$NA_g \rightarrow NA_s + E_b. \tag{1-4}$$

Energy E_b must be supplied to reverse the equation and decompose the solid. The more stable the solid, the higher is its binding energy. It has become customary to picture the process of bonding by considering the energetics between atoms as the interatomic distance progressively shrinks. In each isolated atom the electron energy levels are discrete as shown on the right-hand side of Fig. 1-8a. As the atoms approach one another, the individual levels split as a consequence of an extension of the Pauli Exclusion Principle to a collective solid, namely, no two electrons can exist in the same quantum state. Level splitting and broadening occurs first for the valence or outer electrons since their electron clouds are the first to overlap. During atomic attraction, electrons populate the lower energy levels, reducing the overall energy of the solid. With further reduction in interatomic spacing, the overlap increases and the inner charge clouds begin to interact. Ion-core overlap now results in strong repulsive forces between atoms, raising the electrostatic energy of the system. A compromise is struck between the attractive and repulsive energies such that at the equilibrium interatomic distance (a_0) the overall energy is minimized.

At equilibrium, some of the levels have broadened into bands of energy levels. The bands span different ranges of energy depending on the atoms and specific electron levels involved. Sometimes, as in metals, bands of high energy overlap. In insulators and semiconductors there are energy gaps of varying width between bands where electron states are not allowed. The whys and hows of energy-level splitting, band structure and evolution, and implications with regard to property behavior are perhaps the most fundamental and difficult questions in solid-state physics. We will briefly return to the subject of electron-band structure after introducing the classes of solids.

1.4.2 Bonding in Solids

An extension of the ideas expressed in Fig. 1-8a is commonly made to a group of atoms, in which case the potential energy of atomic interaction, $V(r)$, is plotted as a function of interatomic distance r in Fig. 1-8b. The generalized behavior shown is common for all classes of solid materials regardless of the type of bonding or crystal structure. Although the mathematical forms of the attractive or repulsive portions may be complex, a number of qualitative features of these curves are not difficult to understand.

For example, the energy at $r = a_0$ is the bonding energy. Solids with high melting points tend to have large values of E_b. The curvature of the potential energy is a measure of the elastic stiffness of the solid. To see this we note that around a_0 the potential energy is approximately harmonic or

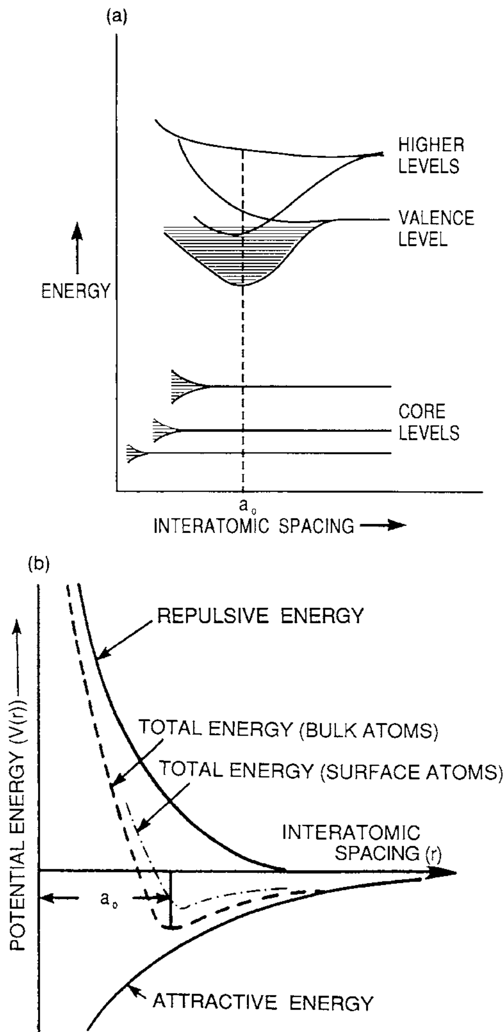


Figure 1-8 (a) Splitting of electron levels and (b) energy of interaction between atoms as a function of interatomic spacing. $V(r)$ vs r shown schematically for bulk and surface atoms.

parabolic. Therefore, $V(r) = 1/2K_s r^2$, where K_s is related to the spring constant (or elastic modulus). Narrow wells of high curvature are associated with large values of K_s, and broad wells of low curvature with small values of K_s. Since the force, F, between atoms is given by $F = -dV(r)/dr$, $F = -K_s r$, which has its counterpart in Hooke's law, i.e., that stress is

linearly proportional to strain where the modulus of elasticity (Y) is the constant of proportionality. Thus, in solids with high K_s or Y values, correspondingly larger stresses develop under straining. Interestingly, a purely parabolic behavior for $V(r)$ implies a material with a coefficient of thermal expansion equal to zero because atoms are equally likely to expand or contract at any temperature or level of energy. In real materials, therefore, some asymmetry or anharmonicity in $V(r)$ exists.

For the most part atomic behavior within a thin solid film can also be described by a $V(r) - r$ curve similar to that for the bulk solid. The surface atoms are less tightly bound, however, which is reflected by the dotted line behavior in Fig. 1-8b. The difference between the energy minima for surface and bulk atoms is a measure of the surface energy of the solid (see Section 7.3.2). From the previous discussion, surface layers would tend to be less stiff and melt at lower temperatures than the bulk. Slight changes in equilibrium atomic spacing or lattice parameter at surfaces may also be expected. Indeed, such effects have been experimentally observed.

An important application of the above ideas is to adhesion between the film and substrate. Bonding occurs because of pairwise interactions across interfaces between different kinds of atoms. This means that a force ($F = -dV(r)/dr$) must be applied to separate the film from the substrate. After the reader differentiates $V(r)$ it will be apparent that the $F(r)$ vs r variation has a similar, but inverted shape relative to the $V(r)$ vs r dependence. By rough analogy to engineering stress–strain curves, the peak value of the force is the theoretical strength of the interface, while the area under the curve to that point is a measure of the energy expended in separating the materials. The important subject of thin-film adhesion is treated at the end of Chapter 12.

1.4.3 The Four Classes of Solids: Bonding and Properties

Despite apparent similarities, there are many distinctions between the four important types of solid-state bonding and the properties they induce. A discussion of these individual bonding categories follows.

1.4.3.1 Metallic

The so-called metallic bond occurs in metals and alloys. In metals the outer valence electrons of each atom form part of a collective free-electron cloud or gas that permeates the entire lattice. Even though individual electron–electron interactions are repulsive, there is sufficient electrostatic

attraction between the free-electron gas and the positive ion cores to cause bonding.

What distinguishes metals from all other solids is the ability of the electrons to readily respond to applied electric fields, thermal gradients, and incident light. This gives rise to high electrical and thermal conductivities as well as high optical reflectivities. Interestingly, comparable properties are observed in liquid metals, indicating that aspects of metallic bonding and the free-electron model are largely preserved even in the absence of a crystal structure. Metallic electrical resistivities typically ranging from 10^{-5} to 10^{-6} ohm-cm should be contrasted with the very much larger values possessed by other classes of solids.

Furthermore, the temperature coefficient of resistivity is positive. Metals thus become poorer electrical conductors as the temperature is raised. The reverse is true for all other classes of solids. Also, the conductivity of pure metals is always reduced with low levels of impurity alloying, a behavior contrary to that of other solids. The effect of both temperature and alloying element additions on metallic conductivity is to increase electron scattering, which in effect reduces the net component of electron motion in the direction of the applied electric field. Interestingly, the electrical properties of metals differ little in film relative to bulk form. Ionic and semiconductor solids behave quite differently in this regard. In them greater charge carrier production and higher electrical conductivity is the result of higher temperatures and increased solute additions. Furthermore, there are generally very large differences in bulk and thin-film electrical behavior.

Bonding electrons in metals are not localized between atoms and nondirectional bonds are said to exist. This causes atoms to slide by each other and plastically deform more readily than is the case, for example, in covalent solids that have directed atomic bonds.

Examples of thin-metal-film applications include contacts and interconnections in integrated circuits, and ferromagnetic alloys for data storage applications. Metal films are also used in mirrors, optical systems and for decorative and protective coatings of packaging materials and various components.

1.4.3.2 Ionic

Ionic bonding occurs in compounds composed of strongly electropositive elements (metals) and strongly electronegative elements (nonmetals). The alkali halides (NaCl, LiF, etc.) are the most unambiguous examples of ionically bonded solids. In other compounds such as oxides and sulfides as well as many of the more complex salts of inorganic chemistry, e.g., nitrates

and sulfates, the predominant, but not necessarily exclusive, mode of bonding is ionic in character. In the rock-salt structure of NaCl, for example, there is an alternating three-dimensional checkerboard array of positively charged cations and negatively charged anions. Charge transfer from the 3s electron level of Na to the 3p level of Cl creates a single isolated NaCl molecule. In the solid, however, the transferred charge is distributed uniformly among nearest neighbors. Thus, there is no preferred directional character in the ionic bond since the electrostatic forces between spherically symmetric inert gas–like ions are independent of orientation.

Much success has been attained in determining the bond energies in alkali halides without resorting to quantum mechanical calculation. The alternating positive and negative ionic charge array suggests that Coulombic pair interactions are the cause of the attractive part of the interatomic potential, which varies simply as $V(r) \simeq 1/r$. Ionic solids are characterized by strong electrostatic bonding forces and thus relatively high binding energies and melting points. They are poor conductors of electricity because the energy required to transfer electrons from anions to cations is prohibitively large. At high temperatures, however, the charged ions themselves can migrate in an electric field resulting in limited electrical conduction. Typical resistivities for such materials can range from 10^6 to 10^{15} ohm-cm.

Perhaps the most important largely ionic thin-film material is SiO_2 because it performs critical dielectric and insulating functions in integrated circuit technology. Other largely ionic film materials of note include MgF_2 and ZnS for use in optical coatings, $YBa_2Cu_3O_7$ high-temperature superconductors, Al_2O_3 for hard coatings, and assorted thin-film oxides such as $Y_3Fe_5O_{12}$ and $LiNbO_3$, used respectively in magnetic and integrated optics applications. Transparent electrical conductors such as In_2O_3–SnO_2 glasses, which serve as heating elements in window defrosters on cars as well as electrical contacts over the light-exposed surfaces of solar cells, have partial ionic character.

1.4.3.3 Covalent

Elemental as well as compound solids exhibit covalent bonding. The outstanding examples are the elemental semiconductors Si, Ge, and diamond, as well as III–V compound semiconductors such as GaAs and InP. Whereas elements at the extreme ends of the periodic table are involved in ionic bonding, covalent bonds are frequently formed between elements in neighboring columns. The strong directional bonds characteristic of the group IV elements is due to the hybridization or mixing of the 2s and 2p electron wave functions into a set of orbitals that have high electron

densities emanating from the atom in tetrahedral fashion. A pair of electrons contributed by neighboring atoms comprises a covalent bond, and four such shared electron pairs complete the bonding requirements.

Covalent solids are strongly bonded hard materials with relatively high melting points. Despite the great structural stability of semiconductors, relatively modest thermal stimulation is sufficient to release electrons from filled valence bonding states into unfilled conduction-electron states. We speak of electrons being promoted from the valence to conduction band, a process that increases the conductivity of the solid. Small dopant additions of group III elements such as B and In as well as group V elements such as P and As take up regular lattice positions within Si and Ge. The bonding requirements are then not quite satisfied for group III elements, which are one electron short of a complete octet. An electron deficiency or hole is thus created in the valence band. Similarly, for each group V dopant an excess of one electron beyond the bonding octet can be promoted into the conduction band. As the name implies, semiconductors lie between metals and insulators insofar as their ability to conduct electricity is concerned. Typical semiconductor resistivities range from $\sim 10^{-3}$ to 10^{6} ohm-cm. Both temperature and level of doping are very influential in altering the conductivity of semiconductors. Ionic solids are similar in this regard.

The controllable spatial doping of semiconductors over very small lateral and transverse dimensions is a critical requirement in processing integrated circuits. Thin-film technology is thus simultaneously practiced in three dimensions in these materials. Similarly, the fabrication of a variety of optical devices such as lasers and light-emitting diodes requires the deposition of covalent compound-semiconductor thin films. Other largely covalent materials such as SiC, TiC, and TiN have found coating applications where hard, wear-resistant or protective surfaces are required. They are usually deposited by the chemical vapor deposition methods discussed in Chapter 6.

1.4.3.4 van der Waals

A large group of solid materials are held together by weak molecular forces. This so-called van der Waals bonding is due to dipole–dipole charge interactions between molecules that, though electrically neutral, have regions possessing a net positive or negative charge distribution. Organic molecules such as methane and inert gas atoms are weakly bound together in the solid by these charges. Such solids have low melting points and are mechanically weak. Thin polymer films used as photoresists or for sealing and encapsulation purposes contain molecules that are typically bonded by van der Waals forces.

1.4.4 ENERGY BAND DIAGRAMS

A common graphic means of distinguishing between different classes of solids makes use of energy band diagrams. Reference to Fig. 1-8a shows how individual energy levels broaden into bands when atoms are brought together to form solids. What is of interest here are the energies of electrons at the equilibrium atomic spacing in the crystal. For metals, insulators, and semiconductors the energy band structures are schematically indicated in Fig. 1-9a, b, c. In each case the horizontal axis can be loosely interpreted as some macroscopic distance within the solid having much larger than atomic dimensions. This distance spans a region within the homogeneous bulk interior where the band energies are uniform from point to point. The uppermost band shown is called the conduction band because once electrons access its levels, they are essentially free to conduct electricity.

Metals have nigh conductivity because the conduction band contains electrons from the outset. One has to imagine that there are a mind-boggling

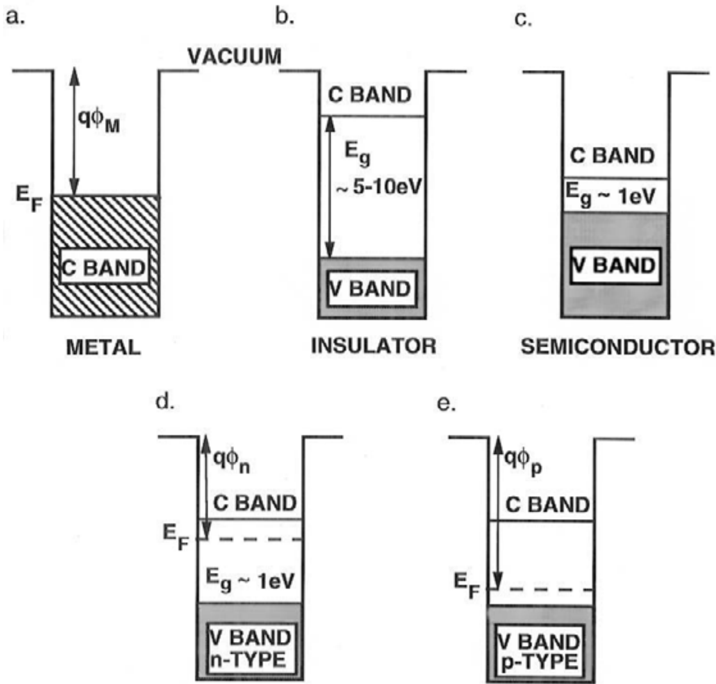


Figure 1-9 Schematic band structure for (a) metal; (b) insulator; (c) semiconductor; (d) *n*-type semiconductor; (e) *p*-type semiconductor.

$\sim 10^{22}$–10^{23} electrons per cubic centimeter (one or more per atom) in the conduction band, all of which occupy different quantum states. Furthermore, there are enormous numbers of states all at the same energy level, a phenomenon known as degeneracy. Lastly, the energy levels are extremely closed spaced and compressed within a typical 5-eV conduction-band energy width. The available electrons occupy states within the band up to a maximum level known as the Fermi energy E_F. Above E_F there are densely spaced excited levels, but they are all vacant. If electrons are excited sufficiently (e.g., by photons or through heating), they can gain enough energy to populate these states or even leave the metal altogether, i.e., by photo- and thermionic emission, and enter the vacuum. As indicated in Fig. 1-9a, the energy difference between the vacuum level and E_F is equal to $q\phi_M$, where ϕ_M is the work function in volts and q is the electronic charge. Values of $q\phi$ for many solids range between 2 and 5 eV. Even when very tiny electric fields are applied, the electrons in states at E_F can easily move into nearby unoccupied levels above it, resulting in a net current flow. This is why metals are such good electrical conductors.

At the other extreme are insulators in which the conduction band normally has no electrons. The valence electrons used in bonding completely fill the valence band. A large energy gap E_g ranging from 5 to 10 eV separates the filled valence band from the empty conduction band. There are normally no states and therefore no electrons within the energy gap. In order to conduct appreciable electricity, many electrons must acquire sufficient energy to span the energy gap, but for practical purposes this energy barrier is all but insurmountable.

Pure (intrinsic) semiconductors at very low temperatures have a band structure like that of insulators, but E_g is smaller, e.g., $E_g = 1.12$ eV in Si and 1.43 eV in GaAs. When the semiconductor is doped, new states are created within the energy gap. The electron (or hole) states associated with donors (or acceptors) are usually but a small fraction of an electron volt away from the bottom of the conduction band (or top of the valence band). It now takes very little thermal stimulation to excite electrons or holes to conduct electricity. The actual location of E_F with respect to the valence and conduction band edges depends on the type and amount of doping atoms present. In an intrinsic semiconductor E_F lies in the middle of the energy gap because E_F is strictly defined as that energy level for which the probability of occupation is 1/2. If the semiconductor is doped with donor atoms to make it n-type, E_F lies above the midgap energy as shown in Fig. 1-9d. When acceptor atoms are the predominant dopants, E_F lies below the midgap energy and a p-type semiconductor results (Fig. 1-9e).

Band diagrams have important implications in electronic devices where structures consisting of different materials in contact are involved. A simple

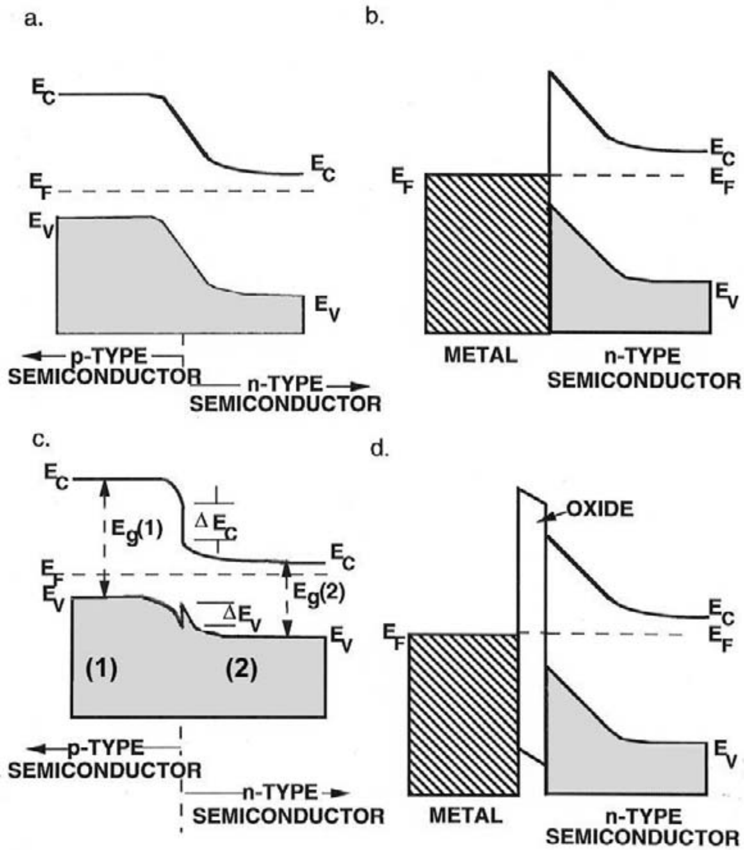


Figure 1-10 (a) p–n semiconductor junction. E_V is the energy at the top of the valence band and E_C is the energy at the bottom of the conduction band; (b) metal–semiconductor contact, e.g., Al–Si; (c) heterojunction between two different semiconductors, (1) and (2), e.g., GaAs–ZnSe. Conduction and valence band offsets or discontinuities, ΔE_C and ΔE_V, occur at the junction due to the different energy gaps, $E_g(1)$ and $E_g(2)$, of the two semiconductors; (d) metal–oxide–semiconductor structure, e.g., Al–SiO_2–Si.

example is the p–n junction, which is shown in Fig. 1-10a without any applied electric fields. A condition ensuring thermodynamic equilibrium for the electrons is that E_F be constant throughout the system. This is accomplished through electron transfer from the n side with high E_F (low ϕ_n) to the p side with low E_F (high ϕ_p). An internal built-in electric field is established because of this charge transfer, resulting in bending of both the valence and conduction bands in the junction region. This ability to support

an electric field reflects the dielectric character of semiconductors. In the bulk of each semiconductor the bands are unaffected as previously noted. Similar band bending occurs over dimensions comparable to the thicknesses of films involved in metal–semiconductor contacts (Fig. 10b), heterojunctions between different semiconductor (Fig. 10c), and in metal–oxide–semiconductor (MOS) transistor structures (Fig. 10d). Since virtually all electronic and optoelectronic semiconductor devices are modeled by these important band-diagram representations, they will be referred to again in later chapters.

1.5 THERMODYNAMICS OF MATERIALS

Thermodynamics is very definite about those events that are impossible. It will say, for example, that reactions or processes are thermodynamically impossible. Thus, gold films do not oxidize and atoms do not normally diffuse up a concentration gradient. On the other hand, thermodynamics is noncommittal about permissible reactions and processes. Thus, even though reactions are thermodynamically favored, they may, in fact, not occur in practice. Films of silica glass should revert to crystalline form at room temperature according to thermodynamics, but the solid-state kinetics are so sluggish that for all practical purposes amorphous SiO_2 is stable. A convenient measure of the extent of reaction feasibility is the Gibbs free-energy function G defined as

$$G = H - TS \tag{1-5}$$

where H is the enthalpy, S the entropy, and T the absolute temperature. Thus, if a system changes from some initial (i) to final (f) state at constant temperature due to a chemical reaction or physical process, a free-energy change $\Delta G = G_f - G_i$ occurs given by

$$\Delta G = \Delta H - T\Delta S \tag{1-6}$$

where ΔH and ΔS are the corresponding enthalpy and entropy changes. A consequence of the Second Law of Thermodynamics is that spontaneous reactions occur at constant temperature and pressure when ΔG is negative, i.e., $\Delta G < 0$. This condition implies that a system will naturally tend to minimize its free energy by successively proceeding from a value G_i to a still lower, more negative value G_f until it is no longer possible to further reduce G. When this happens, $\Delta G = 0$. The system has achieved equilibrium and there is no longer a driving force for change.

On the other hand, for a process that cannot occur, $\Delta G > 0$. It is important to note that neither the sign of ΔH nor that of ΔS taken

individually is sufficient to determine reaction direction; rather, it is the sign of the combined function ΔG that is crucial in this regard. For example, during condensation of a vapor to form a solid film, $\Delta S < 0$. This is so because S is a measure of the "disorder" or number of atomic configuration in a system, and fewer exist in the solid relative to the gas phase. However, the decrease in enthalpy more than offsets that in entropy, and the overall net change in ΔG is negative.

The concept of minimization of free energy, as a criterion for both stability in a system and forward change in a reaction or process, is a central theme in materials science. It is used most notably in the thermodynamic analysis of chemical reactions and in the interpretation of phase diagrams. Subsequent applications will be made to such topics as chemical vapor deposition, interdiffusion, reactions in thin films, and energetics in general.

1.5.1 Chemical Reactions

The general chemical reaction involving three substances A, B, and C in equilibrium is

$$aA + bB \rightarrow cC. \tag{1-7}$$

Correspondingly, the free-energy change of the reaction is given by

$$\Delta G = cG_C - aG_A - bG_B \tag{1-8}$$

where a, b, and c are the stoichiometric coefficients. It is customary to denote the free energy of individual reactant or product atomic or molecular species by

$$G_i = G_i^0 + RT \ln a_i, \tag{1-9}$$

where R is the gas constant and G_i^0 is the free energy of the species in its so-called reference or standard state. For solids this is usually the stable pure material at one atmosphere and 298 K. The activity a_i may be viewed as an effective thermodynamic concentration and reflects the change in free energy of the species when it is not in its standard state. Substitution of Eq. 1-9 into Eq. 1-8 yields

$$\Delta G = \Delta G^0 + RT \ln \frac{a_C^c}{a_A^a a_B^b} \tag{1-10}$$

where $\Delta G^0 = cG^0_C - aG^0_A - bG^0_B$. If the system is now in equilibrium, $\Delta G = 0$ and all a_i values are the equilibrium ones, i.e., $a_{i(eq)}$. Thus,

$$0 = \Delta G^0 + RT \ln \left\{ \frac{a^c_{C(eq)}}{a^a_{A(eq)} a^b_{B(eq)}} \right\} \tag{1-11}$$

or

$$-\Delta G^0 = RT \ln K, \tag{1-12}$$

where the equilibrium constant K is defined by the quantity in braces. Equation 1-12 is one of the most frequently used equations in chemical thermodynamics and will be helpful in analyzing chemical vapor deposition reactions. Combination of Eqs. 1-10 and 1-11 gives

$$\Delta G = RT \ln \left\{ \frac{(a_C/a_{C(eq)})^c}{(a_A/a_{A(eq)})^a (a_B/a_{B(eq)})^b} \right\}. \tag{1-13}$$

Each term $a_i/a_{i(eq)}$ represents a supersaturation of the species if it exceeds 1, and a subsaturation if it is less than 1. Thus, if there is a supersaturation of reactants and a subsaturation of products $\Delta G < 0$. The reaction proceeds spontaneously as written with a driving force proportional to the magnitude of ΔG. For many practical cases the a_i are little different from the standard state activities, which are taken to be unity. Therefore, in such a case Eq. 1-10 yields

$$\Delta G = \Delta G^0. \tag{1-14}$$

Quantitative information on the feasibility of chemical reactions is thus provided by values of ΔG^0, and these can be evaluated from standard references (Ref. E2) and computerized data bases (Ref. E3) dealing with thermodynamic properties of materials. The reader should be aware that although many of the data are the result of measurement, some values are inferred from various connecting thermodynamic laws and relationships. In this way a consistent set of thermodynamic data for a very large number of materials has been generated. Thus, even though the vapor pressure of tungsten at room temperature cannot be directly measured, its value is nevertheless "known." It should be borne in mind that the data refer to equilibrium conditions only, and that many reactions are subject to overriding kinetic limitations despite otherwise favorable thermodynamic indications.

A particularly handy representation of ΔG^0 data for formation of metal oxides as a function of temperature is shown in Fig. 1-11 and known as an Ellingham diagram. To understand its use suppose it is desired to deposit

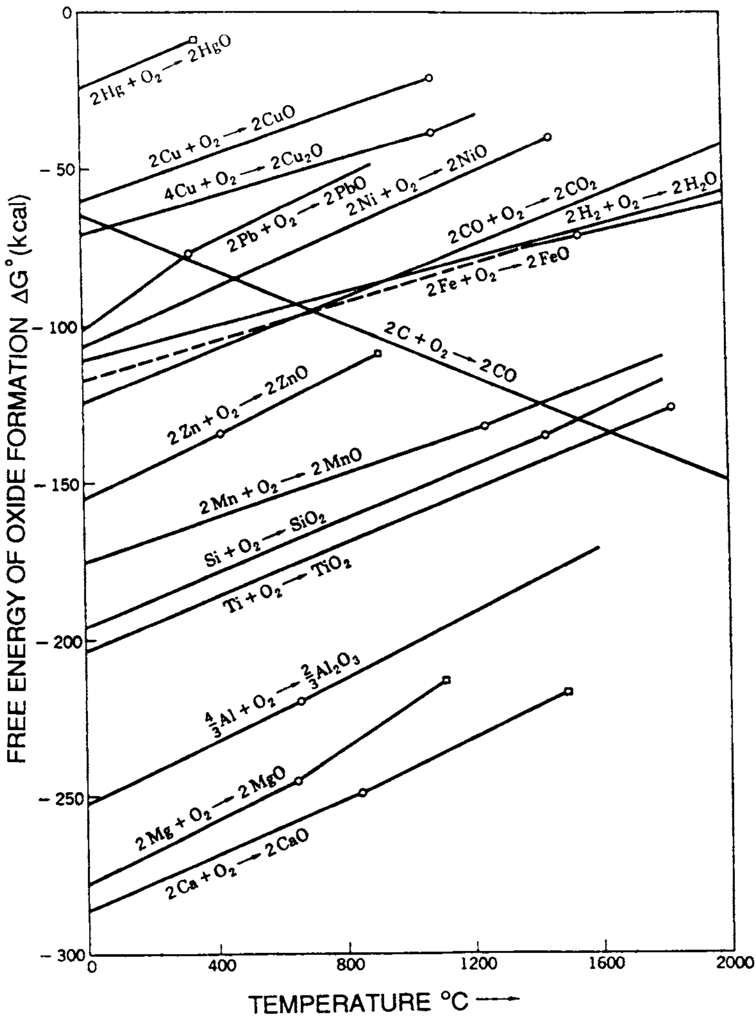
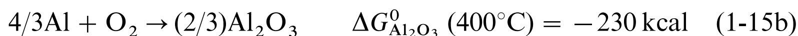


Figure 1-11 Standard free energy of oxide formation vs temperature: ○, melting point of metal; □, boiling point of metal (1 atm). Reprinted with permission from A. G. Guy, *Introduction to Materials Science*, McGraw-Hill, New York, 1972.

thin-film metal interconnections on a SiO_2 substrate for microelectronic purposes. Because of their high conductivity and ease of deposition, Al and Cu are possible candidates. Which metal would be a better choice for this application, other things being equal (which they never are)? Assuming the

deposition temperature is 400°C, the relevant oxidation reactions from Fig. 1-11 are

$$Si + O_2 \rightarrow SiO_2 \qquad \Delta G^0_{SiO_2}\ (400°C) = -175\ \text{kcal} \qquad (1\text{-}15a)$$



$$4/3Al + O_2 \rightarrow (2/3)Al_2O_3 \qquad \Delta G^0_{Al_2O_3}\ (400°C) = -230\ \text{kcal} \qquad (1\text{-}15b)$$

$$2Cu + O_2 \rightarrow 2CuO \qquad \Delta G^0_{CuO}\ (400°C) = -45\ \text{kcal.} \qquad (1\text{-}15c)$$

After elimination of O_2 and algebraic addition of free energies the two reduction reactions are characterized by

$$(4/3)Al + SiO_2 \rightarrow (2/3)Al_2O_3 + Si \qquad \Delta G^0\ (400°C) = -55\ \text{kcal} \qquad (1\text{-}16a)$$

$$2Cu + SiO_2 \rightarrow 2CuO + Si \qquad \Delta G^0\ (400°C) = +130\ \text{kcal.} \qquad (1\text{-}16b)$$

The $\Delta G^0 - T$ curves for Al_2O_3 and CuO are respectively more negative (lower) and more positive (higher) than that for SiO_2; thus reactions 1-16a and 1-16b are respectively thermodynamically possible and impossible, as written. Therefore, Al films tend to reduce SiO_2 films, leaving free Si behind. To minimize this problem practically, TiN or W diffusion barriers are interposed between the Al and SiO_2. However, no such reaction occurs with Cu, and on this basis it is the preferred metallization. In fact Cu is now replacing Al which has long been used for this purpose. As a generalization, the metal of an oxide that has a more negative ΔG^0 than a second oxide will reduce the latter and be oxidized in the process.

Let us now consider the gas ambient required to evaporate pure Al. Use of Eqs. 1-12 and 1-15b indicates that

$$K = \frac{(a_{Al_2O_3})^{2/3}}{(a_{Al})^{4/3} P_{O_2}} = \exp - \Delta G^0/RT. \qquad (1\text{-}17)$$

The Al_2O_3 and Al may be considered to exist in pure standard states with unity activities while the activity of O_2 is taken to be its partial pressure P_{O_2}. Therefore, $\Delta G^0 = RT \ln P_{O_2}$. During evaporation of Al from a crucible to produce a film, the value of P_{O_2} in equilibrium with both Al and Al_2O_3 can be calculated at any temperature once ΔG^0 is known. If the actual oxygen partial pressure exceeds the equilibrium pressure, then Al ought to oxidize. However, if the reverse is true, Al_2O_3 would be reduced to Al. At 1000°C (or 1273 K), $\Delta G^0 = -202$ kcal, and $P_{O_2} = 2 \times 10^{-35}$ atm. Since this value is many orders of magnitude below actual oxygen partial pressures in vacuum systems, Al would be expected to oxidize. It does to some extent and a thin film of alumina forms on the surface of the molten aluminum source. Nevertheless, oxide-free films are effectively deposited in practice. The reader can easily show that oxide-free Cu films are more easily evaporated.

Similar Ellingham plots of free energy of formation versus temperature exist for sulfides, carbides, nitrides, and chlorides. In Chapter 6 we will consider such a diagram for Si–H–Cl compounds because of its utility in the thermodynamic analysis of the chemical vapor deposition of Si.

1.5.2 Phase Diagrams

The most widespread method for representing the conditions of chemical equilibrium for inorganic systems as a function of composition, temperature, and pressure is through the use of phase diagrams. By phases we not only mean the solid, liquid, and gaseous states of pure elements and compounds; also included are materials of variable but homogeneous composition such as alloys. At one level phase diagrams simply appear to provide solubility data as a function of temperature. But at a deeper level they contain a wealth of thermodynamic information on systems in equilibrium that can readily be interpreted without resorting to complex laws, functions, or equations. They have been experimentally determined for many systems by numerous investigators over the years and provide an invaluable guide when synthesizing materials. There are a few simple rules for analyzing phase diagrams. The most celebrated of these is the Gibbs phase rule, which though deceptively simple, is arguably the most important linear algebraic equation in all of physical science. It can be written as

$$F = n + 2 - \psi \tag{1-18}$$

where n is the number of components (i.e., different atomic species), ψ is the number of phases, and F is the number of degrees of freedom or variance in the system. The number of intensive variables that can be independently varied without changing the phase equilibrium is equal to F.

1.5.2.1 One-Component System

As an application of a one-component system consider the $P-T$ diagram given for carbon in Fig. 1-12. Shown are the regions of stability for different phases of carbon as a function of pressure and temperature. Within the broad areas the single phases diamond and graphite are stable. Both P and T variables can be independently varied to a greater or lesser extent without leaving the single-phase field. This is consistent with the phase rule, which gives $F = 1 + 2 - 1 = 2$. Those states that lie on any of the lines of the diagram represent two-phase equilibria. Now $F = 1 + 2 - 2 = 1$. This means, for example, that in order to change but maintain the equilibrium along the diamond–graphite line, only one variable, either T or P, can be

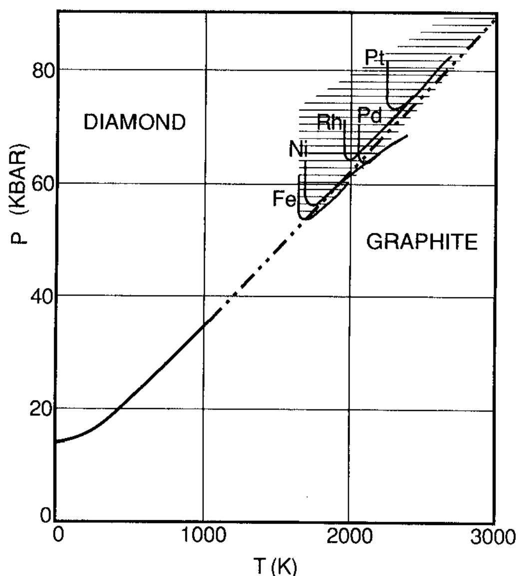


Figure 1-12 Portion of the pressure–temperature diagram for carbon showing stability domains of diamond and graphite. Shaded areas represent regions of diamond formation in the indicated metal solvents. Reprinted with permission from R. C. DeVries, *Ann. Rev. Mater. Sci.* **17**, 161 (1987).

independently varied; the corresponding variables P or T must change in a *dependent* fashion. At a point where three phases coexist (not shown), $F = 0$. Any change of T or P will destroy this unique three-phase equilibrium, leaving instead either one or two phases. The diagram suggests that pressures between 10^4 to 10^5 bars (approximately 10^4 to 10^5 atm) are required to transform graphite into diamond. In addition, excessively high temperatures (2000 K) are required to make the reaction proceed at appreciable rates. It is exciting, therefore, that diamond thin films have been deposited, for example, by decomposing methane in a microwave plasma at low pressure and temperature (Section 6.8.5.2), thus avoiding the almost prohibitive pressure–temperature conditions required for bulk diamond synthesis.

1.5.2.2 Two-Component Systems

When two elements or compounds combine, many very important but less well-known materials than the compounds of inorganic chemistry can be produced. Binary metal alloys (e.g., nichrome, a Ni–Cr alloy) and

compound semiconductors (e.g., GaAs) are examples that have important bulk as well as thin-film uses. The resultant phases that form as a function of initial reactant proportions and temperature are depicted on binary equilibrium phase diagrams. Collections of these have been published for metal, semiconductor, and ceramic systems and are among the most frequently consulted references in the field of materials. Unless noted otherwise, these diagrams hold at atmospheric pressure, in which case the variance is reduced by one. The Gibbs phase rules now states $F = n + 1 - \psi$ or $F = 3 - \psi$. Thus, three phases at most can coexist in equilibrium.

To learn how to interpret binary phase diagrams, let us first consider the Ge–Si system shown in Fig. 1-13. Such a system is interesting because of the possibility of creating semiconductors having properties intermediate to those of Ge and Si. On the horizontal axis the overall composition is indicated. Pure Ge and Si components are represented at the extreme left and right axes, respectively, whereas compositions of initial mixtures of Ge and Si are located in between. Such compositions are given in either weight or atomic percent. The following set of rules will enable a complete equilibrium phase analysis for an initial phase composition X_0 heated to temperature T_0. This entails a simultaneous determination of phase composition and relative phase amount.

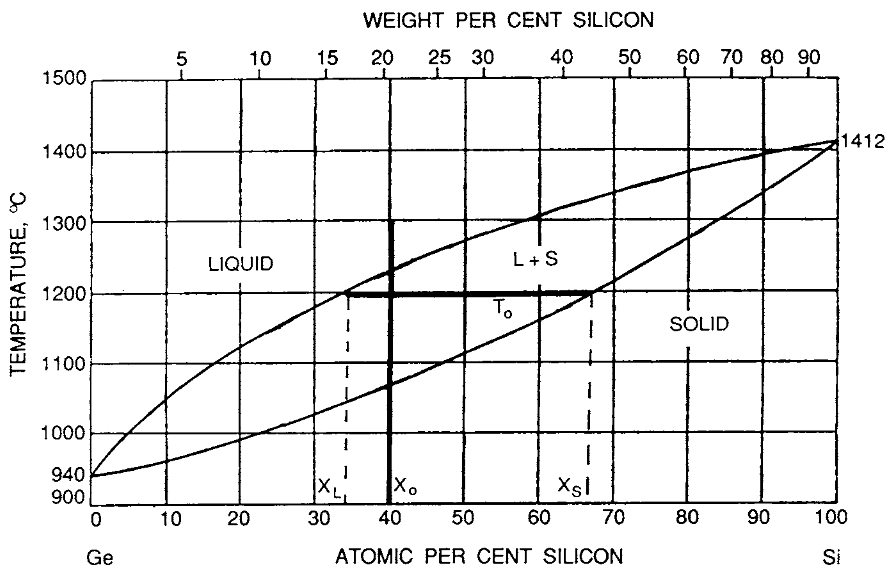


Figure 1-13 Ge–Si equilibrium phase diagram. Reprinted with permission. From M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.

1. Draw a vertical line at composition X_0. Any point on this line represents a state of this system at the temperature indicated on the left-hand scale.

2. The chemical compositions of the resulting phases depend on whether the point lies (a) in a one-phase field, (b) in a two-phase field, or (c) on a sloping or horizontal (isothermal) boundary between phase fields.

(a) For states within a single-phase field, i.e., L (liquid), S (solid), or a compound, the phase composition or chemical analysis is always the same as the initial composition.

(b) In a two-phase region, e.g., L + S, $\alpha + \beta$, a horizontal tie line is first drawn through the state point extending from one end of the two-phase field to the other as shown in Fig. 1-13. On either side of the two-phase field are the indicated single-phase fields (L and S). The compositions of the two phases in question are given by projecting the ends of the tie line vertically down and reading off the values. For example, if $X_0 = 40$ at% Si at $T_0 = 1200°C$, $X_L = 34$ at% Si and $X_S = 67$ at% Si.

(c) State points located exactly on either a sloping or a horizontal boundary cannot be analyzed; phase analyses can only be made above or below the boundary lines according to the rules of (a) and (b) above. Sloping boundaries are known as liquidus or solidus lines for L/L + S or L + S/S phase field combinations, respectively. Such lines also represent solubility limits and are, therefore, associated with the processes of solution or rejection of phases (precipitation) from solution. The horizontal isothermal boundaries indicate the existence of phase transformations involving three phases. The following common reactions occur at these critical isotherms, where α, β, and γ are solid phases:

Eutectic $\quad \mathrm{L} \rightarrow \alpha + \beta \quad$ (1-19a)

Eutectoid $\quad \gamma \rightarrow \alpha + \beta \quad$ (1-19b)

Peritectic $\quad \mathrm{L} + \alpha \rightarrow \gamma \quad$ (1-19c)

3. The relative amounts of phases present depend on whether the state point lies in (a) a one-phase field, or (b) a two-phase field.

(a) Here the one phase in question is homogeneous and present exclusively. Therefore, the relative amount of this phase is 100%.

(b) In the two-phase field the *lever rule* must be applied to obtain the relative phase amounts. Referring to Fig. 1-13, for state X_0 at T_0 and with the indicated tie line, the relative amounts of L and S phases

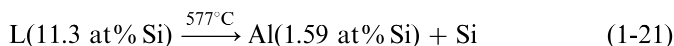
are given by

$$\%L = \frac{X_S - X_0}{X_S - X_L} \times 100; \qquad \%S = \frac{X_0 - X_L}{X_S - X_L} \times 100, \tag{1-20}$$

where %L + %S = 100. Substitution gives %L = (67 − 40)/(67 − 34) × 100 = 81.8, and %S = (40 − 34)/(67 − 34) × 100 = 18.2. Equation 1-20 represents a definition of the lever rule which essentially ensures conservation of mass in the system. It is important to reiterate that the tie line and lever rule can only be applied in a two-phase region; they make no sense in a one-phase region. While such analyses reveal information on phase compositions and amounts, they say nothing about the physical appearance or shape that phases actually assume. Structural morphology is rather dependent on issues related to phase nucleation and growth processes.

Before leaving the Ge–Si system, we should note that L represents a broad liquid solution field where Ge and Si atoms mix in all proportions. Similarly, at lower temperatures, Ge and Si atoms mix randomly but on the atomic sites of a diamond cubic lattice to form a substitutional solid solution. The lens-shaped L + S region separating the single-phase liquid and solid fields occurs in many binary systems including Cu–Ni, Ag–Au, Pt–Rh, Ti–W, and Al_2O_3–Cr_2O_3.

A very common feature on binary phase diagrams is the eutectic isotherm. The Al–Si system shown in Fig. 1-14 is an example of a system undergoing a eutectic transformation at 577°C. Alloy films containing about 1 at% Si have been used to contact silicon in integrated circuits. The insert in Fig. 1-14 indicates that the solid-state reactions for this alloy involve either the formation of an Al-rich solid solution above 520°C, or the rejection of Si below this temperature in order to satisfy solubility requirements. Although this particular alloy cannot undergo a eutectic transformation, all alloys containing more than 1.59 at% Si can. When crossing the critical isotherm from high temperature, the eutectic reaction



$$\text{L(11.3 at\% Si)} \xrightarrow{577^\circ\text{C}} \text{Al(1.59 at\% Si)} + \text{Si} \tag{1-21}$$

occurs. Three phases coexist at the eutectic temperature and therefore $F = 0$. Any change in temperature or phase composition will drive this very precarious three-phase equilibrium into single- (i.e., L) or two-phase fields (i.e., L + Al, L + Si, Al + Si).

The important GaAs system shown in Fig. 1-15 contains two independent side-by-side eutectic reactions at 29.5 and 810°C. For the purpose of analyses one can consider there to be two separate eutectic subsystems,

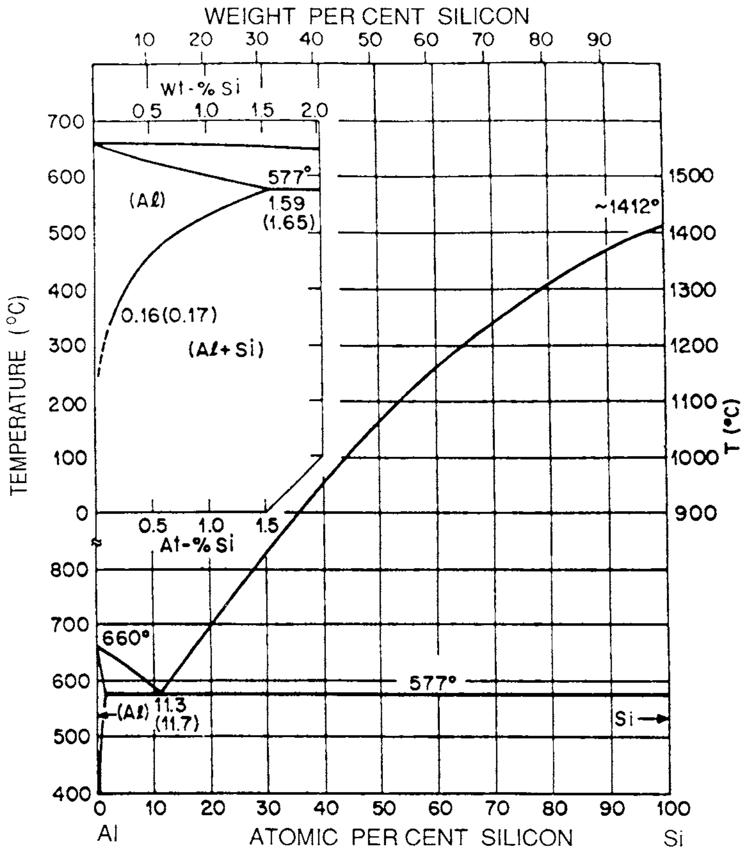


Figure 1-14 Al–Si equilibrium phase diagram. Reprinted with permission. From M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.

Ga–GaAs and GaAs–As. In this way complex diagrams can be decomposed into simpler units. The critical eutectic compositions occur so close to either pure component that they cannot be resolved on the scale of this figure. At the diagram's center, the prominent vertical line represents the stoichiometric GaAs compound, which melts at 1238°C. Phase diagrams for several other important III–V semiconductors, e.g., InP, GaP, and InAs, have very similar appearances. These compound semiconductors are common in other ways. For example, one of the components, e.g., Ga, In, has a low melting point coupled with a rather low vapor pressure, while the other component, e.g., As, P, has a higher melting point and a high vapor pressure. These properties complicate both bulk and thin-film single-crystal growth processes.

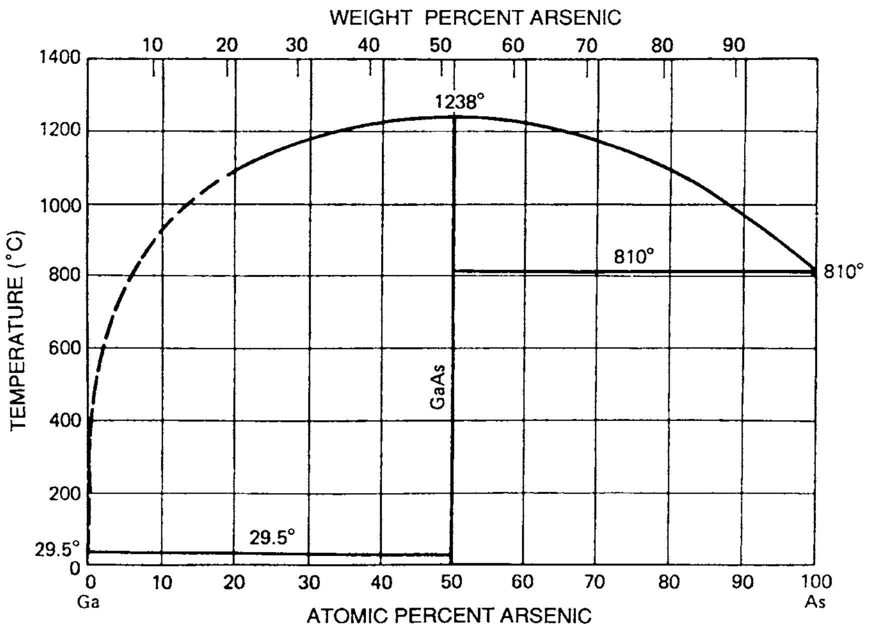


Figure 1-15 Ga–As equilibrium phase diagram. Reprinted with permission. From M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.

We end this section on phase diagrams by reflecting on some distinctions in their applicability to bulk and thin-film materials. High-temperature phase diagrams were first determined in a systematic way for binary metal alloys. The traditional processing route for the bulk metals generally involves melting at a high temperature followed by solidification and subsequent cooling to the ambient. It is a reasonable assumption that thermodynamic equilibrium is attained in these systems especially at elevated temperatures. Atoms in metals have sufficient mobility to enable stable phases to nucleate and grow within reasonably short measurement times. This is not generally the case in metal oxide systems, however, because sluggish atomic motion in viscous melts often results in the formation of metastable glasses.

In contrast, thin films do not generally pass from a liquid phase through a vertical succession of phase fields. For the most part, thin film science and technology is characterized by low-temperature processing where equilibrium is difficult to achieve. Depending on what is being deposited and the conditions of deposition, thin films can be readily produced with varying degrees of thermodynamic stability. For example, single-crystal silicon is the

most stable form of this element below the melting point. Nevertheless, chemical vapor deposition of Si from chlorinated silanes at 1200°C will yield single-crystal films, whereas amorphous films can be produced below 600°C. In between, polycrystalline Si films of varying grain size can be deposited. Since films are laid down an atomic layer at a time, the thermal energy of individual atoms impinging on a massive cool substrate heat sink can be transferred to the latter at an extremely rapid rate. Deprived of energy, the atoms are relatively immobile. It is not surprising, therefore, that metastable and even amorphous semiconductor and alloy films can be evaporated or sputtered onto cool substrates. When such films are heated, they crystallize and revert to the more stable phases indicated by the phase diagram.

Interesting issues related to binary phase diagrams arise with multicomponent thin films that are deposited in layered structures through sequential deposition from multiple sources. For example, the so-called strained-layer superlattices of Ge–Si have been grown by molecular beam epitaxy (MBE) as well as CVD techniques (Section 8.6.3.2). Films of Si and Si + Ge solid-solution alloy, each typically tens of nanometers thick, have been sequentially deposited such that the resultant composite film is a single crystal with strained lattice bonds. The resolution of distinct layers as revealed by the transmission electron micrograph of Fig. 8-29 is suggestive of a two-phase mixture with discernable phase boundaries. On the other hand, a single crystal implies a single phase even if it possesses a modulated chemical composition. Either way, the superlattice is not in thermodynamic equilibrium because the Ge–Si phase diagram unambiguously predicts a stable solid solution at low temperature. Equilibrium can be accelerated by heating resulting in film homogenization by interatomic diffusion. In layered thin-film structures, phases such as solid solutions and compounds are frequently accessed *horizontally* across the phase diagram during an isothermal anneal. This should be contrasted with bulk materials where equilibrium phase changes commonly proceed *vertically* downward from elevated temperatures.

1.6 KINETICS

1.6.1 MACROSCOPIC TRANSPORT

Whenever a material system is not in thermodynamic equilibrium, driving forces arise naturally to push it toward equilibrium. Such a situation can occur, for example, when the free energy of a microscopic system varies from point to point because of compositional inhomogeneities. The resulting atomic concentration gradients trigger time-dependent, mass-

transport effects that reduce free-energy variations in the system. Manifestations of such processes include phase transformations, recrystallization, compound growth and degradation phenomena in both bulk as well as thin-film systems. In solids, mass transport is accomplished by diffusion, which may be defined as the migration of an atomic or molecular species within a given matrix under the influence of a concentration gradient. Fick established the phenomenological connection between concentration gradients and the resultant diffusional transport through the equation

$$J = -D\frac{dC}{dx}. \tag{1-22}$$

Because the vectors representing the concentration gradient (dC/dx) and mass flux J are oppositely directed, there is a minus sign in Eq. 1-22. Thus an increasing concentration in the positive x direction induces mass flow in the negative x direction, and vice versa. The units of C are typically atoms/cm^3, while the magnitude of the diffusion coefficient, D (cm^2/s), which depends on both the diffusing species and matrix, determines the amount of observable mass transport. As we shall later note, D increases in exponential fashion with temperature according to a Maxwell–Boltzmann relation, i.e.,

$$D = D_0 \exp -\frac{E_{\mathrm{D}}}{k_{\mathrm{B}}T} \tag{1-23}$$

where D_0 is a constant, E_{D} is the activation energy for diffusion, and $k_{\mathrm{B}}T$ has the usual meaning.

Even at moderate temperatures solid-state diffusion is generally a slow process with compositional change occurring over a long time; the steady-state condition of time-independent concentrations rarely occurs in bulk solids. Instead, mass flow into and out of regions cause time-varying accumulations or depletions of atomic species governed, in one dimension, by the well-known equation

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2}. \tag{1-24}$$

The non-steady-state heat-conduction equation is identical if temperature is substituted for C and the thermal diffusivity for D. Many solutions for both diffusion and heat-conduction problems exist for media of varying geometries, constrained by assorted initial and boundary conditions (Refs. F3, F4). Since complex applications involving Eq. 1-24 will be discussed subsequently, we introduce simpler solutions here.

Consider an initially pure thick film into which some solute diffuses from the surface. If the film thickness is very large or effectively infinite compared to the extent of diffusion, the situation can be physically modeled by the

following initial and boundary conditions:

$$C(x, 0) = 0 \quad \text{at } t = 0 \qquad \text{for } \infty > x > 0 \tag{1-25a}$$

$$C(\infty, t) = 0 \quad \text{at } x = \infty \qquad \text{for } t > 0. \tag{1-25b}$$

A second boundary condition that must be specified concerns the nature of the diffusant distribution maintained at the film surface $x = 0$. Two simple cases can be distinguished. In the first, a thick layer of diffusant provides an essentially limitless supply of atoms maintaining a constant surface concentration C_0 for all time. In the second case, a very thin layer of diffusant provides an instantaneous source S_0 of surface atoms per unit area. Here the surface concentration diminishes with time as atoms diffuse into the underlying substrate. These two cases are respectively described by

$$C(0, t) = C_0 \tag{1-26a}$$

and

$$\int_0^\infty C(x, t)\, dx = S_0. \tag{1-26b}$$

Expressions for $C(x, t)$ satisfying the above set of conditions are respectively

$$C(x, t) = C_0 \operatorname{erfc} \frac{x}{\sqrt{4Dt}} = C_0 \left(1 - \operatorname{erf} \frac{x}{\sqrt{4Dt}}\right); \tag{1-27a}$$

and

$$C(x, t) = \frac{S_0}{\sqrt{\pi D t}} \exp - \left(\frac{x^2}{4Dt}\right). \tag{1-27b}$$

These represent the simplest mathematical solutions to the diffusion equation and have been employed to determine doping profiles and junction depths in semiconductors. The error function, erf $x/\sqrt{4Dt}$, defined by

$$\operatorname{erf} \frac{x}{\sqrt{4Dt}} = \int_0^{x/\sqrt{4Dt}} e^{-Z^2}\, dZ, \tag{1-28}$$

is a tabulated function of only the upper limit or argument $x/(4Dt)^{1/2}$. Normalized concentration profiles for the erfc and Gaussian (Eq. 1-27) solutions are graphically displayed in Fig. 1-16. Often it is of interest to calculate how the dimensions (x) of an atomic distribution spreads with time; a measure of this is given by the random walk expression,

$$x = 2(Dt)^{1/2}. \tag{1-29}$$

When $(Dt)^{1/2}$ becomes large compared to the film dimensions, the assumption of an infinite matrix is not valid and the solutions do not hold. Film properties may also change appreciably because of interdiffusion. To limit

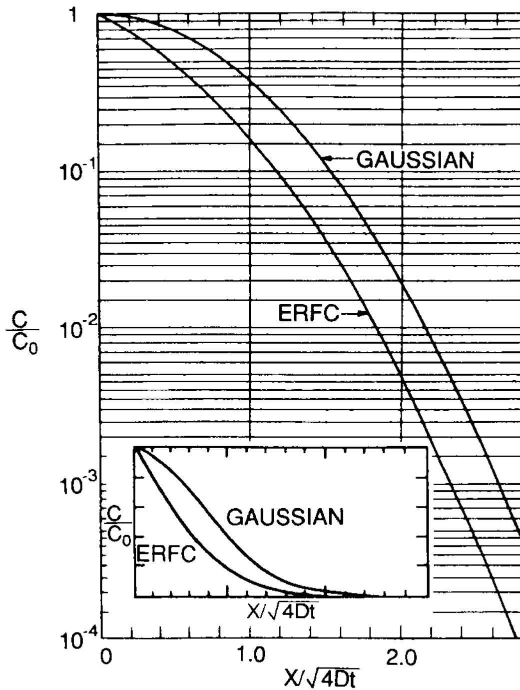


Figure 1-16 Normalized Gaussian and erfc curves of C/C_0 vs $x/(4Dt)^{1/2}$. Both logarithmic and linear scales are shown. Reprinted with permission from W. E. Beadle, J. C. C. Tsai, and R. D. Plummer, *Quick Reference Manual for Silicon Integrated Circuit Technology*, copyright © 1985, Bell Telephone Laboratories Inc. Published by J. Wiley & Sons, New York, 1985.

the latter and ensure the integrity of films, D should be kept small, which in effect means the maintenance of low temperatures. This subject will be discussed further in Chapter 11.

1.6.2 Atom Movements

1.6.2.1 Diffusional Transport

Macroscopic changes in composition during diffusion are the result of the random motion of countless individual atoms unaware of the concentration gradient they have established. On a microscopic level, it is sufficient to explain how atoms execute individual jumps from one lattice site to another because macroscopic changes are manifested by countless repetitions of such

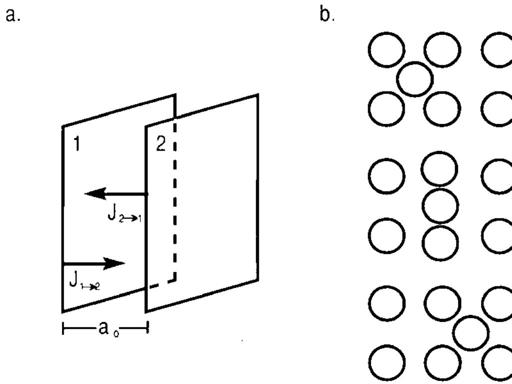


Figure 1-17 (a) Atomic diffusion fluxes between neighboring crystal planes. (b) Atomistic view of atom jumping into a neighboring vacancy.

unit jumps. Consider Fig. 1-17a showing neighboring lattice planes spaced a distance a_0 apart within a region where an atomic concentration gradient exists. If there are n_1 atoms per unit area of plane 1, then in plane 2, $n_2 = n_1 + (dn/dx)a_0$, where we have taken the liberty of assigning a continuum behavior at discrete planes. Each atom vibrates about its equilibrium position with a characteristic lattice frequency, v, typically equal to $10^{13}\ \mathrm{s}^{-1}$. Very few vibrational cycles have sufficient amplitude to cause the atom to actually jump into an adjoining occupied lattice site resulting in a direct atomic interchange. This process would be greatly encouraged, however, if neighboring sites were vacant. The fraction of vacant lattice sites was previously given by $\exp^{-E_f/k_BT}$ (Eq. 1-3). In addition, the diffusing atom must acquire sufficient energy to push the surrounding atoms apart so that it can squeeze past and land in the so-called activated state shown in Fig. 1-17b. This step is the precursor to the downhill jump of the atom into the vacancy. The number of times per second that an atom successfully reaches the activated state is $v \exp^{-E_m/k_BT}$, where E_m is the vacancy jump or migration energy per atom. Here the Boltzmann factor may be interpreted as the fraction of all sites in the crystal that have an activated state configuration. Atom fluxes that pass from plane 1 to 2 and from plane 2 to 1 are respectively given by

$$J_{1\to 2} = \frac{1}{6} v \exp - \frac{E_m}{k_B T} \cdot \exp - \frac{E_f}{k_B T}(Ca_0) \tag{1-30a}$$

$$J_{2\to 1} = \frac{1}{6} v \exp - \frac{E_m}{k_B T} \cdot \exp - \frac{E_f}{k_B T}\left(C + \frac{dCa_0}{dx}\right) a_0, \tag{1-30b}$$

where we have substituted Ca_0 for n and used the factor of 1/6 to account for bidirectional jumping in each of the three coordinate directions. The net flux J_N is the difference or

$$J_N = -\frac{1}{6} a_0^2 v \exp - \frac{E_m}{k_B T} \exp - \frac{E_f}{k_B T} \left(\frac{dC}{dx} \right). \tag{1-31}$$

By association with Fick's law and the expression for D (Eq. 1-23) we see that $D_0 = \frac{1}{6} a_0^2 v$ and $E_D = [E_f + E_m]$.

Although the above model is intended for atomistic diffusion in the bulk lattice, a similar expression for D would hold for transport through grain boundaries or along surfaces and interfaces of films. At such nonlattice sites, energies for defect formation and motion are expected to be lower, leading to high diffusivities. Dominating microscopic mass transport is the Boltzmann factor $\exp - E_D/k_B T$, which is ubiquitous when describing the temperature dependence of many rate processes in thin films. In such cases the kinetics can be described graphically by an Arrhenius plot in which the logarithm of the rate is plotted on the ordinate, and the reciprocal of the absolute temperature on the abscissa. The slope of the resulting line is then equal to $-E_D/k_B$, from which the characteristic activation energy can be extracted.

1.6.2.2 Atom Transport in a Force Field

The discussion to this point is applicable to the motion of both impurity as well as matrix atoms. In the latter case we speak of self-diffusion. For matrix atoms there are driving forces other than concentration gradients that often cause them to migrate. Examples are forces due to stress fields, electric fields, and interfacial energy gradients. To visualize their effect, consider neighboring atomic positions in a crystalline solid where no fields are applied. The free energy of the system has the periodicity of the lattice and varies schematically as shown in Fig. 1-18a. Imposition of an external field now biases or tilts the system such that the free energy is lower in site 2 relative to 1 by an amount $2\Delta G$. A free-energy gradient or force exists in the system that lowers the energy barrier to motion from $1 \rightarrow 2$ and raises it from $2 \rightarrow 1$ (Fig. 1-18b). The rate at which atoms move from 1 to 2 is given by

$$r_{12} = v \exp\left[-\frac{G_D - \Delta G}{RT} \right] \mathrm{s}^{-1}. \tag{1-32a}$$

Similarly,

$$r_{21} = v \exp\left[-\frac{G_D + \Delta G}{RT} \right] \mathrm{s}^{-1} \tag{1-32b}$$

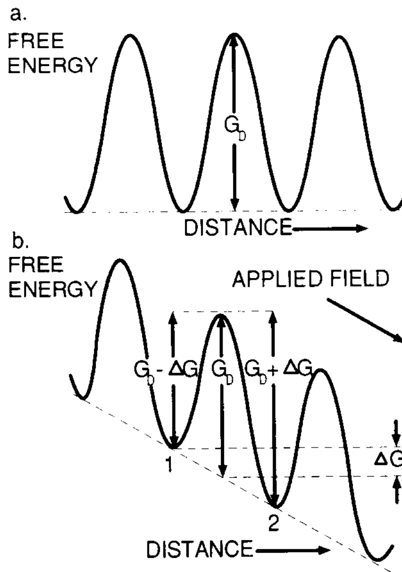


Figure 1-18 (a) Free-energy variation with atomic distance in the absence of an applied field. (b) Free-energy variation with atomic distance in the presence of an applied field.

and the net rate r_N is given by the difference or

$$r_N = \nu \exp - \frac{G_D}{RT}\left\{\exp \frac{\Delta G}{RT} - \exp - \frac{\Delta G}{RT}\right\} = 2\nu \exp - \frac{G_D}{RT} \sinh \frac{\Delta G}{RT} \quad (1\text{-}33)$$

When $\Delta G = 0$, the system is in thermodynamic equilibrium and $r_N = 0$; no net atomic motion occurs. While G_D is typically a few electron volts or so per atom (1 eV = 23,060 cal/mol), ΔG is much smaller in magnitude since it is virtually impossible to impose external forces on solids comparable to interatomic forces. In fact, $\Delta G/RT$ is usually much less than unity so that $\sinh \Delta G/RT \simeq \Delta G/RT$. This leads to commonly observed linear diffusion effects. But when $\Delta G/RT \simeq 1$, nonlinear diffusion effects are possible and are discussed in Section 11.6.2. By multiplying both sides of Eq. 1-33 by a_0, we obtain the atomic velocity v:

$$v = a_0 r_N = [a_0^2 \nu \exp - G_D/RT] \frac{2\Delta G}{a_0 RT}. \quad (1\text{-}34)$$

The term in brackets is essentially the diffusivity D, where G_D is a diffusional activation free energy that for all practical purposes we may equate with E_D.

The term $2\Delta G/a_0$ is a measure of the molar free-energy gradient or applied force F. Therefore, the celebrated Nernst–Einstein equation results:

$$v = DF/RT. \tag{1-35}$$

Application of this equation will be made subsequently to various thin film mass transport phenomena, e.g., electric-field-induced atomic migration (electromigration), stress relaxation, and grain growth. The drift of charge carriers in semiconductors under an applied field can also be modeled by Eq. 1-35. Because of the small dimensions involved, larger generalized forces can often operate in thin films relative to bulk materials.

Chemical reaction-rate theory provides a common application of the above ideas. In Fig. 1-19 the reactants at the left are envisioned to proceed toward the right following the reaction coordinate path. Along the way intermediate activated states are produced upon surmounting the free-energy barrier. Through decomposition of the activated species, products form. If C_R is the concentration of reactants at coordinate position 1 and C_P the concentration of products at 2, then the net rate of reaction is proportional to

$$r_N = C_R \exp - \frac{G^*}{RT} - C_P \exp - \frac{(G^* + \Delta G)}{RT} \tag{1-36}$$

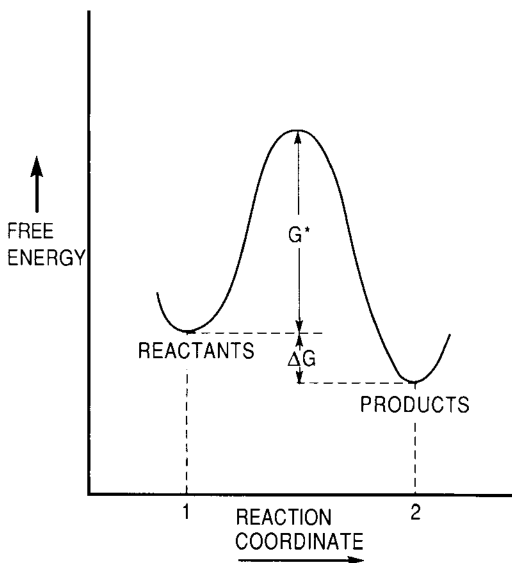


Figure 1-19 Free-energy path for thermodynamically favored reaction $1 \rightarrow 2$.

where G^* is the molar free energy of activation. As before, the Boltzmann factors represent the probabilities of surmounting the respective energy barriers faced by reactants proceeding in the forward direction or products in the reverse direction. When chemical equilibrium prevails, the competing rates are equal and $r_N = 0$. Therefore,

$$\frac{C_P}{C_R} = \exp\frac{\Delta G}{RT} = \frac{\exp - G_P/RT}{\exp - G_R/RT} \tag{1-37}$$

For the reaction to proceed to the right $C_P/C_R > 0$, and $\Delta G = G_R - G_P$ must be positive. By comparison with Eq. 1-12, it is apparent that the left-hand side is the equilibrium constant and ΔG may be associated with $-\Delta G^0$. This expression is perfectly general, however, and applies, for example, to electron energy level populations in semiconductors and lasers, as well as magnetic moment distributions in solids. In fact, wherever thermal energy is a source of activation energy, Eq. 1-37 is valid.

1.7 NUCLEATION

When the critical boundaries separating stable phase fields on equilibrium phase diagrams are crossed, new phases appear. Most frequently a decrease in temperature is involved and this may, for example, trigger vapor-phase condensation, solidification, or solid-state phase transformations from now unstable gases, melts, or solid matrices. When such a transformation occurs, a new phase of generally different structure and composition emerges from the prior parent phase or phases. The process known as nucleation occurs during the very early stages of phase change. It is important in thin films because the grain structure that ultimately develops in a given deposition process is often influenced by what happens during the film nucleation step.

Simple models of nucleation are first of all concerned with thermodynamic questions related to the energetics of forming a single stable nucleus. Once nucleation is possible it is usual to try to specify the nucleation rate ($\dot{N}$) or number of stable nuclei that form per unit volume per unit time. As an example consider the homogeneous nucleation of a spherical solid phase of radius r from a prior supersaturated vapor. Pure homogeneous nucleation is rare but easy to model since it occurs without benefit of the complex heterogeneous sites that exist on an accommodating substrate surface. In such a process the gas-to-solid transformation results in a reduction of the chemical free energy of the system given by $4\pi r^3/3\Delta G_V$ where ΔG_V corresponds to the change in chemical free energy per unit volume. For the condensation reaction, vapor (v) $\rightarrow$ solid (s), Eq. 1-13

suggests that

$$\Delta G_V = \frac{k_B T}{\Omega} \ln \frac{P_S}{P_V} = -\frac{k_B T}{\Omega} \ln \frac{P_V}{P_S} \tag{1-38}$$

where P_S is the vapor pressure above the solid, P_V is the pressure of the supersaturated vapor, and Ω is the atomic volume. A more instructive way to write Eq. 1-38 is

$$\Delta G_V = -k_B T/\Omega \ln(1 + S), \tag{1-39}$$

where S is the vapor supersaturation defined by $(P_V - P_S)/P_S$. Without supersaturation, $S = 0$ and $\Delta G_V = 0$, so that nucleation is impossible. In our example, however, $P_V > P_S$ and ΔG_V is negative, which is consistent with the notion of energy reduction. Simultaneously, new surfaces and interfaces form. This results in an increase in the surface free-energy of the system given by $4\pi r^2 \gamma$, where γ is the solid–vapor interfacial energy per unit area. The total free-energy change in forming the nucleus is thus given by

$$\Delta G = \tfrac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \gamma. \tag{1-40}$$

Minimization of ΔG with respect to r, or $d(\Delta G)/dr = 0$, yields the equilibrium size of r equal to $r^* = -2\gamma/\Delta G_V$. Substitution in Eq. 1-40 gives $\Delta G^* = 16\pi\gamma^3/3(\Delta G_V)^2$. The quantities r^* and ΔG^* are shown in Fig. 1-20, where it is evident that ΔG^* represents an energy barrier to the nucleation process. If a solid-like spherical cluster of atoms momentarily forms by some thermodynamic fluctuation, but with radius less than r^*, the cluster is unstable and will shrink by losing atoms. Clusters larger than r^* have surmounted the nucleation energy barrier and are stable. They tend to grow larger while lowering the overall system energy.

The nucleation rate is essentially proportional to the product of three terms, namely,

$$\dot{N} = N^* A^* \omega \qquad \text{(nuclei/cm}^3\text{-s)}. \tag{1-41}$$

N^* is the equilibrium concentration (per cm^3) of stable nuclei and ω is the rate at which atoms impinge (per cm^2-s) onto the nuclei of critical area, A^* (cm^2). Based on the previous experience of associating the probable concentration of an entity with its characteristic energy through a Boltzmann factor, it is appropriate to take $N^* = n_s \exp - \Delta G^*/k_B T$, where n_s is the density of all possible nucleation sites. The atom impingement flux is equal to the product of the concentration of vapor atoms and the velocity with which they strike the nucleus. In the next chapter we will show that this flux (ω) is effectively given by $\alpha(P_V - P_S)N_A/(2\pi MRT)^{1/2}$ where M is the atomic weight, N_A is Avogadro's number, and α is the sticking coefficient. Since gas

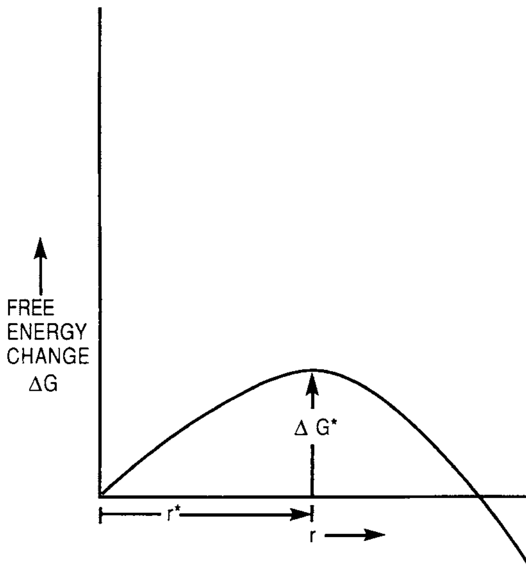


Figure 1-20 Free-energy change (ΔG) as a function of unstable cluster ($r^* > r$) or stable nucleus ($r > r^*$) size (r). r^* is critical nucleus size and ΔG^* is critical free energy barrier for nucleation.

atoms impinge over the entire spherical surface, the nucleus area is simply $4\pi r^2$. Upon combining terms, we obtain

$$\dot{N} = n_s \left[\exp - \frac{\Delta G^*}{k_B T} \right] 4\pi r^2 \frac{\alpha(P_V - P_S)N_A}{\sqrt{2\pi \mathrm{MRT}}}. \tag{1-42}$$

The most influential term in this expression is the exponential factor. It contains ΔG^*, which in turn, is ultimately a function of S. When the vapor supersaturation is sufficiently large, homogeneous nucleation in the gas is possible. This phenomenon is a troublesome one in chemical vapor deposition processes since any solid particles that nucleate may settle on and be incorporated within growing films, thus destroying their integrity.

Heterogeneous nucleation of films is a more complicated subject in view of the added interactions between deposit and substrate. The nucleation surface sites in this case are kinks, ledges, dislocations, etc., which serve to stabilize nuclei of differing size. Capillarity theory will be employed again in Chapter 7 to model heterogeneous nucleation processes. Suffice it to say that when $\dot{N}$ is high during deposition, many crystallites will nucleate and a

fine-grained film results. On the other hand, if nucleation is suppressed, conditions favorable to the growth of a few nuclei or even a single-crystal film are fostered.

1.8 AN INTRODUCTION TO MECHANICAL BEHAVIOR

1.8.1 Stresses, Strains, and Elasticity

Beyond the topics of structure, bonding, thermodynamics, atomic transport, and phase transformations, it is customary to consider materials properties in an elementary review of materials science. Of the many properties we shall only touch on mechanical behavior here because it is prerequisite to a deeper understanding of film structure, stress, and adhesion, all important topics subsequently treated in this volume. Mechanical properties also influence the other physical properties as well. An appropriate way to begin is to consider what is meant by stress. When forces are applied to the surface of a body, they act directly on the surface atoms. The forces are also indirectly transmitted to the internal atoms via the network of bonds which are distorted by the stress field that develops internally. Consider the plate in Fig. 1-21a stretched by equal and opposite axial tensile forces, F. Since the plate is in static equilibrium, if it is cut as shown in Fig. 1-21b, internal forces must act on the exposed surface to keep the isolated section from moving. Regardless of where and at what orientation the cut is made, balancing forces are required to sustain equilibrium. These internal forces distributed throughout the plate constitute a state of stress.

In the example shown, the normal force F divided by the area A defines the tensile stress σ_x. Similarly, normal stresses in the remaining two coordinate directions, σ_y and σ_z can be imagined under more complex loading conditions. If the force is directed into the surface, a compressive stress arises. Convention assigns it a negative sign in contrast to the positive sign for a tensile stress. In addition, mechanical equilibrium on internal surfaces cut at an arbitrary angle will generally necessitate forces resolved in the plane itself; these give rise to shear stresses. It should be noted that the tensile force of Fig. 1-21a produces maximum shear stresses on planes inclined at 45° with respect to the plate axis. If the tensile stress $\sigma_x = F/A$, then the force resolved on these shear planes is $F \cos 45^\circ = \sqrt{2}\,F/2$. The area of the shear planes is $A/\cos 45^\circ = 2A/\sqrt{2}$. Therefore, the shear stress $\tau = F/2A$ and is half of the tensile stress. Shear stresses are extremely important because they are responsible for the plastic deformation of crystalline materials. Two subscripts are generally required to specify a shear

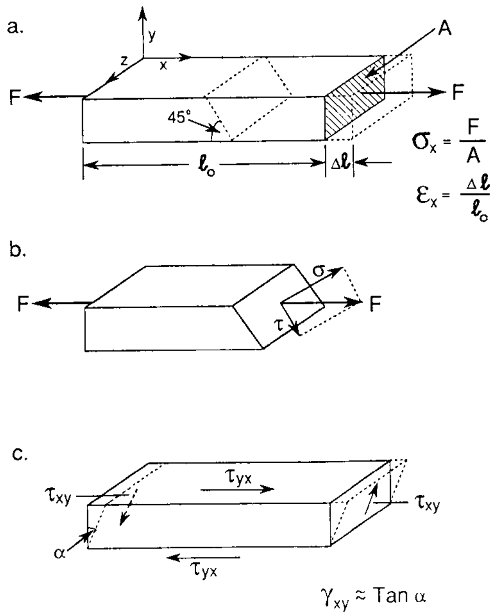


Figure 1-21 (a) Tensile force applied to plate. (b) Arbitrary free-body section revealing spatial distribution of stress through plate. Both tensile and shear stresses exist on exposed plane. (c) Distortion of plate due to applied shear stress.

stress, the first to denote the plane in which shear occurs, and the second to identify the direction of the force in this plane. If, for example, a shear force is applied to the top surface of the plate, it distorts into a prism. In order to balance all forces and moments, a tetrad of shear stresses must act on the horizontal as well as vertical faces. It is left as an exercise to show that $\tau_{xy} = \tau_{yx}$ in equilibrium, and similarly for τ_{xz} and τ_{yz} stresses.

The application of tensile forces extends the plate of Fig. 1-21 by an amount Δl. This results in a normal strain ε_x, defined by $\varepsilon_x = \Delta l / l_0$ where l_0 is the original length. Similarly, in other directions the normal strains are ε_y and ε_z. In the example given, the plate also contracts laterally in both the y and z directions in concert with the longitudinal extension in the x direction. Therefore, even though there is no stress in the y direction, there is a strain given by $\varepsilon_y = -\nu e_x$, and similarly for ε_z. The quantity ν, a measure of this lateral contraction, is Poisson's ratio and for many materials it has a value of about 0.3. Under the action of shear stresses, shear strains (γ) are induced; these are essentially defined by the tangent of the shear distortion angle ϕ in Fig. 1-21c.

The elastic regime, rooted in the theory of elasticity, forms the basis of structural mechanics and much engineering design. In this regime strains are small and Hooke's law dominates the response of the system, i.e.,

$$\sigma_x = Y\varepsilon_x \tag{1-43}$$

where Y is Young's modulus. The strain is thus linearly proportional to the applied stresses and upon unloading, the material snaps back and regains its original shape. Under the action of shear stresses, Hooke's law applies in the form

$$\tau_{xy} = \mu\gamma_{xy} \tag{1-44}$$

where μ is the shear modulus, and similarly for τ_{xz} and τ_{yz}. A quantity of interest, the strain energy per unit volume (E_S), has a magnitude $E_S = 1/2\sigma\varepsilon$, or equivalently

$$E_S = \frac{1}{2}Y\varepsilon^2 = \frac{\sigma^2}{2Y}. \tag{1-45}$$

This energy density, physically identical to the area under the elastic portion of the material's tensile stress–strain curve, is stored in elastically deformed bodies. When a three-dimensional state of stress exists, $\varepsilon_x = 1/Y[\sigma_x - \nu(\sigma_y + \sigma_z)]$, which simplifies to Eq. 1-43 in the absence of σ_y and σ_z. In thin films under plane stress conditions, $\sigma_z = 0$. All of the preceding equations strictly apply to isotropic materials where of the three elastic constants Y, μ, and ν, only two are independent, since they are connected by the relation $\mu = Y/2(1 + \nu)$. In anisotropic media such as the single-crystal quartz plate used to monitor film thickness during deposition (Section 10.2.5.2) the elastic constants reflect the noncubic symmetry of the crystal structure. Although there are more elastic constants to contend with, Hooke's law is still valid.

Elastic stress analysis is generally concerned with specifying the stress and strain distribution throughout the volume of shaped bodies subjected to arbitrary loading conditions; finite element analysis software is now available to solve such problems. Later in Chapter 12 we shall employ concepts of elasticity to derive basic formulas used to determine internal stress in films.

1.8.2 Plastic Behavior

At stress levels above the limit of the elastic response, i.e., yield stress (σ_0), irreversible plastic deformation occurs. Such plastic effects are capitalized upon in metal-forming operations, e.g., rolling, extrusion, drawing, and are

operative in failure phenomena, e.g., creep, fatigue, fracture. Unlike elasticity, plasticity is difficult to model mathematically because the material response is nonlinear and strongly dependent on past thermomechanical processing and treatment history. On a microscopic level, plastic straining of crystalline solids often means the complex motion of large numbers of dislocations and their interaction with other defects, e.g., grain boundaries; in contrast, elastic deformation simply involves interatomic bond stretching. Lastly, plastic deformation effects are often time-dependent, whereas elastic deformation is a time-independent process.

The plastic behavior of bulk solids is often defined by such mechanical properties as yield stress, ultimate tensile strength, percent elongation to failure, and hardness. Only the latter property is commonly cited for thin films and coatings, and therefore it is worthwhile discussing what hardness measures and how it is determined. The hardness H of a material is usually defined as its resistance to local plastic deformation. Since the hardness test consists of pressing a hard indenter into the surface, it is equivalent to performing a highly localized, nonuniform compression test. A correlation between hardness and yield strengh, is therefore expected. Typically $H = 3\sigma_0$ in bulk metals, but the correlation has not been directly verified in films or hard coatings.

Hardness testing is a relatively simple (though deceptively so) measurement to perform. For coatings, the frequently employed Vickers and Knoop hardness tests are based on techniques having long standing in the metallurgical community. The Vickers microhardness test, also known as the diamond pyramid hardness (DPH) test, employs an indenter consisting of a square-based diamond pyramid ground to have a face angle of 136°. Vickers hardness H_v is obtained as the ratio of the applied load, F, to the surface area of the resulting indentation, i.e.,

$$H_v = \frac{2\cos 22^\circ F}{l_v^2} = \frac{1.854F}{l_v^2}\,\frac{\mathrm{kg}}{\mathrm{mm}^2} \tag{1-46}$$

where l_v is the indentation diagonal. The related Knoop microhardness test employs a rhombic-based diamond pyramid indenter where the length ratio of the major to minor indentation diagonals is 7.11; furthermore, the minor diagonal is four times the penetration depth. Knoop hardness is given by

$$H_k = \frac{14.22F}{l_k^2}\,\frac{\mathrm{kg}}{\mathrm{mm}^2} \tag{1-47}$$

where l_k is the length of the major indentation diagonal. Care must be taken in distinguishing H_v and H_k values. Even though the former are some 10–15% lower than the latter, they are sometimes used interchangeably.

Extreme care must be exercised during testing not to indent films or coatings too deeply, because the hardness of the underlying substrate will be measured instead. As we shall see in Chapter 12, nanohardness testing methods have been developed to eliminate such problems.

1.8.3 Stress in Thin Films

What has been sketched above concerning elasticity and plasticity applies both to bulk as well as thin film materials. This is most evident in the hard films and coatings (e.g., TiN, TiC, diamond, Al_2O_3) bonded to substrates such as steel that must resist wear in assorted mechanically functional applications. However, in these as well as electrically and optically functional applications, there are a number of features that distinguish the mechanical behavior of films from that of bulk solids.

1. First and foremost is the paradoxical fact that virtually all films are found to be stressed internally even without the application of external forces!
2. Secondly, the mechanical behavior of films cannot be considered apart from the substrate to which they are attached. Thus, sources of internal film stress include both the mismatch in thermal expansion coefficient and mismatch in lattice parameter between film and substrate.
3. A final distinguishing feature is the often minuscule dimensions of the involved stressed components. This is not only the case on the integrated circuit (IC) device level; on a larger, but nevertheless still small dimensional scale are stresses in assorted films and materials, e.g., contact pads, bonding wires, solder and polymer encapsulents, used in the electronic packaging of IC chips.

The unavoidable consequence of depositing a film at one temperature and using it an another is a major contributor to internal stress in films. To see how thermal stresses arise, consider a film of length l_0 and modulus Y, clamped at both ends. If its temperature is reduced from T_0 to T, it would tend to shrink in length by an amount equal to $\alpha(T - T_0)l_0$, where α is the coefficient of thermal expansion [$\alpha = \Delta l/l_0\Delta T \quad (\mathrm{K}^{-1})$]. But the film is constrained and is, therefore, effectively elongated in tension. The tensile strain magnitude is simply $\varepsilon = \alpha(T - T_0)l_0/l_0 = \alpha(T - T_0)$ and the corresponding thermal stress, by Hooke's law, is

$$\sigma = Y\alpha(T - T_0). \tag{1-48}$$

When a film (f) attached to a substrate (s) strains differentially relative to it, similar thermal stresses arise due both to differences in thermal expansion

coefficients ($\alpha_f - \alpha_s$) and temperature. The analysis of this and other film stress problems will be presented in Chapter 12.

1.9 CONCLUSION

At this point we conclude this introductory sweep through several relevant topics in materials science. If the treatment of structure, bonding, thermodynamics and kinetics, and mechanical properties has introduced the reader to or elevated his prior awareness of these topics, it has served the intended purpose. Threads of this chapter will be woven into the subsequent fabric of the discussion on the preparation, characterization, and properties of thin films.

EXERCISES

1. An FCC film is deposited on the (100) plane of a single-crystal FCC substrate. It is determined that the angle between the [100] directions in the film and substrate is 63.4°. What are the Miller indices of the plane lying in the film surface?
2. Both Au, which is FCC, and W, which is body-centered cubic (BCC), have a density of 19.3 g/cm^3. Their respective atomic weights are 197.0 and 183.9.
   (a) What is the lattice parameter of each metal?
   (b) Assuming both contain hard sphere atoms, what is the ratio of their diameters?
3. Diffraction of 1.5406 Å X-rays from a crystallographically oriented (epitaxial) relaxed bilayer consisting of AlAs and GaAs yields two closely spaced overlapping peaks. The peaks are due to the (111) reflections from both films. The lattice parameters are a_0(AlAs) = 5.6611 Å and a_0(GaAs) = 5.6537 Å. What is the peak separation in degrees?
4. The potential energy of interaction between atoms in an ionic solid as a function of separation distance is given by $V(r) = -A/r + Br^{-n}$, where A, B, and n are constants.
   (a) Derive a relation between the equilibrium lattice distance a_0 and A, B, and n.
   (b) The force constant between atoms is given by $K_s = (d^2V/dr^2)_{r=0}$. If Young's modulus (in units of force/area) is essentially given by K_s/a_0, show that it varies as a_0^{-4} in ionic solids.

5. What is the connection between the representations of electron energy in Figs. 1-8a and 1-9? Illustrate for the case of an insulator. If the material in Fig. 1-8a were compressed, how would E_g change? Would the electrical conductivity change? How?

6. (a) Consider the band diagram representation of the *p–n* junction in Fig. 1-10a. What is the magnitude of the energy step in the conduction-band edge across the junction?
   (b) What is the physical significance of the graded energy at the junction between semiconductors?
   (c) If two metals are placed in contact, would there be a similar step in the conduction-band energy edge? Why?
   (d) In Fig. 1-10b, which has a larger work function, the metal or the semiconductor?

7. Metals Al and Pb are candidate materials for thin-film electrical contacts to a high-temperature superconductor device. The latter is composed of a mixture of oxides, i.e., Y_2O_3, BaO, and Cu_2O, whose enthalpies of oxide formation are $\Delta H^0(Y_2O_3) = -455\,\text{kcal/mol}$, $\Delta H^0(\text{BaO}) = -132\,\text{kcal/mol}$, and $\Delta H^0(Cu_2O) = -40.8$ kcal/mol. Which metal would probably be a better contact material? Why?

8. A 80 at% Ga–20 at% As melt is cooled from 1200°C to 0°C in a crucible.
   (a) Perform a complete phase analysis of the crucible contents at 1200°C, 1000°C, 600°C, 200°C, 30°C, and 29°C. What phases are present? What are their chemical compositions, and what are the relative amounts of these phases? Assume equilibrium cooling.
   (b) A thermocouple immersed in the melt records the temperature as the crucible cools. Sketch the expected temperature–time cooling response.
   (c) Do a complete phase analysis for a 80 at% As–20 at% Ga melt at 1000°C, 800°C, and 600°C.

9. A solar cell is fabricated by diffusing phosphorus (*n* dopant) from a constant surface source of 10^{20} atoms/cm^3 into a *p*-type Si wafer containing 10^{16} boron atoms/cm^3. The diffusivity of phosphorus is 10^{-12} cm^2/s, and the diffusion time is 1 hour. How far from the surface is the junction depth, i.e., where $C_n = C_p$?

10. Indiffusion of dopant atoms from a continuous source (C_0) into a semiconductor wafer was carried out for time t_1 and diffusivity D_1. This was followed by a second drive-in diffusion for time t_2 with diffusivity

D_2. Show that the resulting dopant concentration profile is given by

$$C(x, t) = \frac{2C_0}{\pi}\sqrt{\frac{D_1 t_1}{D_2 t_2}} \exp - \frac{x^2}{4D_2 t_2}$$

if the drive-in diffusion is essentially that from an instantaneous surface source.

11. Measurements on the electrical resistivity of Au films reveal a three-order-of-magnitude reduction in the equilibrium vacancy concentration as the temperature drops from 600 to 300°C.
    (a) What is the vacancy formation energy?
    (b) What fraction of sites will be vacant at 1080°C?
12. The current density (j in A/m^2) through an ionic solid is given by the product of the carrier density (n in number/m^3), velocity (v in m/s), and charge (q in C). If the carriers are charged vacancies that diffusively hop among lattice sites, write an expression for j as a function of an applied electric field ($\mathscr{E}$ in V/m). Specifically, what is the temperature dependence of j?
13. During the formation of SiO_2 for optical fiber fabrication, soot particles 500 Å in size nucleate homogeneously in the vapor phase at 1200°C. If the surface energy of SiO_2 is 1 J/m^2, estimate the value of the supersaturation present.
14. Consider a bimetallic strip consisting of two thin clad sheets of metals A and B with respective thermal expansion coefficients α_A and α_B, where $\alpha_A > \alpha_B$. The strip is clamped at one end and free to move at the other, i.e., a cantilever.
    (a) If the strip is uniformly heated, show how it will deform.
    (b) Explain the geometry of the deformation assuming it is elastic and realizing that mechanical equilibrium must always prevail in the strip.
    (c) Under what conditions will the deformation be elastic? Plastic?
15. A Vickers hardness test on a coating layer yielded square indentations measuring 1 μm $\times$ 1 μm.
    (a) How deeply did the indenter penetrate the coating?
    (b) What load was applied if the Vickers hardness was 2500 kg/mm^2?
16. A thin substrate (s) of thickness d_s was coated on both front and back surfaces with a film (f) of thickness d_f. Young's moduli for the substrate and film are Y_s and Y_f, respectively. A rectangular section of the film/substrate/film composite was tested in tension with load applied parallel to the film plane. Derive an expression for the fraction of the applied load that is supported by the substrate.

REFERENCES

A. General Overview

1. M. Ohring, *Engineering Materials Science*. Academic Press, San Diego, 1995.
2. M. F. Ashby and D. R. H. Jones, *Engineering Materials*, Vols. 1 and 2. Pergamon Press, New York, 1980, 1986.
3. C. Newey and G. Weaver, *Materials in Action Series: Materials Principles and Practice*, 4 volumes. Butterworths, London, 1990.

B. Structure

1. B. D. Cullity, *Elements of X-ray Diffraction*. Addison-Wesley, Reading, MA, 1978.
2. D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*. Plenum, New York, 1996.
3. J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*. Plenum, New York, 1981.

C. Defects

1. D. Hull, *Introduction to Dislocations*. Pergamon Press, New York, 1965.
2. A. H. Cottrell, *Mechanical Properties of Matter*. John Wiley & Sons, New York, 1964.

D. Classes of Solids

a. Metals

1. A. H. Cottrell, *An Introduction to Metallurgy*. St. Martin's Press, New York, 1967.

b. Ceramics

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*. John Wiley & Sons, New York, 1976.

c. Glass

1. R. H. Doremus, *Glass Science*. John Wiley & Sons, New York, 1973.

d. Semiconductors

1. S. M. Sze, *Semiconductor Devices: Physics and Technology*. John Wiley & Sons, New York, 1985.

2. J. M. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*. Macmillan, New York, 1990.

E. Thermodynamics of Materials

1. C. H. Lupis, *Chemical Thermodynamics of Materials*. North Holland, New York, 1983.
2. M. W. Chase, C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald, and A. N. Syverud, *JANAF Thermodynamic Tables*, 3rd ed., American Institute of Physics, New York, 1986.
3. The computerized data base HSC Chemistry (Outokumpu Research Oy, PO Box 60, Fin-28101 Pori, Finland) is recommended.

F. Diffusion, Nucleation, Phase Transformations

1. R. J. Borg and G. J. Dienes, *An Introduction to Solid State Diffusion*. Academic Press, San Diego, 1988.
2. D. A. Porter and K. E. Easterling, *Phase Transformations in Metals and Alloys*, 2nd ed. Van Nostrand Reinhold, London, 1992.
3. H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*. Oxford University Press, London, 1959.
4. J. Crank, *The Mathematics of Diffusion*, Oxford University Press, London, 1964.

G. Mechanical Properties

1. T. H. Courtney, *Mechanical Behavior of Materials*. McGraw Hill, New York, 1990.
2. N. E. Dowling, *Mechanical Behavior of Materials*. Prentice Hall, Upper Saddle River, NJ, 1999.

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 2

Vacuum Science and Technology

2.1 INTRODUCTION

Virtually every thin-film deposition and processing method or technique employed to characterize and measure the properties of films requires either a vacuum or some sort of reduced-pressure ambient. For example, there are plasma discharges, sustained at reduced gas pressures, in which many important thin-film deposition and etching processes occur. Evacuated spaces are usually populated by uncharged gas atoms and molecules, but in addition to these, electrons and ionized gas species are present in the more complex plasmas. To better understand these reduced pressure environments a brief introduction to aspects of vacuum science and technology is the starting point of this book. It is also appropriate in a broader sense because this subject matter is among the most undeservedly neglected in the training of scientists and engineers. This is surprising in view of its broad interdisciplinary implications and the ubiquitous use of vacuum in all areas of scientific research and technological endeavor.

The topics treated in this chapter will, therefore, begin with a discussion of the behavior of gases in a closed system. In such systems gas atoms or molecules undergo an animated motion but their concentration and the pressure they exert is uniform throughout; there is no *net* flow of gas past any arbitrary plane in the system. If the system is opened by attaching pumps to it, net gas transport does occur even though individual gas molecules continue to locally execute the same random motion. With time, however, pressure gradients develop in the system as working chambers are evacuated to low pressures. Limits encountered in achieving high vacuum

levels in such systems will be discussed. Some of the hardware that makes vacuum technology possible, namely the pumps and associated gauges required, will close our treatment of this subject.

For further information on the subject the recent handbook by Hoffman *et al.* (Ref. 1) and text by O'Hanlon (Ref. 2) are recommended references.

2.2 KINETIC THEORY OF GASES

2.2.1 MOLECULAR VELOCITIES

The well-known kinetic theory of gases provides us with an atomistic picture of the state of affairs in a confined rarefied gas (Refs. 3, 4). A fundamental assumption is that the large number of atoms or molecules making up the gas are in a continuous state of random motion that is intimately dependent on their temperature. As they move, the gas particles collide with each other as well as with the walls of the confining vessel. Just how many molecule–molecule or molecule–wall impacts occur depends on the concentration or pressure of the gas. In the perfect or ideal gas approximation there are no attractive or repulsive forces between molecules. Rather, they may be considered to behave like independent elastic spheres separated from each other by distances that are large compared to their size. The net result of the continual elastic collisions and exchange of kinetic energy is that a steady-state distribution of molecular velocities emerges given by the celebrated Maxwell–Boltzmann formula

$$f(v) = \frac{1}{n}\frac{dn}{dv} = \frac{4}{\sqrt{\pi}}\left(\frac{M}{2RT}\right)^{3/2} v^2 \exp - \frac{Mv^2}{2RT}. \tag{2-1}$$

This centerpiece of the kinetic theory of gases states that the fractional number of molecules $f(v)$, where n is the number per unit volume in the velocity range v to $v + dv$, is related to their molecular weight (M) and absolute temperature (T). In this formula the units of the gas constant R are on a per-mole basis. Among the important implications of Eq. 2-1, which is shown plotted in Fig. 2-1, is that molecules can have neither zero nor infinite velocity. Rather, the most probable molecular velocity of the distribution is realized at the maximum value of $f(v)$ and can be calculated from the condition that $df(v)/dv = 0$. Since the net velocity is always the resultant of three rectilinear components v_x, v_y, and v_z, one (or even two) but, of course, not all three of these may be zero simultaneously. Therefore, a similar distribution function of molecular velocities in each of the component

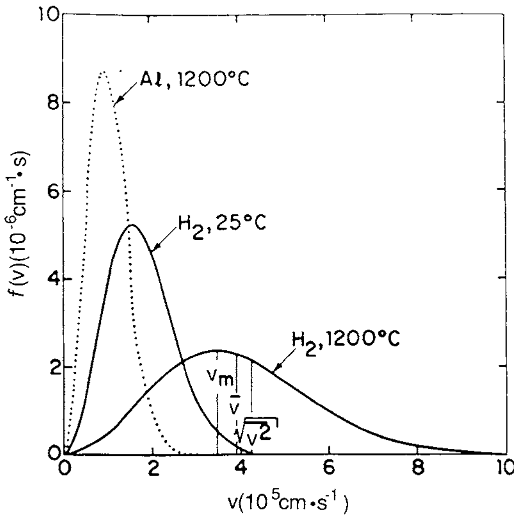


Figure 2-1 Velocity distributions for Al vapor and H_2 gas. (Reprinted with permission from Ref. 3.)

directions can be defined; i.e.,

$$f(v_x) = \frac{1}{n}\frac{dn_x}{dv_x} = \left(\frac{M}{2\pi RT}\right)^{1/2} \exp -\frac{Mv_x^2}{2RT} \tag{2-2}$$

and similarly for the y and z components.

A number of important results emerge as a consequences of the foregoing equations. For example, the most probable (v_m), average ($\bar{v}$) and mean square ($\bar{v}^2$) velocities are given, respectively, by

$$v_m = \sqrt{\frac{2RT}{M}} \tag{2-3a}$$

$$\bar{v} = \int_0^\infty vf(v)dv \Big/ \int_0^\infty f(v)dv = \sqrt{\frac{8RT}{\pi M}} \tag{2-3b}$$

$$\bar{v}^2 = \int_0^\infty v^2 f(v)dv \Big/ \int_0^\infty f(v)dv = 3RT/M; \qquad (\bar{v}^2)^{1/2} = \sqrt{\frac{3RT}{M}} \tag{2-3c}$$

These velocities, which are noted in Fig. 2-1, simply depend on the molecular weight of the gas and the temperature. In air at 300 K, for example, the average molecular velocity is 4.6×10^4 cm/s, which is almost

1030 miles per hour. However, the kinetic energy of any collection of gas molecules is solely dependent on temperature. For a mole it is given by $\frac{1}{2}Mv^2 = \frac{3}{2}RT$ with $\frac{1}{2}RT$ partitioned in each of the three coordinate directions.

2.2.2 PRESSURE

Momentum transfer from the gas molecules to the container walls gives rise to the forces that sustain the pressure in the system. Kinetic theory shows that the gas pressure, P, is related to the mean-square velocity of the molecules and thus, alternately to their kinetic energy or temperature. Thus,

$$P = nM\bar{v}^2/3N_A = nRT/N_A \quad (2\text{-}4)$$

where N_A is Avogadro's number. From the definition of n, n/N_A is the number of moles per unit volume and therefore, Eq. 2-4 is an expression of the perfect gas law.

Pressure is the most widely quoted system variable in vacuum technology and this fact has generated a large number of units that have been used to define it under various circumstances. Basically, two broad types of pressure units have arisen in practice. In what we shall call the scientific system or *coherent* unit system (Ref. 4), pressure is defined as the rate of change of the normal component of momentum of impinging molecules per unit area of surface. Thus, the pressure is defined as a force per unit area, and examples of these units are dynes/cm^2 (CGS) or newtons/meter^2 (N/m^2) (MKS). Vacuum levels are now commonly reported in SI units or pascals; 1 pascal (Pa) = 1 N/m^2. Historically, however, pressure was, and still is, measured by the height of a column of liquid, e.g., Hg or H_2O. This has led to a set of what we shall call practical or *noncoherent* units, such as millimeters and microns of Hg, torr, and atmospheres, which are still widely employed by practitioners as well as by equipment manufacturers. Definitions of some units together with important conversions include

$$1 \text{ atm} = 1.013 \times 10^6 \text{ dynes/cm}^2 = 1.013 \times 10^5 \text{ N/m}^2 = 1.013 \times 10^5 \text{ Pa}$$

$$1 \text{ torr} = 1 \text{ mm Hg} = 1.333 \times 10^3 \text{ dynes/cm}^2 = 133.3 \text{ N/m}^2 = 133.3 \text{ Pa}$$

$$1 \text{ bar} = 0.987 \text{ atm} = 750 \text{ torr.}$$

The mean distance traveled by molecules between successive collisions, called the mean-free path, λ_{mfp}, is an important property of the gas that is dependent on the pressure. To calculate λ_{mfp} we note that each molecule presents a target area πd_c^2 to others where d_c is its collision diameter. A binary collision occurs each time the center of one molecule approaches

within a distance d_c of the other. If we imagine the diameter of one molecule increased to $2d_c$ while the other molecules are reduced to points, then in traveling a distance λ_{mfp} the former sweeps out a cylindrical volume $\pi d_c^2 \lambda_{mfp}$. One collision will occur under the condition that $\pi d_c^2 \lambda_{mfp} n = 1$. For air at room temperature and atmospheric pressure, $\lambda_{mfp} \sim 500$ Å, assuming $d_c \simeq 5$ Å.

A molecule collides in a time given by λ_{mfp}/v, and under the above conditions, air molecules make about 10^{10} collisions per second. This is why gases mix together rather slowly even though the individual molecules are moving at great speeds. The gas particles do not travel in uninterrupted linear trajectories. As a result of collisions, they are continually knocked to and fro, executing a zigzag motion and accomplishing little net movement. Since n is directly proportional to P, a simple relation for ambient air is

$$\lambda_{mfp} = 5 \times 10^{-3}/P, \tag{2-5}$$

with λ_{mfp} given in centimeters and P in torr. At pressures below 10^{-3} torr, λ_{mfp} is so large that molecules effectively collide only with the walls of the vacuum chamber.

2.2.3 Gas Impingement on Surfaces

A most important quantity that plays a role in both vacuum science and vapor deposition is the gas impingement flux Φ. It is a measure of the frequency with which molecules impinge on or collide with a surface, and should be distinguished from the previously discussed molecular collisions in the gas phase. The number of molecules that strike an element of surface, perpendicular to a coordinate direction, per unit time and area is given by

$$\Phi = \int_0^\infty v_x \, dn_x. \tag{2-6}$$

Upon substitution of Eq. 2-2 we get

$$\Phi = n\sqrt{\frac{M}{2\pi RT}} \int_0^\infty v_x \exp - \frac{Mv_x^2}{2RT} dv_x = n\sqrt{\frac{RT}{2\pi M}}. \tag{2-7}$$

The use of Eq. 2-3b yields $\Phi = \frac{1}{4} n\bar{v}$, while substitution of the perfect gas law (Eq. 2-4) converts Φ into the more recognizable form

$$\frac{\Phi}{N_A} = \frac{P}{(2\pi MRT)^{1/2}} \qquad \text{moles/cm}^2\text{-s}. \tag{2-8}$$

A useful variant of this formula is

$$\Phi = 3.513 \times 10^{22} \frac{P}{(MT)^{1/2}} \qquad \text{molecules/cm}^2\text{-s} \tag{2-9}$$

when P is expressed in torr.

As an application of the preceding development, consider the problem of gas escaping the vessel through a hole of area A into a region where the gas concentration is zero. The rate at which molecules leave is given by ΦA and this corresponds to a volume flow per second ($\dot{V}$) given by $\Phi A/n$ cm^3/s. Upon substitution of Eqs. 2-4 and 2-9, we have

$$\dot{V} = 3.64 \times 10^3 (T/M)^{1/2} A \text{ cm}^3\text{/s}. \tag{2-10}$$

For air at 298 K this corresponds to $11.7A$ liters/s, where A has units of cm^2. In essence we have just calculated what is known as the conductance of a circular aperture, a quantity that will be utilized later because of its significance in pumping gases.

As a second application, consider the question of how long it takes for a surface to be coated by a monolayer of gas molecules. This is an issue of great importance when attempting to deposit or grow films under extremely clean conditions. The same concern arises during surface analysis of films, which is performed at very low pressures in order to minimize surface contamination arising from the vacuum chamber environment. Here one must make certain that the analysis time is shorter than that required for impurities to accumulate. This characteristic contamination time, t_c, is essentially the inverse of the impingement flux. Thus, for complete monolayer coverage of a surface containing some 10^{15} atoms/cm^2, the use of Eq. 2-9 yields

$$t_c = \frac{10^{15}}{3.513 \times 10^{22}} \frac{(MT)^{1/2}}{P} = \frac{2.85 \times 10^{-8}}{P} (MT)^{1/2} \text{ s}, \tag{2-11}$$

with P measured in torr. In air at atmospheric pressure and ambient temperature a surface will acquire a monolayer of gas in 3.49×10^{-9} s assuming all impinging atoms stick. On the other hand, at 10^{-10} torr a surface will stay clean for about 7.3 h. Surface exposure to gases is measured in Langmuir (L) units of pressure–time, e.g., torr-s. The conversion is 1 L = 10^{-6} torr-s, which means gas exposure could occur at 10^{-6} torr for 1 s, at 10^{-7} torr for 10 s, etc. Therefore, since a monolayer typically forms after 7.3 h (26,280 s) at 10^{-10} torr (or 2.63 L), 1 Langmuir corresponds to about 0.38 monolayer.

A condensed summary of the way system pressure affects the gas density, mean free path, incidence rates, and monolayer formation times is conveniently displayed in Fig. 2-2. The pressure range spanned in all thin-film

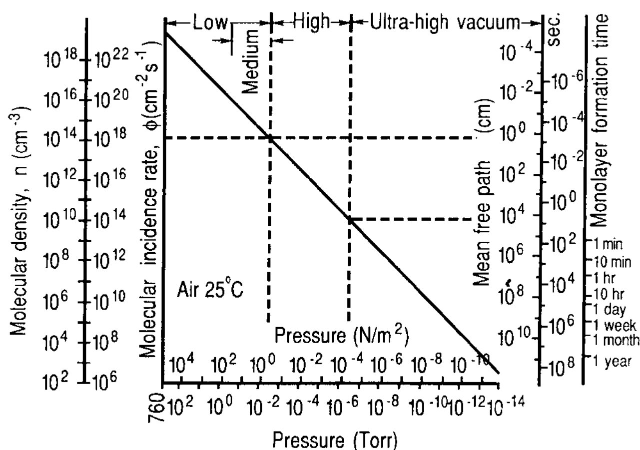


Figure 2-2 Molecular density, incidence rate, mean free path, and monolayer formation time as a function of pressure. (Reprinted with permission from Ref. 4.)

research, development, and technological activities discussed in this book is over 13 orders of magnitude. Since all of the quantities depicted vary directly (on the left-hand axis) or inversely (on the right-hand axis) with pressure, a log–log plot connecting the variables is linear with a slope of 1. The pressure scale is arbitrarily subdivided into corresponding low, medium, high, and ultrahigh vacuum domains, each characterized by different requirements with respect to vacuum hardware (e.g., pumps, gauges, valves, gaskets, feedthroughs). Of the film deposition processes, evaporation requires a vacuum between the high and ultrahigh regimes, whereas sputtering and low-pressure chemical vapor deposition are accomplished at the border between the medium and high vacuum ranges. Of the analytical instruments, electron microscopes operate in high vacuum while surface analytical equipment has the most stringent cleanliness requirements, necessitating ultrahigh vacuum operating conditions.

2.3 GAS TRANSPORT AND PUMPING

2.3.1 Gas Flow Regimes

In order to better design, modify, or appreciate reduced pressure systems, it is essential to understand concepts of gas flow (Refs. 4–6). An incomplete understanding of gas-flow limitations frequently results in less efficient

system performance as well as increased expense. For example, a cheap piece of tubing having the same length and diameter dimensions as the diffusion pump to which it is attached will cut the pumping speed of the latter to approximately one-half of its rated value. In addition to the effectively higher pump cost, a continuing legacy of such a combination will be the longer pumping times required to reach a given level of vacuum each time the system is operated.

Whenever there is a net directed movement of gas in a system under the influence of attached pumps, gas flow is said to occur. Under such conditions the gas experiences a pressure drop. The previous discussion on kinetic theory of gases essentially assumed an isolated sealed system. Although gas molecules certainly move, and with high velocity at that, there is no *net* gas flow and no pressure gradients in such a system. When gas does flow, however, it is appropriate to distinguish between different regimes of flow. These are dependent on the geometry of the system involved as well as the pressure, temperature, and type of gas under consideration. At one extreme we have free *molecular flow*, which occurs at low gas densities. The chambers of high-vacuum evaporators and analytical equipment, such as Auger electron spectrometers and electron microscopes, operate within the molecular-flow regime. Here the mean distance between molecular collisions is large compared to the dimensions of the system. Kinetic theory provides an accurate picture of molecular motion under such conditions. At higher pressures, however, the mean distance between successive molecular collisions is reduced, and they predominate relative to molecule–chamber wall collisions. In this case the so-called *viscous-flow* regime is operative. An important example of such flow occurs in atmospheric chemical vapor-deposition reactors. Compared to molecular flow, viscous flow is quite complex. At low gas velocities the flow is laminar where layered, parallel gas flow lines may be imagined. Under these conditions the laminar flow velocity is zero at the walls of a tube, but it increases to a maximum at the tube axis. For higher flow velocities the gas layers are no longer parallel but swirl and are influenced by any obstacles in the way. In this turbulent flow range, cavities of lower pressure develop between layers. More will be said about viscous flow in Section 6.4.2.

Criteria for distinguishing between the flow regimes are based on the magnitude of the Knudsen number, Kn, which is defined by the ratio of the gas mean free path to a characteristic dimension of the system (e.g., chamber or pipe diameter, D_p), i.e., $\mathrm{Kn} = \lambda_{mfp}/D_p$. Thus, for

Molecular flow	$\mathrm{Kn} > 1$	(2-12a)
Intermediate flow	$1 > \mathrm{Kn} > 0.01$	(2-12b)
Viscous flow	$\mathrm{Kn} < 0.01.$	(2-12c)

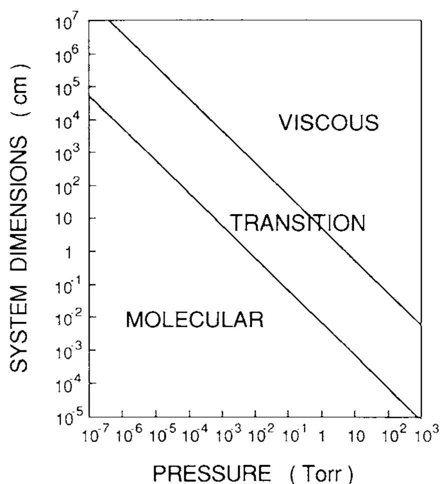


Figure 2-3 Dominant gas flow regimes as a function of system dimensions and pressure.

Through the use of Eq. 2-5, these limits in air can be alternately expressed by $D_pP < 5 \times 10^{-3}$ cm-torr for molecular flow, and $D_pP > 5 \times 10^{-1}$ cm-torr in the case of viscous flow. Figure 2-3 serves to map the dominant flow regimes on this basis. The reader should be aware that flow mechanisms may differ in various parts of the same system. Thus, whereas molecular flow will occur in the high vacuum chamber, the gas may flow viscously in the piping near the exhaust pumps.

2.3.2 CONDUCTANCE

Let us reconsider the molecular flow of gas through an orifice of area A that now separates two large chambers maintained at low pressures, P_1 and P_2. From a phenomenological standpoint, a flow driven by the pressure difference is expected, i.e.,

$$Q = C(P_1 - P_2). \tag{2-13}$$

Here Q is defined as the gas throughput with units of pressure $\times$ volume/s (e.g., torr-liters/s). The constant of proportionality C is known as the conductance and has units of liters/s. Alternately, viewing flow through the orifice in terms of kinetic theory we note that the molecular impingements in each of the two opposing directions do not interfere with each other. Therefore, the net gas flow at the orifice plane is given by the difference or $(\Phi_1 - \Phi_2)A$.

Through the use of Eq. 2-10 it is easily shown that the conductance of the orifice is

$$C = 3.64(T/M)^{1/2}A \quad \text{or} \quad 11.7A \text{ liters/s}$$

for air at 298 K. The reader will undoubtedly note in the choice of terms the analogy to electrical circuits. If $P_1 - P_2$ is associated with electrical pressure or voltage difference, Q may be viewed as a current. Conductances of other components where the gas flow is in the molecular regime can be similarly calculated or measured. Results for a number of important geometric shapes are given in Fig. 2-4 (Ref. 7). It should be noted that conductance is simply a function of the geometry for a specific gas at a given temperature. This is not true of viscous flow where conductance also depends on pressure. As an example, consider air flow at 298 K through a pipe of diameter D and length L connecting the vacuum chamber to the discharge pump. The molecular flow conductance is $12.2D^3/L$. However, for viscous flow in the same pipe the conductance is equal to $184(P_2 + P_1)D^4/2L$ with P in torr. Let us assume $D = 2.5$ cm, $L = 100$ cm, and $(P_2 + P_1)/2 = 380$ torr, which is the average between atmospheric and high vacuum pressure. Substitution and evaluation gives 1.9 liters/s and 27,300 liters/s for the molecular and viscous flow conductances, respectively. This great disparity in conductance means that the geometry of vacuum components becomes increasingly important as the system pressure decreases.

When conductances are joined in series, the system conductance C_{sys} is given by

$$\frac{1}{C_{\text{sys}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots. \tag{2-14}$$

Clearly C_{sys} is lower than that of any individual conductance. When conductances are connected in parallel,

$$C_{\text{sys}} = C_1 + C_2 + C_3 + \cdots. \tag{2-15}$$

As an example (Ref. 8), consider the conductance of the cold trap assembly of Fig. 2-5 which isolates a vacuum system above from the pump below. Contributions to the total conductance come from:

C_1 = conductance of aperture having 10 cm diameter

$= 11.7A = 11.7\pi(5)^2 = 919$ liters/s

C_2 = conductance of pipe 3 cm long

$= 12.2D^3/L = 12.2(10)^3/3 = 4065$ liters/s

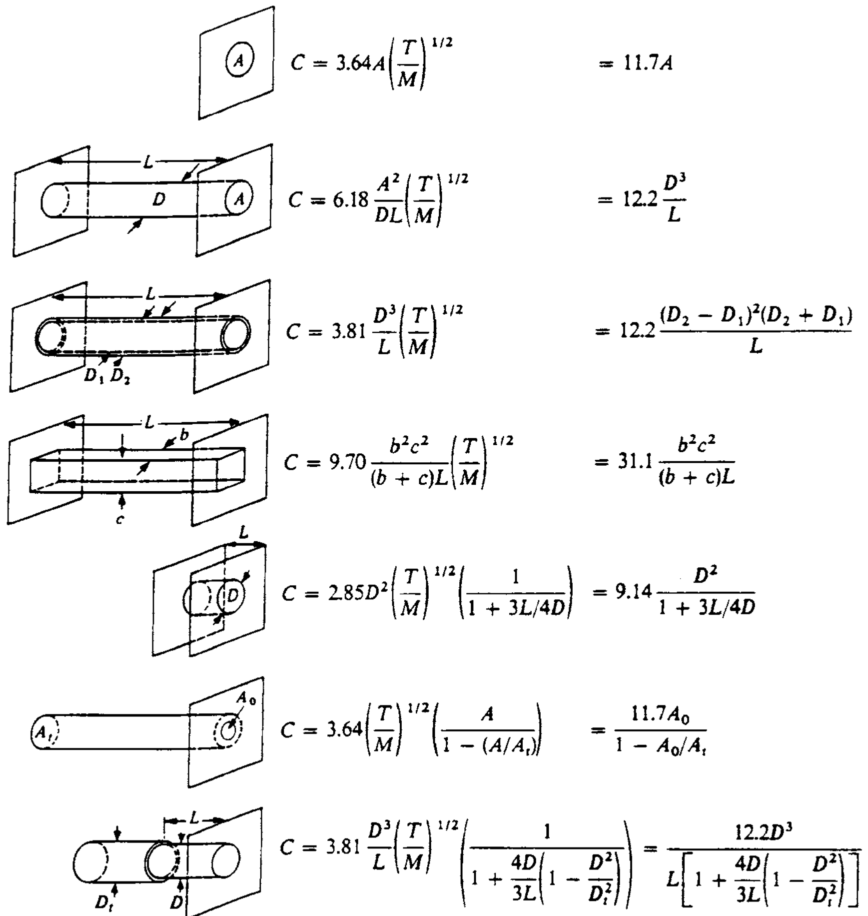


Figure 2-4 Conductances of various geometric shapes for molecular flow of air at 25°C. Units of C are liters/s. (Reprinted with permission from Ref. 7.)

C_3 = conductance of annular aperture

$$= 11.7A_{\text{ann}} = 11.7(0.25)\pi(10^2 - 8^2) = 331 \text{ liters/s}$$

C_4 = conductance of annular pipe

$$= \frac{12.2(D_2 - D_1)^2(D_1 + D_2)}{L} = \frac{12.2(10 - 8)^2(10 + 8)}{15} = 58.6 \text{ liters/s}$$

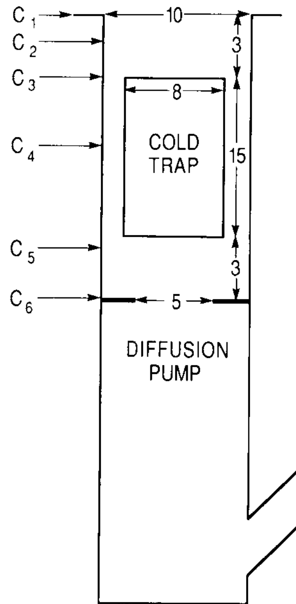


Figure 2-5 Cold trap assembly. (Adapted from R. W. Roberts, *An Outline of Vacuum Technology*, G. E. Report No. 64-RL 3394C, 1964, with permission from General Electric Company.)

$C_5 = C_2 = 4065$ liters/s

C_6 = conductance of aperture in end of pipe/diffusion pump

$$= \frac{11.7AA_0}{A - A_0} = \frac{11.7\pi(2.5)\pi 5^2}{\pi(5^2 - (2.5)^2)} = 303 \text{ liters/s.}$$

Therefore,

$$\frac{1}{C_{\text{Total}}} = \sum_{i=1}^{6} \frac{1}{C_i}, \qquad \text{and} \qquad C_{\text{Total}} = 40 \text{ liters/s}$$

upon evaluation. Strictly speaking, C_3 and C_4 should be multiplied by a correction factor of 1.27, which would have the effect of increasing C_{Total} to 51.1 liters/s. As we shall soon see, it is always desirable to have as large a conductance as possible. Clearly, the overall conductance is severely limited in this case by that of the annular region between the concentric pipes.

2.3.3 Pumping Speed

Pumping is the process of removing gas molecules from the system through the action of pumps. The pumping speed, S, is defined as the volume of gas passing the plane of the inlet port per unit time when the pressure at the pump inlet is P. Thus,

$$S = Q/P, \tag{2-16}$$

and while the throughput Q can be measured at any plane in the system, P and S refer to quantities measured at the pump inlet.

Although conductance and pumping speed have the same units and may even be equivalent numerically, they have different physical meanings. Conductance implies a component of a given geometry across which a pressure differential exists. Pumping speed refers to a given plane that may be considered to be a pump for preceding portions of the system. To apply these ideas consider a pipe of conductance C connecting a chamber at pressure P to a pump at pressure P_p as shown in Fig. 2-6a. Therefore, $Q = C(P - P_p)$. Elimination of Q through the use of Eq. 2-16 yields

$$S = \frac{S_p}{1 + S_p/C} \tag{2-17}$$

where S_p is the intrinsic speed at the pump inlet ($S_p = Q/P_p$) and S is the effective pumping speed at the base of the chamber. The latter never exceeds S_p or C and is, in fact, limited by the smaller of these quantities. If, for example, $C = S_p$ in magnitude, then $S = S_p/2$ and the effective pumping speed is half the rated value for the pump. The lesson, therefore, is to keep

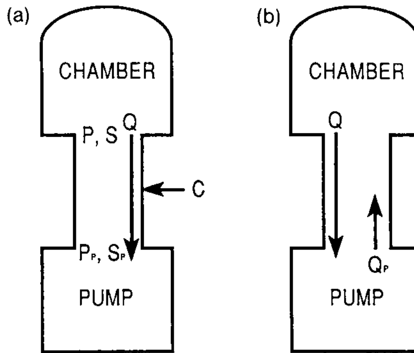


Figure 2-6 Chamber-pipe-pump assembly: (a) no outgassing, (b) with outgassing.

conductances large by making ducts between the pump and chamber as short and wide as possible.

Real pumps outgas or release gas into the system as shown in Fig. 2-6b. Account may be taken of this by including an oppositely directed extra throughput term Q_p such that

$$Q = S_pP - Q_p = S_pP[1 - (Q_p/S_pP)]. \tag{2-18}$$

When $Q = 0$ the ultimate pressure of the pump, P_0, is reached and $Q_p = S_pP_0$. The effective pumping speed is then

$$S = Q/P = S_p(1 - P_0/P) \tag{2-19}$$

and falls to zero as the system pressure reaches the ultimate pump pressure.

An issue of importance in vacuum systems is the time required to achieve a given pressure. The pump-down time may be calculated by noting that the throughput may be defined as the time (t) derivative of the product of volume and pressure, i.e., $Q = -d(VP)/dt = -V\,dP/dt$. Employing Eq. 2-18, we write

$$-V\frac{dP(t)}{dt} = S_pP - Q_p$$

where Q_p includes pump as well as chamber outgassing. Upon integration

$$\frac{P(t) - Q_p/S_p}{P_i - Q_p/S_p} = \frac{P(t) - P_0}{P_i - P_0} = \exp - \frac{S_pt}{V} \tag{2-20}$$

where it is assumed that initially $P = P_i$. During pump-down the pressure thus exponentially decays to P_0 with a time constant given by V/S_p. At high pressures where viscous flow is involved, S_p is a function of P and therefore Eq. 2-20 is not strictly applicable in such cases.

2.4 VACUUM PUMPS

2.4.1 GENERAL CONSIDERATIONS

The vacuum systems employed to deposit and characterize thin films contain an assortment of pumps, tubing, valves and gauges to establish and measure the reduced pressures (Ref. 9). Of these components pumps and gauges are generally the most important and only they will be discussed at any length. Vacuum pumps may be divided into two broad categories: gas-transfer pumps and entrapment pumps. The gas transfer pumps remove gas molecules from the pumped volume and convey them to the ambient in

one or more stages of compression. Entrapment pumps condense or chemically bind molecules to surfaces situated within the chamber being pumped. In contrast to the gas transfer pumps, which remove gas permanently, some of the entrapment pumps are reversible and release trapped or condensed gas back into the system upon warmup.

Gas transfer pumps may be further subdivided into positive displacement and kinetic vacuum pumps. Rotary mechanical and Roots pumps are important examples of the positive displacement variety. Diffusion and turbomolecular pumps are the outstanding examples of the so-called kinetic vacuum pumps. Among the entrapment pumps commonly employed are the adsorption, sputter-ion, and cryogenic pumps. Each pump is used singly or in combination in a variety of pumping-system configurations. It should be noted that pumps do not remove gas molecules by exerting an attractive pull on them. Gas molecules are unaware that pumps exist. Rather, pump action limits, interferes with, and alters natural molecular motion in such a way that gas is transferred from low to high pressure ambients. We start this brief survey of some of the more important pumps with the positive displacement types.

2.4.2 Rotary Mechanical Pump

The rotary-piston and rotary-vane pumps are perhaps the two most common devices used to attain reduced pressure. In the rotary piston pump shown in Fig. 2-7a, gas is drawn into space A as the keyed shaft rotates the eccentric and piston. There the gas is isolated from the inlet after one revolution, then compressed and exhausted during the next cycle. Piston pumps are often employed to evacuate large systems and to back Roots-blower pumps.

The rotary-vane pump contains an eccentrically mounted rotor with spring-loaded vanes. During rotation the vanes slide in and out within the cylindrical interior of the pump, enabling a quantity of gas to be confined, compressed, and discharged through an exhaust valve to the atmosphere. Compression ratios of up to 10^6 can be achieved this way. Oil is employed as a sealant as well as lubricant between components moving within tight clearances of both types of rotary pumps. To calculate the pumping speed let us assume that a volume of gas, V_0 (liters), is enclosed between the rotor and pump stator housing, and swept into the atmosphere for each revolution of the rotor. The intrinsic speed of the pump will then be $S_p = V_0 f_s$ where f_s is the rotor speed in revolutions per second. Typical values for S_p for vane pumps range from 10 to 200 m^3/h (2.8 to 55 liters/s) and from 30 to 1500 m^3/h (8.3 to 416 liters/s) for piston pumps. At elevated pressures the

(a)

Figure 2-7 (a) Schematic of a rotary piston pump: 1, eccentric; 2, piston; 3, shaft; 4, gas ballast; 5, cooling water inlet; 6, optional exhaust; 7, motor; 8, exhaust; 9, oil mist separator; 10, poppet valve; 11, inlet; 12, hinge bar; 13, casing; 14, cooling water outlet. (Courtesy of Stokes Vacuum Inc.)

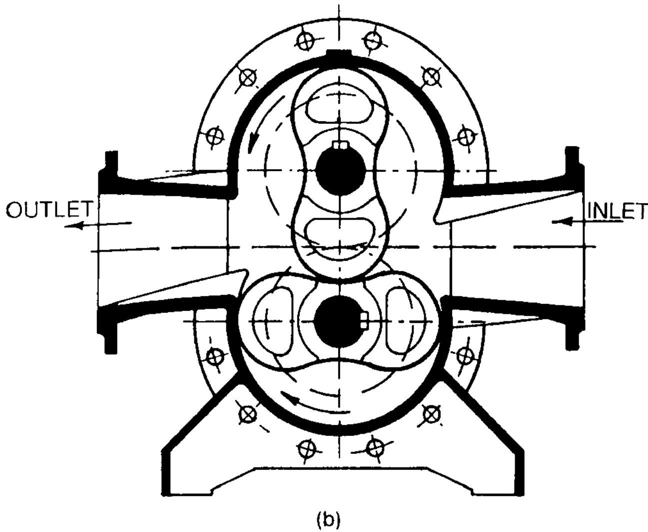


Figure 2-7 (b) Schematic of a Roots pump.

actual pumping speed S is constant but eventually becomes zero at the ultimate pump pressure in a manner suggested by Eq. 2-19. Single-stage vane pumps have an ultimate pressure of 10^{-2} torr, while two-stage pumps can reach 10^{-4} torr. Rotary pumps are frequently used to produce the minimum vacuum level required to operate both oil diffusion and turbomolecular pumps, which can then attain far lower pressures.

2.4.3 Roots Pump

An important variant of the positive displacement pump is the Roots pump, where, as shown in Fig. 2-7b, two figure-eight-shaped lobes rotate in opposite directions relative to each other. The extremely close tolerances eliminate the need to seal with oil. These pumps have very high pumping speeds and although they can attain ultimate pressures below 10^{-5} torr, a forepump, e.g., rotary mechanical, is required. Maximum pumping is achieved in the pressure range of 10^{-3} to 20 torr where speeds of up to several thousand liters/s can be attained. This combination of characteristics has made Roots pumps popular in sputtering as well as low-pressure chemical vapor deposition (LPCVD) systems where large gas volumes continuously pass through reactors maintained at ~ 1 torr.

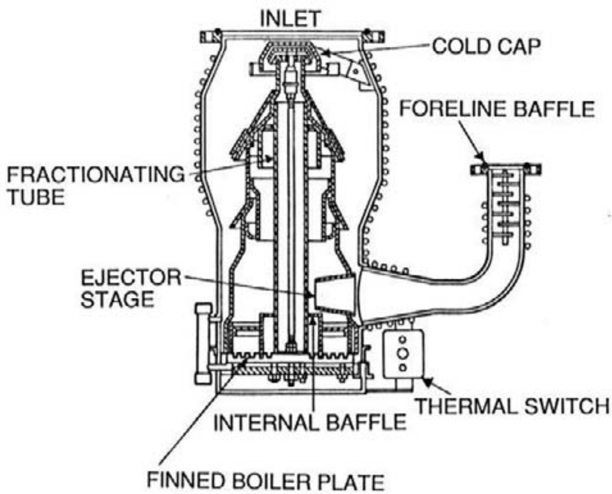
2.4.4 Diffusion Pump

In contrast to mechanical pumps described above, the diffusion pump shown in Fig. 2-8 has no moving parts. Diffusion pumps are designed to operate in the molecular flow regime and can function over pressures ranging from well below 10^{-10} torr to about 5×10^{-2} torr. Because they cannot discharge directly into the atmosphere, a mechanical forepump is required to maintain the latter outlet pressure. Since the pump inlet is essentially like the orifice of Fig. 2-4, a pumping speed of $11.7A$ liter/s would be theoretically expected for air at room temperature. Actual pumping speeds are typically only 0.4 of this value so that a 15-cm diameter diffusion pump would have a pumping speed of $\sim 0.4(11.7)\pi(15)^2/4 = 827$ liters/s.

Diffusion pumps have been constructed with pumping speeds ranging from a few liters per second to over 20,000 liters/s. Pumping is achieved through the action of a fluid medium (typically silicone oil), which is boiled and vaporized in a multistage jet assembly. As the oil vapor stream emerges from the top nozzles, it collides with and imparts momentum to residual gas molecules that happen to bound into the pump throat. These molecules are thus driven toward the bottom of the pump and compressed to the exit side where they are exhausted. A region of reduced gas pressure in the vicinity of the jet is produced and more molecules from the high-vacuum side move into this zone, where the process is repeated. Several jets working in series serve to enhance the pumping action.

A serious problem associated with diffusion pumps is the backstreaming of oil into the chamber. Early during pumpdown there is little backstreaming because viscous flow conditions prevail and the repeated impact between air and oil vapor molecules prevents effective transfer of the latter. However, as the pressure drops into the molecular flow regime these impacts cease and backstreaming now occurs at a rate dependent on the vapor pressure of the oil. Such condensed oil effectively contaminates substrate surfaces and leads to poor adhesion and degraded properties of subsequently deposited films. That is why diffusion pumps have been largely phased out of use in the processing of semiconductor electronic and optoelectronic devices. In addition, oil vapor dissociated on contact with hot filaments or by electrical discharges leave carbonaceous or siliceous deposits which can cause electrical leakage or even high-voltage breakdown. For these reasons diffusion pumps are not used in surface analytical equipment such as Auger-electron and secondary-ion mass spectrometers, nor in ultrahigh-vacuum deposition systems. Nevertheless, diffusion pumped systems are widely used in nonelectronic (e.g., decorative, optical, tool) coating applications. To minimize backstreaming, attempts are made to condense the oil before it enters the high-vacuum chamber. A cold cap on top of the uppermost jet together with

(a)



(b)

Figure 2-8 Diffusion pump: (a) photograph; (b) schematic of pump interior. (Courtesy of Varian Associates, Vacuum Products Division.)

refrigerated traps and optically dense baffles are used for this purpose, but at the expense of somewhat reduced conductance and pumping speed.

2.4.5 TURBOMOLECULAR PUMP

The drive to achieve the benefits of oil-free pumping has spurred the development and use of turbomolecular pumps. Like the diffusion pump, the turbomolecular pump imparts a preferred direction to molecular motion, but in this case the impulse is caused by impact with a rapidly whirling turbine rotor spinning at rates of 20,000 to 30,000 revolutions per minute. The turbomolecular pump of Fig. 2-9 is a vertical axial-flow compressor consisting of many rotor/stator pairs or stages mounted in series. Gas captured by the upper stages is transferred to the lower stages where it is successively compressed to the level of the fore-vacuum pressure. The compression ratio varies exponentially with the product of the circumferential rotor speed and square root of the molecular weight of the gas. Typical compression ratios for hydrocarbons, N_2, and H_2 are 10^{10}, 10^9, and 10^3, respectively. Since the partial pressure of a given gas species on the low-pressure, i.e., vacuum chamber, side of the pump is equal to that on the high-pressure (exhaust) side divided by the compression ratio, only H_2 will fail to be pumped effectively. An important consequence of the very high compression is that oil backstreaming is basically reduced to negligible levels. In fact, no traps or baffles are required and turbomolecular pumps can be backed by rotary pumps and effectively achieve oilless pumping. Turbomolecular pumps are expensive, but are increasingly employed in all sorts of thin-film deposition and characterization equipment. Typical characteristics include pumping speeds of 10^3 liters/s and ultimate pressures below 10^{-10} torr.

2.4.6 CRYOPUMPS

Cryogenic pumps are capable of generating a very clean vacuum in the pressure range of 10^{-3} to 10^{-10} torr. These are gas-entrapment pumps that rely on the condensation of vapor molecules on surfaces cooled below 120 K. Temperature-dependent van der Waals forces are responsible for physically binding or sorbing gas molecules. Several kinds of surfaces are employed to condense gas. They include: (1) untreated bare metal surfaces, (2) a surface cooled to 20 K containing a layer of precondensed gas of higher boiling point, e.g., Ar or CO_2 for H_2 or He sorption, and (3) a microporous surface of very large area within molecular sieve materials such as activated charcoal or zeolite. The latter are the working media of the common sorption

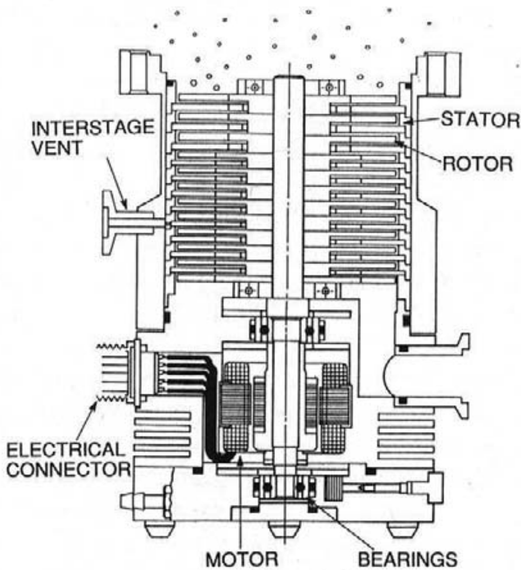


Figure 2-9 Turbomolecular pump: (a) photograph; (b) schematic of pump interior. (Courtesy of Varian Associates, Vacuum Products Division.)

pumps, which achieve forepressures of about 10^{-3} torr by surrounding the steel canister (pump) containing sorbent with a dewar of liquid nitrogen. Cryopumps designed to achieve ultrahigh vacua (Fig. 2-10) have panels that are cooled to 20 K by closed-cycle refrigerators. These cryosurfaces cannot be directly exposed to the room-temperature surfaces of the chamber because of the radiant heat load and, therefore, they are surrounded by liquid-nitrogen-cooled shrouds.

The starting or forepressure, ultimate pressure, and pumping speed of cryopumps are important characteristics. Cryopumps require an initial forepressure of about 10^{-3} torr in order to prevent a prohibitively large thermal load on the refrigerant and the accumulation of a thick ice condensate on the cryopanels. The ultimate pressure (P_{ult}) attained for a given gas is reached when the impingement rate on the cryosurface, maintained at temperature T, equals that on the vacuum chamber walls held at 300 K. Therefore, from Eq. 2-8

$$P_{\text{ult}} = P_s(T)\sqrt{300/\text{T}} \qquad (2\text{-}21)$$

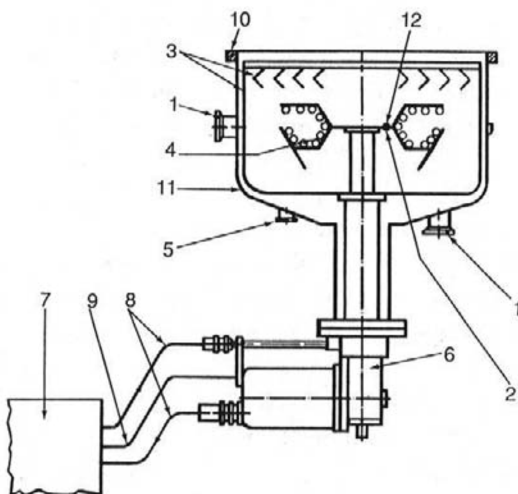
where $P_S(T)$ is the saturation pressure of the pumped gas. As an example, for N_2 at 20 K the P_s value is about 10^{-11} torr, so that $P_{\text{ult}}(N_2) = 3.9 \times 10^{-11}$ torr.

Because of high vapor pressures at 20 K, H_2, as well as He and Ne, cannot be effectively cryopumped. Of all high-vacuum pumps, cryopumps have the highest pumping speeds since they are limited only by the rate of gas impingement. Therefore, the pumping speed is given by Eq. 2-10, which for air at 20 K is equal to 3 liters/s for each square centimeter of cooled surface. Although they are expensive, cryopumps offer the versatility of serving as the main pump or, more frequently, acting in concert with other conventional pumps (e.g., turbopumps). They are, therefore, indispensable in thin-film research and processing equipment.

2.4.7 SPUTTER ION PUMPS

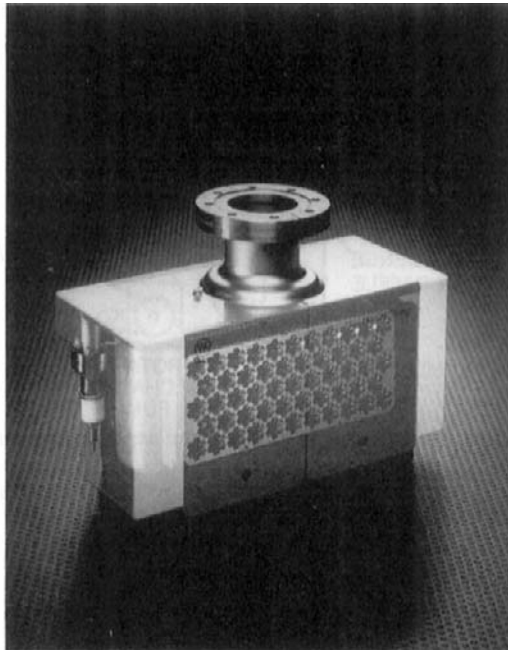
The last pump we consider is the sputter ion pump shown in Fig. 2-11, which relies on sorption processes initiated by ionized gas to achieve pumping. Gas ions are generated in a cold-cathode electrical discharge between stainless-steel anode and titanium cathode arrays maintained at a potential difference of a few kilovolts. Electrons emitted from the cathode are trapped in the applied transverse magnetic field of a few thousand gauss, resulting in a cloud of high electron density, e.g., $\sim 10^{10}$ electrons/cm^3. After impact ionization of residual gas molecules, the ions travel to the cathode and knock out or sputter atoms of Ti. The latter deposit elsewhere within

(a)

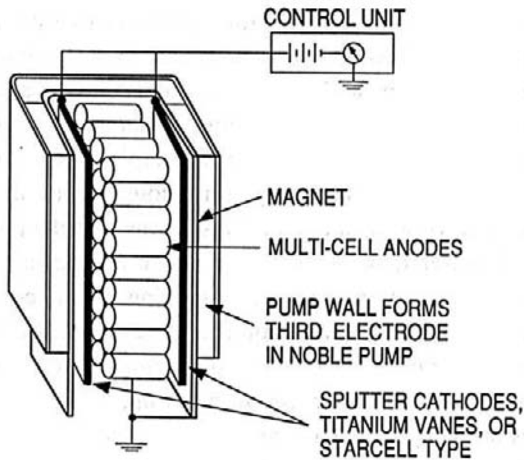


(b)

Figure 2-10 Cryopump: (a) photograph; (b) schematic of pump interior. 1, forevacuum port; 2, temperature sensor; 3, 77 K shield; 4, 20 K condenser with activated charcoal; 5, port for gauge head and pressure relief valve; 6, cold head; 7, compressor unit; 8, helium supply and return lines; 9, electrical supply cable; 10, high vacuum flange; 11, pump housing; 12, to temperature measuring instrument. (Courtesy of Balzers, High Vacuum Products.)



(a)



(b)

Figure 2-11 Sputter ion pump: (a) photograph; (b) schematic of pump interior. (Courtesy of Varian Associates, Vacuum Products Division.)

the pump where they form films that getter or combine with reactive gases such as nitrogen, oxygen, and hydrogen. These gases and corresponding Ti compounds are then buried by fresh layers of sputtered metal.

Similar pumping action occurs in the Ti sublimation pump where Ti metal is thermally evaporated (sublimed) onto cryogenically cooled surfaces. A combination of physical cryopumping and chemical sorption processes then ensues. Sputter ion pumps display a wide variation in pumping speeds for different gases. For example, hydrogen is pumped several times more effectively than oxygen, water, or nitrogen and several hundred times faster than argon. Unlike cryopumping action, the gases are permanently removed. These pumps are quite expensive and have a finite lifetime that varies inversely with the operating pressure. They have been commonly employed in oilless ultrahigh (10^{-10} torr) vacuum deposition and surface analytical equipment, but are being supplanted by turbomolecular and cryopumps.

2.5 VACUUM SYSTEMS

2.5.1 COMPONENTS AND OPERATION

The broad variety of applications requiring a low-pressure environment is reflected in a corresponding diversity of vacuum-system designs. One typical system shown schematically in Fig. 2-12 is employed for vacuum evaporation of metals. The basic pumping system consists of a multistage

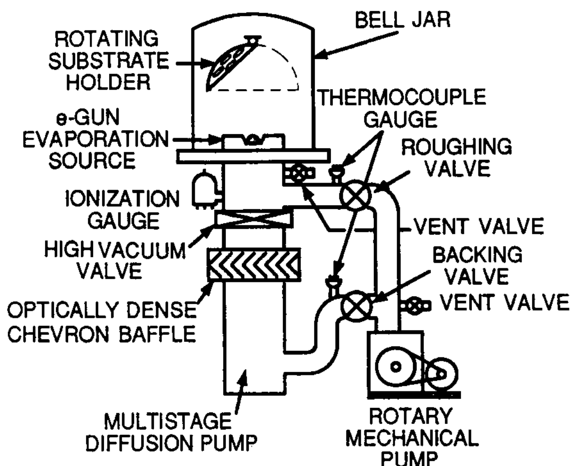


Figure 2-12 Schematic of vacuum deposition system.

oil-diffusion pump backed by a rotary mechanical pump. In order to sequentially coat batch lots, the upper chamber must be vented to air to allow loading and unloading of substrates. To minimize the pumping cycle time, however, it is desirable to operate the diffusion pump continuously, thus avoiding the wait involved in heating or cooling pump oil. This means that the pump must always view a vacuum of better than $\sim 10^{-1}$ torr above and must be backed by a similar pressure at the exhaust. A dual vacuum-pumping circuit consisting of three valves, in addition to vent valves, is required to accomplish these ends.

When starting cold, the high-vacuum and roughing valves are closed and the foreline valve is open. Soon after the oil heats up, a high vacuum is reached above the diffusion pump. The foreline valve is then closed, isolating the diffusion pump, and the roughing valve is opened, enabling the rotary pump to evacuate the chamber to a tolerable vacuum of about 10^{-1} torr. Finally, the roughing valve is shut and both the foreline and high-vacuum valves are opened, allowing the diffusion pump to bear the main pumping burden. By reversing the order of valving the system can now be alternately vented or pumped in a rapid and efficient manner. This same operational procedure is followed in other diffusion-pumped systems such as electron microscopes where ease of specimen exchange is a requirement. In order to eliminate human error, pumpdown cycles are now automated or computerized through the use of pressure sensors and electrically actuated valves. In other oilless vacuum systems a similar valving arrangement exists between the involved fore and main pumps.

Components worthy of note in this evaporator are the high-vacuum valve and the optically dense baffle, both of which are designed to have a high conductance. Cryogenic cooling of the baffle helps prevent oil from back-streaming or creeping into the vacuum chamber. To ensure proper pressure levels for the functioning of pumps, capacitance or thermocouple gauges are located in both the roughing and forelines. They operate in the pressure range 10^{-3} torr to 1 atm. Ionization gauges, on the other hand, are sensitive to vacuum levels spanning the range 10^{-3} to lower than 10^{-10} torr and are, therefore, located so as to measure chamber pressure. Virtually all quoted vacuum pressures in thin-film deposition, processing, and characterization activities are derived from ionization gauges. For this reason vacuum gauges will be further discussed in Section 2.5.4.1.

An actual vacuum deposition system is shown in Fig. 2-13.

2.5.2 System Pumping Considerations

During the pumpdown of a system gas is removed from the chamber by volume pumping as well as pumping of species outgassed from internal

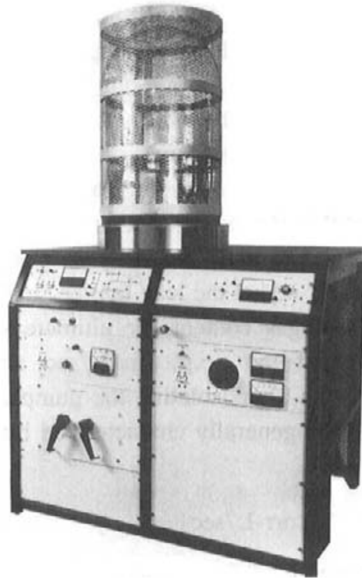


Figure 2-13 Photograph of vacuum deposition system. (Courtesy of Cooke Vacuum Products.)

surfaces. For the case of volume pumping it is a relatively simple matter to calculate the time required to reach a given pressure. As an example, let us estimate the time required to evacuate a cylindrical bell jar, 46 cm in diameter and 76 cm high, from atmospheric pressure to a forepressure of 10^{-1} torr. If an 8 liters/s mechanical pump is used, then substitution of $S_p = 8$ liters/s, $V = \frac{\pi}{4}(46)^2(76)/1000 = 126.3$ liters, $P(t) = 10^{-1}$ torr, $P_i = 760$ torr, and $P_0 = 10^{-4}$ torr in Eq. 2-20 yields a pumpdown time of 2.35 min. This value is comparable to typical forepumping times in clean, tight systems.

It is considerably more difficult to calculate pumping times in the high-vacuum regime where the system pressure depends on outgassing rates. The sources of this gas stem from permeation and diffusion of volatile species through the system walls, followed by desorption from the chamber surfaces and vacuum hardware. Each of the gas sources additively contributes to the overall time dependence of the system pressure decrease as schematically depicted in Fig. 2-14. An equation of the form

$$P = P_0 \exp - (St/V) + Q_{\text{des}}/S + Q_{\text{D}}/S + Q_{\text{pe}}/S \tag{2-22}$$

expresses this behavior, where the first term is due to pumping gas within

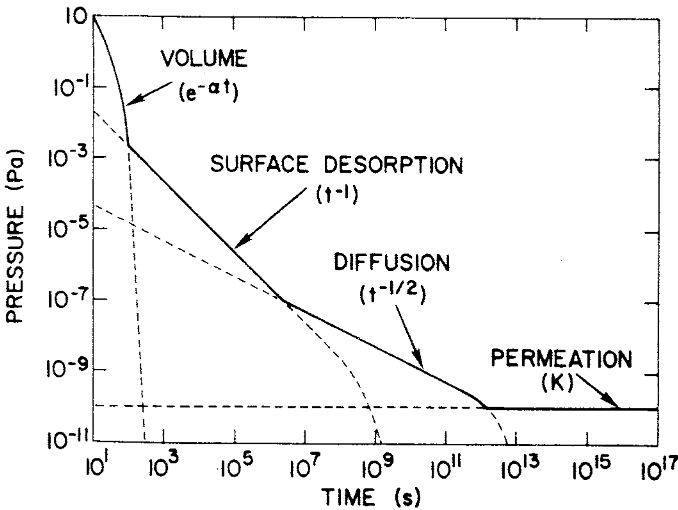


Figure 2-14 Rate limiting pumping processes during evacuation of a vacuum chamber. (From J. F. O'Hanlon, *A User's Guide to Vacuum Technology*, Second Edition. Copyright © 1989 by John Wiley & Sons, Inc. Reprinted with the permission of the publisher.)

the volume (Eq. 2-20), and Q_{des}, Q_D, and Q_{pe} are throughputs associated with surface desorption, diffusion, and permeation, respectively (Ref. 2). Typical time dependencies are indicated in the figure and reveal that most of the time involved in pumping systems to high vacuum is spent in removing gas from surfaces. Gas evolution rates q (in units of (torr-liters/s)/m^2) are strongly dependent on the nature of the materials employed in vacuum systems. Specific vacuum materials, surface conditions (smooth, porous, degree of cleanliness, etc.), and bakeout procedures critically affect the magnitude of q. In general, q for metals (except carbon steel) and glasses is lower than for polymers. For this reason the use of elastomers, hydrocarbon oils and greases should be avoided in vacuum systems.

A practical example will illustrate how to determine the necessary pumping speed at the required operating pressure when significant gas evolution occurs from both the vacuum chamber walls and gasket seals. Suppose a vacuum system has a total wall area (A_w) of 2.5 m^2 with a total gas evolution rate (q_w) of 1.5×10^{-4} (torr-liters/s)/m^2. In addition, Viton A fluoroelastomer gasket seals of total area $A_g = 100\,\text{cm}^2$ evolve gas at a rate $q_g = 3 \times 10^{-3}$ (torr-liters/s)/m^2. What pumping speed is required to maintain a pressure $P = 7.5 \times 10^{-7}$ torr? With the use of Eq. 2-16, $S = [(q_w)(A_w) + (q_g)(A_g)]/P$. Substituting, $S = 540$ liters/s. This value of S should

be regarded as a lower limit for the effective pumping speed of the system.

There are a few effective strategies for stimulating the desorption of adsorbed gases from contaminated surfaces. Adsorbants are either physically bound (physisorbed) to surfaces by weak van der Waals interactions, or chemically bonded to surface atoms (chemisorbed) by stronger ionic or covalent bonds. In either case elevated temperatures facilitate degassing, and this accounts for the widespread practice of baking pumped vacuum chambers and components. The removal of adsorbed water from surfaces of vacuum hardware is a particular objective of such treatments (Ref. 10). Even trace quantities of water vapor are known to assist in the formation of particles or oxide defects in a wide variety of semiconductor processing (film deposition and etching) steps. Beyond thermal methods, electron, ion, and photon (ultraviolet light) beams have also been successfully employed in detaching adsorbants from vacuum hardware. Collision of energetic charged particles with adsorbed gases enhances chemical reactions that lead to their removal. In this way lower ultimate pressures are often attained and in shorter times.

2.5.3 VACUUM LEAKS

It is appropriate to comment on vacuum system leaks since there is scarcely a thin film technologist alive who has not struggled with them. No vacuum apparatus is absolutely vacuum tight and, in principle, it does not have to be. What is important, however, is that the leak rate be small and not influence the required working pressure, gas content, or ultimate system pressure. Leak rates are given in throughput units, e.g., torr-liters/s, and measured by noting the pressure rise in a system after isolating the pumps. The leak tightness of high vacuum systems can be generally characterized by the following leakage rates (Ref. 9):

Very leak tight	$< 10^{-6}$ torr-liters/s
Adequately leak tight	$\sim 10^{-5}$ torr-liters/s
Not leak tight	$> 10^{-4}$ torr-liters/s

One way to distinguish between gas leakage and outgassing from the vessel walls and hardware is to note the pressure rise with time. Gas leakage causes a linear rise in pressure while outgassing results in a pressure rise which gradually diminishes and tends to a limiting value. The effect of leakage throughput on pumping time can be accounted for by inclusion in Eq. 2-20.

2.5.4 Monitoring the Vacuum Environment

2.5.4.1 Gauges to Measure Pressure

From the standpoint of thin-film deposition, properties, and analysis, the total gas pressure and composition are the only vacuum characteristics ever quoted. For this reason it is appropriate to end the chapter by enumerating the techniques to measure system pressure followed by a brief discussion of how the individual gases that contribute to the total pressure are identified. Normally these are separate issues, but they are sometimes coupled because some techniques to measure pressure are sensitive to the nature of the gas sampled. In fact, most gauges record pressure indirectly because they capitalize on measuring properties (e.g., thermal conductivity, ionization) that vary with a particular gas or gas mixture. Pressure values are thus subject to interpretation uncertainty. There are other direct or absolute pressure gauges whose readings are independent of gas composition. Unfortunately they are limited to measurement of relatively high pressures. Representative gauges from each category will be discussed here; treatments of other gauges can be found in all books dealing with vacuum technology (Refs. 1, 2, 11).

2.5.4.1.1 Direct Pressure Gauges

In applications discussed in this book the most widely used direct gauge is probably the capacitance manometer. It consists of a stretched, flexible circular metal diaphragm that is welded in place symmetrically between two fixed electrode plates effectively isolating two regions of space. If, for example, these are two adjacent chambers at different pressures, the diaphragm will bow toward one of the electrodes. This causes an effective capacitance imbalance which is converted to a voltage proportional to the pressure. Alternatively, one chamber may be permanently sealed and evacuated to a pressure of $\sim 10^{-7}$ torr, while the other opens to the vacuum chamber; in this case the capacitance change is a direct measure of the vacuum chamber pressure. Capacitance manometers operate in the range from atmospheric pressure to about 10^{-5} torr. They are absolute gauges because any gas at the same pressure would yield the same capacitance signal.

2.5.4.1.2 Indirect Pressure Gauges

For the relatively high-pressure regime ($\sim 10^{-4}$ torr to 1 atm), thermocouple and to a lesser extent Pirani gauges are the most popular. Their

operating principle is based on the fact that in the range of Knudsen numbers, $10 > \mathrm{Kn} > 0.01$, the rate of heat transfer through gases is linearly proportional to pressure. Both gauges sense the rate of heat transfer between a heated wire and a nearby wall. In the Pirani gauge a metal filament is heated by a precise current and reaches a steady temperature dependent on how many gas molecules are present to transport heat away. The measured filament resistance which varies linearly with temperature is thus proportional to the gas pressure. Rather than measure temperature indirectly through resistance change, the thermocouple gauge employs a thermocouple placed in close thermal but not electrical contact with the heated filament. Thermocouple as well as Pirani gauges are calibrated in systems where the pressure is simultaneously measured by direct gauges, e.g., a capacitance manometer.

In the high and ultrahigh vacuum regimes the gas density is so low that detection of direct momentum or heat transfer to a wall is not practical. Instead gas molecules are ionized and the collection of the ion current is relied upon to yield a measure of the system pressure. Hot-cathode gauges such as the Baird–Alpert ionization tube are based on this principle and are almost universally employed to monitor pressure in the $\sim 10^{-5}$ to 10^{-13} torr range. The tube consists of a filament from which electrons are thermionically emitted and drawn to a positively biased grid. An open coiled grid structure enables electrons to pass through (overshoot) it and return, thus yielding long electron lifetimes that increase the probability of collisional impact with gas molecules. Such impact generates positive ions that are drawn to a fine wire collector positioned coaxially within the grid. The ion current (i^+) detected is a measure of the gas density and hence pressure (P); it is proportional to the electron emission (thermionic) current (i^-) and gas ionization sensitivity factor (S), so that $i^+ = PS\,i^-$. Because different gases have different S values it is apparent that we are dealing with an indirect pressure gauge.

2.5.4.2 Residual Gas Analysis

The difficult task of qualitatively identifying the individual gaseous species within a vacuum system is accomplished with mass spectrometers and residual gas analyzers (RGAs), also known as partial pressure analyzers. These instruments function by measuring the ion mass (M) to charge valence (z) ratio (M/z) of ionized gas particles. As in ionization gauges, positive ions and ion fragments are produced through electron impact with neutral gas molecules that enter the ionization region of the RGA. Ions are generated in numbers proportional to the partial pressure of the particular gas ionized. After extraction into a beam, the ions are normally spatially

separated by means of magnetic or electric fields. In the magnetic sector analyzer (the only one discussed) ions of velocity v are first accelerated through an electric potential of magnitude V and then directed into a uniform magnetic field (B) oriented perpendicular to the ion motion. The well-known Lorentz force, acting at right angles to both v and B, bends the ion trajectory into a circular orbit whose radius r is given by $r = 1/B\ (2MV/zq)^{1/2}$, where q is the electronic charge. A detector located behind a slit situated on the orbit circumference records the impinging ion current. Qualitative analysis is accomplished by sweeping the voltage at fixed r and B as each ion, characterized by its M/z value, is accessed in turn at a specific V value. The signal amplitude is a measure of the specie abundance.

A typical mass spectrum shown in Fig. 12-15 indicates the presence of many gas ion species at different concentration levels. It is beyond our scope to provide an analysis of this spectrum, but the following comments may indicate where to begin as well as where the difficulties await. As expected, common vacuum chamber gases N_2, O_2, H_2O, and CH_4 typically appear at masses of 28, 32, 18, and 16, respectively, when $z = 1$. Mass peaks higher than $\sim$40 are generally due to hydrocarbons from oil, greases and polymers while chlorinated compounds often appear at masses of 35, 36 and 37. Complications arise because two or more species often have the same mass, e.g., a peak mass of 28 could also correspond to CO, whereas 16 could be due to O^+. Quantitative analysis is even more difficult because in addition to mass overlap and peak pileup, the sensitivity factors for ionization of each gas generally influence the signal magnitudes in complex ways. For such cases salvation often comes from comparing actual mass-peak patterns with prerecorded spectral signatures for standard gas mixtures provided by RGA manufacturers.

2.6 CONCLUSION

The kinetic theory of gases is all that is needed to adequately describe isolated reduced-pressure environments and the behavior of gases within them. Thus molecular velocities and their statistical spread, collision rates between gas molecules, and impingement rates at surfaces are all readily calculable in terms of the system temperature and molecular weights and sizes. Summation of these microscopic events over all molecules leads to the perfect gas law and gives significance to the primary measurable variable in a vacuum chamber, namely its pressure.

When the system is now subjected to the action of pumps the behavior of the gas is very much dependent on its pressure. In the low-pressure molecular-flow regime the situation is little changed from that of an isolated

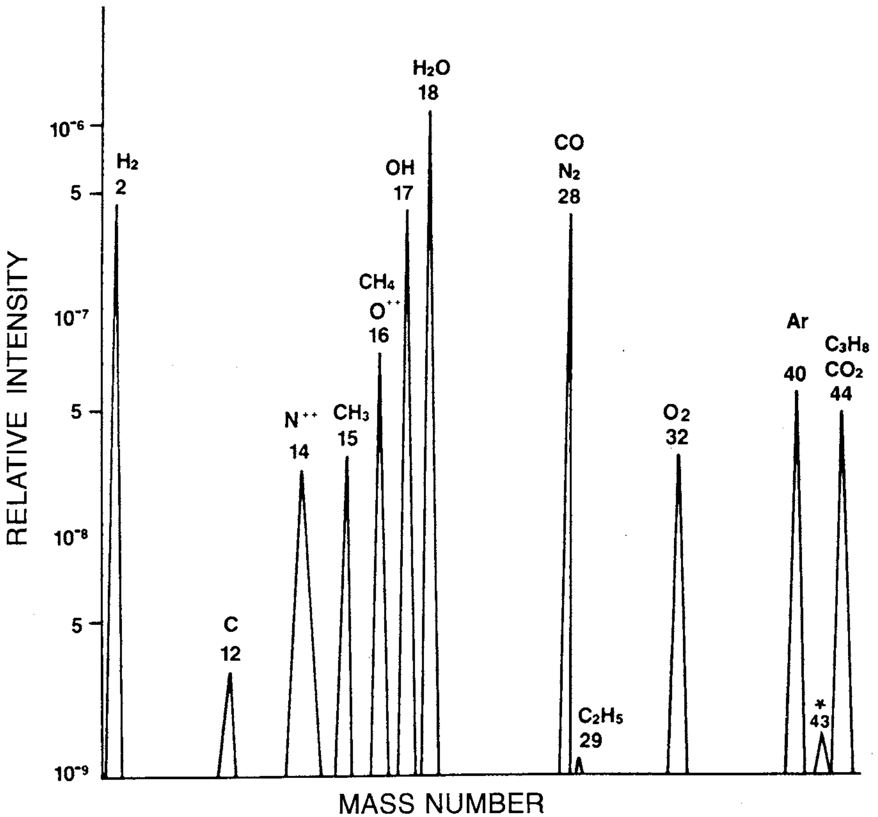


Figure 2-15 Typical residual-gas analysis scan plotted as relative signal intensity (of partial pressures) versus atomic mass.

chamber where the perfect gas law applies. Even though pressure differences develop across components, conductances are calculable from the kinetic theory of gases. However, at the higher pressures in the viscous-flow regime, gas flow is governed by the laws of compressible fluid dynamics. Conductances, throughputs, and pumping speeds are now complicated functions of pressure. With the exception of chemical vapor deposition this book is exclusively concerned with the easier-to-model low-pressure chambers. Attaining low system pressures is largely a matter of connecting a high vacuum–backing pump combination to the chamber via high-conductance ducts. Oil diffusion, turbomolecular, cryo-, and ion pumps produce the low pressures, while rotary, Roots, and sorption pumps are necessary to provide the necessary forepressure for operation of the former. Assorted valves,

cold traps, and gauges round out the complement of required vacuum hardware.

Vacuum concerns are a bit of annoyance and often much worse to the thin-film technologist and surface scientist. Efficient operation of vacuum systems means grappling with leaks, degassing chamber walls and contents, and reducing pumping times. This chapter has attempted to address these issues by presenting the underlying principles as well as practical responses.

EXERCISES

1. Consider a mole of gas in a chamber that is not being pumped. What is the probability of a self-pumping action such that all of the gas molecules will congregate in one-half of the chamber and leave a perfect vacuum in the other half?
2. A 1 m^3 cubical-shaped vacuum chamber contains O_2 molecules at a pressure of 10^{-4} atm at 300 K.
   (a) How many molecules are there in the chamber?
   (b) What is the ratio of maximum potential energy to average kinetic energy of these molecules?
   (c) What fraction of gas molecules has a kinetic energy in the x direction exceeding RT? What fraction exceeds $2RT$?
3. In some vacuum systems there is a gate valve consisting of a gasketed metal plate that acts to isolate the chamber above from the pumps below.
   (a) A sample is introduced into the chamber at 760 torr while the isolated pumps are maintained at 10^{-6} torr. For a 15-cm diameter opening, what force acts on the valve plate to seal it?
   (b) The chamber is forepumped to a pressure of 10^{-2} torr. What force now acts on the valve plate?
4. Supersonic molecular beams have a velocity distribution given by $f(v) = Av^3 \exp - [M(v - v_0)^2/2RT]$, where v_0, the stream velocity, is related to the Mach number.
   (a) What does a plot of $f(v)$ vs v look like?
   (b) What is the average gas speed in terms of v_0, M, and T?
   Note:
   $$\int_0^\infty x^{2n} \exp^{-ax^2} dx = \frac{\Gamma(n + 1/2)}{2a^{n+1/2}}; \qquad \int_0^\infty x^{2n+1} \exp^{-ax^2} dx = \frac{n!}{2a^{n+1}}.$$
5. The trap in Fig. 2-5 is filled with liquid N_2 so that the entire trap surface is maintained at 80 K. What effect does this have on conductance?

6. Two identical lengths of piping are to be joined by a curved 90° elbow section or a sharp right-angle elbow section. Which overall assembly is expected to have a higher conductance? Why?

7. Show that the conductance of a pipe joining two large volumes through apertures of area A and A_0 is given by $C = 11.7AA_0/(A - A_0)$.
Hint: Calculate the conductance of the assembly in both directions.

8. A chamber is evacuated by two sorption pumps of identical pumping speed. In one configuration the pumps are attached in parallel so that both pump simultaneously. In the second configuration they pump in serial or sequential order (one on and one off). Comment on the system pumping characteristics (pressure vs time) for both configurations.

9. It is common to anneal thin films under vacuum in a closed-end quartz tube surrounded by a furnace. Consider pumping on such a cylindrical tube of length L, diameter D, and conductance C that outgasses uniformly at a rate q_0 (torr-liter/cm^2-s). Derive an expression for the steady-state pressure distribution along the tube axis.
Hint: Equate the gas load within length dx to throughput through the same length.

10. After evacuation of a chamber whose volume is 30 liters to a pressure of 1×10^{-6} torr, the pumps are isolated. The pressure rises to 1×10^{-5} torr in 3 minutes.
    (a) What is the leakage rate?
    (b) If a diffusion pump with an effective speed of 40 liters/s is attached to the chamber, what ultimate pressure can be expected?

11. Select any instrument or piece of equipment requiring high vacuum during operation (e.g., electron microscope, evaporator, Auger spectrometer). Sketch the layout of the vacuum-pumping components within the system. Explain the pumping sequence.

12. A system of volume equal to 1 m^3 is evacuated to an ultimate pressure of 10^{-7} torr employing a 200 liters/s pump. For a reactive evaporation process, 100 cm^3 of gas (STP) must be continuously delivered through the system per minute.
    (a) What is the ultimate system pressure in this case?
    (b) What conditions are necessary to maintain this process at 10^{-3} torr?

13. In a tubular low-pressure chemical vapor deposition (LPCVD) reactor, gas is introduced at one end at a rate of 75 torr-liter/min. At the other end is vacuum pump of speed S_p. If the reactor must operate at 1 torr, what value of S_p is required?

14. In systems to metallize or web coat large surface areas of polymer sheet it is often the case that systems contain oil vapor pumps in combination with cryosurfaces. The former are used to pump noncondensable gases while the latter are effective in pumping condensable vapors. Decoupling their respective behaviors is frequently a good assumption. Consider a system with a gas load of 10 mbar-liter/s, of which 80% is water vapor and 20% is air. The system is pumped with a 10,000 liters/s diffusion pump.
    (a) What pressure will be reached?
    (b) If a cryosurface with a pumping speed of 50,000 liters/s is added to the diffusion pump, what system pressure will be reached?
15. (a) Atoms and molecules that do not dissociate upon adsorption on a surface desorb from it when pumped according to $dC/dt = -K_1C$, where K_1 is a rate constant and C is the species concentration. Solve for $C(t)$.
    (b) For molecules that dissociate upon adsorption on a surface, the equation $dC/dt = -K_2C^2$ (K_2 is a constant) describes the rate of desorption during pumping. Solve for $C(t)$.
    (c) Suppose surfaces were contaminated with either type of adsorbent species. In which case do you expect surface cleaning through pumping to be more rapid?
16. Advances in the design of turbomolecular pumps have enabled them to maintain high compression ratios at foreline pressures of about 10 torr, which is about two decades better than previously possible. What effect does this have on the choice of backing pumps?
17. A screw threaded into a blind hole often traps air that can become a virtual leak in the vacuum system. Consider a 1000-liter chamber containing eight 0.635 cm diameter screws, each of which leaves a 0.0794-cm space at the bottom of a blind hole.
    (a) What chamber pressure would be associated with this source of trapped gas?
    (b) Virtual leaks tend to release gas in bursts. If so, describe the pressure gauges you would use to monitor such leaks. Would you recommend digital gauge readouts?

REFERENCES

1. D. M. Hoffman, B. Singh, and J. H. Thomas, eds., *Handbook of Vacuum Technology*. Academic Press, Boston, 1997.
2. J. F. O'Hanlon, *A User's Guide to Vacuum Technology*, 2nd ed. John Wiley & Sons, New York, 1989.

3. R. Glang, in *Handbook of Thin Film Technology*, L. I. Maissel and R. Glang, eds. McGraw-Hill, New York, 1970.
4. A. Roth, *Vacuum Technology*. North Holland, Amsterdam, 1976.
5. S. Dushman, *Scientific Foundations of Vacuum Techniques*. John Wiley & Sons, New York, 1962.
6. R. Glang, R. A. Holmwood, and J. A. Kurtz, in *Handbook of Thin Film Technology*, L. I. Maissel and R. Glang, eds. McGraw-Hill, New York, 1970.
7. G. H. Geiger and D. R. Poirer, *Transport Phenomena in Metallurgy*, Addison-Wesley, Reading, MA, 1973.
8. R. W. Roberts, "An Outline of Vacuum Technology," GE Report No. 64-RL 3394C (1964).
9. *Vacuum Technology: Its Foundations, Formulae and Tables in Product and Vacuum Technology Reference Book*, Leybold-Heraeus Inc., San Jose, CA.
10. M. Li and H. F. Dylla, *J. Vac. Sci. Technol. A* **4**, 1702 (1993).
11. N. Harris, *Modern Vacuum Practice*. McGraw-Hill, London, 1989.

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 3

Thin-Film Evaporation Processes

3.1 INTRODUCTION

This chapter marks the beginning of our discussion on the deposition of thin films, and we start by focusing on *evaporation* in vacuum as the means. The objective of this deposition process is to controllably transfer atoms from a heated source to a substrate located a distance away, where film formation and growth proceed atomistically. Quite simply, thermal energy is imparted to atoms in a liquid or solid source such that their temperature is raised to the point where they either efficiently evaporate or sublime. Evaporation differs from *sputtering*, another method for physically depositing films. In sputtering atoms are ejected from source surfaces usually maintained at room temperature, through the impact of gaseous ions. The earliest experimentation on both of these film deposition techniques can apparently be traced to the same decade of the 19th century. In 1852 Grove (Ref. 1) observed metal deposits sputtered from the cathode of a glow discharge plasma (see Chapter 4). Five years later Faraday (Ref. 2), experimenting with exploding fuselike metal wires in an inert atmosphere, evaporated thin films.

Advances in the development of vacuum pumping equipment and the fabrication of suitable Joule heating sources, first made from platinum and then tungsten wire, spurred the progress of evaporation technology. Scientific interest in the phenomenon of evaporation and the properties of thin metal films was soon followed by industrial production of optical components such as mirrors and beam splitters, and later of antireflection coatings. Simultaneously, sputtering was used as early as 1877 to coat mirrors. Until

the late 1960s evaporation surpassed sputtering as the preferred film deposition technique. Higher deposition rates, better vacuum and cleaner environments for film formation and growth, and general applicability to all classes of materials were among the reasons for the ascendancy of evaporation methods. These advantages still sustain the widespread use of evaporation in the deposition of optical thin films, as well as large-area web coatings used in assorted applications. Furthermore, newer techniques such as pulsed laser deposition also capitalize on thermal evaporation. However, the necessity for alloy films with stringent stoichiometry requirements in microelectronics and magnetic applications spurred the development and widespread use of sputtering. In a parallel vein, chemical vapor deposition (CVD) processes were developed to deposit nonmetallic hard coatings, dielectric films, and single-crystal semiconductor films.

Physical vapor deposition (PVD), the term that includes both evaporation (this chapter) and plasma-assisted sputtering (Chapter 5), and *chemical vapor deposition* (Chapter 6) together with all of their variant and hybrid combinations are the basic film deposition processes treated in this book. These as well as other thin-film processing techniques such as etching and patterning have been broadly reviewed by Vossen and Kern (Ref. 3). Among the factors that distinguish PVD from CVD are the following:

1. Reliance on solid or molten sources, as opposed to generally gaseous precursors in CVD
2. The physical mechanisms (evaporation or collisional impact) by which source atoms enter the gas phase
3. A reduced pressure environment through which the gaseous species are transported
4. The general absence of chemical reactions in the gas phase and at the substrate surface (reactive PVD processes are exceptions)

Nowadays the decision of whether to evaporate, sputter, or chemically deposit thin films for particular applications is not always obvious and has fostered a lively competition among these alternative technologies. In many cases features from each have been forged into hybrid processes possessing added capabilities.

This chapter presents the attributes of evaporation processes, their advantages and limitations, as well as their potential for new uses. Irrespective of particular application, the control of film composition and thickness uniformity are primary concerns addressed. In coping with these issues the science of evaporation, effect of process geometry, and characteristics of heating sources employed all have an influence that will be explored in subsequent sections.

3.2 THE PHYSICS AND CHEMISTRY OF EVAPORATION

3.2.1 EVAPORATION RATE

Early attempts to quantitatively interpret evaporation phenomena are associated with the names of Hertz and Knudsen and, later, Langmuir (see Ref. 4). Based on experimentation on the evaporation of mercury, Hertz in 1882 observed that evaporation rates were:

1. Not limited by insufficient heat supplied to the surface of the molten evaporant.
2. Proportional to the difference between the equilibrium pressure, P_e, of Hg at the given temperature and the hydrostatic pressure, P_h, acting on the evaporant.

He concluded that a liquid has a specific ability to evaporate at a given temperature. Furthermore, the maximum evaporation rate is attained when the number of vapor molecules emitted corresponds to that required to exert the equilibrium vapor pressure while none return. These ideas led to the basic equation for the rate of evaporation from both liquid and solid surfaces, namely,

$$\Phi_e = \frac{\alpha_e N_A (P_e - P_h)}{(2\pi MRT)^{1/2}} \tag{3-1}$$

where Φ_e is the evaporation flux in number of atoms (or molecules) per unit area, per unit time, and α_e is the coefficient of evaporation, which has a value between 0 and 1. When $\alpha_e = 1$ and P_h is zero, the maximum evaporation rate is realized. By analogy with Eq. 2-9 an expression for the maximum value of Φ_e is

$$\Phi_e = \frac{3.513 \times 10^{22}}{(MT)^{1/2}} P_e \qquad \text{molecules/cm}^2\text{-s.} \tag{3-2}$$

When P_e is expressed in torr, a useful variant of this formula is

$$\Gamma_e = 5.84 \times 10^{-2} (M/T)^{1/2} P_e \qquad \text{g/cm}^2\text{-s,} \tag{3-3}$$

where Γ_e is the mass evaporation rate. At a pressure of 10^{-2} torr, a typical value of Γ_e for many elements is approximately 10^{-4} grams per second per cm^2 of evaporant area. The key variable influencing evaporation rates is temperature since it has a profound effect on equilibrium vapor pressures.

3.2.2 Vapor Pressure of the Elements

A convenient starting point for expressing the connection between temperature and vapor pressure is the Clausius–Clapyeron equation, which for both solid–vapor and liquid–vapor equilibria can be written as

$$\frac{dP}{dT} = \frac{\Delta H(T)}{T\Delta V}. \tag{3-4}$$

The changes in enthalpy, $\Delta H(T)$, and volume, ΔV, refer to differences between the vapor (v) and the particular condensed phase (c) from which it originates, while T is the transformation temperature in question. Since $\Delta V = V_v - V_c$, and the volume of vapor normally considerably exceeds that of the condensed solid or liquid phase, $\Delta V \simeq V_v$. If the gas is assumed to be perfect, $V_v = RT/P$, and Eq. 3-4 may be rewritten as

$$\frac{dP}{dT} = \frac{P\Delta H(T)}{RT^2}. \tag{3-5}$$

As a first approximation, $\Delta H(T) = \Delta H_e$, the molar heat of evaporation (a constant), in which case simple integration yields

$$\ln P \simeq -\frac{\Delta H_e}{RT} + I \qquad \text{or} \qquad P = P_0 \exp -\frac{\Delta H_e}{RT} \tag{3-6}$$

where I (or $P_0 = \exp I$) is the constant of integration. Through substitution of the latent heat of vaporization ΔH_V for ΔH_e, the boiling point for T, and 1 atm for P, the value of I can be determined for the liquid–vapor transformation. For practical purposes Eq. 3-6 adequately describes the temperature dependence of the vapor pressure in many materials. It is rigorously applicable over only a small temperature range, however. To extend the range of validity, the temperature dependence of $\Delta H(T)$ must be taken into account. For example, careful evaluation of thermodynamic data reveals that the vapor pressure of liquid Al is given by (Ref. 4)

$$\log P(\text{torr}) = -15{,}993/T + 12.409 - 0.999 \log T - 3.52 \times 10^{-6}T. \tag{3-7}$$

The Arrhenius character of $\log P$ vs $1/T$ is essentially preserved by the first two terms on the right-hand side while the remaining terms are small corrections.

Vapor-pressure data for many other metals have been similarly obtained and conveniently represented as a function of temperature in Fig. 3-1 (Ref. 5). Similarly, vapor-pressure data for elements important in the deposition of semiconductor films are presented in Fig. 3-2 (Ref. 6). Many of the data

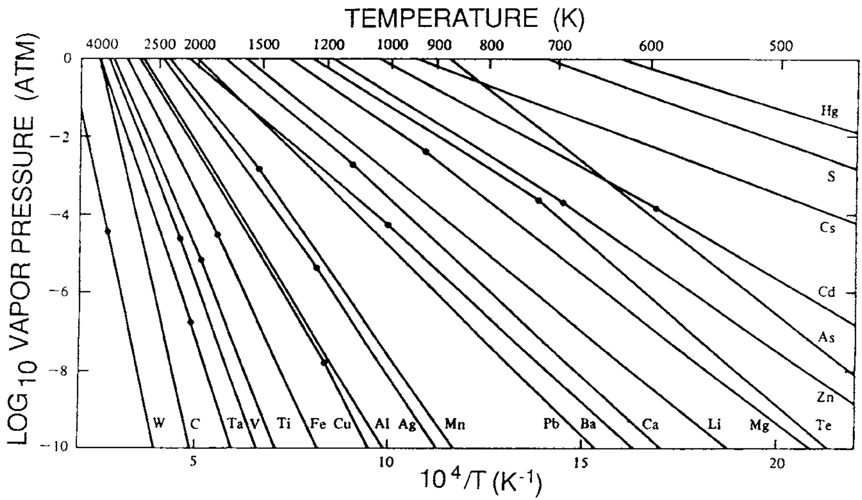


Figure 3-1 Vapor pressures of selected elements. Dots correspond to melting points. (From Ref. 5.)

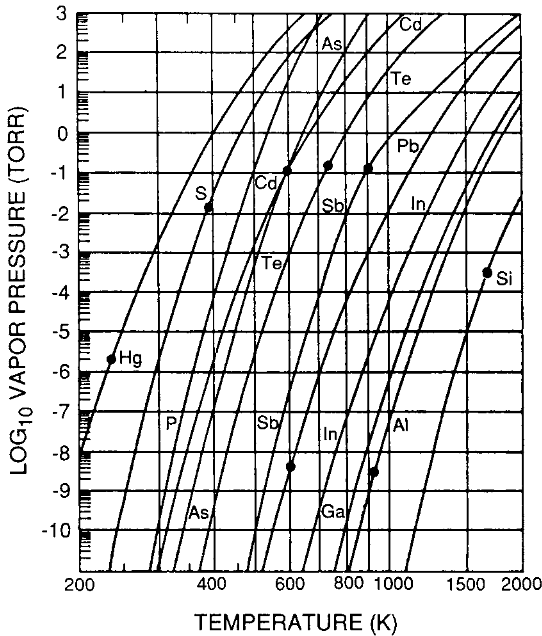


Figure 3-2 Vapor pressures of elements employed in semiconductor materials. Dots correspond to melting points. (Adapted from Ref. 6.)

represent direct measurements of the vapor pressures. Other values are inferred indirectly from thermodynamic relationships and identities using limited experimental data. Thus the vapor pressures of many refractory metals can be unerringly extrapolated to lower temperatures even though it may be impossible to measure them directly. For this to be practical the thermodynamic data that are available must be accurate.

Two modes of evaporation can be distinguished in practice depending on whether the vapor effectively emanates from a liquid or solid source. As a rule of thumb, a melt will be required if the element in question does not achieve a vapor pressure greater than 10^{-3} torr at its melting point. Most metals fall into this category and effective film deposition is attained only when the source is molten. On the other hand, elements such as Cr, Ti, Mo, Fe, and Si reach sufficiently high vapor pressures below the melting point and, therefore, sublime. For example, Cr can be effectively deposited at high rates from a solid metal source because it reaches vapor pressures of 10^{-2} torr some 500°C below its melting point. The operation of the Ti sublimation pump mentioned in Chapter 2 is, in fact, based on the sublimation of Ti from heated Ti filaments. A third example is carbon, which is used to prepare replicas of the surface topography of materials for subsequent examination in the electron microscope. Carbon, which has an extremely high melting point, is readily sublimed from an arc struck between graphite electrodes.

3.2.3 Evaporation of Multielement Materials

3.2.3.1 Ionic Compounds

Whereas metals essentially evaporate as atoms and occasionally as clusters of atoms, the same is not true of compounds. Very few inorganic compounds evaporate without molecular change and, therefore, the vapor composition is usually different from that of the original solid or liquid source. As a consequence the stoichiometry of the film deposit will generally differ from that of the source. Mass spectroscopic studies in the vapor phase have shown that the processes of both molecular association and dissociation frequently occur. A broad range of evaporation phenomena in compounds occurs, and these are categorized briefly in Table 3-1. The troublesome decomposition of multivalent metal oxides to lower oxides can be compensated for by reactive evaporation in an oxygen ambient.

3.2.3.2 Deposition of GaAs: The Growth Window

One might imagine that depositing a compound semiconductor film such as GaAs simply involves pinpointing the desired growth temperature on the

Table 3-1

Evaporation of Compounds

Reaction type	Chemical reaction[a]	Examples	Comments
Evaporation without dissociation	$MX(s\ or\ l) \rightarrow MX(g)$	SiO, B_2O_3, GeO, SnO, AlN, CaF_2, MgF_2	Compound stoichiometry maintained in deposit
Decomposition	$MX(s) \rightarrow M(s) + \frac{1}{2}X_2(g)$ $MX(s) \rightarrow M(l) + \frac{1}{n}X_n(g)$	Ag_2S, Ag_2Se III–V semiconductors	Separate sources are required to deposit these compounds
Evaporation with dissociation			Deposits are metal-rich
(a) Chalcogenides	$MX(s) \rightarrow M(g) + \frac{1}{2}X_2(g)$ $X = S, Se, Te$	CdS, CdSe, CdTe	Separate sources usually required to deposit these compounds
(b) Oxides	$MO_2(s) \rightarrow MO(g) + \frac{1}{2}O_2(g)$	SiO_2, GeO_2, TiO_2, SnO_2, ZrO_2	Metal-rich discolored deposits; dioxides are best deposited in O_2 partial pressure (reactive evaporation)

[a] M = metal, X = nonmetal.

Adapted from Ref. 4.

phase diagram, i.e., Fig. 1-15, and maintaining a Ga:As evaporation ratio of 1:1. The situation is more complex in practice, however. First, the great disparity in vapor pressure between As and Ga means that two separate evaporation sources will be required. Furthermore, because of vacuum cleanliness requirements, phase diagrams at reduced system pressures, not at 1 atmosphere, are pertinent. As a result not all growth temperatures are feasible since other phases normally coexist with the depositing GaAs. For example, at a pressure of 10^{-6} torr the equilibrium temperature–composition diagram shown in Fig. 3-3a reveals the increasing stability of the vapor

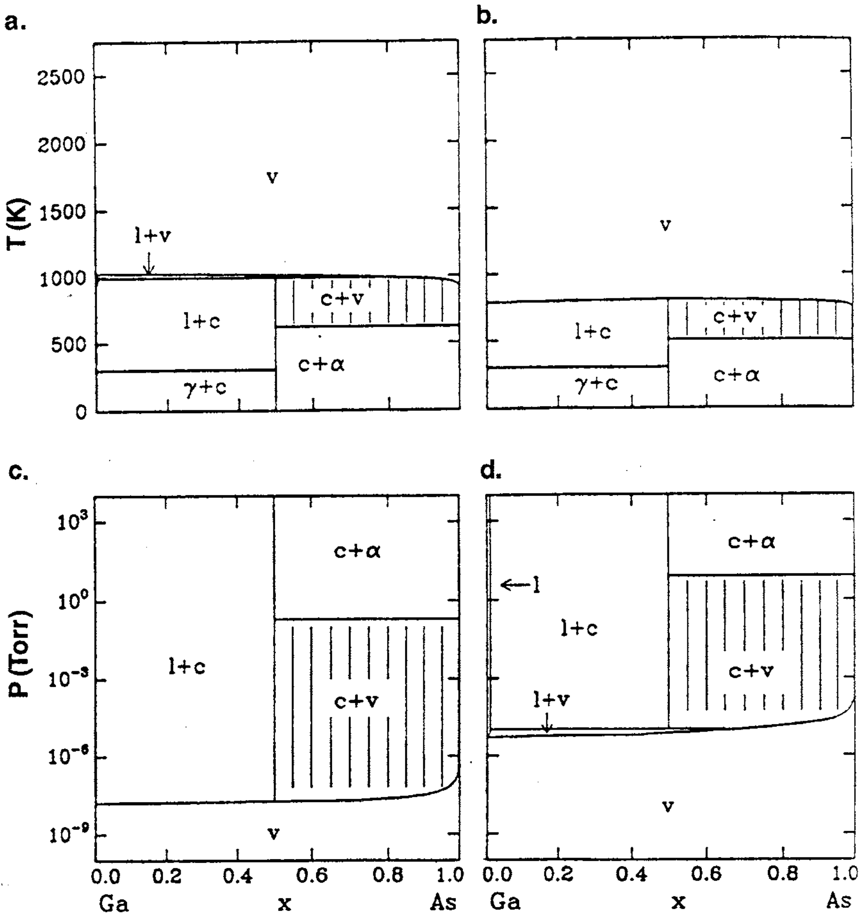


Figure 3-3 Temperature–composition diagram for the Ga–As system at a pressure of (a) 10^{-6} torr and (b) 10^{-9} torr. Pressure–composition diagram for the Ga–As system at a temperature of (c) 850 K and (d) 1000 K. (From J. Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy*. Copyright © 1993 by Academic Press, Inc. Reprinted with the permission of the publisher.)

phase (v) relative to the liquid (l) and solid phases (α, γ, and c). But importantly, there is a growth window that is shaded in consisting of compound c(GaAs) and v. Deposition within this two-phase region will exclusively yield the desired solid GaAs since excess As evaporates; however, outside the window other condensed phases will coexist with c. If deposition conditions are Ga rich, the compound will be contaminated by either Ga-rich liquid droplets or a solid solution (γ) of As in Ga. In either case the low vapor pressure of Ga precludes its removal under vacuum. For these reasons it is clear that an overpressure of volatile As will prevent a Ga-rich environment and promote the stoichiometric growth of GaAs. But even in an As-rich ambient the substrate temperature must not be above $\sim$1000 K or below $\sim$630 K. Under the former conditions GaAs decomposes to 1 + v; in the latter case GaAs will be contaminated by α, an As-rich solid containing some Ga. Physically, at low temperatures the vapor pressure of As in α is less than the impinging As pressure and therefore more As atoms will condense than sublime from the growing film. Operation at a pressure of 10^{-9} torr contracts the c + v field (Fig. 3-3b) and slightly narrows the usable deposition temperature range.

In an alternative representation of GaAs film growth conditions, the equilibrium pressure–composition diagram at 850 K (Fig. 3-3c) now shows v and 1 at the bottom and c + α at the top. Again the desired two-phase c + v shaded region is bounded on the left by the 1 + c region because at elevated pressures Ga condenses rather than reevaporates. The window for GaAs deposition is actually quite wide with possible temperatures ranging from $\sim$350 to 750°C at a typical As overpressure of 10^{-5} torr. If the substrate temperature is raised to 1000 K (Fig. 3-3d) the deposition window contracts slightly and shifts to higher system pressures. Compared to other III–V compounds the GaAs growth window is quite forgiving. For example, in InSb the vapor pressure of Sb is less than that for As, while the vapor pressure of In exceeds that for Ga in the liquid phase. These factors tend to contract the two-phase c(InSb) + v field.

Practical application is made of these thermodynamic fundamentals in the deposition of III–V compound semiconductor films by molecular beam epitaxy techniques (Section 8.6.2).

3.2.3.3 Evaporation of Alloys

Evaporated metal alloy films are widely utilized for a variety of electronic, magnetic, and optical applications as well as for decorative coating purposes. Important examples of such alloys that have been directly evaporated include Al–Cu, Permalloy (Fe–Ni), Nichrome (Ni–Cr), and Co–Cr. Atoms in such alloys are generally less tightly bound than those in

the metal oxides discussed above. Like the III–V compounds, the constituents of metal alloys tend to evaporate nearly independently of each other, entering the vapor phase as single atoms, paralleling the behavior of pure metals. Binary metallic melts are solutions and as such are governed by well-known thermodynamic laws. When the interaction energy between A and B atoms of a binary AB alloy melt are the same as between A–A and B–B atom pairs, then no preference is shown for atomic partners. Such is the environment in an ideal solution. Raoult's law, which holds under these conditions, states that the vapor pressure of component B in solution is reduced relative to the vapor pressure of pure B ($P_B(0)$) in proportion to its mole fraction X_B. Therefore,

$$P_B = X_B P_B(0). \tag{3-8}$$

Metallic solutions usually are not ideal, however. This means that either more or less B will evaporate relative to the ideal solution depending on whether the deviation from ideality is positive or negative, respectively. A positive deviation occurs when B atoms are physically bound more tightly to each other than to the solution, facilitating their tendency to escape or evaporate. In real solutions

$$P_B = a_B P_B(0) \tag{3-9}$$

where a_B is the effective thermodynamic concentration of B known as the activity. The activity is, in turn, related to X_B through an activity coefficient γ_B; i.e.,

$$a_B = \gamma_B X_B. \tag{3-10}$$

Through combination of Eqs. 3-2, 3-9, and 3-10, the ratio of the fluxes of A and B atoms in the vapor stream above the melt is given by

$$\frac{\Phi_A}{\Phi_B} = \frac{\gamma_A X_A P_A(0)(M_B)^{1/2}}{\gamma_B X_B P_B(0)(M_A)^{1/2}} \tag{3-11}$$

where $X_A + X_B = 1$.

Application of this equation is difficult because the melt composition changes as evaporation proceeds. Activity coefficients, which can sometimes be located in the metallurgical literature, but just as frequently not, also change with deposition time making quantitative calculations impractical. Nevertheless, as an example of the use of Eq. 3-11 consider the problem of estimating the approximate Al–Cu melt composition required to evaporate films containing 2 wt% Cu from a single crucible heated to 1350 K. Substituting gives $\Phi_{Al}/\Phi_{Cu} = 98\, M_{Cu}/2\, M_{Al}$. From Fig. 3-1, $P_{Al}(0)/P_{Cu}(0) =$

$10^{-3}/2 \times 10^{-4}$. Furthermore, assuming $\gamma_{Cu} = \gamma_{Al}$,

$$\frac{X_{Al}}{X_{Cu}} = \frac{98(2 \times 10^{-4})(63.7)^{1/2}}{2(10^{-3})(27.0)^{1/2}} = 15.$$

This means that the 2 wt% Cu–Al vapor stream requires a melt with a 15:1 molar ratio of Al to Cu. In order to compensate for the preferential vaporization of Al the original melt composition must be enriched to 13.6 wt% Cu. But the calculation only holds for the first instant of time. With successive loss of the more volatile melt component, the evaporant flux changes in concert, and if nothing is done a graded film of varying composition will deposit, i.e., the desired stoichiometry at the substrate interface, and layers increasingly richer in Cu above it. Clearly, the desired steady-state deposition of alloys having uniform composition is not sustainable, and this fact is a potential disadvantage of evaporation methods.

3.2.4 Maintaining Melt Stoichiometry

Although it may be undesirable from a theoretical standpoint to use a single source to evaporate alloy films, if the melt volume is sufficiently large, fractionation-induced melt composition changes may be small enough to yield acceptable films. There are at least two other ways to cope with fractionation in melts. Even though they are more complicated to deal with than a single melt source, both have been implemented practically. The first is to evaporate from two (or more) independent pure metal melts maintained at different temperatures. This of course means two (or more) of everything, i.e., two sources with separate power supplies, two shutters, two evaporation rate monitors, but one film-thickness monitor. Molecular beam epitaxy capitalizes on such multisource systems to deposit films possessing excellent stoichiometry and high crystalline perfection.

In the second method, the melt composition of a single source is continuously adjusted through external mass additions. This has the effect of replenishing the loss of the more volatile species and maintaining a constant melt height, which would otherwise recede. Eventually the desired steady-state evaporation flux ratio is established. Consider, for example, a solid alloy wire of composition $A_{1-Y}B_Y$ being fed into a melt at a constant volumetric rate $\dot{V}_r$ (cm^3/s) where it is desired to preserve the vapor flux ratio Φ_A/Φ_B given above. Following Smith (Ref. 8) we note that when steady state is reached, the evaporant flux ratio is equal to the feed mass ratio in the wire or $\Phi_A/\Phi_B = (1 - Y)/Y$. If we assume that $\gamma_A/\gamma_B = 1$, then Eq. 3-11 and a bit

of algebra yields a steady-state melt composition of B, $X_S(B)$, given by

$$X_S(B) = \left\{1 + \frac{(1 - Y)P_B(0)(M_A)^{1/2}}{YP_A(0)(M_B)^{1/2}}\right\}^{-1}. \quad (3\text{-}12)$$

At any instant of time a fraction $(X_B/X_S(B))$ of the number of B atoms added per second

$$\frac{\dot{V}_r Y}{\Omega} \quad \text{or} \quad \frac{\dot{V}_r Y}{\Omega}\left[\frac{X_B}{X_S(B)}\right]$$

is lost by evaporation. The remainder accumulates in the melt at a rate given by $V/\Omega \, dX_B/dt$, where Ω is the atomic volume (cm^3/atom) and V is the melt volume (cm^3), which is held constant. As a result, the mass balance can be written as

$$\frac{\dot{V}_r Y}{\Omega} = \frac{\dot{V}_r Y}{\Omega}\left[\frac{X_B}{X_S(B)}\right] + \frac{V}{\Omega}\frac{dX_B}{dt}. \quad (3\text{-}13)$$

This differential equation can be readily integrated assuming the initial condition, at $t = 0$, $X_B = Y$. The result is (see Eq. 2-20)

$$\frac{X_B - X_S(B)}{Y - X_S(B)} = \exp - \frac{\dot{V}_r Yt}{VX_S(B)} \quad (3\text{-}14)$$

and predicts an exponential decay of the melt composition to the steady-state value. Higher wire-feed rates hasten the time for steady-state evaporation, whereas larger melts extend this time. Note that the steady-state melt composition differs from both X_B and Y and is a function of $\dot{V}_r$.

3.3 FILM THICKNESS UNIFORMITY AND PURITY

3.3.1 Deposition Geometry

Deposition of thin films involves consideration of both the source of evaporant atoms and the substrates upon which they impinge. In this section aspects of the deposition geometry including the characteristics of evaporation sources and the orientation and placement of substrates will be discussed. Source–substrate geometry, in turn, influences film uniformity, a concern of paramount importance that will be treated subsequently. Evaporation from a *point source* is the simplest of situations to model. In this case evaporant particles are imagined to originate from an infinitesimally

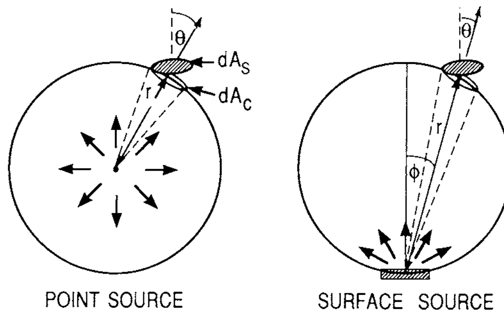


Figure 3-4 Evaporation from (a) point source, (b) surface source.

small region (dA_e) of a spherical source of surface area A_e with a uniform mass evaporation rate as shown in Fig. 3-4a. The total evaporated mass $\bar{M}_e$ is then given by the double integral

$$\bar{M}_e = \int_0^t \int_{A_e} \Gamma_e \, dA_e \, dt. \tag{3-15}$$

Of this amount, mass $d\bar{M}_s$ falls on the substrate of area dA_s. Since the projected area dA_s on the surface of the sphere is dA_c, with $dA_c = dA_s \cos\theta$, the proportionality holds that $d\bar{M}_s : \bar{M}_e = dA_c : 4\pi r^2$. Finally,

$$\frac{d\bar{M}_s}{dA_s} = \frac{\bar{M}_e \cos\theta}{4\pi r^2} \tag{3-16}$$

is obtained where θ is the angle between the vector from the origin to the planar substrate and the vector representing the substrate normal. On a per-unit time basis we speak of the film deposition rate $\dot{R}$ (atoms/cm^2-s), a term that has the same units as Φ and is referred to often in this book. The deposition varies with the geometric orientation of the substrate and with the inverse square of the source–substrate distance. Substrates placed tangent to the surface of the receiving sphere would be coated uniformly irrespective of placement since $\theta = 0$ and $\cos\theta = 1$.

An evaporation source employed in the pioneering research by Knudsen made use of an isothermal enclosure with a very small opening through which the evaporant atoms or molecules effused. These effusion or Knudsen cells are frequently employed in molecular-beam epitaxy deposition systems where precise control of evaporation variables is required. Kinetic theory predicts that the molecular flow of the vapor through the hole is directed according to a cosine distribution law, and this has been

verified experimentally. The mass deposited per unit area is therefore given by

$$\frac{d\bar{M}_s}{dA_s} = \frac{\bar{M}_e \cos\phi \cos\theta}{\pi r^2} \tag{3-17}$$

and is now dependent on two angles that are defined in Fig. 3-4b. These correspond to the evaporant emission angle ϕ and the deposition or receiving angle θ. Evaporation from an extended area or *surface source* is also modeled by Eq. 3-17. Physically, the extended source is a superposition of many point sources that stongly contribute to the vapor stream where ϕ is close to zero degrees, accounting for the vertically directed lobe-shaped emission; however, in the $\phi = 90^\circ$ direction there is no emission. Boat filaments and wide crucibles containing a pool of molten material to be evaporated approximate surface sources in practice.

From careful measurements of the angular distribution of film thickness, it has been found that rather than a $\cos\phi$ dependence, a $\cos^n\phi$ evaporation law is more realistic in many cases. As shown in Fig. 3-5, n is a number that determines the geometry of the lobe-shaped vapor cloud and the angular distribution of evaporant flux from such sources. When n is large, the vapor flux is highly directed. Physically n is related to the evaporation crucible geometry and scales directly with the ratio of the melt depth (below top of crucible) to the melt surface area. Deep, narrow crucibles with large n have been employed to confine evaporated radioactive materials to a narrow angular spread in order to minimize chamber contamination. The corresponding deposition equation for such sources is

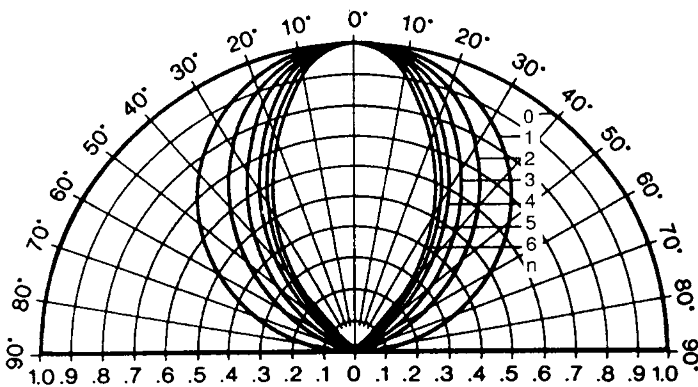


Figure 3-5 Calculated lobe-shaped vapor clouds with various cosine exponents. (From Ref. 9.)

(Ref. 9)

$$\frac{d\bar{M}_s}{dA_s} = \frac{\bar{M}_e(n+1)\cos^n\phi\cos\theta}{2\pi r^2}, \qquad n \geqslant 0. \tag{3-18}$$

As the source becomes increasingly directional, the receiving surface area effectively exposed to evaporant shrinks (i.e., $4\pi r^2$, πr^2, and $2\pi r^2/(n+1)$ for point, $\cos\phi$, and $\cos^n\phi$ sources, respectively).

3.3.2 FILM THICKNESS UNIFORMITY

Maintaining thin-film thickness uniformity is always desirable, but not necessarily required; yet it is absolutely essential for microelectronic and many optical coating applications. For example, thin-film, narrow-band optical interference filters require a thickness uniformity of $\pm 1\%$. This poses a problem particularly if there are many components to be coated or the surfaces involved are large or curved. Utilizing formulas developed in the previous section, the thickness distribution can be calculated for a variety of important source–substrate geometries. Consider evaporation from point and surface sources onto a parallel plane-receiving substrate surface as indicated in the insert of Fig. 3-6. The film thickness d is given by $d\bar{M}_s/\rho\, dA_s$

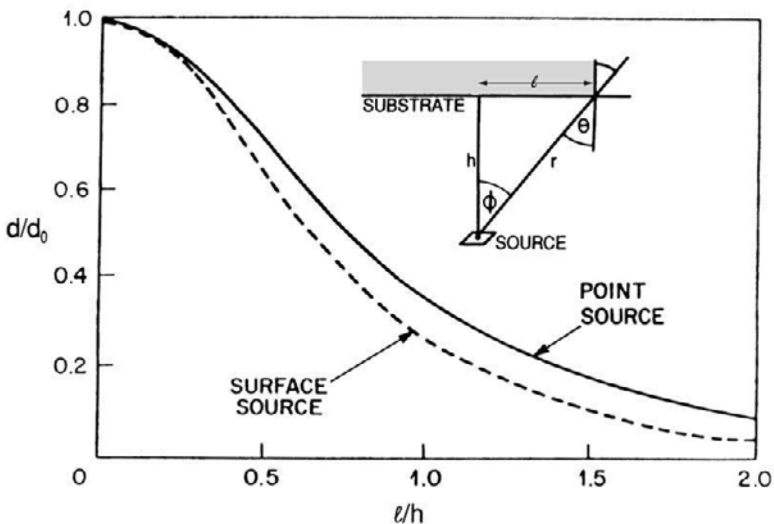


Figure 3-6 Film thickness uniformity for point and surface sources. (Insert) Geometry of evaporation onto parallel plane substrate.

where ρ is the density of the deposit. For the *point* source (Eq. 3-16)

$$d = \frac{\bar{M}_e \cos\theta}{4\pi\rho r^2} = \frac{\bar{M}_e h}{4\pi\rho r^3} = \frac{\bar{M}_e h}{4\pi\rho(h^2 + l^2)^{3/2}}, \tag{3-19}$$

after noting that $r = (h^2 + l^2)^{1/2}$. The thickest deposit (d_0) occurs at $l = 0$ in which case $d_0 = \bar{M}_e / 4\pi\rho h^2$ and thus,

$$\frac{d}{d_0} = \frac{1}{\{1 + (l/h)^2\}^{3/2}}. \tag{3-20}$$

Similarly, for the *surface* source

$$d = \frac{\bar{M}_e \cos\theta \cos\phi}{\pi\rho r^2} = \frac{\bar{M}_e}{\pi\rho r^2}\frac{h}{r}\frac{h}{r} = \frac{\bar{M}_e h^2}{\pi\rho(h^2 + l^2)^2} \tag{3-21}$$

since $\cos\theta = \cos\phi = h/r$. When normalized to the thickest dimensions or $d_0 = \bar{M}_e/\pi\rho h^2$,

$$\frac{d}{d_0} = \frac{1}{\{1 + (l/h)^2\}^2}. \tag{3-22}$$

A comparison of Eqs. 3-20 and 3-22 is made in Fig. 3-6, where it is apparent that less thickness uniformity can be expected with the surface source.

A couple of practical examples (Ref. 10) will demonstrate how these film-thickness distributions are used in designing source–substrate geometries for coating applications. In the first example suppose it is desired to coat a 150-cm-wide strip utilizing two evaporation sources oriented as shown in the insert of Fig. 3-7. If a thickness tolerance of $\pm 10\%$ is required, what should the distance between sources be and how far should they be located from the substrate? A superposition of solutions for two individual surface sources (Eq. 3-22) gives the thickness variation shown graphically in Fig. 3-7 as a function of the relative distance r from the center line for various values of the source spacing, D. All pertinent variables are in terms of dimensionless ratios r/h_v and D/h_v. The desired tolerance requires that d/d_0 stay between 0.9 and 1.1, and this can be achieved with $D/h_v = 0.6$ yielding a maximum value of $r/h_v = 0.87$. Since $r = 150/2 = 75$ cm, $h_v = 75/0.87 = 86.2$ cm. Therefore, the required distance between sources is $2D = 2 \times 0.6 \times 86.2 = 103.4$ cm. Other solutions exist but we are seeking the *minimum* value of h_v. It is obvious that the uniformity tolerance can always be realized by extending the source–substrate distance, but this is wasteful of evaporant.

As a second example consider a composite optical coating where a $\pm 1\%$ film thickness variation is required in each layer. The substrate is rotated to even out source distribution anomalies and minimize preferential film

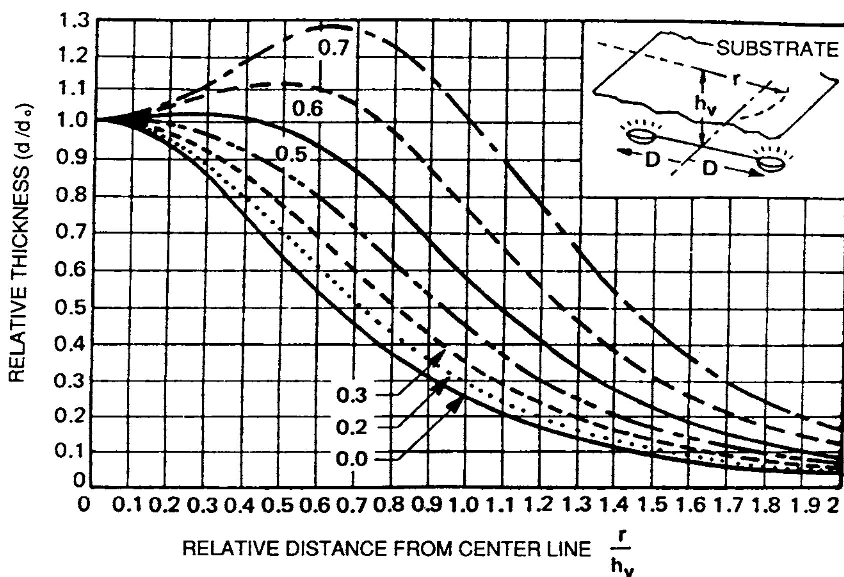


Figure 3-7 Film thickness uniformity across a strip employing two evaporation sources for various values of D/h_v. (From Ref. 10.)

growth, which can adversely affect coating durability and optical properties. Since multiple films of different composition will be sequentially deposited, the necessary fixturing requires that the sources be offset from the axis of rotation by a distance $R = 20$ cm. How high above any given source should a 25 cm diameter substrate be rotated to maintain the desired film tolerance? The film thickness distribution in this case is a complex function of the three-dimensional geometry that, fortunately, has been graphed in Fig. 3-8. Reference to this figure indicates that the curve $h_v/R = 1.33$ in conjunction with $r/R = 0.6$ will generate a thickness deviation ranging from about -0.6 to $+0.5\%$. On this basis, the required distance is $h_v = 1.33 \times 20 = 26.6$ cm.

A clever way to achieve thickness uniformity, however, is to locate both the surface evaporant source and the substrates on the surface of a sphere as shown in Fig. 3-9. In this case the isosceles triangle defining the deposition geometry means that $\theta = \phi$, and $\cos\theta = \cos\phi = r/2r_0$. Therefore, Eq. 3-17 becomes

$$\frac{d\bar{M}_s}{dA_s} = \frac{\bar{M}_e}{\pi r^2}\frac{r}{2r_0}\frac{r}{2r_0} = \frac{\bar{M}_e}{4\pi r_0^2}. \qquad (3\text{-}23)$$

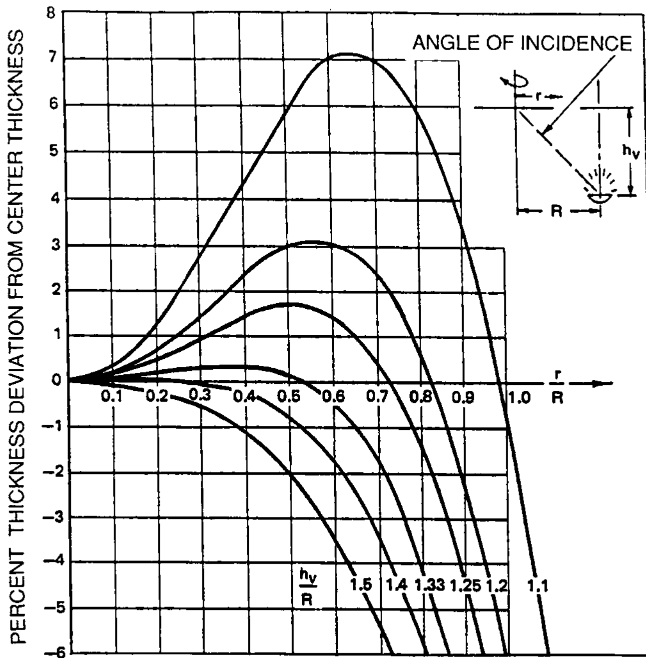


Figure 3-8 Calculated film thickness variation across the radius of a rotating disk. (From Ref. 10.)

The resultant deposit thickness is a constant clearly independent of angle. Use is made of this principle in planetary fixtures that hold circular substrates to be coated with metal (metallized) by evaporation. To further promote uniform coverage the planetary fixture is rotated during deposition. Physically, deposition thickness uniformity is achieved because short source–substrate distances are offset by unfavorably large vapor emission and deposition angles. Alternatively, long source–substrate distances are compensated by correspondingly small emission and reception angles. Uniformity of columnar grain microstructure, e.g., tilt, is not preserved, however, because of variable flux incidence angle (see Section 9.2.2). For sources with a higher degree of directionality (i.e., where $\cos^n \phi$ rather than $\cos \phi$ is involved), the reader can easily show that thickness uniformity is no longer maintained.

Two principal methods for optimizing film uniformity over large areas involve varying the geometric location of the source and interposing static as well as rotating shutters between evaporation sources and substrates.

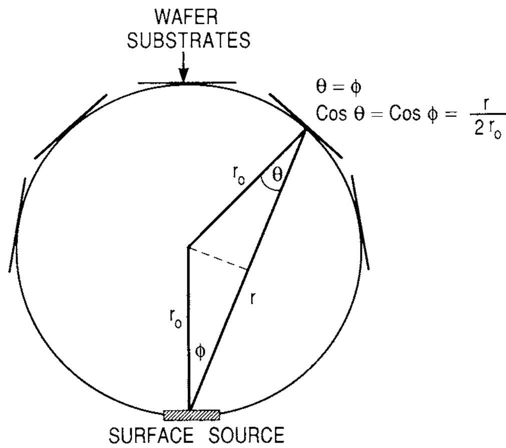


Figure 3-9 Evaporation scheme to achieve uniform deposition. Source and substrates lie on sphere surface of radius r_0.

Computer calculations have proved useful in locating sources and designing shutter contours to meet the stringent demands of optical coatings. Film-thickness uniformity cannot, however, be maintained beyond $\pm 1\%$ because of insufficient mechanical stability of both the stationary and rotating hardware.

In addition to the parallel source–substrate configuration, calculations of thickness distributions have also been made for spherical as well as conical, parabolic, and hyperbolic substrate surfaces (Ref. 9). Similarly, cylindrical, wire, and ring evaporation source geometries have been treated (Ref. 11).

3.3.3 Conformal Coverage of Steps and Trenches

An important issue related to film uniformity is the conformal coverage of nonplanar substrate features. Such situations arise primarily in the fabrication of integrated circuits where semiconductor contact films, interconnection metallizations, and intervening dielectric films are deposited over a terrain of intricate topography where steps, holes, and trenches abound. When a film of the same thickness coats the horizontal as well as vertical surfaces of substrates, we speak of conformal coverage. On the other hand, coverage will not be uniform when physical shadowing effects cause unequal deposition on the top and side walls of steps. Inadequate step coverage can lead to minute cracks in metallizations and has been shown to be a source

of failure in device reliability testing. Such thinned regions on conducting stripes exhibit greater Joule heating, which sometimes fosters early burnout. Step-coverage problems have been shown to be related to the profile of the substrate step as well as to the evaporation source–substrate geometry. The simplest model of evaporation from a point source onto a stepped substrate results in either conformal coverage or a lack of deposition in the step shadow as shown schematically in Fig. 3-10 (top). Line-of-sight motion of evaporant atoms and sticking coefficients of unity can be assumed in estimating the extent of coverage.

Two important needs of integrated-circuit metallization technology are to conformally line deep dielectric substrate trenches, contact holes, and

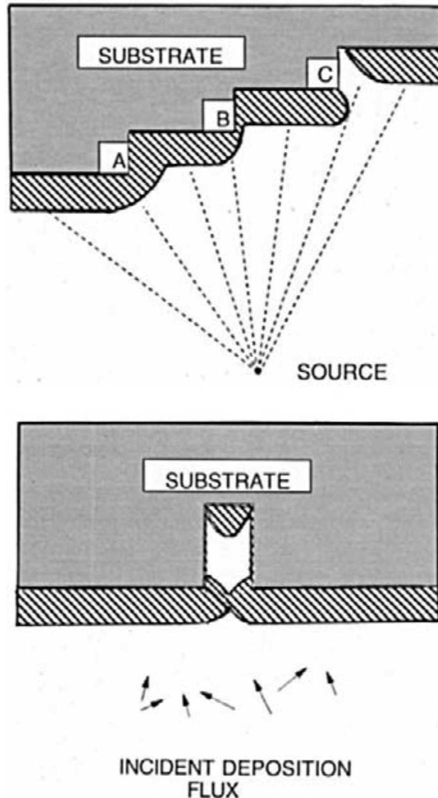


Figure 3-10 (top) Schematic illustration of film coverage of stepped substrate. A, Uniform coverage; B, poor sidewall coverage; C, lack of coverage — discontinuous film. (bottom) Formation of a keyhole void in a high aspect ratio via during sputtering. (From Ref. 12.)

vias, as well as completely fill them with metal (Ref. 13). Even though sputtering and chemical vapor deposition processes are largely employed for such purposes, some of the issues involved also apply to evaporated films. The trenches to be coated have aspect ratios AR (depth to width) greater than unity and are typically a few hundred nanometers in diameter and several times deeper. Achieving conformal coverage and filling of deep narrow channels is a particular challenge particularly if atoms deposit from multiple sources or directions because of scattering from gases (as in sputtering). In such cases arriving atoms coat the channel orifice preferentially as shown in Fig. 3-10 (bottom). The resulting film overhang shadows the deeper recesses of the trench so that while material deposits on the bottom, very little accumulates on the side walls. A "breadloaf" film topography evolves that tends to choke off further deposition in the trench. As a consequence a void may be trapped within, leading to a defective "keyhole" structure. Collimation of the arriving atomic flux and heated substrates favor deeper and more conformal trench penetration, the former by minimizing shadowing and the latter by promoting surface and bulk diffusion of atoms.

Computer modeling of step coverage has been performed for the case in which the substrate is located on a rotating planetary holder (Ref. 13). In Fig. 3-11 coverage of a 1 μm wide, 1 μm high square channel test pattern with 5000 Å of evaporated Al is simulated. For the symmetric orientation (left) the region between the pattern stripes always manages to "see" the source and this results in a small plateau of the full film thickness at the channel bottom. In the asymmetric orientation (right), however, the substrate stripes cast a shadow with respect to the source, biasing the deposition in favor of unequal sidewall coverage. Comparisons with experimentally deposited films are generally in good agreement with the computer models. In generating the simulated film profiles surface migration of atoms was neglected, which is a valid assumption at low substrate temperatures. Heating the substrate increases surface diffusion of depositing atoms, thus promoting coverage by filling potential voids as they form. We shall return to trench lining and filling issues subsequently, particularly by considering strategies to accomplish coverage during sputtering (Chapter 5), and by examining the microstructure of such deposits (Chapter 9).

3.3.4 FILM PURITY

The chemical purity of evaporated films is dependent on the nature and level of impurities that (1) are initially present in the source, (2) contaminate the source from the heater, crucible, or support materials, and (3) originate

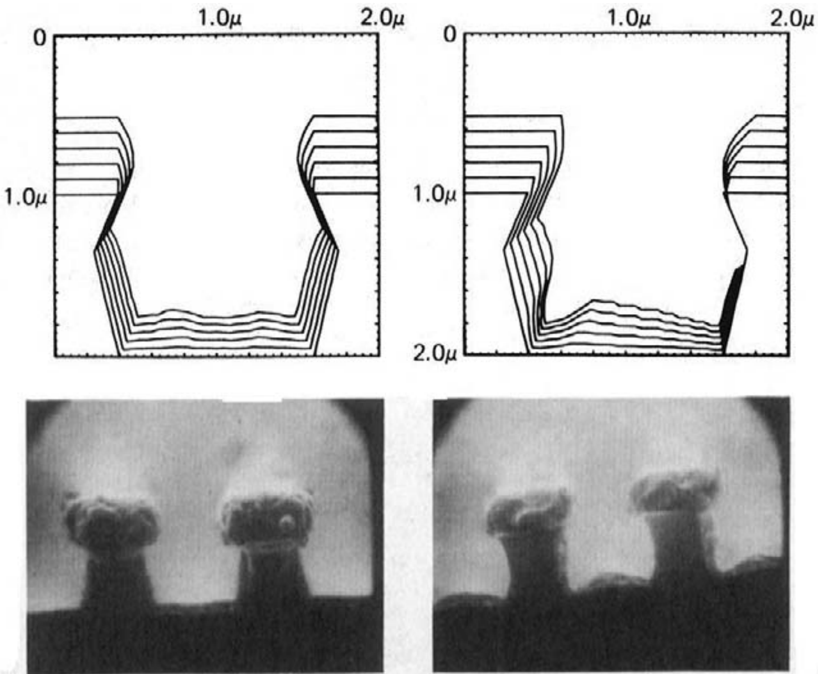


Figure 3-11 Comparison of simulated and experimental Al film coverage of 1 μm line step and trench features. (Left) Orientation of most symmetric deposition. (Right) Orientation of most asymmetric deposition. [Reprinted with permission from Cowan Publishing Co., from C. H. Ting and A. R. Neureuther, *Solid State Technol.* **25**(2), 115 (1982).]

from the residual gases present in the vacuum system. In this section only the effect of residual gases on film purity will be addressed. During deposition the atoms and molecules of both the evaporant and residual gases impinge on the substrate in parallel, independent events. Dimensional analysis shows that the evaporant vapor impingement rate is $\rho N_A \dot{d}/M_a$ atoms/cm^2-s where ρ is the film density and $\dot{d}$ is the deposition rate (cm/s). Simultaneously, gas molecules impinge at a rate given by Eq. 2-9. The ratio of the latter to former impingement rate is the gas impurity concentration C_i or

$$C_i = \frac{5.82 \times 10^{-2} P M_a}{(M_g T)^{1/2} \rho \dot{d}}. \tag{3-24}$$

Terms M_a and M_g refer to evaporant and gas molecular weights, respectively, and P is the residual gas vapor pressure in torr.

Table 3-2
Maximum Oxygen Concentration in Tin Films Deposited at Room Temperature

	Deposition rate (Å/s)			
P_{O_2} (torr)	1	10	100	1000
10^{-9}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-7}	10^{-1}	10^{-2}	10^{-3}	10^{-4}
10^{-5}	10	1	10^{-1}	10^{-2}
10^{-3}	10^{3}	10^{2}	10	1

From Ref. 14.

Table 3-2 illustrates the combined role that deposition rate and residual gas pressure play in determining the oxygen level that can be incorporated into tin films (Ref. 14). Although the concentrations are probably overestimated because the sticking probability of O_2 is of the order of 0.1 or less, the results have several important implications. In order to produce very pure films, it is important to deposit at very high rates while maintaining very low background pressures of residual gases such as H_2O, CO_2, CO, O_2, and N_2. Neither of these requirements is too formidable for vacuum evaporation where deposition rates from electron beam sources can reach 1000 Å/s at chamber pressures of $\sim 10^{-8}$ torr.

On the other hand, in sputtering processes, discussed in Chapter 5, deposition rates are typically about two orders of magnitude lower and chamber pressures four orders of magnitude higher than for evaporation. Therefore, the potential exists for producing films containing high gas concentrations. For this reason sputtering was traditionally not considered to be as "clean" a process as evaporation. Considerable progress has been made in the past two decades, however, with the commercial development of high-deposition-rate magnetron sputtering systems, operating at somewhat lower gas pressures in cleaner vacuum systems. In the case of aluminum films, comparable purities appear to be attained in both processes. Lastly, Table 3-2 suggests that very high oxygen incorporation occurs at residual gas pressures of 10^{-3} torr. Advantage of this fact is taken in reactive evaporation processes where intentionally introduced oxygen serves to promote reactions with the evaporant metal in the deposition of oxide films.

The presence of oxygen and nitrogen impurities within pure metal films sometimes has a pronounced effect in degrading electrical conductivity and optical reflectivity as well as other properties, e.g., hardness.

3.4 EVAPORATION HARDWARE

3.4.1 ELECTRICALLY HEATED EVAPORATION SOURCES

This section describes some of the hardware and techniques used to electrically heat sources for the efficient evaporation of thin films. Discussed are the widely used resistance, induction, and electron-beam heating methods. The overwhelming bulk of evaporated thin films deposited commercially for electrically, optically, and mechanically functional applications are deposited by these methods or variants of them. The first sources used to heat evaporants relied on the Joule heating of metal filaments. Clearly, such heaters must reach the temperature of the evaporant in question while having a negligible vapor pressure in comparison. Ideally, they should not contaminate, react with, or alloy with the evaporant, or release gases such as oxygen, nitrogen, or hydrogen at the evaporation temperature. These requirements have led to the development and use of resistance-heated evaporation sources used singly or with inert oxide or ceramic-compound crucibles. Some of these are shown in Fig. 3-12. They can be divided into the following important categories.

3.4.1.1 Tungsten Wire Sources

These sources are in the form of individual or multiply stranded wires twisted into helical or conical shapes. Helical coils are used for metals that wet tungsten readily; the conical baskets are better adapted to contain poorly wetting materials. In the former case, metal evaporant wire is wrapped around or hung from the tungsten strands and the molten beads of metal are retained by surface tension forces. Tungsten filaments can be operated up to about 2200 K before they begin to fail rapidly.

3.4.1.2 Refractory Metal Sheet Sources

Tungsten, tantalum, and molybdenum sheet metal sources, like the wire filaments, are self-resistance heaters that require low-voltage, high-current power supplies. These sources have been fabricated into a variety of shapes including the dimpled strip, boat, canoe, and deep-folded configurations. Deep-folded boat sources have been used to evaporate MgF_2 and powder mixtures of metals and metal oxides for coating ophthalmic lenses in batch-type evaporators.

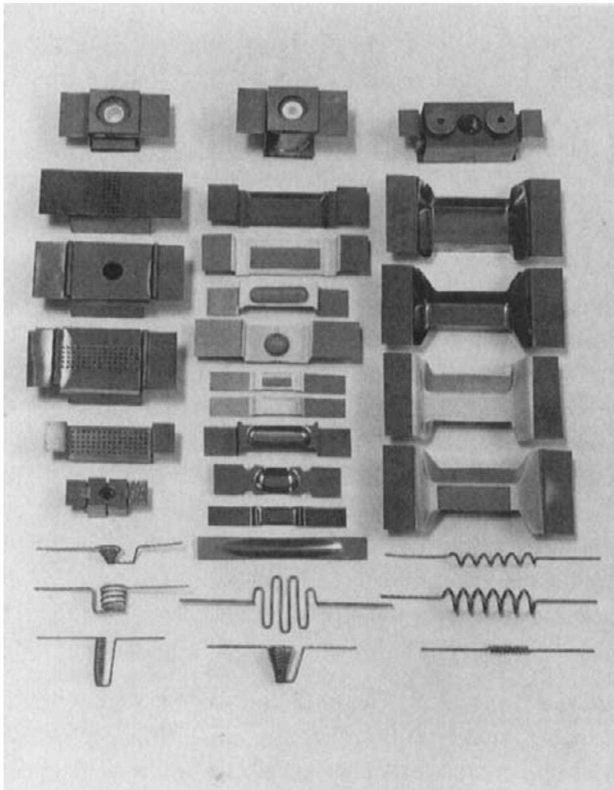


Figure 3-12 Assorted resistance heated evaporation sources. (Courtesy of R. D. Mathis Company.)

3.4.1.3 Sublimation Furnaces

In order to evaporate sulfides, selenides, and some oxides efficiently, sublimation furnaces are employed. The evaporant materials in powder form are pressed and sintered into pellets and heated by surrounding radiant heating sources. To avoid the spitting and ejection of particles caused by evolution of gases occluded within the source compacts, baffled heating assemblies are used. These avoid direct line-of-sight access to substrates, and evaporation rates from such sources tend to be constant over extended periods of time. The furnaces are typically constructed of sheet tantalum that is readily cut, bent, and spot welded to form heaters, radiation shields, supports, and current bus strips.

3.4.1.4 Crucible Sources

Among the common evaporant containers are cylindrical cups composed of oxides, pyrolytic BN, graphite, and refractory metals, fabricated by hot pressing powders or machining bar stock. These crucibles are normally heated by external tungsten-wire resistance heating elements wound to fit snugly around them.

Other crucible sources rely on high-frequency induction rather than resistance heating. In a configuration resembling a transformer, high-frequency currents are induced in either a conducting crucible or evaporant charge serving as the secondary, resulting in heating. The powered primary is a coil of water-cooled copper tubing that surrounds the crucible. As an example of induction heating, aluminum has been commercially evaporated from BN or BN/TiB_2 composite crucibles.

Another category of crucible source consists of a tungsten wire resistance heater in the form of a conical basket that is encased in Al_2O_3 or refractory oxide to form an integral crucible–heater assembly. Such crucibles frequently serve as evaporant sources in laboratory-scale film deposition systems.

3.4.1.5 Estimating the Temperature of Resistance Heaters

In the design of electrical heaters for evaporation systems it is important to estimate the temperature of heated filaments. Under simplifying assumptions it is possible to draw connections between the electrical power ($\mathscr{P}$) supplied and the filament temperature reached. We start by noting that $\mathscr{P}$ is simply given by i^2R or alternately by V^2/R where i, V, and R are the current, voltage, and resistance, respectively. For a wire filament of length L and cross-sectional area A_c, $\mathscr{P}$ can be written as

$$\mathscr{P} = i^2R = i^2\rho(0)[T/T(0)]^n L/A_c. \tag{3-25}$$

This equation assumes that the electrical resistivity, $\rho(T)$, of metals can be approximated over a broad high-temperature (T) range by the relation $\rho(T) = \rho(0)[T/T(0)]^n$, where $\rho(0)$ is the value at reference temperature $T(0)$, and n is a constant generally close to 1. In tungsten, for example, $\rho(0) = 5.5 \times 10^{-8}\,\Omega$-m, $T(0) = 293$ K, and $n = 1.20$. It is therefore apparent that T can be calculated once the resistor dimensions and power delivered are known. In a complementary approach, T may be estimated assuming that all of the input electrical power dissipated is thermally radiated from the filament surface. In this case the Stefan–Boltzmann law yields the defining equation for the radiated power ($\mathscr{P}_r$),

$$\mathscr{P}_r = \varepsilon\sigma A_s(T^4 - T(0)^4), \tag{3-26}$$

where ε is the emissivity, σ is Stefan's constant ($\sigma = 5.67 \times 10^{-8}$ W/m^2-K^4), and A_s is the filament surface area. It should be noted that both Eqs. 3-25 and 3-26 are primarily applicable to straight rather than coiled filaments whose complex geometries promote concentrated heating effects and higher filament temperatures for the same power level.

3.4.2 Electron-Beam Evaporation

3.4.2.1 Hardware and Process Environment

Disadvantages of resistively heated evaporation sources include contamination by crucibles, heaters, and support materials and the limitation of relatively low input power levels. This makes it difficult to deposit pure films or evaporate high-melting-point materials at appreciable rates. Electron-beam (e-beam) heating eliminates these disadvantages and has, therefore, become the preferred vacuum evaporation technique for depositing films. In principle, this type of source enables evaporation of virtually all materials over a wide range of practical rates. As indicated in Fig. 3-13, the evaporant charge is placed in either a water-cooled crucible or in the depression of a water-cooled copper hearth. The purity of the evaporant is assured because only a small amount of charge melts or sublimes so that the effective crucible is the unmelted skull material next to the cooled hearth. For this reason there is no contamination of the evaporant by Cu. Multiple source units are available for either sequential or parallel deposition of more than one material.

In the most common configuration of the gun source, electrons are thermionically emitted from heated filaments that are shielded from direct line of sight of both the evaporant charge and substrate. Film contamination from the heated cathode filament is eliminated this way. The cathode potential is biased negatively with respect to a nearby grounded anode by anywhere from 4 to 20 kilovolts, and this serves to accelerate the electrons. In addition, a transverse magnetic field is applied that serves to deflect the electron beam in a 270° circular arc and focus it on the hearth and evaporant charge at ground potential. The reader can verify the electron trajectory through the use of the left-hand rule. This states that if the thumb is in the direction of the initial electron emission and the forefinger lies in the direction of the magnetic field (north to south), then the middle finger indicates the direction of the force on the electron and its resultant path at any instant.

At the higher evaporation rates, including those usually employed in practice, the vapor just above the hearth approximates a high-pressure

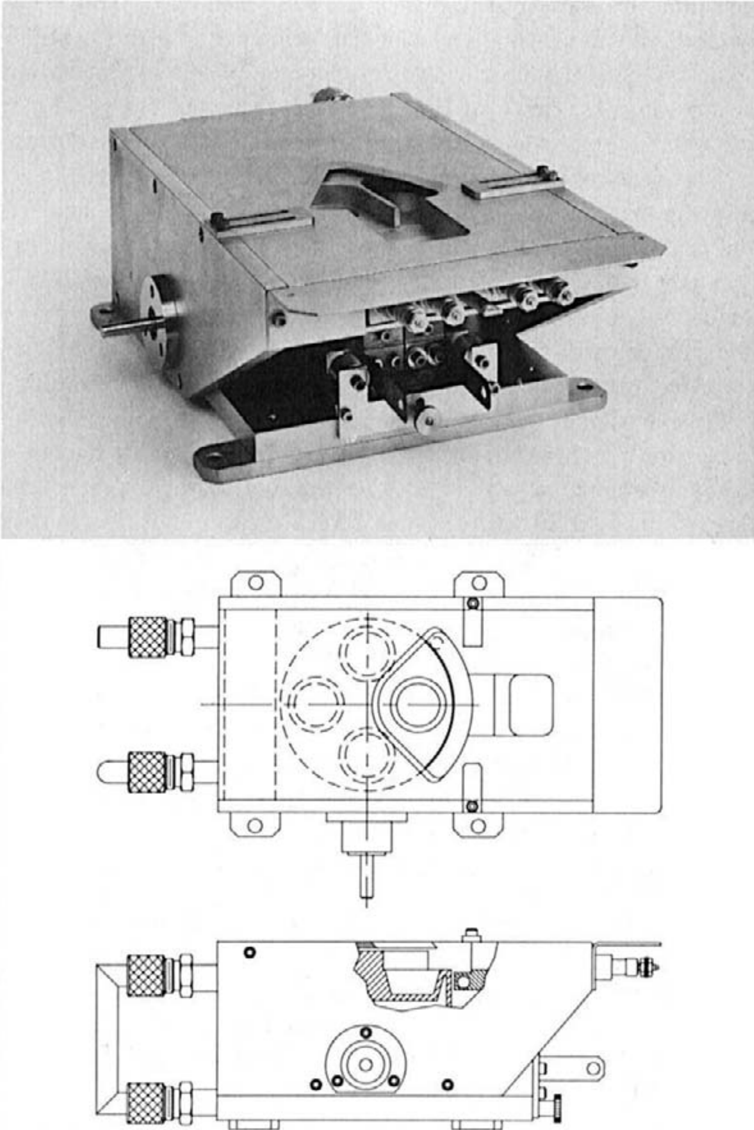


Figure 3-13 Multihearth electron-beam evaporation unit with accompanying top and side view schematics. Courtesy of Temescal unit of Edwards High Vacuum International, a division of the BOC Group, Inc.

viscous cloud of very hot evaporant. The complex energy transfer between electronic excitation and translational motion of vapor atoms in this region, and its effect on flow to the substrate, has been modeled for e-beam evaporated titanium (Ref. 15). As suggested by Fig. 3-14, the region beyond this dense cloud is at much lower pressure and so we may assume molecular flow prevails. Thus, instead of evaporant particles being beamed from various points on the flat source surface, they appear to originate from the perimeter of the viscous cloud. In comparison to the previously considered evaporation geometries the effective or *virtual* source plane has moved away from the melt surface toward the substrate. This is why the source–substrate distance in Figs. 3-7 and 3-8 that must be used in all calculations is the virtual distance h_v. The ratio h/h_v, where h is the actual source–substrate distance, depends on the evaporation rate and a value of 0.7 is not unusual.

Two problems associated with electron-beam sources are "beam curling" and nonuniform beam density. If the magnetic field is improperly designed,

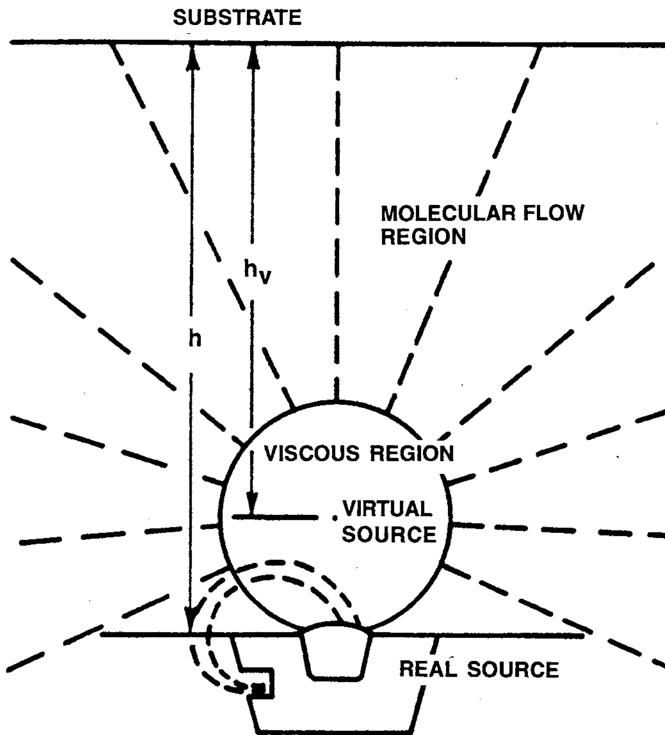


Figure 3-14 Schematic depiction of the regions of viscous and molecular flow around an electron-beam evaporation source. (From *Physical Vapor Deposition*, edited by R. J. Hill. Temescal, BOC Group, 1986. Reprinted with the permission of Russell J. Hill.)

the electron beam may not impinge on the charge surface normally but at some angle to it so that the electron trajectory curls. As a result the emitted vapor distribution shifts with time, and films with unpredictable and variable thicknesses deposit. To optimize evaporation conditions, provision is made for altering the size of the focal spot and for electromagnetically scanning the beam. This prevents spattering of liquid and deep drilling or tunneling into sources that sublime. Sweeping the beam also minimizes the problem of evaporating only a small amount of material before the bottom of the crucible is reached; in this way there is better utilization of materials.

3.4.2.2 Thermal Power and Heating Effects

It is instructive to estimate the total power that must be delivered by the electron beam to the charge in order to compensate for the following heat losses incurred during evaporation of 10^{18} atoms/cm^2-s (Ref. 16).

1. The power density $\bar{\mathscr{P}}_S$ (W/cm^2) that must be supplied to account for the heat of sublimation ΔH_S (eV) is

$$\bar{\mathscr{P}}_S = 10^{18}(1.6 \times 10^{-19})\Delta H_S = 0.16\Delta H_S. \tag{3-27a}$$

2. The kinetic energy of evaporant is $\frac{3}{2}k_B T_S$ per atom so that the required power density, $\bar{\mathscr{P}}_k$, is

$$\bar{\mathscr{P}}_k = 10^{18}(\tfrac{3}{2})(1.38 \times 10^{-23})T_S = 2.07 \times 10^{-5} T_S; \tag{3-27b}$$

where T_S is the source temperature.

3. The radiation heat loss density is

$$\bar{\mathscr{P}}_r = 5.67 \times 10^{-12}\varepsilon(T_S^4 - T_0^4) \tag{3-27c}$$

where ε is the source emissivity at T_S, and $T_0 = 293$ K.

4. Heat conduction through a charge of thickness l into the hearth dissipates a power density, $\bar{\mathscr{P}}_c$, equal to

$$\bar{\mathscr{P}}_c = \kappa\frac{T_S - T_0}{l} \tag{3-27d}$$

where κ is the thermal conductivity of the charge. For the case of Au at $T_S = 1670$ K where $\Delta H_S = 3.5$ eV, $\varepsilon \sim 0.4$, and $l = 1$ cm and $\kappa = 3.1$ W/cm-K, the corresponding values are $\bar{\mathscr{P}}_S = 0.56$ W/cm^2, $\bar{\mathscr{P}}_k = 0.034$ W/cm^2, $\bar{\mathscr{P}}_r = 17.6$ W/cm^2, and $\bar{\mathscr{P}}_c = 4.3$ kW/cm^2. Clearly the overwhelming proportion of the power delivered by the electron beam is conducted through the charge to the hearth. In actuality, power densities of ~ 10 kW/cm^2 are utilized in

melting metals, but such levels would damage dielectrics that require perhaps only 1–2 kW/cm^2. A more practical measure of the energy required for e-beam evaporation is given in units of kW-h/kg. For example, it has been observed that 2.35 kg of a Ti–6Al–4V alloy evaporates each hour for an input power level of 70 kW; thus the specific energy is 29.8 kW-h/kg. Greater heating efficiency is possible by using refractory, thermally insulated liners that reduce the heat flow into the hearth. However, the greater risk of melt contamination is a disadvantage of liners.

Another way to view the energetics of evaporation is to consider heat losses from sources as they are powered (Ref. 10). At low energy-densities and low temperatures, heat loss occurs through conduction and convection in the evaporant. As the energy density and temperature increase, heat loss through radiation and vaporization become more significant. The loss due to radiation, which varies as the fourth power of temperature (Eq. 3-26), is eventually overtaken by that due to the vaporization rate, which varies exponentially with temperature (Eqs. 3-3, 3-7). Thus every material has a temperature where energy transfer by radiation equals that by evaporation. For example, copper with a melting point of 1357 K has a crossover temperature of 1789 K, while chromium with a melting point of 2176 K has a crossover temperature of 1885 K. The fact that Cr sublimes when heated, while Cu must be molten prior to evaporation reflects the opposing trends in crossover temperature for these metals.

3.4.2.3 Modification of Films

Lastly we briefly consider two additional issues related to the modification of film properties induced by e-beam evaporation. The first concerns evaporant ion bombardment of the growing film. Any time electron beams possessing energies of 5–10 keV impact neutral vapor atoms or molecules we may be certain that ionization will occur. This is true of the evaporant atoms, and they acquire a positive charge upon ionization (see Eq. 4-1); typically, the degree of ionization ranges from 0.01 to 0.1. Compared to thermal energies ($\sim$0.1 eV) of evaporant atoms, such ions are perhaps 100-fold more energetic. If the substrate is negatively biased these ions will bombard it and promote impurity removal, film adhesion, and densification. Such beneficial outcomes also arise in assorted plasma-based deposition processes discussed in Chapter 5.

The second issue of interest is the generation of characteristic X-rays from atoms in the evaporant source. This stems from the fact that the electron-beam gun has all of the features of an X-ray tube, i.e., filament for

thermionic emission, anode target (source), high DC voltage supply, and vacuum ambient. The emitted X-rays of relatively low energy ($\sim$10 keV or less) do not generally pose a health hazard because they are too soft to penetrate the chamber walls. But they do impinge on the growing film and while metals are unaffected, sensitive dielectrics can be damaged by X-rays. For example, electronic defects, e.g., broken bonds and electron traps, can be generated in thin gate oxide (SiO_2) films of field-effect transistors as a result of such irradiation.

3.4.3 Deposition Techniques

By now, films of virtually all important materials have been prepared by physical vapor deposition techniques. A practical summary (Refs. 17, 18) of vacuum evaporation methods is given in Table 3-3, where recommended heating sources and crucible materials are listed for a number of metals, alloys, oxides, and compounds. The data on electron-beam evaporated materials from Ref. 17 are noteworthy and invite comparisons because all film depositions were made in a single system (18-inch bell jar), with a constant source–substrate distance (40 cm), a common base pressure (less than 5×10^{-6} torr), and with the same e-gun and power supply (6 kW). Such a system is schematically depicted in Fig. 2-12 and would additionally contain shutters and a film thickness monitor (Section 10.2.5.2) positioned close to the substrate. For evaporation of optical film materials, i.e., oxides, fluorides, sulfides, selenides, and tellurides, the reader will find the data and information listed in Ref. 18 useful. Prior to settling on a particular vapor phase deposition process, both PVD and CVD options together with the numerous hybrid variants of these methods should be weighed. Paramount attention should be paid to film quality and properties and to the equipment costs necessary to achieve them. If, after all, vacuum evaporation is selected, modestly equipped laboratories may wish to consider the resistively heated sources before the more costly electron beam or induction heating alternatives.

The structure of a chromium film, electron-beam evaporated under conditions comparable to those noted above, is shown in Fig. 3-15. Irrespective of class of material, whether deposited by PVD or CVD methods, or deposition temperature (within limits), it is quite remarkable that a columnar-like film structure prevails. Physical reasons for the structural and morphological evolution of depositing films will be subsequently addressed in later chapters. Until then it should be borne in mind that film grains produced by the evaporation processes we now consider will for the most part usually possess a columnar morphology.

Table 3-3

Evaporation Characteristics of Materials

Material	Minimum evaporation temperature[a]	State of evaporation	Recommended crucible material	Deposition rate (Å/s)	Power (kW) (e-beam[b])
Aluminum	1010	Melts	BN	20	5
Al_2O_3	1325	Semimelts		10	0.5
Antimony	425	Melts	BN, Al_2O_3	50	0.5
Arsenic	210	Sublimes	Al_2O_3	100	0.1
Beryllium	1000	Melts	Graphite, BeO	100	1.5
BeO		Melts		40	1.0
Boron	1800	Melts	Graphite, WC	10	1.5
BC		Semimelts		35	1.0
Cadmium	180	Melts	Al_2O_3, quartz	30	0.3
CdS	250	Sublimes	Graphite	10	0.2
CaF_2		Semimelts		30	0.05
Carbon	2140	Sublimes		30	1.0
Chromium	1157	Sublimes	W	15	0.3
Cobalt	1200	Melts	Al_2O_3, BeO	20	2.0
Copper	1017	Melts	Graphite, Al_2O_3	50	0.2
Gallium	907	Melts	Al_2O_3, graphite		
Germanium	1167	Melts	Graphite	25	3.0
Gold	1132	Melts	Al_2O_3, BN	30	6.0
Indium	742	Melts	Al_2O_3	100	0.1
Iron	1180	Melts	Al_2O_3, BeO	50	2.5
Lead	497	Melts	Al_2O_3	30	0.1
LiF	1180	Melts	Mo, W	10	0.15
Magnesium	327	Sublimes	Graphite	100	0.04
MgF_2	1540	Semimelts	Al_2O_3	30	0.01
Molybdenum	2117	Melts		40	4.0
Nickel	1262	Melts	Al_2O_3	25	2.0
Permalloy	1300	Melts	Al_2O_3	30	2.0
Platinum	1747	Melts	Graphite	20	4.0
Silicon	1337	Melts	BeO	15	0.15
SiO_2	850	Semimelts	Ta	20	0.7
SiO	600	Sublimes	Ta	20	0.1
Tantalum	2590	Semimelts		100	5.0
Tin	997	Melts	Al_2O_3, graphite	10	2.0
Titanium	1453	Melts		20	1.5
TiO_2	1300	Melts	W	10	1.0
Tungsten	2757	Melts		20	5.5
Zinc	250	Sublimes	Al_2O_3	50	0.25
ZnSe	660	Sublimes	Quartz		
ZnS	300	Sublimes	Mo		
Zirconium	1987	Melts	W	20	5.0

[a] Temperature (°C) at which vapor pressure is 10^{-4} torr.
[b] For 10 kV, copper herth, source–substrate distance of 40 cm.
Adapted from Refs. 17 and 18.

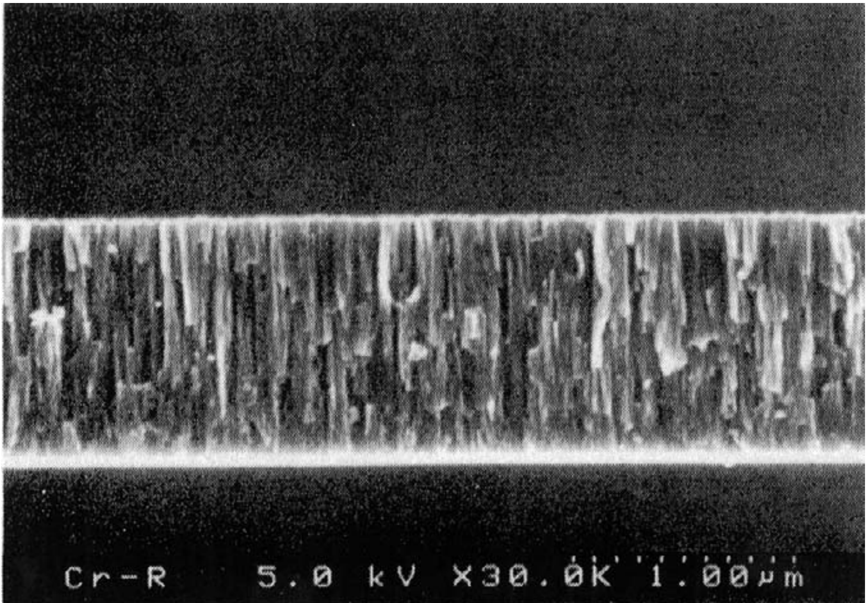


Figure 3-15 Cross-sectional scanning electron micrograph of a 1-μm-thick electron-beam evaporated Cr film. The substrate was rotated effectively ensuring normal vapor incidence. (From N. Kuratani, A. Ebe, and K. Ogata, *J. Vac. Sci. Technol.* **A19**, 153 (2001), Fig. 7. Reprinted by permission.)

3.5 EVAPORATION PROCESSES AND APPLICATIONS

3.5.1 SCOPE

The chapter closes by focusing on a few rather different evaporation processes. Pulsed laser deposition (PLD), a new unconventional evaporation technique, is the first. Although it has been largely limited to laboratory investigations of small-area films, it has the very desirable capability of producing stoichiometric multicomponent films. Because of its ability to deposit ceramic films it is given considerable attention here. In contrast, web coating is commercially used to coat extremely large substrate areas. Because large amounts of evaporant are required, the high deposition rates needed are generally provided by electron-beam heating sources. This section also introduces the relatively recent ion-assisted evaporation process

that has been widely adopted in optical coating applications. A discussion of hybrid evaporation techniques that additionally incorporate the use of plasmas, ion beams, and reactive gases is, however, deferred to Chapter 5, where the role of the latter can be better appreciated.

3.5.2 Pulsed Laser Deposition

3.5.2.1 Introduction

One of the newer techniques for depositing thin films makes use of the interaction of laser beams with material surfaces (Ref. 19). Lasers were used in assorted applications involving materials processing (e.g., welding, drilling) and surface modification (e.g., annealing, hardening) before techniques were developed to capitalize on them as a heat source for the flash evaporation of thin films. Early experimentation with laser-evaporation sources in the 1970s culminated in the successful deposition of stoichiometric, mixed-oxide films by the late 1980s. High-temperature superconductor films of good quality, in particular, fueled much of this activity, which continues virtually undiminished to the present day.

3.5.2.2 PLD Process Details

The pulsed laser deposition process is schematically depicted in Fig. 3-16. In its simplest configuration, a high-power laser situated outside the vacuum deposition chamber is focused by means of external lenses onto the target surface, which serves as the evaporation source. Most nonmetallic materials that are evaporated exhibit strong absorption in the ultraviolet spectral range between 200 and 400 nm. Absorption coefficients tend to increase at the shorter wavelengths meaning reduced penetration depths. Correspondingly, lasers that have been most widely used for PLD center around the solid state Nd^{3+}:YAG (1064 nm) and gas excimer types. In the case of the former, which can deliver up to $\sim$2 J/pulse at a pulse repetition rate of $\sim$30 Hz, the 1064-nm radiation is frequency doubled twice and mixed so that outputs of 355 and 266 nm are produced. Although attenuated in power relative to the fundamental output, they are sufficiently intense for PLD work. Included among the popular gas excimer lasers are the ArF (193 nm), KrF (248 nm), and XeCl (308 nm) types. Commercial versions of these deliver outputs of $\sim$500 mJ/pulse at pulse repetition rates of several hundred Hz.

Irrespective of laser used, the absorbed beam energy is converted into thermal, chemical, and mechanical energy, causing electronic excitation of

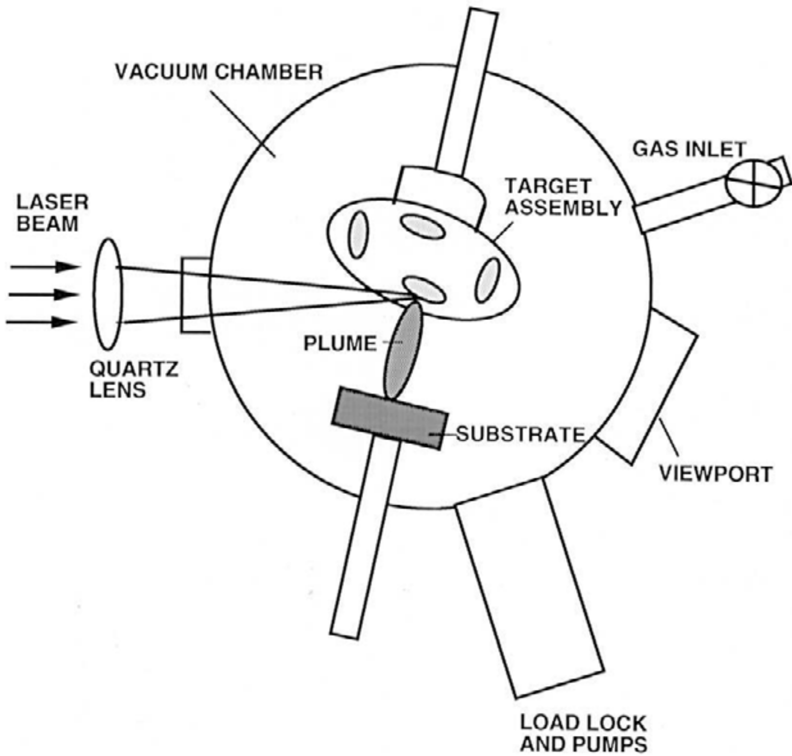


Figure 3-16 Schematic of PLD system for the deposition of metal oxide films. (From R. Ramesh, O. Auciello, V. G. Keramidas, and R. Dat in *Science and Technology of Electroceramic Thin Films*, O. Auciello and R. Waser, eds. Kluwer, Dordrecht, The Netherlands, 1995. Reprinted with the permission of the publisher.)

target atoms, ablation and exfoliation of the surface, and plasma formation. Evaporants form a plume above the target consisting of a motley collection of energetic neutral atoms, molecules, ions, electrons, atom clusters, micron-sized particulates, and molten droplets. The plume is highly directional, i.e., $\cos^n \phi$, where $8 < n < 12$, and its contents are propelled to the substrate where they condense and form a film. Gases, e.g., O_2, N_2, are often introduced in the deposition chamber to promote surface reactions or maintain film stoichiometry. A single homogeneous, multielement target is usually sufficient for the deposition of individual films, e.g., a sintered powder compact target to deposit mixed oxide films. However, for the deposition of layered film structures, multiple sources must be vaporized by the laser beam. This can be achieved through use of a single laser and beam splitters, two or more independent lasers emitting simultaneously, or a single laser

sequentially focused on different multielement targets mounted on a rotating carousel.

Provision is usually made to minimize the number of gross particulates ejected as a result of *splashing* from being incorporated into the depositing film. Splashing of macroscopic particles during laser-induced evaporation is a major concern and one of the two significant drawbacks of PLT; the other stems from the highly directed plume, which makes it difficult to uniformly deposit films over large substrate areas. The generation of particulates during splashing is believed to have several origins (Ref. 20). These include the rapid expansion of gas trapped beneath the target surface, a rough target surface morphology whose mechanically weak projections are prone to fracturing during pulsed thermal shocks, and superheating of subsurface layers before surface atoms vaporize. A common strategy for dealing with splashing effects is to interpose a rapidly spinning pinwheel-like shutter between the target and substrate. Acting much like the compressor vanes of a turbomolecular pump, the slower moving particulates can be batted back, allowing the more mobile atoms, ions, and molecules to penetrate this mechanical mass filter.

Window materials, an important component in PLD systems, must generally satisfy the dual requirement of optical transparency to both visible and ultraviolet light. Relatively few materials are suitable for this demanding role, but MgF_2, sapphire, CaF_2, and UV-grade quartz have served as suitable window materials.

3.5.2.3 Modeling the PLD Process

Although we can appreciate that a laser of sufficient output power can, in principle, vaporize any material, we begin by roughly estimating the minimum laser power needed. Such a calculation properly requires a detailed analysis of macroscopic heating effects in surfaces modified by laser irradiation. This subject has been treated analytically (Ref. 21), and it turns out that two physical lengths, both properties of the irradiated material, play a critical role in the analysis. The first is the optical-absorption depth α^{-1}, a quantity that determines how incident light of intensity I_0 (J/cm^2-s) is attenuated with depth x beneath the surface. Readers will recognize Beer's law, which states that the light intensity at z is given by

$$I(z) = I_0 \exp - \alpha x. \tag{3-28}$$

The second is the thermal-diffusion distance, $2(Kt)^{1/2}$, a measure of how far a pulse of heat spreads in time t. By analogy to diffusional spreading (Eq. 1-29) the key materials constant here is the thermal diffusivity, K (cm^2/s), which in turn, is a composite of other constants, i.e., $K = \kappa/\rho c_p$, where κ is

the thermal conductivity (W/cm-K), ρ is the density (g/cm^3), and c_p is the heat capacity (J/g-K). Importantly, both α^{-1} and K are temperature dependent, and this fact complicates any heating analysis.

A simple boundary-value problem enables an estimate of the temperature distribution at any point (x) beneath the laser-irradiated surface ($x = 0$). The basic heat-conduction equation (note the analogy to Eq. 1-24) and associated initial and boundary conditions are given by

$$\frac{\partial T(x,t)}{\partial t} = K\frac{\partial^2 T(x,t)}{\partial x^2}, \tag{3-29}$$

and $T(x, 0) = T_0$, $T(\infty, t) = 0$, and $\kappa(\partial T(0, t))/\partial x = I_0(1 - R)$. Only the last of these conditions requires comment. It states that the heat flux (W/cm^2) delivered to the surface is provided by the incident laser beam but reduced by the fraction (R) of the radiation that is reflected. The solution is given by

$$T(x,t) = T_0 + I_0\frac{(1-R)}{\kappa}\{(4Kt/\pi)^{1/2}\exp(-x^2/4Kt) - x\,\mathrm{erfc}[x/2(Kt)^{1/2}]\}. \tag{3-30}$$

For a laser pulse of time τ we may assume that Eq. 3-30 holds only for $\tau > t > 0$.

Returning to our objective, the minimum laser power density $\bar{\mathscr{P}}$ (J/cm^2-s) required for evaporation is proportional to the sublimation energy U_S (J/g) and the size of the heat-affected zone, i.e., $2(K\tau)^{1/2}$. Specifically, the associated energy density is $2U_S(K\tau)^{1/2}\rho$ (J/cm^2). Dividing by the absorption time,

$$\bar{\mathscr{P}} = 2U_S(K\tau)^{1/2}\rho/\tau = 2U_S K^{1/2}\tau^{-1/2}\rho \qquad (\mathrm{W/cm^2}) \tag{3-31}$$

and is seen to depend strongly on τ. For $U_S = 10^4$ J/g, $\rho = 10$ g/cm^3, $K = 0.1$ cm^2/s, and $\tau = 10^{-9}$ s (characteristic of Q-switched lasers), $\bar{\mathscr{P}} = 2 \times 10^9$ W/cm^2. This is an extraordinarily large instantaneous power density; actual densities depend on the laser wavelength and target and are generally lower. For continuous wave (cw) radiation where the effective dwell time is longer, a power density as low as 10^4 W/cm^2, similar to that for e-guns, is sufficient to cause evaporation. Alternatively, thermal modeling (Ref. 22) suggests that the target begins to melt when the laser fluence reaches $\sim$0.1 J/cm^2, while the onset of ablation (i.e., when a forward directed plume appears) requires an energy density of $\sim$0.4 J/cm^2.

Another issue of interest is the thickness of material evaporated per laser pulse. Theoretical estimates are not difficult to make if it can be assumed that vaporization occurs at the target boiling point, T_B. Making use of Eq. 3-3, we note that by dividing by the target density the thickness of material

evaporated is given by

$$\dot{d} = \frac{5.84 \times 10^{-2}}{\rho}(M/T_B)^{1/2}P_e \qquad \text{cm/s.} \tag{3-32}$$

In the case of aluminum at $T_B = 2793$ K where $P_e = 760$ torr, $\dot{d} = 0.0436$ m/s. Similarly for other metals the linear evaporation rate is typically 0.05 m/s. Therefore, for a 10 ns laser pulse the vaporization depth is 0.5 nm and this estimate compares with typical values of 1–10 nm/pulse.

A more rigorous analysis (Ref. 23) relates this so-called "thermal sputtering" yield to the vaporizing flux during laser heating. By integrating over the pulse period where a distribution of evaporation events at different temperatures (Eq. 3-6) contributes to the overall sputtering, the depth of material removed (vaporized) per pulse is

$$\text{Depth/pulse} = \frac{P_0}{n_s\sqrt{2\pi M k_B}}\int_0^{\infty} T^{-1/2}\exp(-\Delta H_V/k_B T)\,dt, \tag{3-33}$$

with n_s the atomic density (atoms/cm^3). An approximate solution given by the authors for the depth (nm) per pulse is $1.53 \times 10^6\ P_{atm}T_s^{1/2}\tau/M^{1/2}\Delta H_V$, where T_s is the maximum surface temperature and ΔH_V, the heat of vaporization, is measured in eV. (Note that U_S and ΔH_V are comparable in magnitude.) Calculation shows that boiling temperatures are required for vaporization rates of 1 nm/pulse. Relative to standard thermal evaporation where Γ_e is typically 10^{-3} g/cm^2-s, the linear material removal rate using lasers is roughly a factor of 10^4 greater. During flash evaporation the rapid rate of material removal provides very little time for atomic segregation; because of the congruent melting, good stoichiometry is achieved.

Assorted metal and semiconductor surfaces have been ablated using femtosecond (10^{-15} s) laser pulses (Ref. 24). Through time-resolved optical microscopy capable of snapping frames every tenth of a picosecond, Newton's rings were observed during the ultrafast ablation process. We recall from elementary optics that such rings are the result of interference phenomena, in this case arising from light reflecting off the boiling ablated material and the solid material underneath. Apparently the metastable ablating surface becomes nearly transparent (very low absorption) and has a high index of refraction relative to the solid.

3.5.2.4 Ceramic Films Deposited by PLD

By now virtually every material that has been deposited in thin-film form from the gas phase by other PVD techniques has also been evaporated by PLD methods. The most recent and complete compilation of films deposited

Table 3-4

Ceramic Films Deposited by Pulsed Laser Methods

Property	Applications	Materials
High-temperature superconductivity	Microwave filters and delay lines, digital electronics, sensors	$YBa_2Cu_3O_7$, $Tl_2Ca_2Sr_2Cu_3O_x$, $Nd_{1.85}Ce_{0.15}CuO_4$
Ferroelectricity	DRAM capacitors, nonvolatile RAMS, optoelectronics, microwave devices	$Pb(Zr)TiO_3$, $(Sr, Ba)TiO_3$, $(Sr, Ba)Nb_2O_6$, $LiNbO_3$
Ferrimagnetism	Circulators, phase shifters, magnetic recording, antennas	$BaFe_{12}O_{19}$, $Y_3Fe_5O_{12}$, $(Mn, Zn)Fe_2O_4$, Li_2FeO_4
Electrochromic effects	Optical modulators, sunroofs, sensor protection	WO_3, MoO_3, V_2O_5
Electro-optical effects	Transparent conductors, solar energy, photovoltaics	F-doped ZnO_2, In_2O_3/SnO_2, $(La, Sr)CoO_3$
Piezoelectricity	Microelectrical-mechanical (MEM) devices	$Pb(Zr)TiO_3$
Giant magnetoresistance	Magnetic recording head field sensors	$(La, Ca)MnO_3$
Thermal and corrosive stability	Oxidation and thermal protection coatings for turbine blades	Y-ZrO_2, $MgAl_2O_4$
Friction and wear	Hard, low-friction, wear-resistant coatings	MoS_2, BN, SiC, diamond-like carbon
Biocompatibility	Prostheses, hip/knee implants	Hydroxylapatite, Al_2O_3

From D. B. Crissey, J. S. Horwitz, P. C. Dorsey, and J. M. Pond, *Laser Focus World*, p. 155, May (1995).

by PLD is found in the bibliography prepared by Saenger (Ref. 25). In it the materials evaporated, reactive gases introduced, lasers used, and literature references are listed. Included is information on elements, inorganic compounds, oxides, organic compounds and polymers, semiconductors, assorted layered structures, and mixtures.

Since the deposition of stoichiometric ceramic films is not easily achieved by other means, it is worth elaborating further on this unique feature of PLD. Unlike chemical vapor deposition where synthesis of the film is the result of reaction between precursor gases, stoichiometric bulk sintered ceramics are the target sources in PLD. In this respect the PLD target is like a sputtering cathode target. Because of the shallow melting that occurs, stoichiometric films require a highly homogeneous target. Table 3-4 includes a selection of laser-deposited ceramic films together with compositions and

applications. We shall encounter some of these same materials later in the book where they will be discussed in other contexts.

3.5.3 WEB COATING

3.5.3.1 Introduction to the Process

The development of commercial web-coating processes (Ref. 26) was a response to the widespread need for large surface areas of metallized flexible polymer film and paper sheet. In web-coating systems provision is made to supply a flexible substrate by uncoiling it from one roll and winding it on another in much the same way that audio or video tapes are unwound and wound. Maintenance of a vacuum environment for coating occasionally means movement of the web from air to vacuum to air through successive differentially pumped sections with sliding seals. Within the vacuum region an exposed section of web located between the two coiling rolls is stretched around a cooled roll that lies above the evaporant (or sputtering) source as shown in Fig. 3-17. The web travels at a preset speed to enable it to be sufficiently metallized during its brief exposure to the source. Webs as wide as 2.5 m are in use today and coating speeds in excess of 10 m/s have been attained. Resistance and induction heated boat or crucible sources are employed, but for deposition rates of $\sim 5\,\mu$m/s and above, electron-beam guns are used. Typically, web-coating operations involving aluminum require that it be continuously fed in wire form to replenish what is evaporated.

3.5.3.2 Scope of the Industry and Products

Primary applications of polymer (polyester, PET, polyethylene, polypropylene) web coatings include packaging for snacks, lidding, and films for capacitors, windows, and decorative purposes. Paper and boards are also coated in the manufacture of labels and gift wrap. Well over 10 billion square meters (some 4000 square miles) of polymer film and paper are coated annually worldwide, primarily with evaporated aluminum. Packaging consumes $\sim 60\%$ of this while capacitors account for most of the remainder.

To gain a broader perspective of web-coating possibilities, some of the more unusual applications require many different metal, alloy, oxide, nitride, and halide materials deposited on assorted polymers, metals, paper, and fabrics as single or multilayer films. For example, in beverage containers and see-through, microwavable packaging it is important that the polymers used

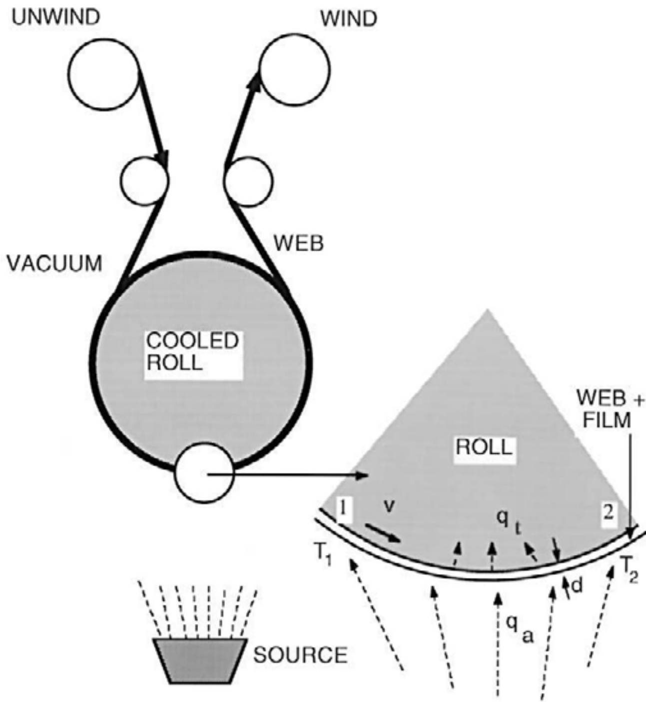


Figure 3-17 Schematic of web coating system including heat-transfer model describing web temperature.

be impervious to gas transport. For these purposes thin SiO, SiO_x, and Al_2O_3 gas barrier coatings are evaporated and serve to seal the polymer porosity (Ref. 27). Additional web coating applications include optical interference filters (Ref. 28), flexible printed circuits, films for magnetic recording purposes (Ref. 29), and transparent conductors for displays.

3.5.3.3 Process Issues

Among the important variables that limit web speed and deposition rates are the temperature of the web and the source evaporation rate. The two are related since hotter sources mean greater radiant heating of the web. Furthermore, increased levels of latent heat are deposited onto the web by condensing atoms derived from such sources. Because webs are often heat-sensitive substrates they cannot tolerate excessive temperatures. In

traveling along the roll with velocity v between points 1 and 2 of Fig. 3-17, the web temperature T rises by an amount ΔT that depends on how much thermal energy (q_a) is absorbed per unit area and time (t), and how much heat (q_t) is transferred or lost to the chilled roll. It is not difficult to estimate the rate of temperature rise dT/dt in any web element having a thickness d and heat capacity c_p. Assuming q_a is constant, and q_t is proportional to $h(T - T_1)$, where h is the heat transfer coefficient and T_1 is the roll temperature, the heat balance is expressed by

$$c_p d\rho \frac{dT}{dt} = q_a - h(T - T_1). \tag{3-34}$$

Integrating between the temperature limits T_1 and T_2 associated with the time interval between 0 and t, the equation

$$\Delta T = T_2 - T_1 = \frac{q_a}{h}\left[1 - \exp - \frac{hL}{vc_p d\rho}\right] \tag{3-35}$$

emerges, where $t = L/v$ and L is the exposed web arc length. Several simple methods to minimize the maximum temperature (T_2) reached by the web are suggested by this analysis, including:

1. Shorten the chill-roll circumference since this effectively reduces L.
2. Operate at high web velocities. This is an obvious conclusion, but it may not be practical, however, because excessive thinning of the coating and stressing of the web may result.
3. Employ polished chill rolls because more intimate contact between web and roll raises h. Depending on web tension and roll surface conditions, values for h have been measured to range from 100 to 300 W/m^2-K (Ref. 30).
4. Reduce both q_a and T_1. The former will no doubt lower the deposition rate and the latter may involve additional cooling costs.

Clearly, optimizing the operating variables within web coating systems involves trade-offs.

Attainment of the vacuum necessary for the deposition of good films is a major processing issue of concern. The scale and nature of web coating operations means that very large pumping systems will be required and that contamination and sources of gas must be dealt with. A typical value of the system pressure during high-rate evaporation of Al is $\sim 5 \times 10^{-4}$ mbar (4×10^{-4} torr). Because of the lower deposition rates during sputtering of alloys even better vacuum is necessary to prevent film contamination. To

intelligently design web coaters, one must know the sources of the gas load or throughput. Gases purposely introduced or released in the process and outgassing of chamber walls, hardware, and web rolls are the sources that have been identified. Of these, the largest gas load arises from the web itself where trapped gas is released during unwinding and then desorbed from the roll ends as well as flexible film surfaces. More gas is evolved from paper and cardboard than plastic film webs because of the rough, fibrous nature of the former and smoother, less permeable surfaces of the latter. Since most of the gas consists of water vapor, extensive use is made of cryopanels for pumping.

Sizing the pumps needed requires an estimate of the gas load or throughput due to roll-trapped gases (Ref. 31). We assume this is given by $Q_{tg} = P_{tg}\delta wv$, where P_{tg} is taken as 1.01 bar (the pressure of atmospheric gas trapped during web coiling), δ is the effective thickness of the gas layer, and w and v are the web width and velocity, respectively. Further, assuming that $\delta = 1\ \mu$m, $w = 1$ m, and $v = 10$ m/s, $Q_{tg} = 10.1$ mbar-liters/s. If we consider high rate evaporation of Al at a process pressure P_p of 5×10^{-4} mbar, a pumping speed of Q_{tg}/P_p or 20,200 liters/s will be required to pump this source of gas.

3.5.4 Ion Beam Assisted Evaporation

An impression may have been left with the reader that during evaporation processes relatively little is done to modify depositing films other than altering substrate temperatures. However, many commercial evaporation processes presently employ ion bombardment of the substrate as a means of improving film properties. A simple example is shown in Fig. 3-18 where an ion gun, typically generating ions with energies of a few keV, is used in conjunction with an evaporation source (Ref. 32). This process known as ion-beam-assisted deposition (IBAD), combines the benefits of high film-deposition rate and substrate ion bombardment in a relatively clean vacuum environment. Central to the technique is the generation of inert (e.g., Ar^+) and/or reactive (e.g., O_2^+) ion beams and their subsequent interaction with surfaces to modify film structure and composition. These subjects are treated in the next two chapters and further discussion of ion-beam properties and their effects will therefore be deferred until then. Here it is simply noted that ions (~ 1 keV) impacting the growing film are considerably more energetic than arriving evaporant atoms (~ 0.1 eV). They serve to peen the surface and compact the film, densifying it in the process. Enhancing the density and index of refraction of assorted oxide films used for optical coatings is a major application of IBAD processes.

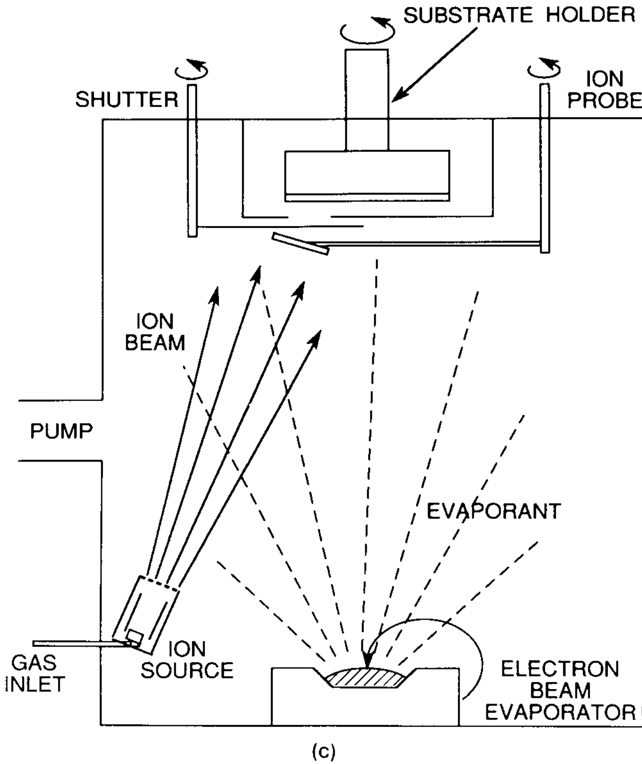


Figure 3-18 Ion-beam-assisted deposition employing an e-gun evaporator. (From Ref. 32.)

3.6 CONCLUSION

Evaporation methods are perhaps the simplest means of producing thin films and coatings comprised of sequentially deposited atoms. In the evaporation process, events at the heated source, at the substrate and in the intervening vacuum space all play roles in the successful deposition of thin films. Thermodynamics, particularly source temperature–vapor pressure relationships, govern rates of evaporation and the nature of the vapor stream. In this regard the complexities involved in evaporating multicomponent metal and semiconductor sources have been addressed. The important issue of film thickness uniformity is governed by the geometric placement of source and substrate. As long as the substrates are planar, the film thickness distribution can usually be modeled in terms of the laws of

emission from point or surface sources, and substrate reception based on the inverse square separation distance between the two. Film uniformity and coverage become more problematical over substrates that are stepped or contain trenches and other complex topographies. In addition to line-of-sight shadowing, atomic diffusional effects determine the extent of film coverage. Lastly, the vacuum space and partial pressures of residual gases within it have an influence on film purity and properties such as resistivity, reflectivity, and hardness.

Evaporation techniques for thin-film deposition have been superseded in many instances by sputtering and chemical vapor deposition methods; difficulties in maintaining stoichiometry and achieving conformal coverage are among the reasons for this. Paradoxically, however, pulsed-laser deposition enables excellent stoichiometry to be attained in complex oxide films. Despite shortcomings, evaporation methods are readily scalable to large operations such as web coating. However, in addition to conventional applications, the versatility of evaporation techniques is exploited in the demanding arena of molecular beam epitaxy as well as in hybrid deposition processes that employ plasmas, ion beams, and reactive gases. These processes and applications will be discussed in Chapters 5 and 8.

EXERCISES

1. (a) Employing vapor-pressure data, calculate values for the molar heat of vaporization of Si and Ga.
   (b) Design a laboratory procedure to experimentally determine the value of the heat of vaporization of a metal employing common thin-film deposition and characterization equipment.

2. Suppose Fe satisfactorily evaporates from a surface source, 1 cm^2 in area, which is maintained at 1550°C. Higher desired evaporation rates can be achieved by raising the temperature 100°C, but this burns out the source. Instead, the melt area is increased without raising its temperature. By what factor should the source area be enlarged to achieve equivalent deposition?

3. A molecular-beam epitaxy system contains separate Al and As effusion evaporation sources of 4 cm^2 area, located 10 cm from a (100) GaAs substrate. The Al source is heated to 1000°C, and the As source is heated to 300°C. What is the growth rate of the AlAs film in Å/s? *Note:* AlAs basically has the same crystal structure and lattice parameter (5.661 Å) as GaAs.

4. How far from the substrate, in the illustrative problem on p. 110, would a single surface source have to be located to maintain the same deposited film thickness tolerance?
5. An Al film was deposited at a rate of ~1 μm/min in vacuum at 25°C, and it was estimated that the oxygen content of the film was 10^{-3}. What was the partial pressure of oxygen in the system?
6. In order to deposit films of the alloy YBa_2Cu_3, the metals Y, Ba, and Cu are evaporated from three point sources. The latter are situated at the corners of an equilateral triangle whose side is 20 cm. Directly above the centroid of the source array, and parallel to it, lies a small substrate; the deposition system geometry is effectively a tetrahedron, each side being 20 cm long.
   (a) If the Y source is heated to 1740 K to produce a vapor pressure of 10^{-3} torr, to what temperature must the Cu source be heated to maintain film stoichiometry?
   (b) Rather than a point source, a surface source is used to evaporate Cu. How must the Cu source temperature be changed to ensure deposit stoichiometry?
   (c) If the source configuration in part (a) is employed, what minimum O_2 partial pressure is required to deposit stoichiometric superconducting oxide $YBa_2Cu_3O_7$ films by a reactive evaporation process?

   The atomic weights are Y = 89, Cu = 63.5, Ba = 137, and O = 16.
7. One way to deposit a thin metal film of known thickness is to heat an evaporation source to dryness (i.e., until no metal remains in the crucible). Suppose it is desired to deposit 5000 Å of Au on the internal spherical surface of a hemispherical shell measuring 30 cm in diameter.
   (a) Suggest two different evaporation source configurations (source type and placement) that would yield uniform coatings.
   (b) What weight of Au would be required for each configuration, assuming evaporation to dryness?
8. (a) Consider the heating of a room-temperature polymer substrate during deposition of 10^{18} aluminum atoms per cm^2 per minute. The temperature of the evaporant is 1400 K, the heat of condensation of Al is 310 kJ/mol, and the emissivity of the Al source is assumed to be 0.2. Calculate the thermal power delivered to the substrate through condensation, evaporant kinetic energy, and radiation.
   (b) If the polymer density, heat capacity, and thickness are 1.3 g/cm^3, 1.5 J/g-K, and 0.1 mm, respectively, to what temperature would the substrate rise after 1 minute?
   (c) This polymer is now used in a web coating operation where it will experience the same Al flux. For an exposed web length of 30 cm,

how fast should the web travel for the temperature not to rise above that calculated in part (b)?

Assume the heat transfer coefficient between web and roll is 200 W/m^2-K.

9. A tungsten evaporation source is rated at 1000 W and operates at 120 V. If the filament heater wire is 20 cm long and 0.75 mm in diameter estimate the temperature (T) the source will reach when powered. Compare your answer with an alternative estimate of the temperature assuming all of the input electrical power dissipated is thermally radiated from the filament surface.

10. The measured specific energy required for the evaporation of zirconium is reported to be 61.5 kW-h/kg. If the metal was electron-beam melted in a 100-mm-diameter water-cooled copper crucible, how does this energy compare with that predicted in the text? Assume the thermal conductivity of Zr is 30 W/K-m, and the average charge thickness is 1 cm.

11. Calculate the crossover temperature for silicon, i.e., where the energy transfer by radiation equals that by evaporation. Assume the emissivity of Si = 0.7.

Questions 12 to 16 are related to the evaporation geometry shown in Fig. 3-6.

12. The collection efficiency for a given deposition process is the ratio of the amount of material that falls on a substrate of given size relative to the total amount evaporated. Derive a formula for the collection efficiency from a planar source as a function of $1/d$ and plot the results. Repeat for the point source. Which source yields a greater collection efficiency?

13. Assume that an evaporation source is highly directional such that

$$\frac{d\bar{M}_s}{dA_s} = \frac{11\bar{M}_e \cos^{10}\phi \cos\theta}{2\pi r^2}.$$

Plot the deposited film profile and compare it to those for the point and surface sources.

14. Two small area evaporation sources are 100 cm apart and situated 50 cm below a planar substrate. The line between the sources is parallel to that of the substrate plane. One source evaporates material A while the second source evaporates material B. Suppose the vapor pressure of A is 10 times that of B at the evaporation temperature of 1300 K.
    (a) At what distance along the substrate will the film composition be 60 at.%A–40 at.%B?
    (b) If the vapor pressure of A is 15 times that of B at an evaporation temperature of 1500 K, what is the *difference* between the heats of vaporization for A and B?

15. For a new application it is desired to continuously coat a 1-m-wide steel strip with a 2-μm-thick coating of Al. The $x-y$ configuration of the steel is such that an array of electron-beam gun evaporators lies along the y direction and maintains a uniform coating thickness across the strip width. How fast should the steel be fed in the x direction past the surface sources, which can evaporate 20 g of Al per second? Assume that Eq. 3-21 holds for the coating thickness along the x direction, that the source–strip distance is 30 cm, and that the steel sheet is essentially a horizontal substrate 40 cm long on either side of the source before it is coiled.
16. The level of molten metal in a crucible that behaves like a surface evaporation source is initially a distance h from a planar substrate surface. During deposition the level of the metal recedes a distance Δh. Derive an expression for the fractional change in film thickness at any point along the substrate as a function of $\Delta h/h$.
17. Verify that Eq. 3-30 is the solution to the boundary value problem described in the text. Graph the evolution of the temperature profiles predicted by this solution.

REFERENCES

1. W. R. Grove, *Phil. Trans. Roy. Soc. London* **A142**, 87 (1852).
2. M. Faraday, *Phil. Trans.* **147**, 145 (1857).
3. J. L. Vossen and W. Kern, *Thin Film Processes II*. Academic Press, New York, 1991.
4. R. Glang, in *Handbook of Thin Film Technology*, L. I. Maissel and R. Glang, eds. McGraw-Hill, New York, 1970.
5. C. H. P. Lupis, *Chemical Thermodynamics of Materials*. North-Holland, Amsterdam, 1983.
6. R. E. Honig, *RCA Rev.* **23**, 567 (1962).
7. J. Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy*. Academic Press, Boston, 1993.
8. D. L. Smith, *Thin-Film Deposition*. McGraw-Hill, New York, 1995.
9. H. K. Pulker, *Coatings on Glass*, 2nd ed. Elsevier, New York, 1999.
10. Examples taken from *Physical Vapor Deposition*, R. J. Hill, ed. Temescal, BOC Group, 1986.
11. L. Holland, *Vacuum Deposition of Thin Films*. John Wiley & Sons, New York, 1956.
12. R. A. Powell and S. M. Rossnagel, *PVD for Microelectronics: Sputter Deposition Applied to Semiconductor Manufacturing*. Academic Press, San Diego, 1999.
13. C. H. Ting and A. R. Neureuther, *Solid State Technol.* **25**(2), 115 (1982).
14. H. L. Caswell, in *Physics of Thin Films*, Vol. 1, G. Hass, ed. Academic Press, New York, 1963.
15. J. Balakrishnan and I. D. Boyd, *J. Vac. Sci. Technol.* **A18**, 907 (2000).
16. W. D. Westwood, in *Microelectronic Materials and Processes*, R. A. Levy, ed. Kluwer Academic, Dordrecht, The Netherlands, 1989.
17. E. B. Graper, *J. Vac. Sci. Technol.* **A**(5), 2718 (1987); **8**, 333 (1971).
18. *The Photonics Design and Applications Handbook*, Book 3, p. H-82. Laurin Publ., 1999.

19. D. B. Chrisey and G. K. Hubler, eds., *Pulsed Laser Deposition*. John Wiley & Sons, New York, 1994.
20. J. T. Cheung and H. Sankur, *CRC Crit. Rev. Solid State Mat. Sci.* **15**, 63 (1988).
21. M. Ohring, *The Materials Science of Thin Films*, p. 595. Academic Press, Boston, 1992.
22. T. J. Jackson and S. B. Palmer, *J. Phys. D: Appl. Phys.* **27**, 1581 (1994).
23. R. Kelley and J. E. Rothenberg, *Nucl. Instrum. Meth.* **B7/8**, 755 (1985).
24. R. W. Hardin, *Laser Focus World* **34**(8), 15 (1998).
25. K. L. Saenger, in *Pulsed Laser Deposition*, D. B. Chrisey and G. K. Hubler, eds. John Wiley & Sons, New York, 1994.
26. See annual issues of the *Proceedings of the International Conference on Vacuum Web Coating*, R. Bakish, ed. Bakish Materials Corporation, P.O. Box 148, Englewood, NJ 07631.
27. C. Hayashi, in *Proceedings of the Tenth International Conference on Vacuum Web Coating*, R. Bakish, ed., p. 2 (1996).
28. J. D. Affinito, M. E. Gross, G. L. Graff, S. Eufinger, E. N. Greenwell, and P. M. Martin, in *Proceedings of the Tenth International Conference on Vacuum Web Coating*, R. Bakish, ed., p. 207 (1996).
29. P. Lippens and H. te Lintelo, in *Proceedings of the Ninth International Conference on Vacuum Web Coating*, R. Bakish, ed., p. 154 (1995).
30. S. Schiller, G. Beister, M. Neumann, and G. Jaesch, *Thin Solid Films* **96**, 199 (1982).
31. W. Schwarz, in *Proceedings of the Tenth International Conference on Vacuum Web Coating*, R. Bakish, ed., p. 51 (1996).
32. J. M. E. Harper and J. J. Cuomo, *J. Vac. Sci. Technol.* **21**(3), 737 (1982).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 4

Discharges, Plasmas, and Ion–Surface Interactions

4.1 INTRODUCTION

Evaporation caused by absorption of thermal energy is not the only way to induce atoms to leave a liquid or solid surface. Atoms can also be ejected or sputtered from solids at room temperature by bombarding their surfaces with energetic ions. In either case the emitted atoms traverse a reduced pressure ambient and deposit atomistically on a substrate to form a film. Because physical means are primarily involved in producing films, both are known as physical vapor-deposition (PVD) processes. Despite some superficial similarities, it is immediately apparent that evaporation and sputtering are quite different if we consider Fig. 4-1a depicting a simplified sputtering system capable of depositing metals films. Inside is a pair of parallel metal electrodes, one of which is the cathode or target of the metal to be deposited. It is connected to the negative terminal of a DC power supply and typically, several kilovolts are applied to it. Facing the cathode is the substrate or anode, which may be grounded, biased positively or negatively, heated, cooled, or some combination of these. After evacuation of the chamber, a working gas, typically argon, is introduced and serves as the medium in which an electrical discharge is initiated and sustained. Gas pressures usually range from a few to a hundred millitorr. After a visible glow discharge is maintained between the electrodes, it is observed that a current flows, and metal from the cathode deposits on the substrate.

Microscopically, positive gas ions in the discharge strike the cathode and physically eject or sputter target atoms through momentum transfer to them. These atoms enter and pass through the discharge region to eventually deposit on the growing film. In addition, other particles (secondary

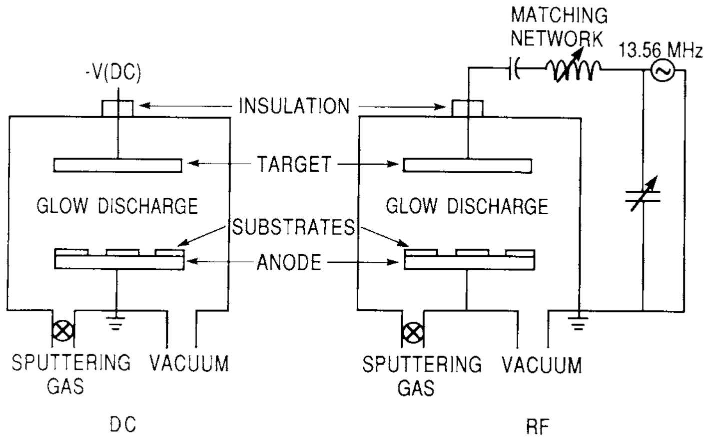


Figure 4-1 Schematics of simplified sputtering systems: (a) DC, (b) RF.

electrons, desorbed gases, and negative ions) as well as radiation (X-rays and photons) are emitted from the target. The electric field accelerates electrons and negatively charged ions toward the anode substrate where they impinge on the growing film. An ionized gas or plasma rather than a vacuum environment, active electrodes that participate in the deposition process, and low-temperature processing are among the features that distinguish sputtering from evaporation. From this simple description, it is quite clear that compared to the predictable rarefied-gas behavior in an evaporation system, the glow-discharge plasma is a very busy and not easily modeled environment. Similar effects occur even when the electrodes are AC powered as shown in Fig. 4-1b.

In the past few decades, advances in our understanding of the physics and chemistry of ionized gases has led to the widespread adoption of plasma technology for the deposition and removal (etching) of thin films as well as the modification of surfaces in a diverse variety of technologies. Microelectronics applications have been the main technological driver in this regard; presently, upward of a third of integrated circuit fabrication steps are associated with the use of plasmas. In addition, there are critical plasma processing operations in the automotive, optical coating, biomedical, information recording, waste management, and aerospace industries.

Regardless of the plasma process, however, roughly similar discharges, electrode configurations, and gas/solid interactions are involved. The purpose of this chapter is to introduce fundamental *scientific* issues common to all glow discharge systems. These include topics related to initiating and sustaining discharges, the dynamical behavior of the charged and neutral

species, and their interactions within the plasma. Both inert gas discharges and the more complex chemically reactive plasmas will be discussed in this regard. Ion interactions with the cathode and film or substrate surfaces are particularly important in plasma processing. Therefore, coverage of the fundamental physics of sputtering and ion-induced modification of growing films rounds out this chapter. Practical *engineering* issues related to the deposition and etching of metal and insulator films and descriptions of assorted plasma PVD and ion-beam processes are the subjects of the next chapter. Our treatment of plasma processing does not end then, but continues in Chapter 6 with a discussion of plasma-assisted chemical vapor deposition processes.

4.2 PLASMAS, DISCHARGES, AND ARCS

4.2.1 PLASMAS

The term plasma was apparently coined by Irving Langmuir in 1929 (Ref. 1) to describe the behavior of ionized gases in high-current vacuum tubes. It was soon realized that plasmas exhibited a behavior different from simple ideal (or nonideal) un-ionized gases, and were obviously distinct from condensed liquid and solid states of matter. For these reasons plasmas were termed a rare fourth state of matter. On a cosmic scale, however, considering the northern lights, stars, and interstellar hydrogen, it has been claimed that 99% of matter in the universe exists in plasma form; on this basis solids and liquids are actually the rare states of matter. Science fiction movies containing spectacular lightning discharges and glittering advertising signs employing neon lighting provided early vehicles for displaying the mysterious attributes of plasmas. Nitriding of steel surfaces in order to harden them, and fluorescent lighting based on mercury discharges, were among the early widespread applications capitalizing on this strange state of matter.

A plasma may be broadly defined as a *quasineutral* gas that exhibits a *collective* behavior in the presence of applied electromagnetic fields. Plasmas are weakly ionized gases consisting of a collection of electrons, ions, and neutral atomic and molecular species. This definition is broad enough to encompass the spectrum of space and man-made plasmas extending from stars, solar winds and coronas, and the earth's ionosphere to the regime of high-pressure arcs, shock tubes, and fusion reactors. These space and laboratory created plasmas broadly differ in the density n (number per cm^3) of charged species. In the former rarified environments, n is typically less than 10^7 cm^{-3}, whereas experimentally, densities approaching 10^{20} cm^{-3} in magnitude have been realized in the latter man-made high-pressure plasmas.

In between these extremes are the *glow discharges* and *arcs* with which this book is primarily concerned. These plasmas with ion densities ranging from $\sim 10^8$ cm^{-3} to $\sim 10^{14}$ cm^{-3} are the ones exploited in the industrial plasma-processing applications we shall consider.

4.2.2 The Townsend Discharge

We have seen that application of a sufficiently high DC voltage between metal electrodes immersed in a low-pressure gas initiates a discharge. What is the mechanism that converts the initially insulating gas into an electrically conducting medium? This is the first question that must be addressed in order to understand the nature of such gaseous conducting media or plasmas. In essence, the discharge reflects a gaseous breakdown that may be viewed as the analog of dielectric breakdown in insulating solids; there dielectrics conduct electricity at critical applied voltages. In gases the process begins when a stray electron near the cathode carrying an initial current i_0 is accelerated toward the anode by the applied electric field ($\mathscr{E}$). After gaining sufficient energy the electron collides with a neutral gas atom (A) converting it into a positively charged ion (A^+). During this impact ionization process, charge conservation indicates that two electrons are released, i.e.,

$$e^- + A \rightarrow 2e^- + A^+. \tag{4-1}$$

These are accelerated and now bombard two additional neutral gas atoms, generating more ions and electrons, and so on. Meanwhile, the electric field drives ions in the opposite direction where they collide with the cathode, ejecting, among other particles, secondary electrons. These now also undergo charge multiplication. The effect snowballs until a sufficiently large avalanche current ultimately causes the gas to breakdown.

In order for breakdown to occur, the distance (d) between electrodes must be large enough to allow electrons to incrementally gain the requisite energy for an ionization cascade. Also, the electrodes must be wide enough to prevent the loss of electrons or ions through sideways diffusion out of the interelectrode space. The Townsend equation, whose derivation will be left for the reader in Exercise 1, is written as

$$i = i_0 \frac{\exp \alpha d}{[1 - \gamma_e(\exp \alpha d - 1)]}. \tag{4-2}$$

This equation reveals that the discharge current (i) rises dramatically from i_0 because of the combined effects of impact ionization and secondary-electron generation. These processes are respectively defined by constants α

and γ_e. Known as the Townsend ionization coefficient, α represents the probability per unit length of ionization occurring during an electron–gas atom collision. Quantity γ_e is the Townsend secondary-electron emission coefficient and is defined as the number of secondary electrons emitted at the cathode per incident ion. For an electron of charge q traveling a distance λ, the probability of reaching the ionization potential V_i is $\exp-(V_i/q\mathscr{E}\lambda)$, so that

$$\alpha = \frac{1}{\lambda}\exp^{-\frac{Vi}{q\mathscr{E}\lambda}}. \tag{4-3}$$

We may associate λ with the intercollision distance or mean free path in a gas. Since Eq. 2-5 reveals that $\lambda \sim P^{-1}$, we expect α to be a function of the system pressure.

Breakdown is assumed to occur when the denominator in Eq. 4-2 is equal to zero, i.e., $\gamma_e(\exp \alpha d - 1) = 1$, for then the current is infinite. From this condition plus Eqs. 4-3 and 2-5, the critical breakdown field ($\mathscr{E} = \mathscr{E}_B$) and voltage ($V_B = d\mathscr{E}_B$) can be calculated with a bit of algebra and expressed in terms of a product of pressure and interelectrode spacing. The result, known as Paschen's Law, is expressed by

$$V_B = \frac{APd}{\ln(Pd) + B}, \tag{4-4}$$

where A and B are constants.

The Paschen curve, a plot of V_B vs Pd, is shown in Fig. 4-2 for a number of gases. At low values of Pd there are few electron–ion collisions and the secondary electron yield is too low to sustain ionization in the discharge. On the other hand, at high pressures there are frequent collisions, and since electrons do not acquire sufficient energy to ionize gas atoms, the discharge is quenched. Thus at either extreme, ion generation rates are low and high voltages are required to sustain the discharge. In between, at typically a few hundred to a thousand volts, the discharge is self-sustaining. This means that for each electron at the cathode, $\exp(\alpha d)$ electrons reach the anode, and the net effect of the collisions is to produce a new electron at the cathode. Practically, however, in most sputtering discharges the Pd product is well to the left of the minimum value.

4.2.3 Types and Structures of Discharges

It is instructive to follow the progress of a glow discharge in a low-pressure gas using a high-impedance DC power supply (Ref. 2). In the regime just considered, known as the Townsend discharge, a tiny current flows initially due to the small number of charge carriers in the system. With

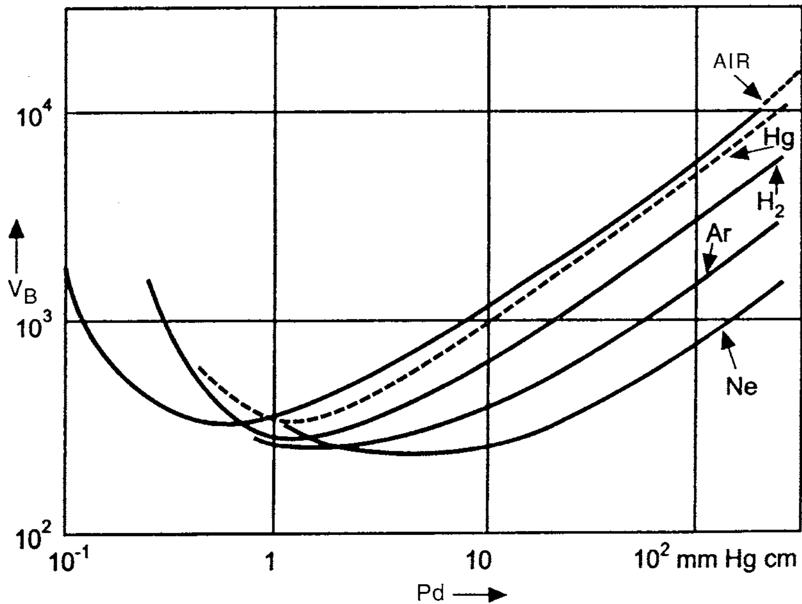


Figure 4-2 Paschen curves for a number of gases. (From A. von Engel, *Ionized Gases*. Oxford University Press, Oxford, 1965. Reprinted with permission.)

charge multiplication, the current increases rapidly, but the voltage, limited by the output impedance of the power supply, remains constant. Eventually, when enough electrons produce sufficient ions to regenerate the same number of initial electrons, the discharge becomes self-sustaining. The gas begins to glow now and the voltage drops accompanied by a sharp rise in current. At this point *normal glow* occurs. Initially, ion bombardment of the cathode is not uniform but concentrated near the cathode edges or at other surface irregularities. As more power is applied, the bombardment increasingly spreads over the entire surface until a nearly uniform current density is achieved. A further increase in power results in both higher voltage and cathode current-density levels. The *abnormal discharge* regime has now been entered and this is the operative domain for sputtering and other discharge processes such as plasma etching.

At still higher currents, the cathode gets hotter. Now thermionic emission of electrons exceeds that of secondary-electron emission and low-voltage arcs propagate. Arcs have been defined (Ref. 3) as gas or vapor discharges where the cathode voltage drop is of the order of the minimum ionizing or excitation potential. Furthermore, the arc is a self-sustained discharge that supports high currents by providing its own mechanism for electron emission from negative or positive electrodes. A number of commercial

PVD processes rely on arcs. This subject will therefore be deferred to the end of Chapter 5 so that the intervening treatment of plasma physics and processing can form the basis for a discussion of these arc-deposition methods.

Returning to the DC discharge we note that there is a progression of alternating dark and luminous regions between the cathode and anode, as shown in Fig. 4-3. Although the general structure of the discharge has been known for a long time the microscopic details of charge distributions, behavior, and interactions within these regions are not totally understood. The *Aston dark space* is very thin and contains both low energy electrons and high energy positive ions, each moving in opposite directions. Beyond it the *cathode glow* appears as a highly luminous layer that envelops and clings to the cathode. De-excitation of positive ions through neutralization is the probable mechanism of light emission here.

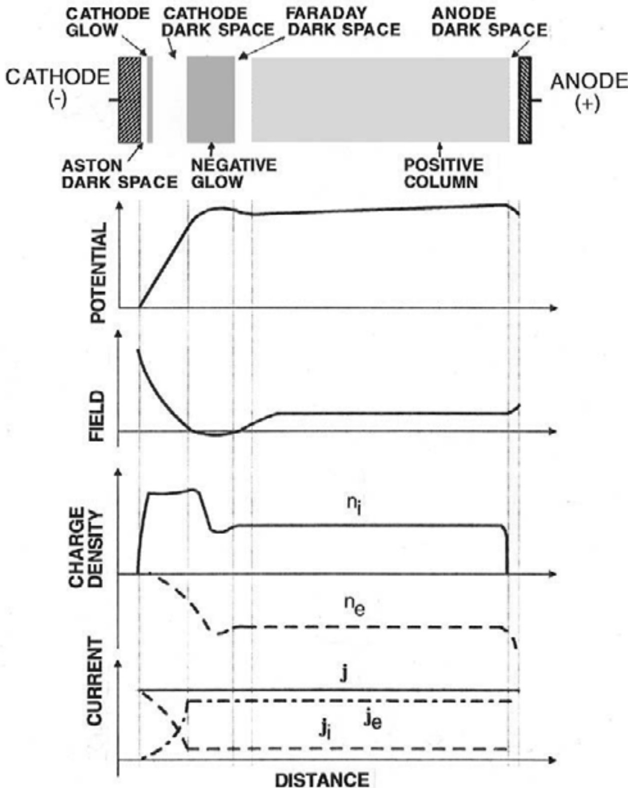


Figure 4-3 Structure of a DC glow discharge with corresponding potential, electric field, charge, and current distributions.

Next to appear is the important Crookes or *cathode dark space* where some electrons are energized to the point where they begin to impact-ionize neutrals; other lower energy electrons impact neutrals without ion production. Because there is relatively little ionization this region is dark. Most of the discharge voltage is dropped across the cathode dark space, also commonly referred to as the cathode sheath. The resulting electric field serves to accelerate ions toward their eventual collision with the cathode. Next in line is the *negative glow*. Here the visible emission is apparently due to interactions between assorted secondary electrons and neutrals with attendant excitation and de-excitation. Beyond lie the *Faraday dark space*, the *positive column*, and finally the *anode*. During sputtering the substrate is typically placed inside the negative glow before the Faraday dark space so that the latter as well as the positive column do not normally appear.

When a DC voltage V is applied between the anode and cathode the electric potential distribution, unlike the case for a simple vacuum capacitor, is highly nonlinear with distance x; similarly, the electric field ($\mathscr{E} = -dV/dx$) is not constant. Furthermore, the deviations are most pronounced near the electrodes. These characteristics stem from the complex distribution of charge near electrodes and within the plasma, and the resultant currents they produce. After considering aspects of plasma physics contained in Section 4.3, some causes of these puzzling electrical responses may, hopefully, be clarified.

4.3 FUNDAMENTALS OF PLASMA PHYSICS

In this section readers will find a concise treatment of several important issues related to plasmas and their interaction with surfaces placed in their midst. The discussion here is largely distilled from the more complete treatments of the subject that can be found in the readily accessible books by Chapman (Ref. 4), Grill (Ref. 5), Mahan (Ref. 6), and Lieberman and Lichtenberg (Ref. 7). They are all recommended for their integration of the fundamental principles of glow-discharge plasmas in applications to thin-film processing.

4.3.1 PLASMA SPECIES

Let us now consider the interior of the plasma, i.e., a partially ionized gas composed of respective densities of electrons (n_e), ions (n_i), and neutral gas species (n_0). Electrons and ions have more or less independent velocity distributions with electrons possessing far higher velocities than ions. The

plasma is electrically neutral when averaged over all the particles contained within so that $n_e = n_i = n$. Collisions between neutral gas species essentially cause them to execute random Brownian motion. However, the applied electric field disrupts this haphazard motion because of ionization. If the density of charged particles is high enough compared with the dimensions of the plasma, significant coulombic interaction exists among particles. This interaction enables the charged species to flow in a fluid-like fashion that determines many of the plasma properties.

The degree of gas ionization (f_i) is defined by

$$f_i = n_e/(n_e + n_0) \tag{4-5}$$

and typically has a magnitude of $\sim 10^{-4}$ in the glow discharges used in thin-film processing. Therefore, at pressures of ~ 10 millitorr, Fig. 2-2 based on the ideal gas law indicates a gas density of $n_0 \sim 10^{14}\,\text{cm}^{-3}$; hence the electron and ion densities will be about $10^{10}\,\text{cm}^{-3}$ each at 25°C. In high density plasmas, f_i can reach 10^{-2} and charge densities more than $10^{12}\,\text{cm}^{-3}$.

4.3.2 Particle Energies and Temperatures

Measurements on glow discharges yield electron energies (E_e) that span the range 1 to 10 eV with 2 eV being a typical average value for calculation purposes. The effective or characteristic temperature T associated with a given energy E is simply given by $T = E/k_B$, where k_B is the Boltzmann constant. Substituting $E_e = 2$ eV, we find that electrons have an astoundingly high temperature T_e of some 23,000 K. However, because there are so few of them, their heat content is small and the chamber walls do not heat appreciably. Neutral gas atoms or molecules and ions are far less energetic; the former have energies of only 0.025 eV (or $T_0 = 293$ K) and the latter, energies of ~ 0.04 eV (or $T_i = 500$ K). Ions have higher energies than neutrals because they acquire energy from the applied electric field.

In addition, there may be excited species at temperature T_{ex} with energy E_{ex}. Neutral molecules may become excited by virtue of acquired energy that is partitioned into translational as well as internal vibrational and rotational modes of motion; for each of these modes there is a corresponding characteristic temperature. For example, in a nitrogen gas plasma at several torr, T_e may be over 12,000 K and T_0 due to molecular translation is ~ 1000 K, while equivalent temperatures for vibrational (T_v) and rotational (T_{ro}) modes are ~ 3800 K and 2800 K, respectively (Ref. 7). Higher plasma pressures tend to narrow this overall disparity in temperature.

Thermodynamic equilibrium in the system implies that all of the temperatures are equilibrated, i.e., $T_e = T_0 = T_i = T_{ex} = T_r = T_w$, where T_r and

T_w are the radiation and chamber wall temperatures, respectively. Since this condition is never met in our low-pressure glow discharges, we speak of a nonequilibrium or *cold* plasma. Plasmas types are often differentiated on the basis of the electron energy and temperature. For example, T_e for glow discharges is greater than that for flames but considerably less than that for fusion plasmas.

4.3.3 Motion of Plasma Species: Currents and Diffusion

Since surfaces (e.g., targets, substrates) are immersed in the plasma, they are bombarded by the species present. Simple kinetic theory of gases helps us understand what happens. The neutral particle flux can be calculated from Eq. 2-8. Unlike neutrals however, charged particle impingement results in an effective electrical current density (j) given by the product of the particle flux ($\frac{1}{4}n\bar{v}$) and the charge (q) transported, where the factor of $\frac{1}{4}$ reflects that fraction of the random motion that is directed at the planar surface. Therefore,

$$j = n\bar{v}q/4, \tag{4-6}$$

where n and $\bar{v}$ are the charged species concentration and mean velocity. To compare the behavior of different species we take $\bar{v} = (8k_BT/\pi m)^{1/2}$ (Eq. 2-3b). In the case of electrons, $m_e = 9.1 \times 10^{-28}$ g, and if we assume $T_e =$ 23,000 K, $v_e = 9.5 \times 10^7$ cm/s; similarly, for typical Ar ions $v_i = 5.2 \times 10^4$ cm/s. Furthermore, if $n_e = n_i = 10^{10}/\text{cm}^3$, $j_e \sim 38\ \text{mA/cm}^2$ and $j_i = 21\ \mu\text{A/cm}^2$. The implication of this simple calculation is that an isolated surface within the plasma charges negatively initially because of the greater electron bombardment. Subsequently, additional electrons are repelled while positive ions are attracted. Therefore, the surface continues to charge negatively at a decreasing rate until the electron flux equals the ion flux and there is no net current.

We now consider the mobility (μ) of charged species in the presence of an applied electric field ($\mathscr{E}$). The mobility is defined as the velocity per unit electric field or $\mu = v/\mathscr{E}$. Using Newton's law,

$$m\,dv/dt = |q|\mathscr{E} + m[\delta v/\delta t]_{\text{coll}} \tag{4-7}$$

where q is the species charge. The second term on the right reflects a kind of frictional drag particles experience during motion because of their collisions with other particles. When the particle collides, it essentially loses

its directed motion. It is common to set $[\delta v/\delta t]_{\text{coll}} = vv$, where v is the collision frequency, a factor assumed for simplicity to be constant. In the steady state, $dv/dt = 0$ and $\mu = |q|/mv$. Typical mobilities for gaseous ions at 1 torr and 273 K range from $\sim 4 \times 10^2$ cm^2/V-s (for Xe^+) to 1.1×10^4 cm^2/V-s (for H^+).

A second kinetic effect involving species motion in plasmas is diffusion, a phenomenon governed by Fick's Law (Eq. 1-22). When migrating species move under the simultaneous influence of two driving forces, i.e., diffusion in a concentration gradient (dn/dx) and drift in the applied electric field, we may write for the respective electron and ion particle fluxes,

$$J_e = -n_e \mu_e \mathscr{E} - D_e \, dn_e/dx \tag{4-8}$$

$$J_i = n_i \mu_i \mathscr{E} - D_i \, dn_i/dx. \tag{4-9}$$

To maintain charge neutrality in the region under consideration it is assumed that $J_e = J_i = J$, and $n_e = n_i = n$. By equating Eqs. 4-8 and 4-9,

$$\mathscr{E} = \frac{(D_i - D_e)}{n(\mu_e + \mu_i)} \frac{dn}{dx}. \tag{4-10}$$

Therefore, it is apparent that an electric field develops because the difference in electron and ion diffusivities produces a separation of charge. Physically, more electrons than ions tend to leave the plasma, establishing an electric field that hinders further electron loss but at the same time enhances ion motion. Because of the coupled electron and ion motions we can assign (see Exercise 2) an effective *ambipolar* diffusion coefficient D_a to describe the effect, i.e.,

$$D_a = \frac{(D_i \mu_e + D_e \mu_i)}{(\mu_e + \mu_i)}. \tag{4-11}$$

The magnitude of D_a lies somewhere between those of D_i and D_e so that both ions and electrons diffuse faster than intrinsic ions do.

4.3.4 Electron Motion in Combined Electric and Magnetic Fields

4.3.4.1 Parallel Fields

Let us now examine what happens when a magnetic field of strength B is superimposed parallel to the electric field $\mathscr{E}$ between the target and substrate. Charged particles within the dual field environment experience

the well-known Lorentz force in addition to electric field force, i.e.,

$$\vec{\mathbf{F}} = \frac{m\,d\vec{\mathbf{v}}}{dt} = -q(\vec{\mathscr{E}} + \vec{\mathbf{v}} \times \vec{\mathbf{B}}) \tag{4-12}$$

where q, m, and v are the electron charge, mass, and velocity, respectively. First consider the case where B and $\mathscr{E}$ are parallel as shown in Fig. 4-4a. Only electrons will be considered because as we have already seen, their dynamical behavior controls glow-discharge processes. When electrons are emitted exactly normal to the target surface or parallel to both B and $\mathscr{E}$, then $v \times B$ vanishes; electrons are only influenced by the $\mathscr{E}$ field and simply accelerate toward the anode, gaining kinetic energy in the process. If, however, $\mathscr{E} = 0$, and the electron is launched with velocity v at an angle θ with respect to the uniform B field between electrodes (Fig. 4-4b), it experiences a force $qvB \sin\theta$ in a direction perpendicular to B. The electron now executes a circular motion whose radius r is determined by a balance of the centrifugal ($m(v \sin\theta)^2/r$) and Lorentz forces involved, i.e.,

$$r = \frac{mv \sin\theta}{qB}.$$

A spiral electron motion ensues and in corkscrew fashion the electron returns to the same radial position around the axis of the field lines. If the magnetic field were not present, such off-axis electrons would tend to migrate out of the discharge and be lost at the walls.

The case where electrons are launched at an angle to parallel and uniform $\mathscr{E}$ and B fields is somewhat more complex (Fig. 4-4c). Helical motion with constant radius occurs, but because of electron acceleration in the $\mathscr{E}$ field the pitch of the helix lengthens with time. Time-varying fields complicate matters further and electron spirals of variable radius can occur. Clearly magnetic fields prolong the electron residence time in the discharge and enhance the probability of ion collisions.

4.3.4.2 Perpendicular Fields

Through the application of perpendicular electric and magnetic fields even greater electron confinement is achieved. The geometry is shown in Fig. 4-4d, where $\mathscr{E}$ is still normal to the cathode while B, which is directed into the page ($+z$ direction), lies parallel to the cathode plane. Electrons emitted normally from the cathode ideally do not even reach the anode but are trapped near the electrode where they execute a periodic hopping motion over its surface. Physically, the emitted electrons are initially accelerated

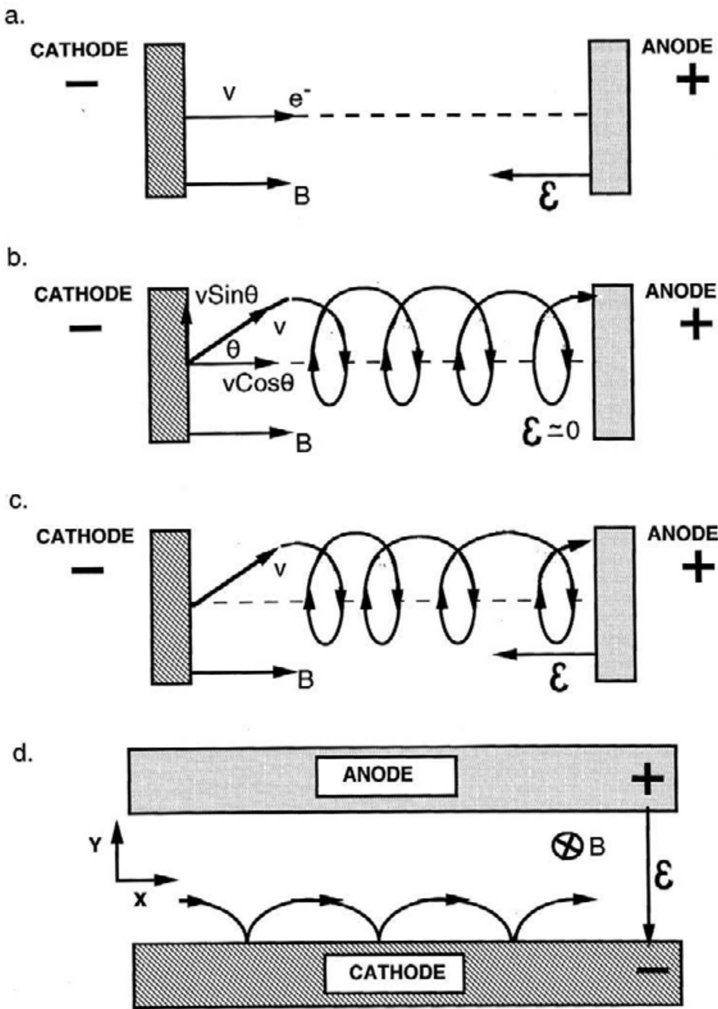


Figure 4-4 Effect of $\mathscr{E}$ and B on electron motion. (a) Linear electron trajectory when $\mathscr{E} \| B(\theta = 0)$. (b) Helical orbit of constant pitch when $B \neq 0$, $\mathscr{E} = 0(\theta \neq 0)$. (c) Helical orbit of variable pitch when $\mathscr{E} \| B(\theta \neq 0)$. (d) Cycloidal electron motion on cathode when $\mathscr{E} \perp B(\theta = 0)$.

toward the anode, executing a helical motion in the process; but when they encounter the region of the parallel magnetic field, they are bent in an orbit back to the target in very much the same way that electrons are deflected toward the hearth in an e-gun evaporator. The analysis for this behavior is not difficult and starts with the equations for electron motion in the three

perpendicular directions. Coordinate positions of the electron above and along the cathode are y and x, respectively. Applying the Lorentz equation we have

$$m_e\, dx^2/dt^2 = qB\, dy/dt \tag{4-13a}$$

$$m_e\, dy^2/dt^2 = q\mathscr{E} - qB\, dx/dt \tag{4-13b}$$

$$m_e\, dz^2/dt^2 = 0. \tag{4-13c}$$

By solving these coupled differential equations it is readily shown that the parametric equations of motion are

$$y = -\frac{q\mathscr{E}(1 - \cos \omega_c t)}{m_e \omega_c^2} \tag{4-14a}$$

and

$$x = \frac{\mathscr{E}t}{B}\left(1 - \frac{\sin \omega_c t}{\omega_c t}\right), \tag{4-14b}$$

where $\omega_c = qB/m_e$. Known as the cyclotron frequency, ω_c has a value of $2.8 \times 10^6 B$ Hz with B in gauss. Physically, these parametric equations describe a cycloidal motion where electrons repeatedly return to the cathode at time intervals of π/ω_c. The same motion is traced out by a point on the circumference of a circle rolling on a planar surface. Electron motion is strictly confined to the cathode dark space where both fields are present; if, however, electrons stray into the negative glow region where $\mathscr{E}$ is small, they describe a circular orbit before collisions may drive them either back into the dark space or forward toward the anode. Confinement in crossed fields prolongs the electron lifetime over and above that in parallel fields, enhancing the ionizing efficiency near the cathode. A denser plasma and larger discharge currents result. As we shall see in the next chapter, these effects are very widely capitalized upon in magnetron-sputtering processes.

4.3.5 Collective Charge Effects

4.3.5.1 The Debye Length

The behavior of plasmas derive largely from the Coulombic interactions among the charged species within them. Properly accounting for these electrostatic interactions is complicated but we can appreciate a bit of what is involved by considering the radial electric potential $V(r)$ around an isolated positive ion. This ion repels other ions and attracts a cloud of

electrons with a density given by

$$n_e(r) = n_i \exp qV(r)/k_B T. \tag{4-15}$$

The Boltzmann factor reflects the probability that electrons will acquire the energy needed to establish the electric potential at temperature T. Because n_e cannot deviate much from its average value (which is equal to the ion density n_i), V must be small. Therefore, by expanding the exponential, $n_e(r) = n_i(1 + qV(r)/k_B T)$. Furthermore, $V(r)$ must satisfy Poisson's equation, which in spherical coordinates takes the form

$$\frac{1}{r^2}\left[\left(\frac{d}{dr}\frac{[r^2\, dV(r)]}{dr}\right)\right] = -\frac{q(n_i - n_e)}{\varepsilon_0} = \frac{n_i q^2 V(r)}{\varepsilon_0 k_B T}, \tag{4-16}$$

where ε_0 is the permittivity of free space. Physically, the Poisson equation expresses a self-consistency condition that the potential due to the net electron density reproduces its potential energy. The Boltzmann term reflects the balance between the Coulombic attraction of electrons to the ion, and charge dispersal due to the thermal or kinetic energy of the electrons. Direct substitution shows that

$$V(r) = q/r \exp - (r/\lambda_D) \tag{4-17}$$

satisfies Eq. 4-16 where $\lambda_D = (\varepsilon_0 k_B T/n_i q^2)^{1/2}$. This solution, which has the form of an exponentially attenuated or screened Coulomb potential, also satisfies the boundary value $V = 0$ (the plasma potential) far from the point charge. The same sort of calculation can be performed for a charged planar electrode immersed in the plasma. Poisson's equation in one dimension (x) then yields a solution for the potential $V(x)$ that essentially varies as $\exp - (x/\lambda_D)$.

Known as the Debye length, λ_D is an important characteristic dimension in plasmas. If the plasma potential is perturbed by the point charge, λ_D is a measure of the size of the mobile electron cloud required to reduce V to 0.37 (i.e., $1/e$) of its initial value. Assuming $n_i = 10^{10}\,\text{cm}^{-3}$, and $k_B T = 2\,\text{eV}$, $\lambda_D = 1 \times 10^{-2}\,\text{cm}$. Outside of a sphere of radius λ_D there is effectively no interaction between the ion and the rest of the plasma. In the case of an inserted electrode, λ_D is a measure of the plasma sheath dimension (Section 4.3.7).

4.3.5.2 Electron Plasma Frequency

By evaluating its response to a perturbation, the ability of a plasma to protect its charge neutrality can be assessed. Consider that an external

electric field is suddenly turned on, displacing plasma electrons over some length. If it is just as suddenly turned off, the electron displacement induces a field that pulls the electrons back to their original position. But the inertia of the electrons will cause them to overshoot the mark and harmonically oscillate about the equilibrium site. This electron plasma (angular) frequency (ω_e), which has a magnitude of

$$\omega_e = (q^2 n_e / m_e \varepsilon_0)^{1/2} = 8.98 \times 10^3 n_e^{1/2} \text{ Hz}, \tag{4-18}$$

is a measure of the time required to restore charge equilibrium. The product of λ_D and ω_e is essentially equal to the electron velocity.

If the plasma is thought of as a dielectric medium analogous to a solid dielectric, then at frequencies less than ω_e the dielectric constant is high and the plasma appears opaque to such radiation. On the other hand the plasma becomes transparent to radiation at frequencies greater than ω_e where the dielectric constant drops. For $n_e = 10^{10}$ cm^{-3}, $\omega_e = 9 \times 10^8$ Hz, a frequency much larger than that typically used in AC (RF) plasmas.

4.3.5.3 Plasma Criteria

Ionized gases can be characterized as plasmas if they meet three criteria:

1. The system dimensions $\bar{D}$ must considerably exceed λ_D, i.e., $\bar{D} \gg \lambda_D$. Only in this way can the quasineutrality of the bulk of the plasma be ensured.
2. The total number (N_D) of shielding electrons drawn into the Debye sphere must be large; at the very least N_D should be greater than unity. By definition, $N_D = 4\pi\lambda_D^3 n_e/3$, and under the plasma conditions noted above, $N_D \sim 4 \times 10^4$.
3. Electrons should interact more strongly with each other than with the neutral gas. Under these conditions, particle motion in the plasma will be controlled by electromagnetic forces rather than by gas fluid dynamics.

4.3.6 AC Effects in Plasmas

It is instructive to analyze the kinetic behavior of electrons in an AC discharge in order to appreciate how plasmas are sustained. After all, it is not obvious that, in their to and fro motion in the field, electrons would absorb and gain sufficient energy to cause enhanced ionization of neutrals. Assuming no collisions with neutrals, the resultant harmonic motion of the electrons resembles the oscillations of a spring. Therefore we may write

$$m_e \, dx^2/dt^2 = -q\mathscr{E}_0 \sin \omega t \tag{4-19}$$

where x, m_e, and q are the electronic displacement, mass, and charge, respectively, and t is the time. Furthermore, the electric field $\mathscr{E}$ is equal to $\mathscr{E}_0 \sin \omega t$, with $\mathscr{E}_0$ and ω the field amplitude and circular frequency, respectively. From this basic equation and its solution it is a simple matter to show that the maximum electron displacement amplitude x_0 and energy E_0 are given by

$$x_0 = q\mathscr{E}_0/m_e\omega^2 \tag{4-20}$$

and

$$E_0 = (q\mathscr{E}_0)^2/2m_e\omega^2. \tag{4-21}$$

We may now estimate what field strength is required to ionize argon, whose ionization energy (E_0) is 15.7 eV. For the commonly employed radio frequency 13.56 MHz ($\omega = 2\pi \times 13.56 \times 10^6$ Hz), $\mathscr{E}_0$ is calculated to be 11.5 V/cm, an easily attainable field in typical plasma reactors. No power is absorbed in the collisionless harmonic motion of electrons, however. But when the electrons undergo inelastic collisions their motion is randomized and power is effectively absorbed from the RF source. Even smaller values of $\mathscr{E}_0$ can produce ionization if, after electron–gas collisions, the reversal in electron velocity coincides with the changing electric-field direction. Through such effects RF discharges are more efficient than their DC counterparts in promoting ionization.

The question arises of how to generate AC discharges. Interestingly, provided the frequency is high enough, reactors can be built without interior plate electrodes as shown schematically in Fig. 4-5. For example, a coil wrapped around a tubular reactor can inductively couple power to the gas inside ionizing it. So, too, can capacitor plates on the outside; in such a case we speak of capacitive coupling. Such electrodeless reactors have been used for etching films. However, for the deposition of films by RF sputtering, internal cathode targets are required (Section 5.2.4).

4.3.7 Electrode Sheaths

We have already seen that immersion of a floating electrode into a plasma causes it to charge negatively because of the disparity (in mass, velocity, and energy) between electrons and ions. As a consequence, in a glow discharge we can expect that both the anode and cathode surfaces will be at a negative floating potential (V_f) relative to the plasma potential (V_p). Of course, application of the large external negative potential alters the situation, but the voltage distribution in the DC glow discharge shown in Fig. 4-6 (also Fig. 4-3) can be qualitatively understood in these terms. In

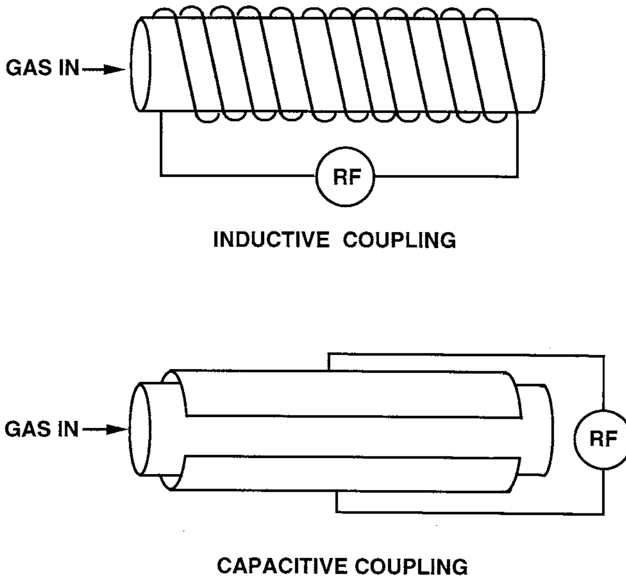


Figure 4-5 Inductively and capacitively coupled tubular RF plasma reactors.

essence a Debye-like, positive space charge layer shields the negative surface; we now speak of a plasma *sheath* of potential V_s ($V_s = V_p - V_f$) that envelops the electrode and repels electrons. As noted earlier, the lower electron density in the sheath means less ionization and excitation of neutrals. Hence, there is less luminosity there than in the glow itself and the sheath appears dark. Large electric fields are restricted to the sheath regions. It is at the

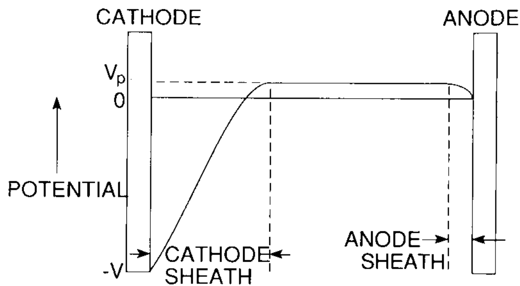


Figure 4-6 Voltage distribution across DC glow discharge. Note cathode sheath is wider than anode sheath.

sheath–plasma interface that ions begin to accelerate on their way to the target during sputtering. The plasma itself is not at a potential intermediate between that of the electrodes but is typically some ~15 volts positive with respect to the anode. In essence the chamber walls charge negatively by the same mechanism that the electrodes do, leaving the plasma at positive potential V_p.

It is not difficult to quantitatively sketch the magnitude of the potential energy barrier $q(V_p - V_f)$ electrons face in moving from the plasma to the cathode surface through the sheath. The number of electrons (n_e^*) that can gain enough energy to surmount this barrier is given by

$$\frac{n_e^*}{n_e} = \exp -\frac{q(V_p - V_f)}{k_B T_e}. \tag{4-22}$$

A Maxwell–Boltzmann-type expression of this kind is ubiquitous in describing the probability that a species will exceed a given energy barrier; thus, n_e^*/n_e represents the fractional probability of success in acquiring the requisite energy. After accounting for the electrical flux balance between electrons and ions, the equation

$$V_p - V_f = k_B T_e/2q \ln(m_i/2.3m_e) \tag{4-23}$$

has been derived. Since m_i is 3–4 orders of magnitude higher than m_e, the sheath potential will be several times the electron temperature in eV, e.g., ~10 eV.

Ion current flow through the cathode sheath is an important issue because all thin-film processing in plasmas depends on it. In this regard the interesting question arises as to whether ion motion in the sheath occurs in a "free-fall," collisionless manner, or through "mobility limited" motion involving repeated collisions with other gas species. Thus for voltage V applied across a sheath of thickness d_s, two different formulas govern the current density (j) through it, namely,

$$j = \frac{4\mathscr{E}_0}{9d_s^2} \cdot \left(\frac{2q}{m}\right)^{1/2} V^{3/2} \qquad \text{(free fall or space charge limited)} \tag{4-24}$$

and

$$j = \frac{9\mathscr{E}_0}{8d_s^3} \cdot V^2 \qquad \text{(mobility limited).} \tag{4-25}$$

It turns out that Eq. 4-24, also known as the Child–Langmuir equation, better describes the measured cathode-current characteristics; this is certainly true at low pressures where few collisions are likely.

The sheath dark space is sometimes visible with the unaided eye and is therefore considerably larger than calculated Debye lengths of $\sim 100\,\mu$m. This means that a large planar surface behaves differently from a point charge when both are immersed in plasmas. Instead of electrons shielding a point charge, bipolar diffusion of both electrons and ions is required to shield an electrode, and this physically broadens the sheath dimensions. A useful formula (Ref. 5) suggests that the relation between d_s and λ_D is

$$d_s \sim [q(V_p - V_f)/k_B T_e]^a \lambda_D \tag{4-26}$$

where constant a ranges between $\frac{2}{3}$ at higher pressures and $\frac{3}{4}$ at lower pressures. Thus, d_s is typically tens of times larger than λ_D.

4.4 REACTIONS IN PLASMAS

To initiate and sustain plasma reactions, collisions between involved species are required. We may think of reactions as having both physical and chemical attributes. An example of the former is the collision between an electron and an Ar atom. In this case the physical processes of ionization and ion multiplication within the plasma dominate the properties of the discharge. However, when we consider discharges in reactive rather than inert gases, chemical reactions often occur as a result of collisions involving ions, atoms, molecules, and assorted excited and metastable variants of these. An elementary description of the assorted physical and chemical interactions and reactions between and among the assorted species within the plasma is the substance of this section. In particular, we are interested in the energies that are exchanged and the rates at which these collisions occur.

4.4.1 COLLISION PROCESSES

Collisions are either *elastic* or *inelastic* depending on whether the internal energy of the colliding species is preserved or not. In an elastic collision, exemplified by the billiard-ball analogy of elementary physics and depicted in Fig. 4-7a, only kinetic energy is exchanged; we speak of conservation of both momentum and translational *kinetic energy*. On the other hand the *potential energy* basically resides within the electronic structure of the colliding entities; increases in potential energy are manifested by ionization or other excitation processes. In an elastic collision no atomic excitation occurs and potential energy is conserved. This is why only kinetic energy is considered in the calculation. The well-known result for the elastic binary

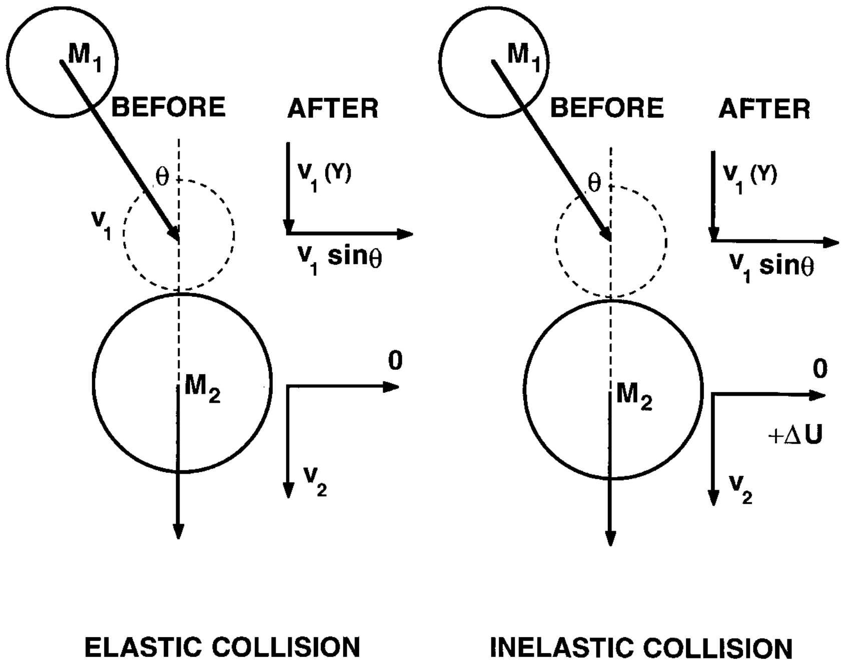


Figure 4-7 Models of (a) elastic and (b) inelastic collisions between moving (1) and stationary (2) particles of masses M_1 and M_2, respectively.

collision between a moving particle of mass M_1 and an initially stationary particle of mass M_2 is

$$\frac{E_2}{E_1} = \frac{\frac{1}{2}M_2 v_2^2}{\frac{1}{2}M_1 v_1^2} = \frac{4M_1M_2}{(M_1 + M_2)^2}\cos^2\theta. \tag{4-27}$$

We assume collision occurs at an angle θ defined by the initial trajectory of M_1 and the line joining the mass centers at contact. The quantity $4M_1M_2/(M_1 + M_2)^2$, known as the energy transfer function γ, represents the ratio of the kinetic energy (E_2) acquired by M_2 relative to the kinetic energy (E_1) of M_1. When $M_1 = M_2$, γ has a value of 1, i.e., after collision the moving projectile is brought to a halt and all of its energy is efficiently transferred to M_2, which speeds away. When, however, $M_1 \ll M_2$ reflecting, say, a collision between a moving electron and a stationary nitrogen molecule, then the energy transfer function is $\sim 4M_1/M_2$ and has a typical value of $\sim 10^{-4}$. Very little kinetic energy is transferred in the elastic

collision between the electron and nitrogen molecule. This same formula albeit with modification is incorporated in theories used to describe ion collisions with surface atoms that ultimately result in the ejection of atoms (sputtering).

Now consider inelastic collisions (Fig. 4-7b). The change in internal energy, ΔU, of the struck particle must now be accounted for under the condition requiring conservation of total energy. It is left as an exercise for the reader to demonstrate that the maximum fraction of kinetic energy transferred is given by (Ref. 4)

$$\frac{\Delta U}{\frac{1}{2}M_1 v_1^2} = \frac{M_2}{M_1 + M_2}\cos^2\theta, \tag{4-28}$$

where v_1 is the initial velocity of particle 1. For the inelastic collision between an electron and nitrogen molecule, $\Delta U/[\frac{1}{2}M_1 v_1^2] \sim 1$, when $\cos\theta = 1$. Therefore, in contrast to an elastic collision, virtually all of an electron's kinetic energy can be transferred to the heavier species in the inelastic collision.

4.4.2 Cross-Sections

In Section 2.2.2 the collision diameter d_c was introduced in connection with mean free paths (λ_{mfp}) of colliding gas atoms or molecules in a reduced-pressure environment. The plasmas we will deal with have a sufficient number of gas-phase atoms, molecules, and ions so that collisions and reactions involving these species occur with some frequency. To quantitatively deal with these processes we first define the collision cross-section σ_c, a circular area of magnitude πd_c^2, that reflects the probability of interaction or collision between particles (Refs. 4, 5). The larger σ_c is, the greater is the chance that other particles will encounter it. If the concentration of the gas species is n (number/cm^3), then the preceding quantities are related by

$$1/\lambda_{mfp} = n\sigma_c. \tag{4-29}$$

Although both λ_{mfp} and σ_c characterize collisions, λ_{mfp} is usually reserved for elastic collisions. On the other hand σ_c has broader applicability because it characterizes inelastic collisions as well.

As an example of a collision process let us consider ionization of an inert noble gas atom due to electron impact. Ionization cross sections, σ_e, for such gases are plotted as a function of E in Fig. 4-8. Units of σ_e are in 8.88×10^{-17} cm^2, which corresponds to the circular area associated with the Bohr radius (a_0), i.e., $a_0 = 0.53 \times 10^{-8}$ cm. The ionization energy thresh-

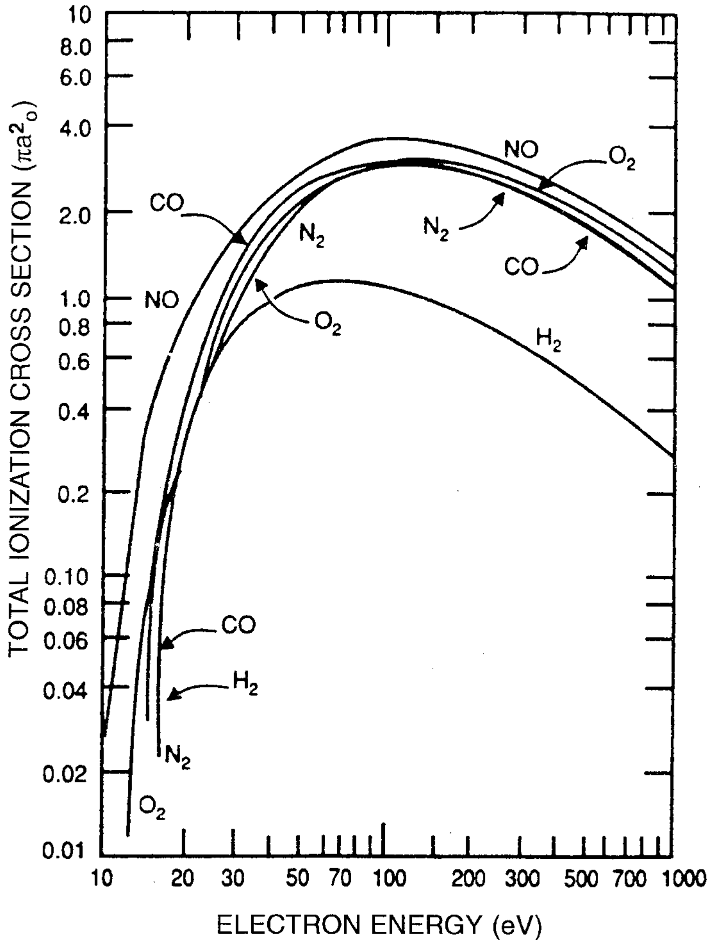


Figure 4-8 Total ionization cross sections for various gases plotted as a function of energy. (From S. C. Brown, *Basic Data of Plasma Physics*, 2nd ed. MIT, Cambridge, MA, 1967. Reprinted with the permission of The MIT Press.)

old (E_{th}) is the minimum energy required to eject the most weakly bound electron; typically, E_{th} is 15–20 eV. Because no ionization occurs for electrons impacting with energy (E) below E_{th}, the ionization cross section σ_e is zero. As E rises so does σ_e, because a greater number of accessible electron levels means an increasingly greater ionization probability. A maximum is reached at E values of $\sim$100 eV, after which σ_e declines.

In addition to ionization of atoms or molecules, inelastic collisions by electrons may also lead to internal vibrational and rotational excitation of

these species with cross sections given by σ_v and σ_r, respectively (Refs. 4, 7). Attachment and dissociation reactions characterized by σ_a and σ_d may also occur. It is common to add the various σ to σ_e and define a total cross-section σ_T for the reaction process, i.e., $\sigma_T = \sigma_e + \sigma_v + \sigma_r + \sigma_a + \sigma_d + \cdots$. Each constituent cross section contributes its particular energy dependence so the overall variation of σ_T vs E is very complex. The overall value σ_T is applicable when describing plasma reactions in a macroscopic sense.

4.4.3 Plasma Chemistry

Thus far we have primarily considered inert-gas plasmas and physical interactions. Now we turn our attention to far more complex plasmas that contain multicomponent species in assorted activated states. These undergo the kinds of chemical reactions that occur in the plasma-enhanced etching and chemical vapor deposition processes discussed in Chapters 5 and 6, respectively. A brief summary of the rich diversity of inelastic collisions and reactions that occur in the gas phase is given in Table 4-1, where both generic examples and actual reactions are noted. In addition to electron collisions listed first, ion–neutral as well as excited or metastable-ion–excited and excited atom–neutral collisions also occur. Evidence for these uncommon gas-phase species and reactions has accumulated through real-time monitoring of discharges by mass as well as light-emission spectroscopy. As a result, a remarkable picture of plasma chemistry has emerged. For example, a noble gas such as Ar when ionized loses an electron and resembles Cl electronically as well as chemically.

Simple associations and comparisons with traditional gas-phase chemistry disguise the complexity of plasma reactions. Our first inclination may be to think that plasmas represent a tractable perturbation on traditional gas or gas–solid chemistry. But unlike the several hundred degrees K characteristic of ordinary or free atoms and molecules, electron temperatures are tens of thousands of kelvins. Under these conditions activated and charged atomic and molecular species and compounds are created. These participate in *homogeneous* gas-phase chemical reactions that are not driven thermally but rather by the energetics of the discharge; this means they occur by nonthermal and nonequilibrium processes. In addition, there are the plasma-modified gas-solid or *heterogeneous* reactions to contend with. These possess their peculiar collection of active reactants, modified surfaces, and resultant metastable and stable products.

In conclusion, plasma reactions are not in equilibrium and react in complex ways that confound thermodynamic and kinetic descriptions of them.

Table 4-1
Chemical Reactions in Plasmas

A. Electron collisions

Type	Generic reaction	Example reaction
Ionization	$e^- + A \rightarrow A^+ + 2e^-$ $e^- + A_2 \rightarrow A_2^+ + 2e^-$	$e^- + O \rightarrow O^+ + 2e^-$ $e^- + O_2 \rightarrow O_2^+ + 2e^-$
Recombination	$e^- + A^+ \rightarrow A$	$e^- + O^+ \rightarrow O$
Attachment	$e^- + A \rightarrow A^-$ $e^- + AB \rightarrow AB^-$	$e^- + F \rightarrow F^-$ $e^- + SF_6 \rightarrow SF_6^-$
Excitation	$e^- + A_2 \rightarrow A_2^* + e^-$ $e^- + AB \rightarrow (AB)^* + e^-$	$e^- + O_2 \rightarrow O_2^* + e^-$
Dissociation	$e^- + AB \rightarrow A^* + B^* + e^-$	$e^- + CF_4 \rightarrow CF_3^* + F^* + e^-$
Dissociative ionization	$e^- + AB \rightarrow A + B^+ + 2e^-$	$e^- + CF_4 \rightarrow F + CF_3^+ + 2e^-$
Dissociative attachment	$e^- + A_2 \rightarrow A^+ + A^- + e^-$	$e^- + N_2 \rightarrow N^+ + N^- + e^-$

B. Atom–ion–molecule collisions

Type	Generic reaction
Symmetrical charge transfer	$A + A^+ \rightarrow A^+ + A$
Asymmetric charge transfer	$A + B^+ \rightarrow A^+ + B$
Metastable–neutral (Penning ionization)	$A^* + B \rightarrow B^+ + A + e^-$
Metastable–metastable ionization	$A^* + B^* \rightarrow B + A^+ + e^-$

4.4.4 Chemical Reaction Rates

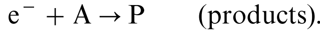
For convenience we have divided chemical reactions in plasmas into two categories, i.e., electron–atom (or –molecule) and molecule–molecule. Of the two, the latter reactions are perhaps more readily understood in terms of classical chemical reaction rate theory. Consider, for example, the bimolecular reaction, $A + B \rightarrow P$ (products). We may then write (Ref. 5)

$$d\,n_P/dt = k_{AB}(T)n_A n_B \tag{4-30}$$

where concentrations of the involved species are denoted by n. Rate constant $k_{AB}(T)$ is expected to be thermally activated, i.e., $k_{AB}(T) = k_0 \exp - (E/k_B T)$, with k_0 and E characteristic constants of the reaction. Plasma-etching reactions have been analyzed (Ref. 8) employing these concepts and the

Arrhenius dependence of the rate constant conclusively demonstrated (see Section 5.4.4).

Now consider electron collision reactions generically denoted by



$$e^- + A \rightarrow P \qquad \text{(products)}.$$

For this case, ionization, excitation, attachment, etc., products form at a rate given by

$$dn_P/dt = k(e, T)n_e n_A \tag{4-31}$$

where $k(e, T)$ in typical units of cm^3/s is the rate constant. Unlike reactions characterized by Eq. 4-30 where the neutrals or ions are close to translational equilibrium, the electron energies are well above thermal equilibrium values. Because of this $k(e, T)$ depends strongly on the electron energy (E_e) as well as the electron-energy distribution function $f(E_e)$. Like the velocity probability distribution function of Eq. 2-1, it is common to approximate the temperature dependence of $f(E_e) = (1/n_e)\, dn_e/dE_e$ by a similar Maxwell–Boltzmann-like expression. But $k(e, T)$ is also proportional to the particular collision cross section ($\sigma_T(E)$) and electron velocity ($v_e(E)$) so that by integrating over the range of energies we obtain (Ref. 5)

$$k(e, T) = \int_0^\infty f(E_e)v_e(E)\sigma_T(E)\, dE. \tag{4-32}$$

It is well beyond the scope of this book to continue much further except to note that rate constants for many plasma reactions have been theoretically estimated. In view of the complexities involved, values of $k(e, T)$ so determined may illuminate trends but often have limited quantitative value. Suffice it to say that plasma generated reactions generally enhance chemical-vapor deposition and film etching processes. Thus gas-phase chemical reactions will usually occur more rapidly and at lower tempertatures with benefit of plasma assist.

4.5 PHYSICS OF SPUTTERING

4.5.1 AN INTRODUCTION TO ION–SURFACE INTERACTIONS

Aside from occasional references to plasma sheaths, this chapter has been almost totally concerned with events occurring within the plasma gas phase. However, films and surfaces immersed in plasmas or exposed to impinging ions are subjected to one or more of the interactions shown schematically in Fig. 4-9. Critical to the implementation of thin-film

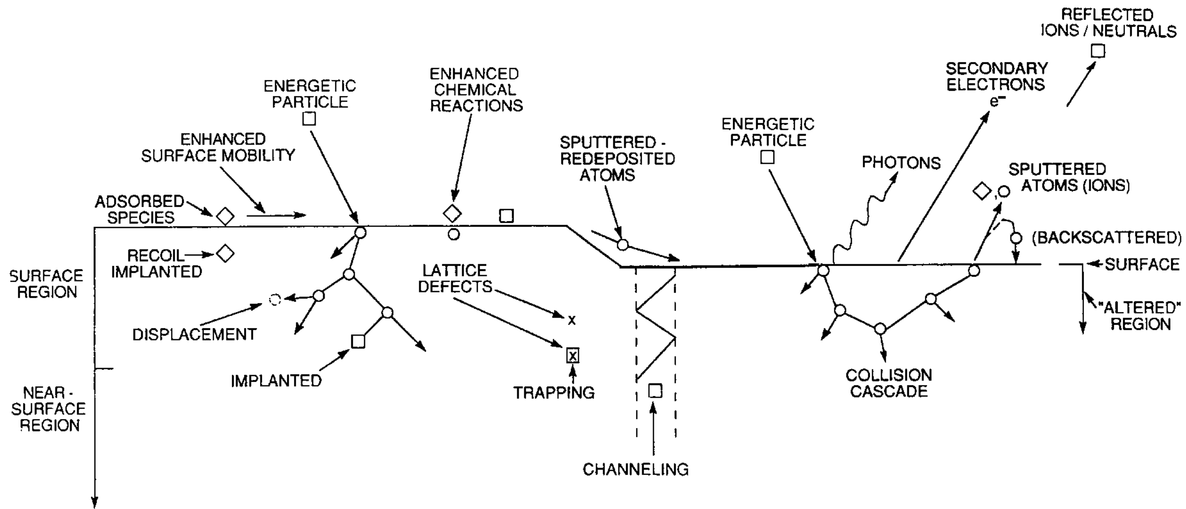


Figure 4-9 Depiction of energetic-particle bombardment effects on surfaces and growing films. (From Ref. 9.)

processing, characterization of resulting films, and modification of their properties is an understanding of such ion–surface interactions. This is a large subject and the following comments and distinctions may prove helpful in guiding the reader through our treatment of it.

1. Ions that bombard surfaces can arise from both plasmas and ion beams.

2. Upon bombarding a surface incoming ions may be reflected back, stick or adsorb, scatter, eject or sputter surface atoms, or get buried in subsurface layers (ion implantation). Surface heating, chemical reactions, atom mixing, and alteration of surface topography are other manifestations of ion bombardment.

3. Ion beam energy is critical in defining the nature of the interaction with surfaces by changing the probability of surface sticking and reaction (Ref. 10). Thus at kinetic energies less than $\sim 10^{-2}$ eV (the thermal energy $k_B T$ at room temperature), the sticking probability, defined as the ratio of the number of product or deposited atoms to the number of impinging ions, is usually unity; therefore, condensation as well as chemisorption (Section 7.2.5) occurs readily. As shown in Fig. 4-10, from $\sim 10^{-2}$ eV to $\sim 10^{4}$ eV the ion sticking probability typically drops, reaching a minimum of ~ 0.2 at 20 eV, but thereafter it rises with increasing energy to about 0.6; important sputtering processes occur in this ion-energy range. In the regime of ion

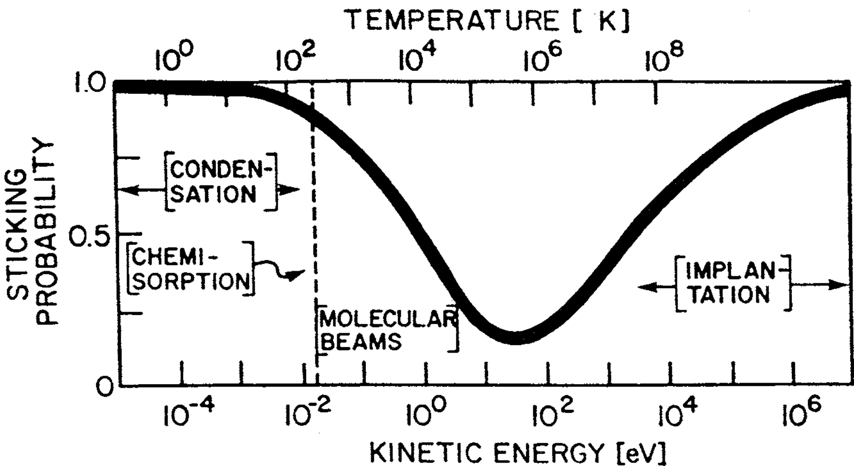


Figure 4-10 Particle-sticking probability as a function of energy. The dashed vertical line corresponds to room-temperature thermal energy. (From S. R. Kasi, H. Kang, C. S. Sass, and J. W. Rabalais, *Surface Science Reports* **10**, Nos. 1/2, p. 1 (1989). Reprinted with the permission of Elsevier Science Publishers and Professor J. W. Rabalais.)

implantation from roughly 10^4 eV and above (up to $\sim 10^6$ eV), the sticking probability again rises to near unity as ions are buried beneath the surface. In this range of energies the sputtering probability is small. In addition to ion energy, other important variables include type of ion (mass, charge), the nature of surface as well underlying atoms, and the film or substrate crystallography and texture.

4. When generated within typical glow discharges, ion energies range from a few to a hundred electron volts. On the other hand, ion beams possessing well-defined, generally higher energies are usually intended for processing in vacuum. For low ion energies (~ 1 keV) specially designed broad beam ion guns (see Section 5.5.4.2) are employed, while ion accelerators generate high-energy beams. In general, plasmas and low-energy ion beams are utilized during film deposition and etching processes. However, high-energy ion beams are primarily used for ion implantation, and to a lesser extent for the surface modification of both bulk solids and previously deposited thin films.

5. Ion bombardment of surfaces is exploited in two different ways during the sputter deposition of thin films. Sputtering occurs at cathodes as a result of ion transport through the dark space and impact with the target. However, where substrates are located, ion bombardment by plasma species simultaneously serves to modify the properties of the depositing sputtered film.

6. In many applications where they are used, e.g., ion implantation and ion milling, the ion beams are generally "pure" with respect to mass and charge, monoenergetic, and well focused, and they impinge with a controlled geometry on surfaces maintained under vacuum. In contrast, ions emanating within plasmas possess a broader energy distribution, travel along more random trajectories, and coexist with other particles in an elevated-pressure environment.

7. As a result of ion bombardment, assorted charged particles (e.g., electrons, ions), neutrals, and photons of varying energies and abundances are emitted from the surface. Contained within them is a rich source of compositional and structural information on surface properties. Therefore, films can be characterized by detecting and analyzing these emitted signals. Several ion-beam techniques for achieving these ends, e.g., Rutherford backscattering (RBS) and secondary ion mass spectroscopy (SIMS), are described in Chapter 10.

In keeping with an earlier stated objective, the remainder of this chapter will focus on the scientific principles underlying ion–surface interaction phenomena as they relate to the deposition of thin films. We focus on sputtering in this section and conclude with film modification effects in Section 4.6. The practical benefits of ion bombardment during film growth

as well as etching will be better appreciated in the context of plasma processing and the issues they raise; these subjects are therefore deferred to the next chapter.

4.5.2 SPUTTERING

4.5.2.1 The Moment of Impact

It is instructive to record the sequence of events that occurs as an energetic ion approaches and impinges on a surface (Ref. 11). The first thing that happens is electron exchange when they are angstroms apart. Such processes are extremely fast, occurring within $\sim 10^{-15}$ s, and result in electronic excitations such as Auger-electron tunneling transitions. Because work functions of most solids are less than the ionization potentials of most gases, the latter capture electrons from the former. Thus, the scattered and recoiled species with keV energies are largely neutral. When the ion–solid encounter distance further decreases, the separate atoms of atomic number Z_1 (ion) and Z_2 (surface) evolve into molecular orbitals of a quasimolecule and finally into the atomic orbitals of an unstable but united atom of atomic number $Z_1 + Z_2$. As the encounter distance shrinks even more, electronic repulsion and the Pauli exclusion principle begin to dominate, resulting in atomic separation and collisional reionization of neutrals. This may be viewed as the moment of collision. During the collision step several processes are possible depending on what is impacted and with what energy. For example, if an ion strikes an atom of a molecule, the latter may dissociate.

Reflection of incoming ions from surfaces is also a possibility. Reversal of ion motion and return to the vapor phase becomes more probable the closer the angle θ in Eq. 4-27 is to zero, i.e., normal impact, and the larger M_2 is relative to M_1. Thus, bombarding ions of low energy often get adsorbed on the surface.

4.5.2.2 Sputter Yields

When the ion impact establishes a train of collision events in the target, leading to the ejection of a matrix atom, we speak of sputtering. Since sputtering is the result of momentum transfer it has been aptly likened to "atomic pool" where the ion (cue ball) breaks up the close-packed rack of atoms (billiard balls), scattering some backward (toward the player). The sputter yield S is defined as

$$S = \frac{\text{Number of sputtered atoms}}{\text{Incident particle}} \qquad (4\text{-}33)$$

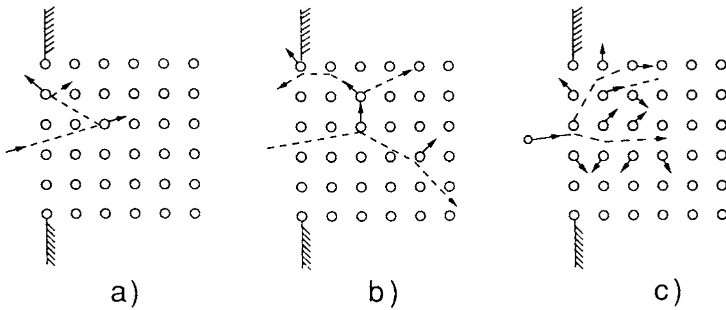


Figure 4-11 Three energy regimes of sputtering. (a) Single knock-on (low energy), (b) linear cascade, (c) spike (high energy). (After P. Sigmund.)

and is a measure of the efficiency of sputtering. Experimental values of S ranging from 10^{-5} to as high as 10^3 have been reported (Ref. 11). However, a narrower two orders of magnitude range in S from 10^{-1} to 10 characterizes most practical sputtering processes. Three regimes of sputtering have been identified, and they are schematically depicted in Fig. 4-11; much has been written on all three of these, but we shall only consider the first two regimes here.

4.5.2.2.1 Single Knock-On

In this case ion–surface collisions set target atoms in motion and may simply give rise to separate knock-on events. If enough energy is transferred to target atoms, they overcome forces that bind them and sputter. The *threshold energy*, E_{th}, is the minimum energy required to do this. Typical values for E_{th} range from 5 to 40 eV and depend on the nature of the incident ion, and on the mass and atomic number of the target atoms. Most important, however, is the binding energy of atoms to the surface (U_S), a quantity that appears in all theoretical estimates of E_{th}. Typically, U_S may be assumed to be the heat of sublimation or vaporization and ranges between 2 and 5 eV. A chronological summary of these estimates is given by Malherbe (Ref. 11) together with ranges of applicability. Two of the simplest approximations include $E_{th} = 4U_S$, for $0.08 < M_1/M_2 < 1$, and $E_{th} = U_S/\gamma$, where γ, the energy transfer function, was defined earlier (Section 4.4.1). In the last expression γ essentially magnifies the value of U_S by accounting for the fraction of the ion energy transferred in the collision. Experimentally measured values for E_{th} are entered in Table 4-2 for a number of metals and semiconductors.

Table 4-2
Sputtering Yield Data for Metals (Atoms/Ion) and Semiconductors (Molecules/Ion)

Sputtering gas energy (keV) →	He 0.5	Ne 0.5	Ar 0.5	Kr 0.5	Xe 0.5	Ar 1.0	Ar threshold voltage (eV)
Ag	0.20	1.77	3.12	3.27	3.32	3.8	15
Al	0.16	0.73	1.05	0.96	0.82	1.0	13
Au	0.07	1.08	2.40	3.06	3.01	3.6	20
C	0.07	—	0.12	0.13	0.17		
Co	0.13	0.90	1.22	1.08	1.08		25
Cu	0.24	1.80	2.35	2.35	2.05	2.85	17
Fe	0.15	0.88	1.10	1.07	1.0	1.3	20
Ge	0.08	0.68	1.1	1.12	1.04		25
Mo	0.03	0.48	0.80	0.87	0.87	1.13	24
Ni	0.16	1.10	1.45	1.30	1.22	2.2	21
Pt	0.03	0.63	1.40	1.82	1.93		25
Si	0.13	0.48	0.50	0.50	0.42	0.6	
Ta	0.01	0.28	0.57	0.87	0.88		26
Ti	0.07	0.43	0.51	0.48	0.43		20
W	0.01	0.28	0.57	0.91	1.01		33
GaAs		0.10	0.83			1.52	20–25
InP			1.00			1.4	25
GaP			0.87				36
SiC		0.13	0.40				17
InSb			0.50				

From Refs. 4 and 6. Compound semiconductor data for normal ion incidence (Ref. 11).

4.5.2.2.2 Linear Collision Cascade

At higher ion energies, one or more so-called linear collision cascades are initiated. When this happens the density of recoils is sufficiently low so that most collisions involve one moving and one stationary particle, instead of two moving particles. The result of such processes is sputtering, i.e., the ejection of target atoms. Sputtering in the linear collision-cascade regime has been theoretically modeled by many investigators, but the theory due to Sigmund (Ref. 12) is the most widely accepted and used to describe this process. This theory views S as a product of two terms, namely,

$$S = \Lambda F_{\mathrm{D}}(E), \tag{4-34}$$

where the first term, denoted by Λ, is a materials constant. It reflects properties such as U_S, the range of displaced target atoms, and the number of recoil atoms that overcome the surface barrier of the solid and escape. The second term, F_D, accounts for the energy deposited at the surface and depends on type, energy, and incident angle of the ion, as well as on target parameters. In particular, the energy loss ions suffer through repeated nuclear collisions as they traverse the target is the key factor. Specifically, $F_D = \alpha N S_n(E)$, where $S_n(E)$ is defined as the nuclear stopping power and $NS_n(E)$ is the nuclear energy loss, $(dE/dz)_n$; furthermore, N is the atomic density of the solid, and $\alpha(M_2/M_1, \theta)$ can range from 0.1 to 1.4 depending on mass ratio and angle of impact, but often assumes a value between 0.2 and 0.4. The Sigmund theory provides the specific dependence of S on E for both low and high energies. At low energy

$$S = \frac{3\alpha 4 M_1 M_2 E}{4\pi^2 (M_1 + M_2)^2 U_S} \qquad (E < 1\,\text{keV}). \tag{4-35}$$

However, for energies above 1 keV, $\Lambda = 0.042/NU_S$ (Å/eV) and therefore,

$$S = 0.042\alpha S_n(E)/U_S. \tag{4-36}$$

As an exercise (Ref. 13), let us calculate S for argon incident on copper in the approximation that $S_n(E)$ is independent of energy. For Cu, $NS_n =$ 124 eV/Å and $N = 8.47 \times 10^{-2}$ atoms/Å^{-3}, yielding $S_n = 1464$ eV-Å^2/atom. Further, assuming $U_S = 3$ eV and $\alpha = 0.25$, substitution in Eq. 4-36, gives $S = 5.1$, a value that compares with the measured value of ~ 2.6.

When the energy dependence of S is required, $S_n(E)$ must be known and in one approximation it takes the form

$$S_n = 4\pi a \frac{Z_1 Z_2 q^2 M_1 s'_n(E)}{M_1 + M_2}. \tag{4-37}$$

Here Z_1 and Z_2 are the atomic numbers of the projectile and target atoms, respectively, a is the effective radius (0.1 to 0.2 Å) over which nuclear charge is screened by electrons during the collision, q is the electronic charge, and $s'_n(E)$ is a *reduced* nuclear stopping cross section, whose values have been tabulated (Refs. 12, 14).

The sputter yields for a number of metals and semiconductors are entered in Table 4-2. Readers should be aware that there is much scatter in these determinations among different investigators. For the metals, values at two different energies (0.5 keV and 1.0 keV) as well as five different inert gases (He, Ne, Ar, Kr, and Xe) are listed. It is apparent that S values typically span a range from 0.01 to 4, and increase with the mass and energy of the sputtering gas. Data on the energy dependence of the sputter yield of Al by

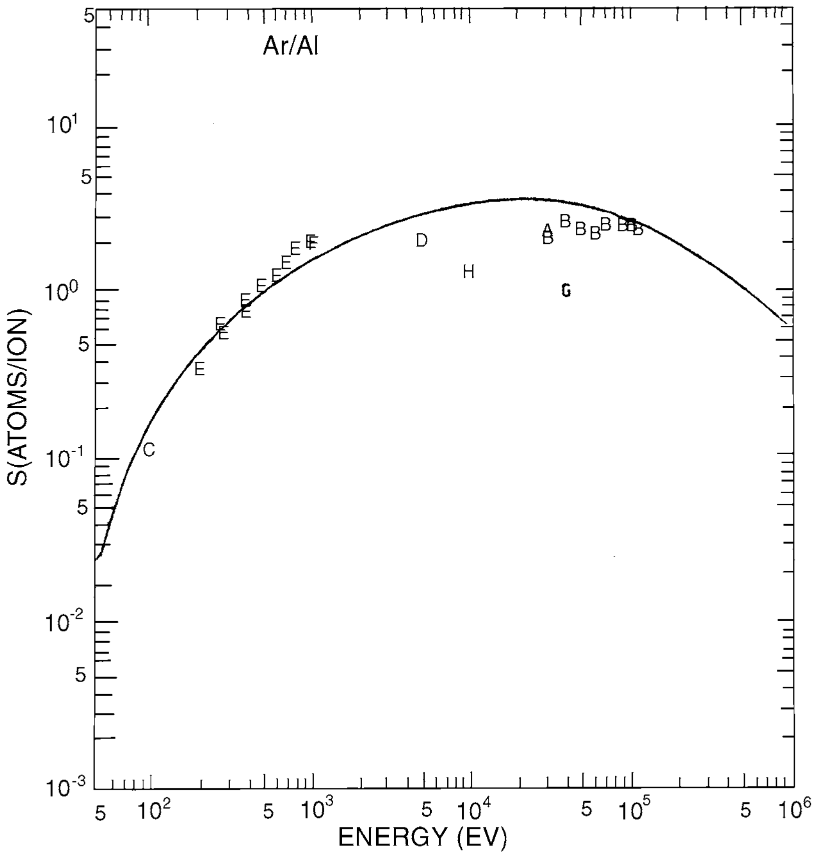


Figure 4-12 Sputter-yield values for Al as a function of energy. Letters on the plot refer to data from the following investigators: A. Yonts, Normand, and Harrison (1960); B. Fert, Colombie, and Fagot (1961); C. Laegreid and Wehner (1961); D. Robinson and Southern (1967); E. Weijsenfeld (1967); F. Oechsner (1973); G. Braun, Emmoth, and Buchta (1976); H. Okajima (1981). (From N. Matsunami, *et al.*, *AT. Data. Nucl. Data Tables* **31**, 1 (1984). Reprinted with the permission of Academic Press, Inc.)

Ar, obtained by eight different investigators over a 20-year span, are plotted in Fig. 4-12. The roughly linear rise in S with E, the peaking at approximately 10 keV, and the subsequent decline at energies above 100 keV is typical for many metals and basically reflects the energy dependence of $S_n(E)$.

One of the recent attempts to model sputtering in the linear collision cascade regime is due to Mahon and Vantomme (Ref. 15). They considered

S to be a product of three factors given by

$$S = (E/E_{avg})(R_e/R_p)(1/4). \tag{4-38}$$

This expression reflects the respective energy, spatial, and directional distribution factors that characterize the sputtering process. The first term E/E_{avg} is the effective number of recoiling target atoms created per incident ion penetration, where E_{avg} is the average energy at the practical endpoint of the cascade. Factor R_e/R_p essentially represents the ratio of the range of recoiling atoms at the surface (R_e) to the range of the projectile ion (R_p) in the target; R_p is dependent on energy loss with distance. Finally, factor $\frac{1}{4}$ reflects the fraction of atoms close to the surface, possessing the requisite energy to escape, that are traveling in the right direction. Spatial averaging of ion emission obeying a cosine distribution law is used to obtain this value.

The theory was applied to aluminum and tungsten sputtered with $E =$ 1 keV argon ions. Upon substitution of appropriate values for the factors in Eq. 4-38, a value of $S = 1.34$ was obtained for Al. Proceeding in this manner the calculated values of S as a function of E were obtained and plotted in Fig. 4-13. A comparison with measured sputter yields for Al (Fig. 4-12)

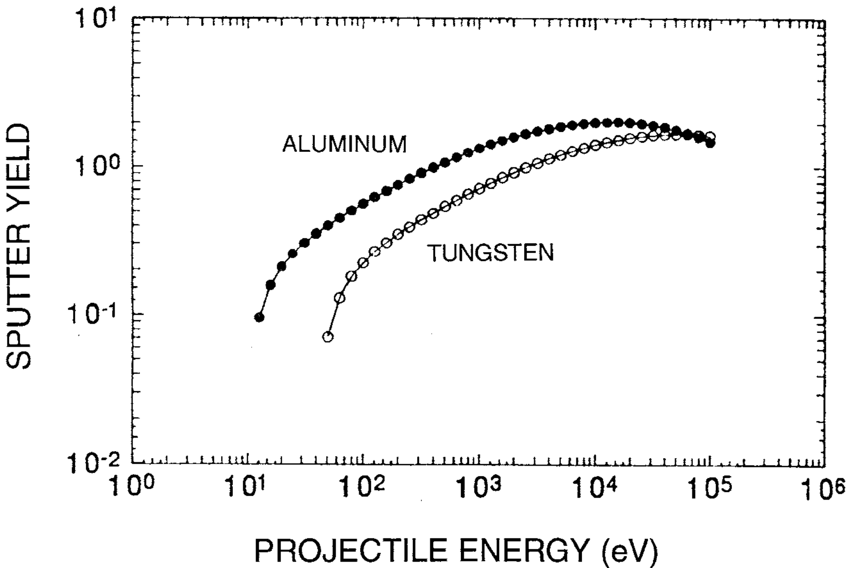


Figure 4-13 Calculated sputter yields for aluminum and tungsten as a function of projectile energy for argon. (From J. E. Mahan and A. Vantomme, *J. Vac. Sci. Technol.* **A15**, 1976 (1997). Reprinted with the permission of Professor J. E. Mahan.)

generally reveals good agreement. Significantly, this simplified collisional model reproduces trends in experimental data for the projectile energy, mass, and target dependence of the sputter yield.

4.5.2.3 Sputtering of Alloys

In contrast to the fractionation of alloy melts during evaporation, with subsequent loss of deposit stoichiometry, sputtering allows for the deposition of films having the same composition as the target source. This is a primary reason for the widespread use of sputtering to deposit metal alloy films. We note, however, that each alloy component evaporates with a different vapor pressure and sputters with a different yield. Why, then, is film stoichiometry maintained during sputtering and not during evaporation? One reason is the generally much greater disparity in vapor pressures compared to the difference in sputter yields under comparable deposition conditions. Secondly, and more significantly, melts homogenize readily because of rapid atomic diffusion and convection effects in the liquid phase; during sputtering, however, minimal solid-state diffusion enables the maintenance of the required altered target surface composition.

Consider now sputtering effects (Ref. 16) on a binary-alloy target surface containing a number of A atoms (n_A) and B atoms (n_B) such that the total number is $n = n_A + n_B$. The target concentrations are $C_A = n_A/n$ and $C_B = n_B/n$, with sputter yields S_A and S_B. Initially, the ratio of the sputtered atom fluxes (ψ) is given by

$$\frac{\psi_A}{\psi_B} = \frac{S_A C_A}{S_B C_B}. \tag{4-39}$$

If n_g sputtering gas atoms impinge on the target, the total number of A and B atoms ejected are $n_g C_A S_A$ and $n_g C_B S_B$, respectively. Therefore, the target surface concentration ratio is modified to

$$\frac{C'_A}{C'_B} = \frac{C_A}{C_B}\frac{(1 - n_g S_A/n)}{(1 - n_g S_B/n)} \tag{4-40}$$

instead of C_A/C_B. If $S_A > S_B$, the surface is enriched in *B* atoms, which now begin to sputter in greater profusion, i.e.,

$$\frac{\psi'_A}{\psi'_B} = \frac{S_A C'_A}{S_B C'_B} = \frac{S_A C_A(1 - n_g S_A/n)}{S_B C_B(1 - n_g S_B/n)}. \tag{4-41}$$

Progressive change in the target surface composition lowers the sputtered flux ratio to the point where it is equal to C_A/C_B, which is the same as the original target composition. Simultaneously, the target surface reaches the value $C'_A/C'_B = C_A S_B / C_B S_A$, which is maintained thereafter. A steady-state transfer of atoms from the bulk target to the plasma ensues resulting in stoichiometric film deposition. This state of affairs persists until the target is consumed. Conditioning of the target by sputtering a few hundred layers is required to reach steady-state conditions. As an explicit example, consider the deposition of Permalloy films having the atomic ratio 80 Ni–20 Fe from a target of this same composition. Using 1 keV Ar, the sputter yields are $S_{Ni} = 2.2$ and $S_{Fe} = 1.3$. The target surface composition is altered in the steady state to $C_{Ni}/C_{Fe} = 80(1.3)/20(2.2) = 2.36$, which is equivalent to 70.2 Ni and 29.8 Fe.

4.5.2.4 A Potpourri of Sputtering Results and Effects

Over the years a large number of interesting experimental observations have been made with regard to sputtering effects. The influence of sputtering gas and ion energy have already been discussed. Other phenomena and results are listed next in no particular order.

1. *Effect of periodic table.* Sputter yields were measured for metal elements in given rows of the periodic table using 400 eV Ar ions (Ref. 17). In the sequence Zr, Nb, Mo, Ru, Pd, and Ag, there was a continuous rise in S from $\sim$0.5 to about 2.7. Similar, although smaller, increases in S were observed for those elements lying in the row between Ti and Cu, as well as the row between Ta and Au. The well-known strong inverse variation between S and sublimation energy is apparent in these results. In a similar vein, the previously noted correlation between threshold energy (E_{th}) and sublimation energy (U_S) has been roughly verified for many metals, i.e., $E_{th} \simeq 5U_S$.

2. *Crystallographic effects.* Studies of ion-bombarded single crystals reveal that atom emission reflects the lattice symmetry. In FCC metals it has been demonstrated that atoms are preferentially ejected along the [110] direction, but ejection in [100] and [111] directions also occurs to lesser extents (Ref. 18). For BCC metals [111] is the usual direction for atom ejection. These results are consistent with the idea that whenever a beam sees a low density of projected lattice points the ions penetrate more deeply, thus reducing S. Such observations on single crystals confirmed momentum transfer as the mechanism for atomic ejection; the notion of ion-induced melting and evaporation of atoms was dispelled because preferred directions for sublimation of atoms are not observed.

3. *Energy distribution of sputtered atoms.* Sputtered atoms have neither zero nor very high energies; instead the distribution peaks in between at typical energies 2 to 7 eV. The number distribution of sputtered atoms as a function of energy is reminiscent of the Boltzmann distribution for gas or evaporated particle energies whose peak magnitudes cluster about the much smaller thermal value of ~ 0.1 eV.

4. *Angular distribution of sputtered atoms.* As is the case for evaporation, a cosine law distribution for the emission of sputtered atoms is generally observed (Ref. 18). Slight deviations from this law have been observed depending on ion energy, metal, and degree of crystallographic texture in the target. The emission lobe is generally extended (i.e., overcosine) normal to the target for high ion energies and compressed (i.e., undercosine) at low energies.

5. *Angle of ion incidence.* The entire discussion of sputtering until now has assumed that ion impingement is normal to the target surface. It has been observed, however, that the sputter yield ($S(\theta)$) depends on the angle (θ) defined between the directions of ion incidence and the target normal. Furthermore, as shown in Fig. 4-14, $S(\theta)$ is enhanced relative to $S(\theta = 0)$ such that the ratio $S(\theta)/S(\theta = 0)$ is found to vary as $(\cos\theta)^{-1}$ for values of θ up to $\sim 70^\circ$ (Ref. 19). Physically, shallower collision cascades create a greater density of displaced surface atoms that can be potentially sputtered. However, the inverse $\cos\theta$ dependence obviously fails at glancing angles approaching 90° because the number of ions penetrating the surface drops precipitously.

6. *Sputtering of compound semiconductors.* Most of the previous discussion applies to metals, and with the exception of item (2) earlier, refers largely to polycrystalline targets. Sputtering effects in compound semiconductors display similar features but there are interesting differences (Ref. 11). Based on emission patterns there is evidence that these normally single-crystal targets sputter preferentially along crystalline directions at elevated temperatures, but when the targets are cooled sputtering is isotropic. Apparently, at lower temperatures the ion bombardment and radiation damage create sufficient structural defects to amorphize surface layers. At higher temperatures, amorphous regions do not form because the defects anneal out; the target remains crystalline and sputters accordingly. A transition temperature separating these sputtering regimes has been suggested in GaAs. Wide variations in S may arise from deep ion penetration or channeling along certain crystallographic directions and a complex damage distribution as a function of ion energy and flux.

7. *Sputtering of molecules.* Sputtering is not limited to the world of inorganic materials. Very complex organic molecules have been ejected intact into the vapor phase when electronically excited by incident photons,

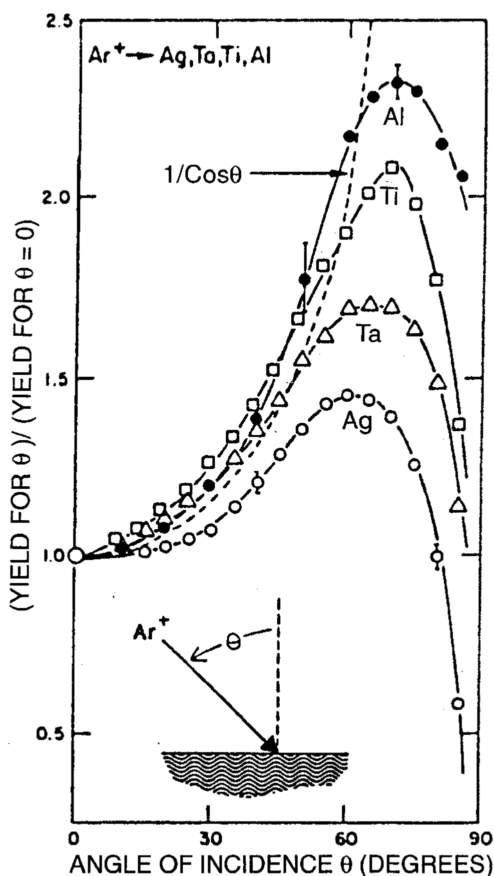


Figure 4-14 Sputter-yield dependence of Ag, Ta, Ti, and Al on angle of incident 1.04 keV Ar. (From H. Oechsner, *Appl. Phys.* **8**, 195 (1975). Reprinted with the permission of Professor H. Oechsner.)

electrons, and ions. The process has been called "electronic" sputtering (Ref. 20). For example, with fast heavy ions (~ 1 MeV) bovine insulin molecules ($C_{254}H_{377}N_{65}O_{75}S_6$) were emitted from a solid sample and C^{60} buckyballs were ejected from a $(C_2H_2F_2)_n$ polymer. In the latter case it has been suggested that ion-beam interaction yields a highly ionized region 0.4 nm in diameter where molecules are formed. Surrounding this out to a 2-nm diameter, large intact ions are ejected. And large neutral molecules are sputtered within a 4-nm diameter. Beyond this out to a diameter of 10 nm there is impact damage.

4.6 ION BOMBARDMENT MODIFICATION OF GROWING FILMS

4.6.1 INTRODUCTION

In the previous section our primary interest was ion bombardment of cathodes, and particularly the *removal* of atoms by sputtering. Now we are concerned with how ion-bombardment effects modify the composition, structure, and properties of the atoms that *remain* in the growing film. At the outset it is useful to enumerate (Ref. 21) the various energetic particles that impinge on surfaces immersed in plasmas or exposed to ion bombardment. First there are the sputtered *neutral* species that leave the target with translational kinetic energies of $\sim$3–10 eV. In addition, some ions striking the target become charge neutralized and are reflected back toward the substrate as neutrals, retaining much of their initial ion energy. For a given ion mass the reflection probability is greater the higher the target mass. The energy both kinds of neutrals (reflected atoms and sputtered atoms) carry to the growing film depends strongly on system pressure because the latter controls gas-phase scattering. Plasma *ions* also strike the film surface. These arrive with energies that largely depend on the potential gradient across the anode plasma sheath and gas pressure. Unlike neutrals, ions can enhance their energy by accelerating through the sheath. In DC sputtering systems the plasma potential is typically 5–10 V above ground, but can reach values 10 times higher in certain RF configurations. A negative bias voltage placed on the substrate can substantially increase the energy of bombarding ions, and this is the basis of bias sputtering (Section 5.2.3.3).

During plasma-deposition processes, it is common for atoms that will eventually comprise the film to deposit together with a flux of energetic, bombarding inert or reactive gas ions. As long as more incident atoms deposit than are sputtered away, the film thickens. Just as the simultaneous impingement of (neutral) evaporated metal and residual gas atoms affects the resultant film purity (see Section 3.3.4), we may expect a potentially richer assortment of property modification effects in plasma deposited films. Many more variables are at play now, e.g., type of ions and their energy, impingement flux, and direction of impact in addition to the nature of the film, substrate, and deposition temperature. Most important perhaps is the ion energy, whose release promotes the atomic kinetic activity that modifies composition, film structure, and many derivative properties. Examples of the latter include the film grain size and orientation, defect concentrations, stress, film adhesion, and topography of the film surface, all of which alter mechanical as well as electrical, magnetic, and optical properties.

The details of ion-enhanced film modification are governed by the complex interplay of microscopic events that flow from the involved process variables. Specific ion-assisted deposition processes, film materials, and properties modified will be treated again in later chapters, which may occasionally have to be consulted in advance. With this caveat, it is instructive to begin by considering the bonding of an incident low-energy atom to a film surface.

4.6.2 Temperature Spikes

Even at the thermal energies ($k_B T$) involved in the condensation of atoms from the vapor phase, the release of latent heat induces a temperature spike in the film at the point of impingement. Machlin (Ref. 22) has estimated that the maximum thermal-diffusion distance (r) for such a depositing adatom is $r = 0.4a_r(E/E_S)^{1/3}$, where E is the energy transferred, E_S is the activation energy for surface diffusion, and a_r is the atomic radius. For ion energies at the sputtering threshold where $E/E_S \simeq 30$, r is calculated to be only $1.2a_r$. Thus, adatoms do not move much more than an atomic distance away from where they impinge.

At much higher energies, however, a spike regime (see Fig. 4-11c) is entered where all of the atoms are simultaneously in motion within a local volume of the bulk. Atoms within the spike resemble a high-temperature, pressurized gas bubble. For example, if an ion impinges at 100 eV, calculation shows that within 7×10^{-11} s, $T \simeq 3300$ K and $P = 10^5$ atm (Ref. 23). Ejection of atoms in such a thermal spike can be expected as a result of vaporization rather than a collisional mechanism. Localized defects, structural rearrangements, and radiation damage are generated in the process, each at a characteristic critical ion-impact energy.

From the examples given it is clear that in the energy window of a few to a few tens of electron volts, surface diffusion occurs; but at higher energies subsurface atom motion becomes more important.

4.6.3 Structural Modification

Provided an elastic collision approximates what happens between an incident ion and the surface of a depositing film, considerable sub- and near-surface atomic shuffling may occur. Because of upward momentum transfer, this eventually leads to rearrangement of surface atoms. Thus we may expect modification of both film structure and film composition. These

two important issues are now addressed in the remainder of this and the next section. Ion bombardment of growing films modifies at least four measurable characteristics of its structure: (1) surface topography and roughness, (2) crystallography and texture, (3) grain structure, including grain size and morphology, and (4) defects and stress.

4.6.3.1 Surface Topography Modification

In evaporative deposition processes the evolution of surface features on growing films is very much dependent on the substrate temperature and statistical fluctuations in the impinging particle flux and subsequent atomic surface diffusion. At high temperatures, diffusion is rapid, and a typically rough, hillocky film topography that is controlled by surface energy considerations emerges. Adatom diffusion is frozen at lower temperatures, however, and film surfaces are less rough. A smoothing of the surface is also observed under simultaneous ion bombardment, as comparisons between thermally evaporated and sputtered films typically reveal. This is particularly true for low-energy (1–10 eV) ions in a low pressure plasma. For ion bombardment at higher energy we may expect breakup of surface clusters upon impingement, a process that would tend to planarize deposits.

Film smoothing and improved coverage of depressions, corners, and steps are enhanced by *resputtering*. The phenomenon of resputtering occurs when energetic ions cause atoms in the film deposit to sputter. It is less tightly bound atoms at atomic projections and regions of high curvature that are particularly vulnerable to resputtering under angular ion impact. The involved atoms generally land nearby where they further energize atomic diffusion and promote planarization of the film surface.

While ion flux induced surface smoothing occurs during film growth, an interesting roughening of surface topography is observed under certain conditions. Ions of moderate energy, i.e., less than 1 keV on up to tens of keV, are necessary. Under ion bombardment various micron-sized surface-structures such as cones, pyramids, ridges, ledges, pits, and faceted planes form. Sometimes quasi liquid-like and microtextured labyrinth-like features also develop. An example of cone formation on a Cu single crystal after 40 keV Ar bombardment is shown in Fig. 4-15. The term “cone formation” is used to generically categorize this class of topological phenomena. Wehner (Ref. 24) and Banks (Ref. 25) have extensively reviewed aspects of ion-beam-induced surface topography and texturing effects, including materials systems that are susceptible, processing conditions, and theories for the observed effects. Known for a half-century, cone formation has been interpreted in terms of either *left-standing* or *real-growth* models. The former model views conical projection arising from sputter-resistant impurities or

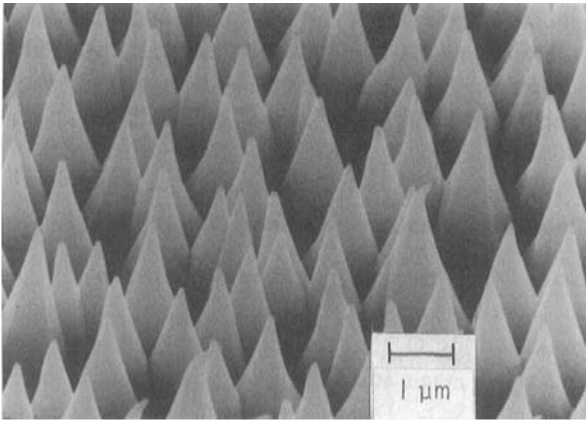


Figure 4-15 Pyramid structures on a single-crystal Cu surface after 40 keV Ar bombardment. From J. L. Whitton, G. Carter, and M. J. Nobes, *Radiation Effects* **32**, 129 (1977).

intentionally deposited seed atoms (e.g., Mo) that etch or sputter at a lower rate than the surrounding surface (e.g., Cu). Sputter-yield variations as a function of ion incidence angle, nonuniform redeposition of atoms on oblique cones, and surface-diffusion effects are operative in this mechanism. The latter model likewise requires the presence of impurity seed atoms. In this case, however, they serve as nucleating sites for growth of genuine single-crystal whiskers that sprout in all crystallographic directions relative to ion incidence. The subsequent interplay among whisker growth, surface diffusion, and sputtering results in the observed cones. Cone formation is likely to be more visible on impure sputtering targets than on depositing thin films. Nevertheless, crystallographic features may be etched into a film surface if strongly anisotropic sputtering occurs.

4.6.3.2 Crystallography and Texture

Changes in interplanar spacings and film orientations are among the effects in this category of structural modification. Distortions from cubic to tetragonal crystal structures also occur. Expansions in the (111) lattice spacing of close to 1% have been observed in several ion-bombarded metal films (Fig. 4-16a). Interestingly, with increasing ion energy the interplanar spacing sometimes peaks and then diminishes, suggesting that gas and associated defects are first incorporated, but then partially anneal out. Similar effects were reported in an extensive study by Kuratani *et al.* (Ref. 26) on ion beam assisted deposition of Cr under energetic Ar ion

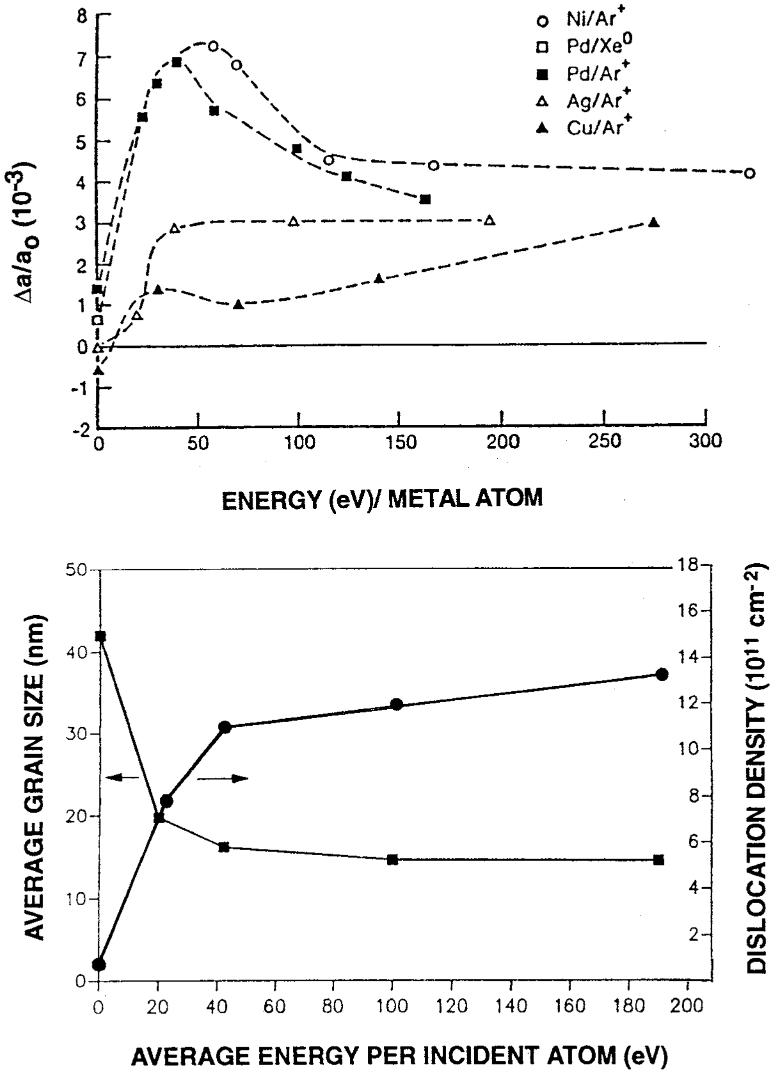


Figure 4-16 (a) Lattice distortion of Ni, Pd, Ag, and Cu films grown under Ar ion bombardment of indicated energy. (b) Grain size and dislocation density of Ag films deposited at room temperature as a function of the average energy per deposited atom. The average energy is the weighted sum over ions plus atoms. Film thickness is approximately the same in all cases. (From Ref. 21.)

bombardment. By either increasing ion energies or the ratio of arriving Ar (ions) to Cr (atoms), the Cr lattice constant was found to increase from 0.28845 to 0.2910 nm. The observed film surface roughening at the maximum lattice constants is apparently due to the very large internal compressive stress levels generated and the microdestruction of the lattice as a result.

Films frequently display preferred orientation, or preponderance of certain planes lying parallel to the substrate plane, in the presence of ion bombardment. For example, in metal films grown on amorphous substrates, low-index, high atomic density planes, e.g., (111), (110), often lie parallel to the film surface. A better appreciation of why such preferred orientation develops in films will emerge with an understanding of ion-beam channeling effects, a subject discussed in Section 4.6.4.2.

4.6.3.3 Grain Structure

That the microstructure of thin films can vary widely is a theme that pervades the book, and in particular, Chapter 9. When observed in plan view, typical polycrystalline thin films (if there is such a thing) appear to be roughly equiaxed with considerable variation in grain size. In cross section such films have a tilted, voided columnar structure that reflects low atom mobility and the shadowing effects of previously deposited material. It is well documented that ion bombardment generally causes a reduction of film grain size. In bulk solids, it is also known that impurity atoms segregate to grain boundaries (GBs) and associate with matrix atoms and defects located there. Incorporated gas atoms and precipitates constitute such impurities in sputtered films. Since their presence effectively pins GBs, migration of the latter are prevented and small grains result. As Fig. 4-16b shows, the grain size of Ag falls as the average energy per depositing atom rises to 40 eV/atom, but then remains constant above this value. Apparently, some structural annealing occurs at the higher energies where there is more subsurface atom penetration.

A more revealing look at ion-beam modified grain structures is evident in Fig. 4-17. Shown is a sequence of Cr films exposed to successively greater Ar/Cr ratios while maintaining a 2-keV Ar beam energy. Regardless of ratio the film structure is columnar and the columns appear to grow at various angles. At low Ar/Cr ratios the columns are narrow and short, but they widen and lengthen at large Ar/Cr values where the structure is rough and contains voids between columns.

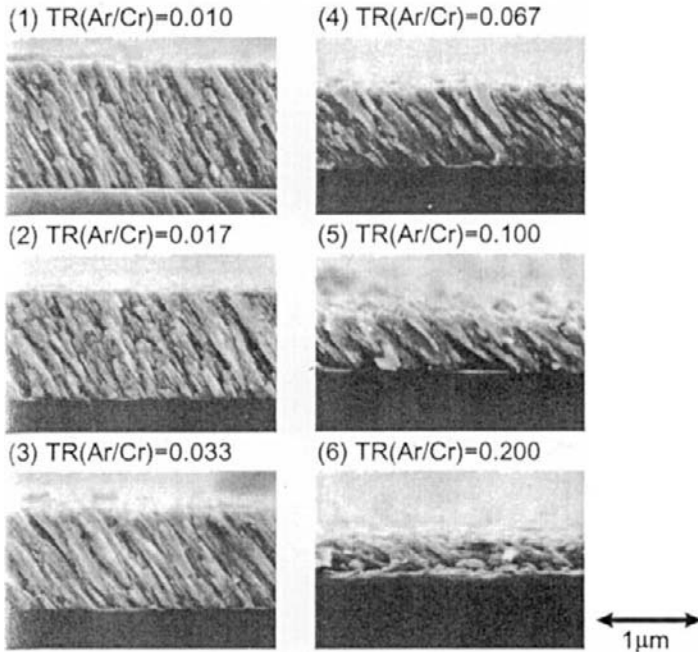


Figure 4-17 Cross-sectional scanning electron micrographs of electron beam evaporated Cr films bombarded by 2-keV Ar ions during deposition. TR(AR/Cr) values refer to the concentration ratio of the species transported to the film surface. (From N. Kuratani, A. Ebe, and K. Ogata, *J. Vac. Sci. Technol.* **A19**(1), 153 (2001): 155. Reprinted by permission.)

4.6.3.4 Defects and Stress

One might imagine that films would become more defective as energetic ions insert themselves in and forcibly displace atoms from equilibrium sites. This is why the dislocation density generally increases in concert with the change in grain size as shown in Fig. 4-16b. However, there are trade-offs in the net defect densities (C_d) in the presence of both ion-energy (E_i) and temperature (T) variables. Thus C_d usually increases as E_i rises and T falls, and vice versa. Here as with other properties (e.g., grain size, stress), threshold ion energies and critical temperatures often delineate the limits of behavior that reflect a competition between ion-beam lattice damage and defect annealing effects.

Considering the extent of gas impingement it is not surprising that some gas will be trapped within the solid film. When such atoms as well as energetic reflected neutrals are stuffed into interstitial lattice positions we can expect residual compressive film-stress. Furthermore, theory (Ref. 27) suggests that the stress should scale as the product of the ion/atom flux ratio

and the square root of the ion energy. Stress reversals in sputtered films as a function of gas pressure have also been observed and will be discussed in Section 12.5.3.

4.6.3.5 Compositional Modification of Surfaces

In addition to sputtering, inert gas ion–surface interactions also promote removal of gas atoms or molecules adsorbed on the surface, thus preventing them from being buried in the bulk of the growing film. Even before atoms condense into clusters, such ion bombardment is effective in cleaning substrates and promoting subsequent film adhesion. (Because film adhesion is such an important subject, it is treated again in Chapters 5 and 12.) Reactive-gas ions are often more effective cleaning agents by activating and decomposing organic impurities, or by promoting formation of volatile species through chemical reaction. However, the same gases used to rid the surface of contaminants may also react with the substrate. For example, oxygen plasmas that remove hydrocarbons from an aluminum substrate will coat it with a film of Al_2O_3. Similar reactions are relied on in the reactive sputtering of dielectric films, e.g., the deposition of nonvolatile oxide or nitride films. Through reaction between trace amounts of reactive gas ions and metal contaminants, unwanted precipitate compounds can be incorporated into growing films.

4.6.4 Ion Implantation

Ion implantation, an extremely important processing technique, is primarily used to modify or alter the subsurface structure and properties of previously deposited films. Since there is a large and accessible literature on its major use in doping semiconductors, this subject will not be treated here. Similarly, high-energy ion implantation has beneficially modified the surfaces of mechanically functional components such as dies and surgical prostheses (Ref. 28). Projectile ions impinging on such components that are not reflected or adsorbed, and do not cause sputtering, are implanted. At ion energies between tens and hundreds of keV, the probability is great that ions will be buried hundreds to thousands of angstroms deep beneath the surface.

Although such energies are normally beyond the range of common ion-assisted film-deposition processes (plasma-immersion ion-implantation (Section 5.5.6) is an exception), we may, nevertheless, profitably extrapolate ion-implantation phenomena to lower energy regimes. During implantation, ions lose energy chiefly through two mechanisms, namely electronic and nuclear interactions.

1. *Electronic losses* are due to coulomb interactions between the moving ion and substrate electrons. As a result, the excited electrons access higher bound levels or generate a continuum of ionization states with their eventual relaxation products. The latter are manifested by emitted photons, photoelectrons, and Auger electrons.

2. *Nuclear losses* occur when bombarding ions, having lost energy, i.e., because of electronic excitations, slow sufficiently until they begin to set in motion violent nuclear collision cascades along their trajectories. These cascades, the result of displaced atoms that displace yet other atoms, leave a jagged branched trail of matrix damage in their wake.

Through electronic and nuclear interactions the ion energy (E) continuously decreases in a very complicated way with distance (z) traversed beneath the surface. For simplicity, the energy loss is expressed by (see Section 4.5.2.2)

$$-\frac{dE}{dz} = N[S_e(E) + S_n(E)] \tag{4-42}$$

where $S_e(E)$ and $S_n(E)$ are the respective electronic and nuclear stopping powers (in units of eV-cm^2), and N is the target atom density. The magnitudes of both stopping powers depend on the atomic numbers and masses of the ions as well as matrix atoms. Typically, electronic stopping results in energy losses of 5–10 eV/Å as opposed to the higher losses of 10–100 eV/Å for nuclear stopping. When comparing these values with typical electronic and lattice energies in solids, film modification over many angstroms can be expected.

4.6.4.1 Subsurface Compositional Change

As a result of implantation it is clear that no two ions will execute identical trajectories but will, rather, participate in some admixture of nuclear and electronic collision events. Furthermore, the collective damage and zigzag motion of ions within the matrix cause them to deviate laterally from the surface entry point. When summed over the huge number of participating ions, these factors lead to the statistical distribution of ions as a function of depth (z) shown in Fig. 4-18. The concentration of implanted ions ideally has a Gaussian depth profile given by

$$C(z) = \frac{\phi}{\sqrt{2\pi}\,\Delta R_P} \exp - \left(\frac{z - R_P}{\sqrt{2}\,\Delta R_P}\right)^2 \tag{4-43}$$

with a peak magnitude varying directly as the fluence or dose ϕ of incident ions. Dose has units of number (of ions) per cm^2 and is related to the

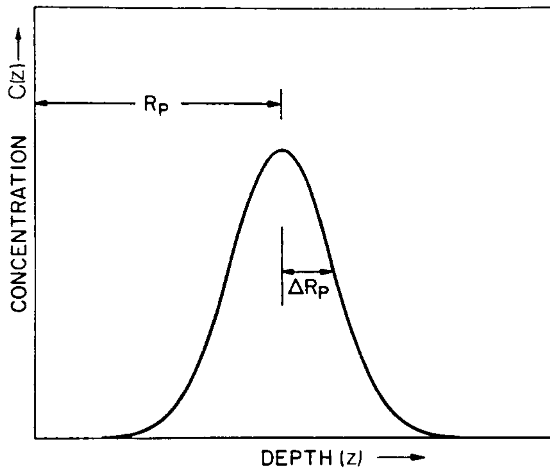


Figure 4-18 Gaussian distribution of implanted ions as a function of depth beneath the surface.

measured time-integrated current or charge Q deposited per unit surface area A. Specifically,

$$\phi = \frac{Q}{nqA} \tag{4-44}$$

where n is the number of electronic charges, q, per ion. The projected range, R_P, is the depth most ions are likely to come to rest at yielding a peak concentration of

$$C(R_P) = \frac{\phi}{\sqrt{2\pi}\,\Delta R_P}.$$

Note that the actual distance an ion travels is greater than the depth projected normal to the target surface. This is analogous to the total distance executed in atomic random-walk jumps exceeding the net diffusional displacement. Spread in the ion range is accounted for by the term ΔR_P, the standard deviation or longitudinal "straggle" of the distribution. Similarly, ΔR_L or the lateral ion straggle is a measure of the spread in the transverse direction.

4.6.4.2 Channeling

An interesting phenomenon known as *channeling* occurs in single-crystal matrices, such that along certain crystallographic directions the depth of ion

penetration is extended, greatly altering the profile shape predicted by Eq. 4-43. Channeling has been most studied in silicon and can be understood by viewing a ball-and-stick model of the diamond crystal structure along various crystal directions. In virtually all orientations the model appears impenetrable to impinging ions. But along the [110] direction a surprisingly large open tunnel is exposed through which ions can deeply penetrate by undergoing glancing, zigzag collisions with the tunnel-wall atoms. The ion trajectories simply do not bring them close enough to target atoms where they can undergo the nuclear collisions that are particularly effective in slowing them down. Rather, these channeled ions lose energy primarily by electronic excitation of the lattice and therefore range further than if the matrix were, say, amorphous.

Since channeling of ions tends to exaggerate anisotropic properties of specially aligned grains, we may suspect it plays a role affecting the preferred orientation of depositing crystalline films. In this regard it makes a difference whether ion impingement is normal or off-normal with respect to the film plane (Ref. 29). Actually, what is important is the ion-channeling direction in the film relative to the direction of incoming ions. Since sputter yields are generally less along channeling directions, larger film growth rates and survival of grains so aligned may be expected; the unaligned grains would be preferentially sputtered instead.

An alternative explanation for the development of preferred film orientations is based on the role of thermal spikes. In crystallites that do not channel the beam, thermal spikes resulting from nuclear stopping lead to highly damaged regions. On the other hand, spikes are rare in crystallites that channel the beam because ions lose energy primarily by electronic stopping. The latter crystallites are then the seeds for recrystallization of the damaged lattice. In FCC metals the channeling direction is [110], so it is not surprising that the (110) plane lies parallel to the substrate of ion-irradiated copper films.

4.6.4.3 Ion-Beam Mixing

During ion bombardment of two-or multiple-component film systems, atoms tend to mix causing both compositional and structural change. The effect is known as ion mixing. As an example, consider thin film A on substrate B bombarded by a beam of inert gas ions. Typically, the ion range (R) exceeds the escape depth of the sputtered A atoms. If R does not exceed the thickness of A, then only A atoms sputter. If, after some sputtering R extends into the substrate region, the atomic displacements and enhanced diffusional effects that occur within collision cascades will cause A and B to intermix. The local mixing at the interface eventually links with other

similarly intermixed zones to create a continuous ion-beam mixed layer. Now B atoms also enter the stream of sputtered atoms because the combination of continued surface erosion and interfacial broadening, due to ion mixing, has brought them closer to the surface. Kelly and Miotello (Ref. 30) have reviewed the mechanisms of these effects and concluded that ballistic mixing, random motion of defects, and chemically guided defects played important roles, but that thermal-spike (Section 4.6.2) mixing was unimportant.

It is of interest to estimate the extent of interfacial broadening due to ion mixing. To do this we set range R equal to the half width of the broadened layer and assume that nuclear energy loss dominates with a stopping power independent of energy. Therefore, from Eq. 4-42,

$$R = \int_{E_0}^{0} \frac{dE}{NS_{\mathrm{n}}(E)} = \frac{E_0}{\left(\frac{dE}{dz}\right)_{\mathrm{n}}} \tag{4-45}$$

where $(dE/dz)_{\mathrm{n}} = -NS_{\mathrm{n}}(E)$. In the case of Ar in Cu, $(dE/dz)_{\mathrm{n}} \simeq 100\,\mathrm{eV/A}$. Thus, for every keV of ion energy the altered layer extends about 10 Å (Ref. 13).

Through the use of high-energy ion beams, mixing reactions can occur over substantial dimensions. Films can, therefore, be effectively alloyed with substrates, and layered, normally immiscible films can be partially homogenized with the assistance of ion implantation. Such effects are extremely efficient in promoting adhesion of films to substrates, probably the most significant benefit ion bombardment confers on growing films.

4.6.4.4 Amorphization of Films

Perhaps the most extreme form of structural modification is converting a crystalline film into an amorphous one. As we shall see in Section 9.6, materials that are crystalline in bulk form (e.g., metals) can sometimes be deposited as amorphous thin films, but not easily. One route to producing amorphous thin films is to expose crystalline films to an ion beam of appropriate flux and energy. The dose of ions required for amorphization can be roughly estimated assuming that the energy density is essentially the same as that needed for melting. In the case of Si this amounts to about 20 eV/atom, or $\sim 10^{24}\,\mathrm{eV/cm^3}$. For ions of energy E_0, the dose is therefore

$$\phi = \frac{10^{24} R_{\mathrm{P}}}{E_0}\ \mathrm{ions/cm^2}. \tag{4-46}$$

Assuming $E_0 = 1\,\mathrm{keV}$ and $R_{\mathrm{P}} = 10\,\text{Å}$, $\phi = 10^{14}\,\mathrm{ions/cm^2}$. In practice a dose

greater than 10^{16} ions/cm^2 is required, indicating that lattice damage and recrystallization effects occur simultaneously. Greater depths of film amorphization can be achieved through overlapping implants with successively higher-energy ions. Amorphous films are frequently deposited during plasma and ion-beam assisted film-deposition processes. In such cases the critical dose-ion energy requirements for amorphization are exceeded.

4.7 CONCLUSION

This chapter invites comparisons with the previous one and attempts to establish a base for better understanding plasma-assisted film deposition and etching processes in the next two chapters. Such a backward as well as forward perspective can be gleaned from Table 4-3 and Fig. 4-19. Differences in the plasma discharge environment relative to that of a vacuum are compared in Table 4-3. Both are low-pressure environments, but introduction of cathode and anode electrodes and a means of coupling electromagnetic energy to the system partially ionizes the discharge gas. The disproportionate role played by this relatively small concentration of electrons and ions is responsible for the gas-phase discharge structure, sustained ionization, and the varied dynamical motion of these species in electric and magnetic fields. Many of the effects described in the chapter arise from the fact that electrons travel faster and are more energetic than ions. Furthermore, electron collisions with reactive gases create metastable species that promote plasma-assisted chemical reactions. Central to thin-film deposition processes is the ion bombardment of cathodes. Upon impact, a train of events is initiated resulting in the ejection or sputtering of atoms from them. Critical in this regard is the sputter yield, the property that determines how efficient the process is. Once ejected, these atoms fly through the intervening plasma where they deposit sequentially at the substrate (anode). Here again ion bombardment plays a beneficial role, this time to modify the structure and composition of the growing films.

Consideration of Fig. 4-19 is a way to view these latter ion–surface interactions. In this figure the range over which particular processes dominate are mapped as a function of (charged) ion/(neutral) atom ratio and ion energy. When the ion/atom ratio and ion energy values are both below threshold levels defined by the line with slope equal to -1, ion bombardment simply does not modify film properties. Irrespective of ion energy, desorption of surface impurities occurs at the lowest ion fluxes. However, as the ion/atom, flux ratio increases, film densification and ion implantation become more probable; in general densification proceeds at lower ion energies than those required for implantation. Resputtering effects dominate

Table 4-3
Vacuum/Evaporation vs Plasma/Sputtering

	Vacuum/evaporation	Plasma/sputtering
A. Source attributes		
1. Phase	Melt or solid	Solid target
2. Mechanism of atom removal	Thermal evaporation (hot source)	Ion bombardment and collisional momentum transfer (cool target)
3. Energy supplied to source	Thermal energy ~ 0.1 to $0.2\,\text{eV/atom} + \Delta H_v$	$> 20\,\text{eV/atom}$
4. Atom removal rate	$\sim 1.3 \times 10^{17}$ atoms/cm^2-s for $M = 50$, $T = 1500\,\text{K}$, $P_e = 10^{-3}$ torr (Eq. 3-2)	$\sim 10^{16}$ atoms/cm^2-s at $1\,\text{mA/cm}^2$ and $S = 2$
5. Atom emission geometry	$\cos\phi$ and $\cos^n\phi$	$\cos\phi$ as well as directional according to crystallography
6. Applicability, availability	All materials, generally high purity	Targets of all materials, variable purity
B. Gas phase attributes		
1. Composition	Evaporant atoms, associated and dissociated compound fragments, residual gases	Sputtered atoms, assorted metastable ionized and excited species, sputtering gas, ions, electrons, residual gases
2. Pressure	High to ultrahigh ($\sim 10^{-5}$ to 10^{-10} torr)	~ 1–100 mtorr discharge
3. Species energy	~ 0.1–0.2 eV for evaporants	3–10 eV for sputtered atoms 2–5 eV for electrons
4. Atomic mean free path	Larger than evaporant–substrate spacing. No gas collisions in vacuum	Less than target–substrate spacing. Many gas collisions in the discharge
C. Condensed film attributes		
1. Energy of condensing atoms	Low (0.1 to 0.2 eV)	High ($\sim$ a few eV); higher with substrate bias
2. Gas incorporation	None	Some
3. Adhesion to substrate	—	Generally good
4. Film stoichiometry	Generally different from multicomponent alloy and compound sources	Same as the target composition

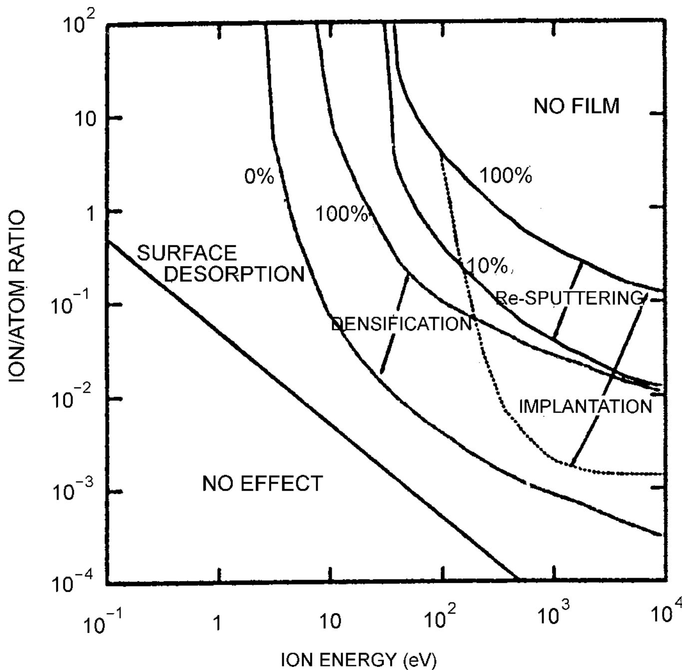


Figure 4-19 Regions of dominance for various ion-bombardment processes as a function of ion/atom ratio and ion energy. Experimental data for different material systems typically fall within the indicated fields. After J. M. E. Harper, J. J. Cuomo, R. J. Gambino, and H. R. Kaufman, in *Ion Bombardment Modification of Surfaces: Fundamentals and Applications*, eds. O. Auciello and R. Kelley. Elsevier, Amsterdam, 1984.

at the highest flux levels and ion energies. Finally, no film forms if the atom flux is simply too low; in such a case all that happens is ion implantation and sputtering or etching of the substrate.

EXERCISES

1. Consider a system containing an ionizable gas that surrounds parallel electrodes with the cathode at $x = 0$ and anode at $x = d$, where $d > x > 0$. A discharge is initiated and the following is assumed:
   (a) The total current $i = i_-(x) + i_+(x)$, a sum of electron and ion components, is constant, i.e., independent of x and time.
   (b) At the cathode $i_-(0) = i_0 + \gamma_e i_+(0)$, with i_0 a constant.

(c) $d[i_-(x)]/dx = \alpha i_-(x)$, i.e., $i_-(x) = i_-(0) \exp \alpha x$.
(d) At the anode, $i_+(d) = 0$.
From this information derive the Townsend equation.

2. (a) With the aid of the Nernst–Einstein equation (Chapter 1), show that

$$J = -\frac{(D_i \mu_e + D_e \mu_i)}{(\mu_e + \mu_i)} \frac{dn}{dx}$$

(b) Demonstrate that the ambipolar diffusion coefficient is given by Eq. 4-11.
(c) Verify that D_a is effectively equal to $D_i[1 + (T_e/T_i)]$.

3. (a) Obtain solutions to Eq. 4-13a, b, and c for the x, y, and z components of electron motion when confined by the perpendicular electric and magnetic fields.
(b) Show that Eqs. 4-14a and b follow from the solutions of part (a).

4. Suppose the electron velocity distribution function in a plasma is essentially given by the Maxwell–Boltzmann form of Eq. 2-1, or

$$f(v_e) = \frac{1}{n_e}\frac{dn_e}{dv_e} = 4/\pi^{1/2}(m_e/2k_B T_e)^{3/2} v_e^2 \exp - (m_e v_e^2/2k_B T_e).$$

(a) Recast this distribution function in terms of the electron energy (E_e), i.e., $f(E_e)$, noting that $E_e = \frac{1}{2} m_e v_e^2$.
(b) Write an expression that represents the fraction of electrons in the distribution with E_e greater than E_0.
(c) The threshold energy for ionization of argon is 15.76 eV. Determine the fraction $f(E > 15.76\,\text{eV})$ if the electron temperature T_e is given by $k_B T_e = 2$ eV. Repeat the calculation for $k_B T_e = 0.25$ eV and $k_B T_e = 8$ eV. *Hint:* See Ref. 4, p. 118.

5. Rather than a point ionic charge immersed in a plasma consider a planar sheet of positive ionic charge surrounded by plasma. If Poisson's equation for this case is

$$\frac{d^2 V(x)}{dx^2} = -q(n_i - n_e)/\varepsilon_0 = n_i q^2 V(x)/\varepsilon_0 k_B T$$

obtain a solution for $V(x)$.

6. In a 13.56 MHz RF plasma operating at a pressure of 250 mtorr, assume $T_e = 20{,}000$ K.
(a) What is the typical mean free path between electron–gas collisions?

(b) Calculate the collision frequency. How many electron collisions will occur during each applied field cycle?

7. Derive an expression for the time it takes a positive ion to traverse the cathode sheath assuming collisionless motion in a uniform electric field. Calculate this time for an Ar^+ ion if the sheath thickness is 0.5 mm and the target voltage is 800 V.

   Why is the energy of ion bombardment of the cathode expected to be higher for DC rather than high frequency operation?

8. The electrical conductivity (σ in units of $(\Omega\text{-cm})^{-1}$), or reciprocal of the resistivity of a medium is given by $\sigma = \sum_i n_i q \mu_i$ where n_i, q, and μ_i are the density, charge, and mobility of the charge carriers. Making appropriate assumptions, estimate the electrical conductivity of a 100 mtorr Ar plasma whose electron density is $10^{10}/\text{cm}^3$ and energy is 2.5 eV. How does this value compare to the conductivity of typical metals and semiconductors?

9. (a) An electron-beam gun with a filament–anode spacing of 0.5 cm operates at 10 kV and delivers a current of 1 A. For a reactive evaporation process the vacuum chamber contains oxygen at a pressure of 10^{-4} torr. If the first ionization potential of O_2 is 12.1 V, what is the probability of ionizing O_2 in the vicinity of the electron gun?

   (b) Electrons pass between electrodes that are spaced 5 cm apart and support a voltage difference of 1000 V. At what O_2 pressure will 50% ionization occur?

10. A simple approximation by Sigmund for the sputter yield when the projectile energy (E) is less than 3 keV is

$$S(E) = \frac{5}{3U_S}\sqrt{\frac{Z_t E}{1000}}.$$

   This relation holds when $\frac{1}{5} < Z_p/Z_t < 5$, where Z_p and Z_t are the atomic numbers of the projectile and target, respectively. Calculate S for Si bombarded with 500-eV Ar^+ ions. Assume the escape barrier energy for Si is 7.8 eV. How does the value of S compare with that of Eq. 4-35?

11. A parallel plate plasma reactor with 60 cm diameter electrodes contains argon at a pressure of 20 Pa. The electron temperature is 2 eV while the ion and neutral temperature is 0.03 eV. If the plasma density is $10^{16}/\text{m}^3$ calculate:

   (a) The mean electron velocity.

   (b) The Debye length.

(c) The electron mobility.
(d) The degree of ionization.
Are the criteria for a true plasma met in this case?

12. The diode plasma configuration of Fig. 4-1 outwardly resembles the electrochemical cell that is widely used to electrodeposit thin metal films on metallic substrates. What are the differences in the operation of each of these systems? In particular, focus on the distinctions between ionized gases and aqueous electrolytes containing ions, and the dynamics of charge carrier motion between the electrodes.

13. Theory indicates that the kinetic energy (E) and angular spread of neutral atoms sputtered from a surface are given by the distribution function

$$F(E, \phi) = \mathrm{CSE} \cos \phi / (E + U_S)^3$$

where U_S = binding energy of surface atoms, ϕ = angle between sputtered atoms and the surface normal, and C = constant.
(a) Sketch the dependence of $f(E, \phi)$ vs E for two values of U_S.
(b) Show that the maximum in the energy distribution occurs at $E = U_S/2$.

14. (a) Using the conservation of both energy and momentum, prove Eq. 4-27.
(b) Argon and helium ($^4He^+$) ions make head-on collisions with tin and oxygen nuclei at rest. What fraction of the kinetic energy of Ar^+ and $^4He^+$ is transferred to each of these atoms?
(c) Suppose a 2 MeV $^4He^+$ ion makes a head-on collision with a Sn atom at rest which then collides head-on with a resting O atom. What is the resulting energy of the oxygen atom?

15. Implantation of 10^{16} phosphorus dopant ions/cm^2 into silicon at a energy of 160 keV yields a Gaussian profile with a peak dopant concentration of 4.8×10^{20}/cm^3, recorded at a subsurface distance of 0.18 μm. At the Si surface the dopant concentration was 3×10^{19}/cm^3. What is the dose, projected range, and projected straggle of this distribution?

16. The critical ion/atom (I/A) ratio for the annealing of stress in Ge films by Ar ion bombardment was experimentally determined to obey the relation $I/A = 150E^{-1.59}$, for values of energy E (eV) between 200 and 2000 eV. At higher energies it was found that $I/A = 4760E^{-2.04}$ for 2000 eV $< E <$ 5000 eV. Plot these results on Fig. 4-19. How do you reconcile the location of these plots within the overall map?

REFERENCES

1. I. Langmuir, *Phys. Rev.* **33**, 954 (1929).
2. J. L. Vossen and J. J. Cuomo, in *Thin Film Processes*, eds. J. L. Vossen and W. Kern. Academic Press, New York, 1979.
3. D. M. Sanders, *J. Vac. Sci. Technol.* **A7**, 2339 (1989).
4. B. Chapman, *Glow Discharge Processes*. John Wiley & Sons, New York, 1980.
5. A. Grill, *Cold Plasma in Materials Fabrication*. IEEE Press, New York, 1994.
6. J. E. Mahan, *Physical Vapor Deposition of Thin Films*, John Wiley & Sons, New York, 2000.
7. M. A. Lieberman and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*. John Wiley & Sons, New York, 1994.
8. D. L. Flamm, and D. L. Flamm and G. K. Herb, in *Plasma Etching — An Introduction*, ed. D. M. Manos and D. L. Flamm. Academic Press, Boston, 1989.
9. D. M. Mattox, *J. Vac. Sci. Technol.* **A7**(3), 1105 (1989).
10. J. W. Rabalais and D. Marton, *Nucl. Instrum. Method* **B67**, 287 (1992).
11. J. B. Malherbe, *Crit. Rev. Solid State Mat. Sci.* **19**(2), 55 (1994).
12. P. Sigmund, *Phys. Rev.* **184**, 383 (1969).
13. L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and Thin Film Analysis*. North Holland, New York, 1986.
14. J. Lindhard, M. Scharff, and H. E. Schiott, *Mat. Fys. Medd. Dan. Vid. Selsk.* **33**, 14 (1963).
15. J. E. Mahon and A. Vantomme, *J. Vac. Sci. Technol.* **A15**, 1976 (1997).
16. W. D. Westwood, in *Microelectronic Materials and Processes*, ed. R. A. Levy. Kluwer Academic, Dordrecht, The Netherlands, 1989.
17. N. Langreid and G. K. Wehner, *J. Appl. Phys.* **32**, 365 (1961).
18. G. K. Wehner, in *Handbook of Thin Film Technology*, eds. L. I. Maissel and R. Glang. McGraw-Hill, New York, 1970.
19. H. Oechsner, *Appl. Phys.* **8**, 185 (1975).
20. R. E. Johnson and B. U. R. Sundqvist, *Physics Today* **45**(3), 28 (1992).
21. E. Kay and S. M. Rossnagel, in *Handbook of Ion Beam Processing Technology*, ed. J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman. Noyes Publications, Park Ridge, NJ, 1989.
22. E. S. Machlin, *Materials Science in Microelectronics*, *The Relationships between Thin Film Processing and Structure*. GIRO Press, Hastings on Hudson, NY, 1995.
23. H. Tsai and D. B. Bogy, *J. Vac. Sci. Technol.* **A5**, 3287 (1987).
24. G. K. Wehner, *J. Vac. Sci. Technol.* **A3**, 1821 (1985).
25. B. A. Banks, in *Handbook of Ion Beam Processing Technology*, eds. J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman. Noyes Publications, Park Ridge, NJ, 1989.
26. N. Kuratani, A. Ebe, K. Ogata, I. Shimizu, Y. Setsuhara, and S. Miyake, *J. Vac. Sci. Technol.* **A19**, 153 (2001).
27. H. Windischmann, *Crit. Rev. Solid State Mat. Sci.* **17**, 547 (1992).
28. M. Ohring. *The Materials Science of Thin Films*. Academic Press, Boston, 1992.
29. R. M. Bradley, in *Handbook of Ion Beam Processing Technology*, eds. J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman. Noyes Publications, Park Ridge, NJ, 1989.
30. R. Kelly and A. Miotello, in *Materials and Processes for Surface and Interface Engineering*, ed. Y. Pauleau. Kluwer, The Netherlands, 1995.

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 5

Plasma and Ion Beam Processing of Thin Films

5.1 INTRODUCTION

This chapter details the ways in which plasmas and ion beams have been harnessed to yield efficient methods for both deposition and etching of thin films used in a variety of technologies. Historically, the challenges of microelectronics, virtually single-handedly, thrust a long-standing but largely scientific interest in the physics and chemistry of plasmas and glow discharges into the main arena of thin-film technology. As we shall see, it was a combination of the need for improved thin-film metallizations and insulating films, in the face of the inexorable shrinkage of device feature dimensions, that drove the growing adoption of plasma processing (Refs. 1, 2). A brief sketch of the incremental adoption of plasma usage in solid-state electronics applications will provide some perspective on reasons for the treatment and organization of this chapter's subject matter.

In the early days of integrated-circuit (IC) technology, metallizations for contacts to the doped-silicon regions as well as for interconnections between devices were primarily thermally evaporated aluminum films. But as device feature dimensions continued to shrink in the manner suggested by Moore's law (i.e., a doubling of the transistor count every 18 months), ever greater demands were made on materials. For example, pure Al conductors were a reliability risk because they occasionally suffered fuselike open circuiting or electromigration failure at large current densities (see Section 11.6.3). Aluminum–copper interconnects were found to be more resistant to such damage but such alloys could not be easily evaporated (Section 3.2.3.3). However, film stoichiometry was maintained by glow-discharge sputtering, which in the early 1970s became an integral part of metallization processing.

A parallel need arose to deposit insulating film materials such as SiO_2 and silicon nitride in order to passivate and environmentally protect devices. In order to maintain film compositional integrity, low-temperature deposition was essential to prevent unwanted diffusion and chemical reactions between and among surrounding materials. In the form of RF (radio frequency) sputtering and plasma-assisted or enhanced chemical vapor-deposition (PECVD), plasmas facilitated the deposition of insulating films.

In another challenge, decreasing device dimensions made it more difficult to sharply delineate features after lithographic patterning and etching of the involved photoresist, metal, and masking films. Instead, the desired features were rounded and undercut, a consequence of employing wet etchants to remove unwanted films. This difficulty, due to isotropic etching attack in fluids, was overcome with the introduction of dry plasma and ion-beam etching methods. Largely because of ion bombardment normal to the film surface, material was anisotropically removed, enabling vertical wall topographies of high aspect ratio to be sculpted in films that covered an increasingly nonflat IC chip terrain. One of the early triumphs was the ability of CF_4–O_2 plasmas to etch windows through silicon-nitride passivating films, thus enabling underlying metallizations to be bonded for packaging. Prior to plasmas, silicon nitride could not be etched with wet chemicals. In addition to improved feature definition, the benefits of plasma etching relative to wet etching include better rection-product removal and elimination of the environmental problem caused by liquid waste disposal. Plasma etching began to be widely adopted in the early 1980s.

An appreciation of the impact of plasmas in semiconductor processing is shown in Fig. 5-1, where eight levels of interconnects and surrounding insulation above the device plane of integrated circuits are exposed in cross section (also see Fig. 6-1). Fabrication of this remarkable structure would have been impossible without plasmas to deposit and etch the metal and dielectric films; this includes the corresponding thin adhesion and barrier layers over steps, trenches, and vias, i.e., contact holes and vertical links between metallization levels. Since the very speed of IC chips hinges on the RC (resistance–capacitance) time constant in these metal–dielectric interlayers, monumental efforts are currently underway to reduce the magnitude of this product. This accounts for the introduction of copper metallizations with lower resistivity than aluminum, and the almost frantic search for new low dielectric constant intermetal fillings.

Advances in plasma-processing technology that originated in microelectronics have now broadly permeated other coating applications as well. Industries producing magnetic films for information storage, hard coatings for wear-resistant applications, and optical coatings on glass spring to mind as beneficiaries of these advances. In varying degrees they share a common

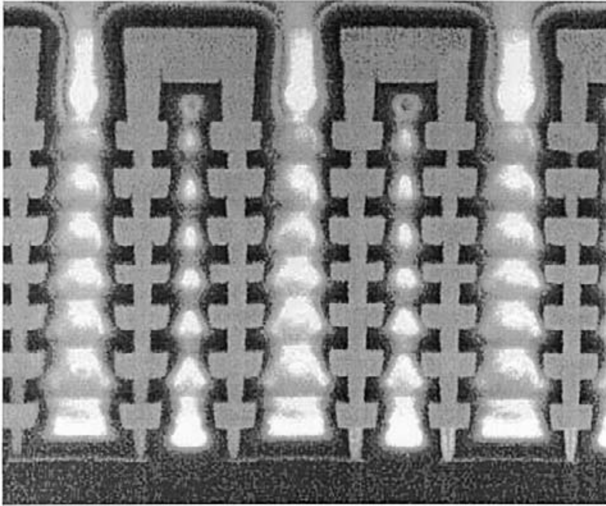


Figure 5-1 Scanning electron micrograph cross section of an advanced IC chip with eight levels of copper interconnections plus vias (the gray, connected vertical structures with seven crosses per stack) surrounded by low dielectric constant ("black diamond") insulation (black and white). At the bottom are unresolved 0.13 μm CMOS devices. (Courtesy of L. Liu and D. J. Holden, TSMC.)

interest in sputtering, PECVD, and ion beam technologies. For this reason, this chapter builds on the fundamental knowledge base of the two previous chapters to illustrate how plasmas and ion beams are practically exploited in deposition and etching processes. System configurations (e.g., DC, AC, magnetron, microwave) and hardware, film deposition issues (e.g., rates, step coverage, incorporation of defects), plasma etching, and a treatment of novel hybrid PVD processes are among the topics that will be addressed in this chapter. Recommended reviews that treat much of this chapter's subject matter can be found in Refs. 3–9.

5.2 DC, AC, AND REACTIVE SPUTTERING PROCESSES

5.2.1 Introduction

For convenience we will divide sputtering processes into four major categories: namely, DC, AC (mostly RF), reactive, and magnetron. We

recognize, however, that there are important variants within each category (e.g., DC bias) and even hybrids between categories (e.g., reactive magnetron). In this section the first three sputtering methods will be discussed. Magnetron sputtering is practiced in DC, AC, and RF as well as reactive variants and has significantly enhanced the efficiency of these processes. Because magnetron sputtering is now the dominant method of physically depositing films by plasma methods, it merits a section all its own.

In common, all of these sputtering processes are basically configured as planar diodes with facing anode and cathode electrodes (Fig. 4-1). The thin films that deposit are derived from target cathodes which play an active role in the plasma.

5.2.2 Targets

At present, targets of virtually all important materials are commercially available for the production of sputtered thin films. They include compositions representing all of the important classes of solids, e.g., metals, semiconductors, oxides, fluorides, borides, carbides, nitrides, silicides, and sulfides. Furthermore, targets come in a variety of shapes (e.g., rectangular, triangular, and circular plates, conical disks, toroids) and in assorted sizes. For example, targets range in size from $\sim 10\,\text{cm}$ dimensions for laboratory purposes to several meters for coating architectural glass.

In general, the metal and alloy targets are fabricated by melting either in vacuum or under protective atmospheres, followed by thermomechanical processing involving some combination of forging, extrusion, rolling, and annealing, and ending with machining to final dimensions. Refractory alloy targets, e.g., Ti–W, are hot-pressed via the powder-metallurgy route. Similarly, ceramic and oxide targets are generally prepared by hot pressing of powders. The elemental and metal targets generally have purities of 99.99% or better, whereas those for the nonmetals have typical upper purity limits of 99.9%. For microelectronic purposes purities an order of magnitude higher are demanded together with stringent stoichiometries. While the metal targets usually attain theoretical densities, lower densities are achieved during powder processing. These metallurgical processing realities are particularly evident in lower sintered-density targets, which are prone to greater arcing rates, emission of particulates, release of trapped gases, nonuniform target erosion, and deposition of generally inferior films (Ref. 10). Prior to use, targets must be bonded to a cooled backing plate to avoid thermal cracking during sputtering. Metal-filled epoxy cements of high thermal conductivity are employed for this purpose.

5.2.3 DC SPUTTERING

Virtually everything mentioned about sputtering to this point has dealt with DC sputtering, also known as diode or cathodic sputtering. There is no need to further discuss the system geometry (Fig. 4-1), the discharge environment (Section 4.2), the ion surface interactions (Section 4.5.1) or sputter yields (Section 4.5.2.2). It is worthwhile, however, to note how the relative film deposition-rate depends on sputtering pressure and current. At low pressures, the cathode sheath is wide, ions are produced far from the target, and their chances of being lost to the walls is great. The mean free electron path between collisions is large, and electrons collected by the anode are not replenished by ion-impact-induced secondary-electron emission at the cathode. Therefore, ionization efficiencies are low and self-sustained discharges cannot be maintained below about 10 mtorr. As the pressure is increased at a fixed voltage, the electron mean free path is decreased, more ions are generated, and larger currents flow. But if the pressure is too high, the sputtered atoms undergo increased collisional scattering and are not efficiently deposited. Trade-offs in these opposing trends are shown in Fig. 5-2, where optimum operating conditions are shaded in and include the relatively high operating pressure of ~ 100 mtorr. At a given pressure the discharge current–voltage characteristic is steep, typically varying as $i \sim K(P)V^m$, with $m \simeq 5$ or more, and $K(P)$ a constant that increases with pressure.

A relation that has been suggested for the film-growth rate ($\dot{G}$) during DC sputtering is (Ref. 11)

$$\dot{G}\left(\frac{\text{cm}}{\text{s}}\right) \approx \frac{\bar{\mathscr{P}}_{\text{d}}\langle x_{\text{th}}\rangle}{g\rho(1+\gamma_{\text{e}})E}, \tag{5-1}$$

where $\bar{\mathscr{P}}_{\text{d}}$ is the discharge power-density (W/cm^2), $\langle x_{\text{th}}\rangle$ is the mean distance from the cathode sputtered atoms travel before they become thermalized, g is the cathode–anode gap distance, ρ is the atomic density (atoms/cm^3), γ_{e} is the Townsend secondary-electron emission coefficient and E is the average sputtering energy (~ 1 keV). This expression reflects the important experimental observations that the deposition rate scales directly with applied power, and inversely with gap spacing, all at fixed pressure and cathode voltage. However, it is the $\langle x_{\text{th}}\rangle$ term that is interesting. During DC-diode sputtering the atoms that leave the target with typical energies of 5 eV undergo gas scattering events in passing through the plasma gas; this is so even at low operating pressures. As a result of repeated energy-reducing collisions they eventually *thermalize* or reach the kinetic energy of the surrounding gas. This happens at the distance $\langle x_{\text{th}}\rangle$, where their initial

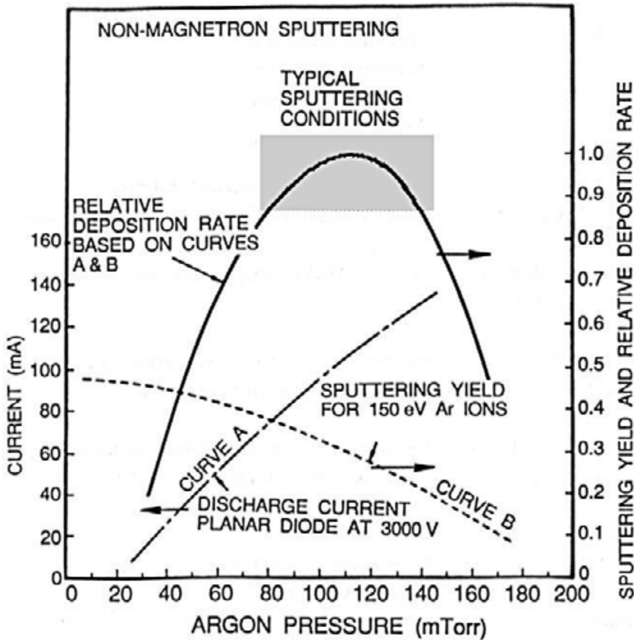


Figure 5-2 Influence of working pressure and current on film deposition rates in nonmagnetron sputtering. (From Ref. 3.)

excess kinetic energy, so necessary to provide bombardment of the depositing film, has dissipated. No longer directed, such particles now diffuse randomly. Not only is there a decrease in the number of atoms that deposit, but there is little compaction or modification of the resulting film structure. By virtue of lower operating pressures magnetron sputtering removes these disadvantages.

Despite the fact that simple DC sputtering was historically the first to be used and has an appealing simplicity, it is no longer employed in production environments. The reasons are not hard to see. For example, film deposition rates are simply too low, e.g., a few hundred Å/min at most for many metals. These rates cannot be appreciably raised at higher operating pressures, because in addition to more gas scattering, increased contaminant levels of O_2 and H_2O in chamber gases can oxidize cathodes. Thin insulating layers that form on the target further reduce the current and deposition rates. A fundamental problem is the small ionization cross sections at typical electron energies in the plasma. Even at optimum operating conditions, secondary electrons emitted from the cathode have an

appreciable probability of reaching the anode or chamber walls without making ionizing collisions with the sputtering gas.

5.2.3.1 Triode Sputtering

Over the years several variants of DC sputtering have been proposed to enhance the efficiency of the process. In one known as triode sputtering and shown in Fig. 5-3a, a filament cathode and anode assembly is installed close to the target, parallel to the plane of the target–substrate electrodes. When the filament is heated to high temperatures, thermionic emission and injection of electrons into the plasma increase the gas-ionization probability. The resulting ions are then extracted by the negative target potential. A disadvantage of triode sputtering is the nonuniform plasma density over

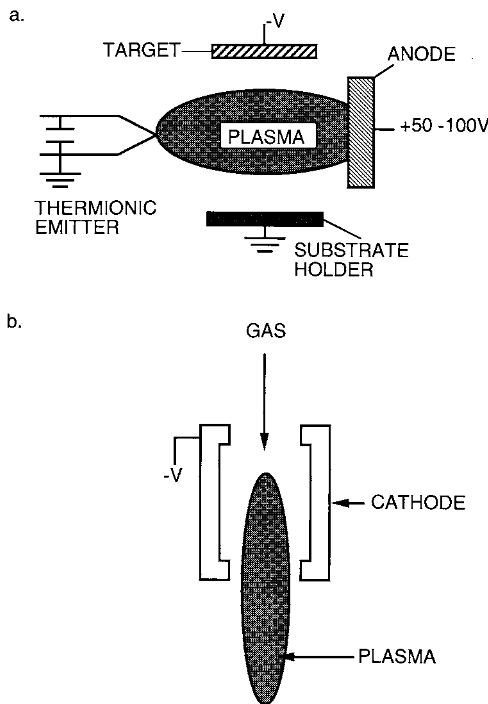


Figure 5-3 (a) Triode sputtering configuration utilizing a thermionic electron emitter. A magnetic field may be applied along the emitter–anode axis. (After L. I. Maissel in *Handbook of Thin Film Technology*, L. I. Maissel and R. Glang, eds., McGraw-Hill, New York, 1970.) (b) Hollow cathode source. An axial magnetic field may also be applied.

target surfaces, which leads to uneven erosion of metal. However, the discharge can be maintained at low pressures, and this is an advantage.

5.2.3.2 Hollow Cathode

The use of a hollow cathode is another way to enhance ionization in a DC discharge. Hollow-cathode sources are little more than a tube, or cylindrical hole in an electrode, through which ionized gas flows (Fig. 5-3b). Operation of hollow cathodes is related to the "pendel-electron" effect. The discharge cavity consists of a negative glow region separated from opposite cathode surfaces by two dark spaces. In a pendulum-like motion, electrons emitted from the cavity surface are accelerated toward the negative glow, penetrate the opposite dark space, turn around, and enter the negative glow again. This to-and-fro electron motion leads to an enhanced gas-ionization rate and substantially increases the plasma density within the cavity. Compared to glow discharges contained between parallel-plate electrodes, the current densities in hollow cathodes are one to two orders of magnitude higher. A highly ionized plasma is generated and issues as a jet into the main discharge region where it increases the plasma density. In this sense, hollow cathodes are outwardly similar to magnetron discharges.

5.2.3.3 Sputtering in the Presence of Substrate Bias

Negative DC or RF electric fields are often applied to bias the substrate in order to vary the flux and energy of depositing charged species. With target voltages of -1000 to -3000 V, bias voltages of -50 to -300 V are typically used. Because of charge-exchange processes in the anode dark space, very few discharge ions strike the substrate with full bias voltage. Rather, a broad low-energy distribution of ions and neutrals bombards the growing film. The technique of applying bias has been utilized in all sputtering configurations, e.g., DC, magnetron, RF, reactive, and has been effective in modifying a broad range of properties in deposited films. Examples include (Ref. 3) improved film adhesion during initial stages of film formation; substantial improvement in step coverage (e.g., Al); increased film density (e.g., Cr); significant reduction in resistivity of metal films (e.g., Ta, W, Ni, Au, and Cr); change in film hardness and residual stress (either increase or decrease); enhanced optical reflectivity (e.g., W, Ni, and Fe films); change in electrical properties of dielectric films (e.g., increase in resistivity, decrease in relative dielectric constant of SiO_2 films); reduction in etch rate of silicon-nitride films; and alteration of film morphology (e.g., columnar microstructure of Cr replaced by a compacted, fine-grained structure).

Although the details are not always clearly understood, there is little doubt that bias controls the film gas content. Thus it was found that the fraction of gas atoms incorporated is proportional to V_b^2, where V_b is the bias voltage (Ref. 12). Also, chamber gases (e.g., Ar, O_2, N_2) sorbed on the growing film surface may be resputtered during low-energy ion bombardment. In such cases both weakly bound physisorbed gases (e.g., Ar) and strongly attached chemisorbed species (e.g., O or N on Ta) apparently have larger sputtering yields and lower sputter threshold voltages. In other cases sorbed gases may have anomalously low sputter yields and will be incorporated within the growing film. In addition, energetic-particle bombardment prior to and during film formation and growth promotes numerous changes and processes at a microscopic level. These include removal of contaminants, alteration of surface chemistry, enhancement of nucleation and renucleation (due to generation of nucleation sites via defects and implanted and recoil-implanted species), higher surface mobility of adatoms, and elevated film temperatures with attendant acceleration of atomic reaction and interdiffusion rates.

Film properties are then modified through roughening of the surface, elimination of interfacial voids and subsurface porosity, creation of a finer, more isotropic grain morphology, and the elimination of columnar grains—in a way that strongly dramatizes structure–property relationships in practice.

In summary, there are few ways to broadly influence such a wide variety of thin-film properties, in so simple and cheap a manner, than by application of substrate bias. Later in Section 5.5.4 we shall revisit the subject of film modification due to ion bombardment, this time under controlled conditions where ion beams of well-defined energies are employed.

The last and most important variant of DC sputtering incorporates perpendicularly oriented electric and magnetic fields. These have the desirable effect of extending electron lifetimes, promoting electron–neutral collisions, and generally creating more intense plasmas (Section 4.3.4.2). How these effects are capitalized upon and implemented in practical magnetron sputtering systems is considered in Section 5.3.

5.2.4 AC (RF) SPUTTERING

In contrast to the earlier discussion in Section 4.3.6 dealing with the issue of sustaining AC discharges in an electrodeless environment, our present concern is with sputter deposition. Building on our experience with sputtering metal films, suppose we now wish to deposit thin SiO_2 films by using a quartz disk of thickness *d* as the target in a conventional DC diode-

sputtering system. For quartz the resistivity ρ is $\sim 10^{16}\,\Omega$-cm. In order to draw a current density j of 1 mA/cm^2, a voltage $V = \rho j d$ would have to be dropped across the target. Assuming $d = 0.1$ cm, substitution gives an impossibly high value of 10^{12} V; the discharge extinguishes and DC sputtering will not work. If a convenient level of $V = 100$ V is set, it means that a target with a resistivity exceeding $10^6\,\Omega$-cm could not be DC sputtered. This impasse can be overcome, however, if we recall that impedances of dielectric-filled capacitors drop with increasing frequency. Therefore, high-frequency plasmas ought to pass current through dielectrics the way DC plasmas do through metal targets. And since the sputter yields are essentially similar for both target materials, sputtering of a dielectric cathode in an AC plasma should be feasible. The trick is to sustain the plasma while ensuring positive-ion bombardment of the cathode. How this is accomplished is the subject of the next two sections.

5.2.4.1 AC Discharges Revisited

Consider what happens when an AC signal is applied to electrodes within a plasma. Smith (Ref. 13) has presented a description of plasma behavior as a function of operating frequency that is worth reviewing. It appears that 1 MHz is a very rough dividing line between low- and high-frequency plasma behavior. In the former range of 60 Hz, audio, and low-frequency RF, ions are sufficiently mobile to establish a complete discharge at each electrode in each half cycle. DC sputtering conditions essentially prevail at both electrodes, which alternate as cathodes and anodes. Above $\sim$1 MHz several important effects occur. First, electrons oscillating in the glow region acquire enough energy to cause ionizing collisions. The mechanism for electron-energy gain resembles electronic ping-pong, where the pulsating plasma sheaths effectively reflect electrons back and forth between the sheath and plasma. And if the timing is right, electrons returning to the plasma may gain an additional energy boost from the simultaneously expanding sheath. The net result is that there is no need to rely on secondary electrons emitted from the cathode to sustain the discharge. Secondly, at radio frequencies, voltage can be coupled through any kind of impedance so that the electrodes need not be conductors. This makes it possible to sputter any material irrespective of its resistivity. A third important effect is the reduction of ion energy above $\sim$1 MHz. If for the same applied power, more is partitioned to electrons in raising their energies, less is available to ions. Essentially more cathodic cycles are now required for ions to traverse an effectively lower sheath potential. A clear manifestation of the lower ion-bombardment energy is the observed stress transition in deposited silicon-nitride films (Ref. 14). At frequencies below $\sim$1 MHz, energetic ion bombardment peens the film surface, stressing it residually in compression.

Above 1 MHz, ion bombardment is effectively absent and tensile film stresses develop.

5.2.4.2 Target Self-Bias

Although we have established that RF discharges are sustainable, it is not clear how they can be used for sputtering. RF sputtering essentially works because the target *self*-biases to a negative potential. Once this happens, it behaves like a DC target where positive ion-bombardment sputters away atoms for subsequent deposition. Negative target bias is a consequence of the fact that electrons are considerably more mobile than ions and have little difficulty in following the periodic change in the electric field. The disparity in electron and ion mobilities means that isolated positively charged electrodes draw more electron current than comparably isolated negatively charged electrodes draw positive ion current. For this reason the discharge current–voltage characteristics are asymmetric and resemble those of a leaky rectifier or diode as indicated in Fig. 5-4. Even though it applies to a DC discharge, it helps to explain the concept of self-bias at AC (RF) powered electrodes.

As the pulsating AC signal is applied to the target, a large initial electron current is drawn during the positive half of the cycle. However, only a small ion current flows during the second half of the cycle. This enables a net current averaged over a complete cycle to be different from zero; but this cannot happen because no charge can be transferred through the capacitor. Therefore, the operating point on the characteristic shifts to a negative voltage — the target bias — where no net current flows.

It is not too difficult (Ref. 15) to quantitatively estimate the time interval (Δt) that an AC powered cathode is positively charged if the total cycle has a period τ. During the positive portion of the cycle the total electron charge drawn to the cathode from the sheath is $qn_e v_e \Delta t$, where q, n_e, and v_e are the electronic charge, density, and velocity, respectively. When the cathode goes negative the impinging ion current $qn_i v_i$ induces a total secondary electron charge emission of $qn_i v_i(1+\gamma_e)(\tau-\Delta t)$, where γ_e is the Townsend secondary-electron coefficient. Because motion of an ion toward the cathode is equivalent to an electron moving away, the term 1 is included. Equating these two charges and noting that $n_e = n_i$, we have

$$\frac{\Delta t}{\tau} = \frac{1}{1 + (1 + \gamma_e)^{-1}(v_e/v_i)}. \tag{5-2}$$

Assuming $\gamma_e = 0.1$, and using typical v values for electrons and ions (Section 4.3.3), $\Delta t/\tau \simeq 10^{-3}$. Therefore, the cathode is negatively charged most of the time.

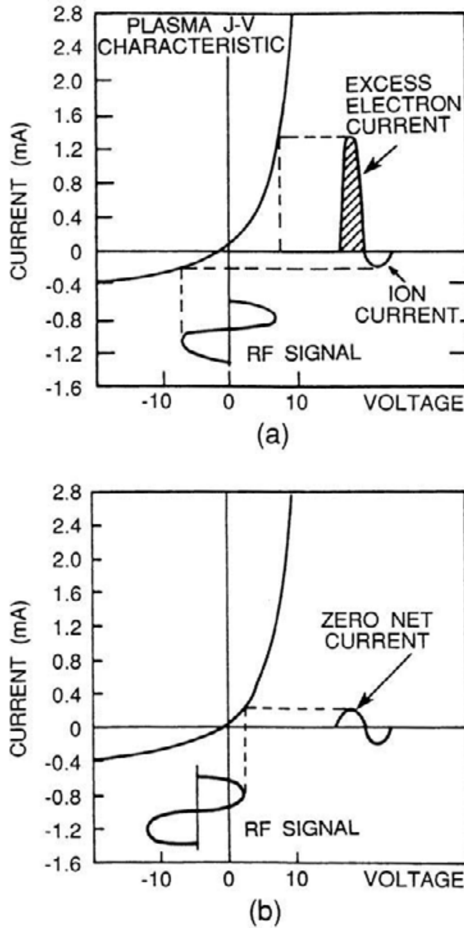


Figure 5-4 Formation of pulsating negative voltage on the capacitively coupled cathode in an RF discharge. (a) Net current/zero self-bias voltage. (b) Zero current/nonzero self-bias voltage. (From Ref. 3.)

5.2.4.3 Some RF Sputtering Issues

5.2.4.3.1 Systems

Since there is a rough equivalence between sputtering metals and dielectrics under respective DC and RF powering, we may expect similarities in electrode configurations. In fact, parallel-plate RF sputter deposition chambers schematically depicted in Fig. 4.1b physically resemble DC sputter systems. Electrically, the target is usually the powered electrode and it must

be similarly insulated from the grounded substrate and the rest of the chamber. Because the plasma discharge has capacitive as well as resistive characteristics, efficient coupling of the RF power supply to the target is not trivial. What is required is an impedance-matching network consisting of some combination of variable and fixed capacitors and inductors to ensure maximum power delivery. Practical RF frequencies employed range from 5 to 30 MHz. However, the frequency of 13.56 MHz has been reserved for plasma processing by the Federal Communications Commision and is most widely used. Peak-to-peak voltages of greater than 1000 V, current densities of 1 mA/cm^2, electron densities ranging from 10^9 to 10^{11} cm^{-3}, and discharge pressures of 0.5 to 10 mtorr are typical.

5.2.4.3.2 Electrode-Size Effects

In contrast to DC discharges where the plasma is confined to the cathode, RF plasmas tend to fill the chamber volume. The astute reader will realize that since AC electricity is involved, both electrodes should sputter. This presents a potential problem because the resultant film may be contaminated as a consequence. The equivalent circuit of the sputtering systems can be thought of as two series capacitors — one at the target sheath region (c_{RF}), the other at the grounded substrate (c_G) with the applied peak RF sheath voltage (V) divided between them (Ref. 16). Since capacitance is proportional directly to the electode area (A) and inversely to the sheath thickness d_s,

$$V_{RF}/V_G = c_G/c_{RF} = (A_G/A_{RF})[d_s(RF)/d_s(G)]. \tag{5-3}$$

Respective values for the voltages and sheath thicknesses at the indicated electrodes are V_{RF}, V_G and d_s(RF), d_s(G). Furthermore, for collision-free ion motion in the sheath, the current and its DC self-bias potential are related by the Child–Langmuir equation (Eq. 4-24). Assuming the same current density at each sheath, $d_s(RF)/d_s(G) = (V_{RF}/V_G)^{3/4}$. Therefore, elimination of $d_s(RF)/d_s(G)$ in Eq. 5-3 yields

$$V_{RF}/V_G = (A_G/A_{RF})^4. \tag{5-4}$$

This fourth-power dependence between variables means that a large value of A_G is very effective in raising the target sheath potential while minimizing ion bombardment of grounded fixtures. For efficient sputtering the area of the powered target electrode should be small compared to the total area of the other electrode. In practice the latter consists not only of the substrate stage and system ground but also of base plates, chamber walls, etc. Electrode area effects are also important in plasma-etching processes.

5.2.4.3.3 Concerns

All kinds of insulating as well as conductive films have been RF sputtered over the years, but the deposition rates are generally very low. Film stoichiometries are problematical as metal-rich films tend to be deposited. In addition, it is difficult to make high-purity, dense compound (e.g., oxide, nitride) targets. Since ceramic and oxide insulators are simultaneously mechanically brittle and poor thermal conductors, they are susceptible to thermal-stress cracking when subjected to ion bombardment at high power levels. For these reasons as well as the expense of power supplies, which limits process scale-up, simple inert-gas RF sputtering, though common in laboratory settings, has seen limited industrial use. The alternative process of reactive sputtering, considered in the next section, eliminates many of the disadvantages of RF sputtering.

In conclusion, it should be noted that besides the RF sputtering systems treated here, simple electrodeless, tubular AC reactors were introduced earlier (Fig. 4-5). But this by no means exhausts our discussion of the subject. In fact, RF plasmas are ubiquitous in thin-film processing, and the two most important applications have yet to be discussed, namely plasma etching (Section 5.4) and plasma-enhanced chemical-vapor deposition of films (Section 6.7). What distinguishes these chemical plasma processes from RF sputtering is the presence of reactive gases and the absence of targets.

5.2.5 Reactive Sputtering

In reactive sputtering, thin films of compounds are deposited on substrates by sputtering from *metallic* (not nonmetallic) *targets* in the presence of a *reactive gas* usually mixed with an inert working gas (invariably Ar). The most common compounds reactively sputtered and the reactive gases employed are briefly listed:

1. Oxides (oxygen)—Al_2O_3, In_2O_3, SnO_2, SiO_2, Ta_2O_5.
2. Nitrides (nitrogen, ammonia)—TaN, TiN, AlN, Si_3N_4, CN_x.
3. Carbides (methane, acetylene, propane)—TiC, WC, SiC.
4. Sulfides (H_2S)—CdS, CuS, ZnS.
5. Oxycarbides and oxynitrides of Ti, Ta, Al, and Si.

Irrespective of which of the above materials is being considered, during reactive sputtering the resulting film will usually be a solid solution alloy of the target metal doped with the reactive element (e.g., $TaN_{0.01}$), a compound (e.g., TiN), or some mixture of the two.

5.2.5.1 Hysteresis Effects

Westwood (Ref. 17) has provided a way of visualizing the conditions required to yield either alloys or compounds. These two regimes are distinguished in Fig. 5-5a illustrating the generic hysteresis curve for the total system pressure (P) as a function of the flow rate of *reactive* gas (Q_r) in the system. First, however, consider the dotted line representing the variation of P with flow rate of an *inert* sputtering as (Q_i). As Q_i increases, P increases because of the constant pumping speed (Eq. 2-16). An example of this characteristic occurs during Ar gas sputtering of Ta. Now consider what happens when reactive N_2 gas is introduced into the system. As Q_r increases from $Q_r(0)$, the system pressure essentially remains at the initial value, P_0, because N_2 reacts with Ta and is removed from the gas phase. But beyond a critical flow rate, Q_r^*, the system pressure jumps to the new value P_1. If no reactive sputtering took place, P would be somewhat higher (i.e., P_3). Once the equilibrium value of P is established, subsequent changes in Q_r cause P to increase or decrease linearly as shown. As Q_r decreases sufficiently, P again reaches the initial pressure.

The hysteresis behavior represents two stable states of the system with a rapid transition between them. In state A there is little change in pressure, whereas for state B the pressure varies linearly with Q_r. All of the reactive gas is incorporated into the deposited film in state A — the doped metal, and the atomic ratio of reactive gas dopant to sputtered metal increases with Q_r. The transition from state A to state B is triggered by compound formation on the metal target. Since ion-induced secondary-electron emission is usually much higher for compounds than for metals, Ohm's law suggests that the plasma impedance is effectively lower in state B than in state A. This effect is reflected in the hysteresis of the target voltage with reactive gas flow rate as schematically depicted in Fig. 5-5b. By substituting deposition rate for cathode voltage, this same hysteresis behavior also describes the reactive sputtering rate for metals and compounds as a function of Q_r.

5.2.5.2 Modeling Reactive Sputtering

Many of the observed features of reactive sputtering have been successfully modeled by Berg *et al.* (Ref. 18). A schematic of the transport of sputtered material from both target and substrate and the involved gas and ion fluxes is shown in Fig. 5-6a. In particular, both the target (t) and substrate (s) are partially covered with unreacted metal as well as a compound film produced through reaction with the reactive gas.

Assumptions and definitions of the model include (1) an elemental metal target has sputter yield S_m; (2) target sputtering is due only to the inert

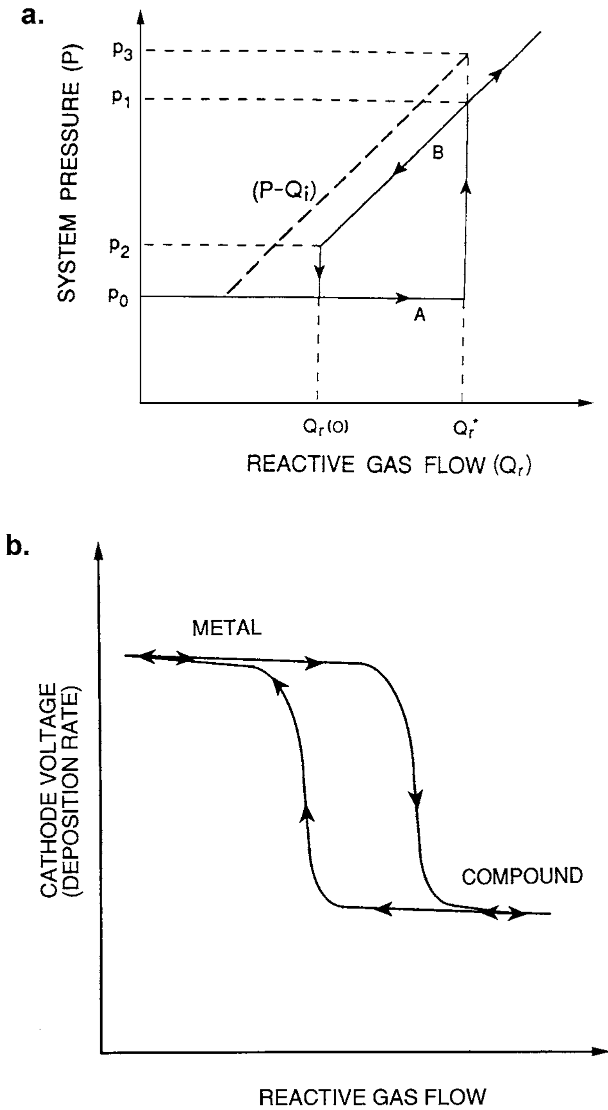


Figure 5-5 (a) Generic hysteresis curve for system pressure vs reactive-gas flow rate during reactive sputtering. Dotted line represents behavior with inert gas. (From Ref. 17.) (b) Schematic hysteresis curve for both cathode voltage and deposition rate vs reactive gas flow rate at constant discharge current.

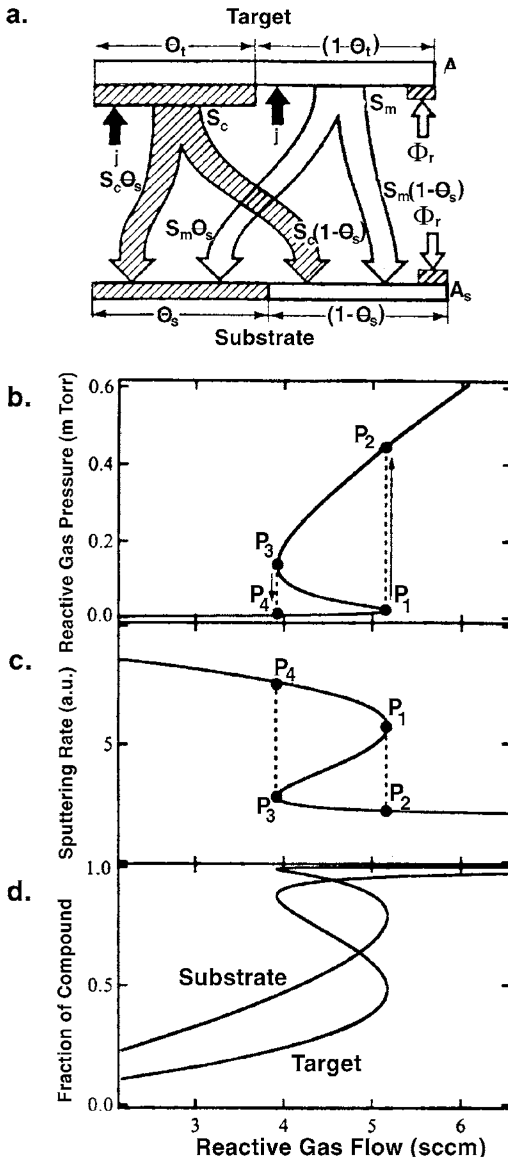


Figure 5-6 (a) Model of reactive sputtering. (b) Simulation of reactive gas pressure vs reactive gas flow. As Q increases, pressure values follow the path $P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow P_4$, tracing out a hysteresis loop. (c) Simulated sputtering rate vs reactive gas flow rate. (d) Simulation of target and substrate composition as a function of reactive gas flow. (From S. Berg *et al.* in *Handbook of Thin Film Process Technology*, edited by D. A. Glocker and S. I. Shah, 1998. Reprinted with the permission of the Institute of Physics Publishing and the authors.)

working gas; (3) compounds sputtered from the target with sputter yield S_c are deposited as molecules; (4) a uniform ion-current density (j) flows over the target area (A_t); (5) the collecting substrate surface area is A_s; (6) the fraction of the target area covered by the compound is θ_t, so that $1 - \theta_t$ represents the fraction of unreacted metal; (7) similarly, the fraction of the substrate covered by the compound is θ_s; (8) the reactive gas flux (Φ_r) and pressure (P_r) are related through Eq. 2-8; (9) reactive gas molecules do not stick to compound covered areas, but the sticking coefficient to metal surfaces is α_t. Three fundamental coupled mass-balance equations at the metal target, the substrate, and in the gas phase govern the reactive-sputtering process.

5.2.5.2.1 Target

At the target, the steady-state compound film formation rate is equal to the compound sputter rate, or

$$\Phi_r \alpha_t (1 - \theta_t) A_t a = (j/q) \theta_t A_t S_c, \tag{5-5}$$

where q is the electronic charge and a is the number of compound molecules formed by one reactive gas molecule. In addition, by summing over both metal and compound contributions, the total target erosion rate (R_t) is

$$R_t = (j/q)[S_c \theta_t + S_m (1 - \theta_t)] A_t. \tag{5-6}$$

5.2.5.2.2 Substrate

A similar mass balance between steady-state compound formation and "compound elimination" occurs at the substrate. No compound sputtering is assumed at substrates, but nevertheless compound elimination occurs because metal sputtered from the target covers the prior deposited compound film and thus effectively reduces θ_s. Therefore,

$$(j/q)[S_c \theta_t A_t](1 - \theta_s) + \Phi_r \alpha_s (1 - \theta_s) A_s b = (j/q)[S_m (1 - \theta_t) A_t] \theta_s / b, \tag{5-7}$$

where b is the number of metal atoms in the compound. Terms on the left side reflect the two contributions to the compound formation rate on the substrate. The first is due to sputter deposition of compound from the target onto the metal fraction of the substrate; the second is due to reaction of this metal with the reactive gas. This is set equal to the metal sputtered from the target, the term on the right. Once θ_t is calculated from Eq. 5-5, θ_s can be determined from Eq. 5-7.

5.2.5.2.3 Gas Kinetics

Lastly, the total input reactive gas flow (Q) must equal the loss due to the gettering rate of this gas at the surfaces of target (Q_t) and the substrate (Q_s), plus the throughput of reactive gas pumped out of the system (Q_p). Thus,

$$Q = Q_t + Q_s + Q_p, \tag{5-8}$$

where explicit expressions for Q_t and Q_s are $\Phi_r \alpha_t(1-\theta_t)A_t$ and $\Phi_r \alpha_s(1-\theta_s)A_s$, respectively. Also, by virtue of Eq. 2-16, $Q_p = P_r S$, where S is the system pumping speed.

Using the foregoing equations, computer simulations for a standard reactive sputtering process have been carried out. Results for the dependence of the reactive gas pressure (P_r), the sputtering rate (R), and the fraction of compound coverage (θ_t, θ_s) vs gas flow rate (Q) are shown in Figs. 5-6b, c, and d, respectively. Not surprising are the multivalued solutions, hysteresis, and switching between two stable states. Thus the behavior of Fig. 5-5a parallels that of Fig. 5-6b, while that of Fig. 5-6c has its counterpart in Fig. 5-5b. In concert, compound formation on the target and substrate also displays multivalued behavior as a function of Q. Importantly, modeling has revealed the presence of intermediate metastable states between metallic and nonmetallic sputtering; these lie along the line connecting P_1 and P_3 operating points in Fig. 5-6c. Through closed-loop control, increased rates of insulator sputtering have been achieved by operating at these intermediate states.

5.2.5.3 Film Deposition Rates, Stoichiometry, and Properties

In practice, sputter rates of metals drop dramatically when compounds form on the targets. Corresponding decreases in deposition rate occur because of the lower sputter yield of compounds relative to metals. The effect is very much dependent on reactive gas pressure. Sputtering effectively halts at high gas pressures in DC discharges, but the limits are also influenced by the applied power. Conditioning of the target in pure Ar is essential to restore the pure-metal surface and desired deposition rates.

Considerable variation in the composition and properties of reactively sputtered films is possible depending on operating conditions. The case of tantalum nitride is worth considering in this regard. One of the first electronic applications of reactive sputtering involved deposition of TaN resistors employing DC power at voltages of 3 to 5 kV, and pressures of 30 mtorr. The dependence of the resistivity of "tantalum nitride" films is shown in Fig. 5-7, where Ta, Ta_2N, TaN or combinations of these form as

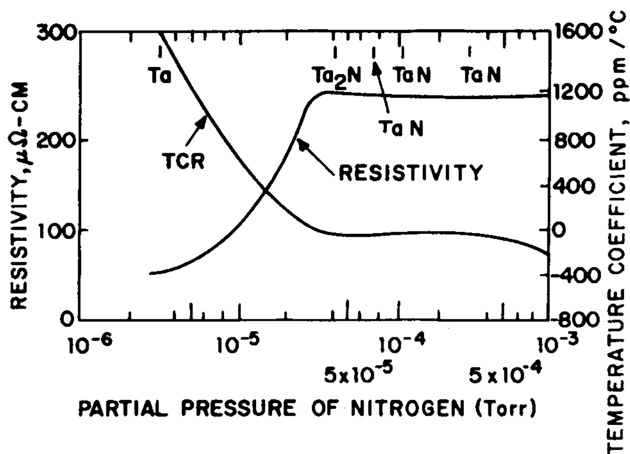


Figure 5-7 Influence of nitrogen on composition, electrical resistivity, and temperature coefficient of resistivity of reactively sputtered Ta films. (From Ref. 19.)

a function of N_2 partial pressure (Ref. 19). Color changes accompany the varied film stoichiometries. In the case of titanium nitride films, for example, the metallic color of Ti gives way to a light gold, then a rose, and finally a brown color with increasing nitrogen partial pressure.

In deciding whether to employ compound targets and sputter directly or to sputter reactively, it should be noted that it is easier to manufacture high-purity (ductile) metal targets than (brittle) compound targets. Therefore, compound film purity using the former is expected to be higher, particularly since high-purity reactive gases are now commercially available. Although reactive sputtering can be carried out in RF diodes, much higher deposition rates are possible in magnetrons employing pulsed-DC power (Section 5.3.3.1).

5.3 MAGNETRON SPUTTERING

5.3.1 INTRODUCTION

For reasons given earlier as well as others that follow, magnetron sputtering is the most widely used variant of DC sputtering. One to two orders of magnitude more current is typically drawn in magnetron than simple DC discharges for the same applied voltage. Important implications of this are higher deposition rates (e.g., $\sim 1\,\mu$m per minute for Al metal-

lization alloys) or alternatively, lower voltage operation than for simple DC sputtering.

Another important advantage is reduced operating pressures. At typical magnetron-sputtering pressures of a few millitorr, sputtered atoms fly off in ballistic fashion to impinge on substrates. Avoided are the gas phase collisions and scattering at high pressures which randomize the directional character of the sputtered-atom flux and lower the deposition rate. A fundamental reason for these beneficial effects in magnetrons is the displacement of the Paschen curve to lower Pd values (see Section 4.2.2) relative to simple discharges. Therefore, for the same electrode spacing and minimum target voltage a stable discharge can be maintained at lower pressures. These attributes of magnetron sputtering should be borne in mind as we now proceed to discuss three popular target geometries employed in magnetron sputtering.

5.3.2 Magnetron Configurations

5.3.2.1 Planar Magnetron

Of the three magnetron configurations, the planar one with parallel target and anode electrode surfaces is most common. In this geometry (Fig. 5-8) a typical DC electric field of $\mathscr{E} \sim 100\,\text{V/cm}$ is impressed between the target (cathode metal to be sputtered) and anode (substrate) plates. Small permanent magnets are arranged on the back of the target in either ellipse-like or circular rings depending on whether the targets are

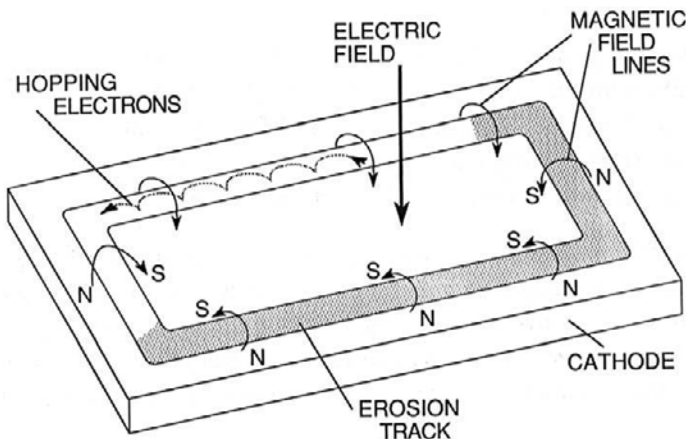


Figure 5-8 Applied fields and electron motion in the planar magnetron.

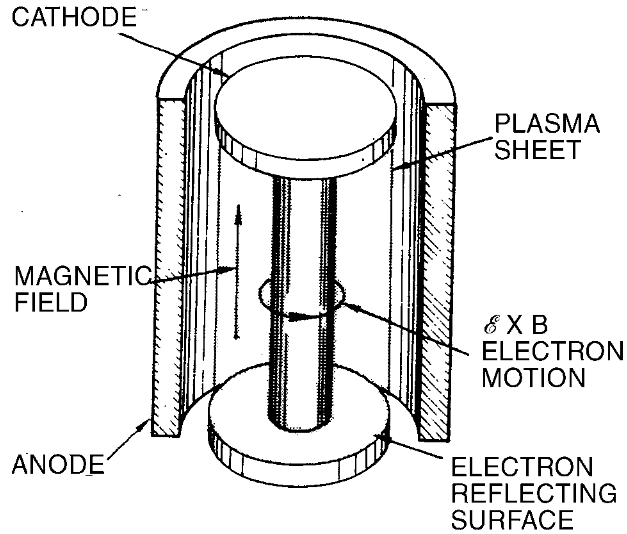
rectangular or circular in shape. To conceptually understand magnetron operation, consider the simple case of a single bar magnet, whose north and south poles are spaced far apart, mounted parallel to the back plane of the target. Roughly half of the total magnetic field (B) distribution emerges from the north pole roughly perpendicular to the front target face into the interelectrode space. Then the field lines arch over with a portion parallel to the target surface, i.e., the magnetron component. Finally B returns roughly normal to the target surface into the south pole, completing the closure of field lines. If we now imagine a linear array of such bar magnets, a B field tunnel is created on the front surface of the target. Electrons launched slightly off the target normal will initially spiral along the B field lines emanating normal to the target. In the region where $\mathscr{E}$ and B fields are perpendicular to one another, electrons are forced to drift in a cycloidal hopping motion along the tunnel track length (Section 4.3.4.2).

In actual planar magnetrons, small but powerful NdFeB bar magnets, ring the back of the target and are capable of creating fields of about ~ 0.5 G on the target front surface. Together with a strong central magnet plus soft-iron plates to complete the magnetic circuit, electron motion is confined to a so-called "racetrack." Because ionization of the working gas is most efficient above the racetrack, the plasma is most intense there. What has been described here is the conventional or *balanced* magnetron. Later in Section 5.3.3.2 we shall consider *unbalanced* magnetrons and the advantages associated with them.

5.3.2.2 Cylindrical-Post Magnetron

In contrast to the planar target–anode configuration, the cylindrical-post magnetron (Fig. 5-9, left) consists of electrodes arranged in a cylindrical geometry (Ref. 5). The target, fabricated as a cylinder, is surrounded in a 2π geometry by anode substrates so that a radial electric field is established between these electrodes. Simultaneously, a uniform magnetic field is applied parallel to the z direction by means of permanent or electromagnets, or established via a coaxial solenoid. Electrons emitted radially from the cathode immediately find themselves trapped between perpendicular electric and magnetic fields where they hop around the target circumference in closed orbits defined by the Lorentz-force geometry. As in the case of the planar magnetron, an intense cylindrical plasma column is created near the cathode by the enhanced gas-ionization efficiency there. Although uniform target erosion over its surface area is desired, and achieved over a restricted length, some barreling distortion of the cylindrical geometry is inevitable.

Cylindrical-post magnetron sputtering has been commercially employed for many years in coating razor-blade cutting edges with platinum/chromium. The sputter deposition of diamond-like carbon on stainless-steel



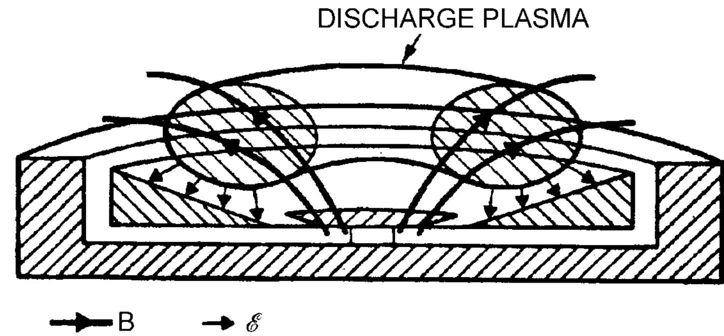


Figure 5-9 Nonplanar magnetron sputtering configurations. (Left) Cylindrical-post magnetron geometry. (From J. A. Thornton and A. S. Penfold, in J. L. Vossen and W. Kern, eds., *Thin Film Processes*. Academic Press, New York, 1978. Reprinted with the permission of Academic Press and A. S. Penfold.) (Right) Sputter-gun geometry. (Reprinted with the permission of S. M. Rossnagel.)

blades coated with a niobium glue layer is the latest advance in blade technology (Ref. 20). In hollow cathodes, the natural cylindrical sputtering geometry is essentially inverted with substrates placed interior to the target. Fibers can be coated continuously by feeding them along the hollow-cathode axis. Similarly, large numbers of small three-dimensional parts can be uniformly coated because the deposition flux impinges from all directions.

5.3.2.3 Sputter Gun

Also known as the S-gun, this sputtering source (Fig. 5-9, right) uses a toroidal-conical (triangular or trapezoidal cross section) target. With magnets bonded to the backside, the outwardly spreading axial B field lines perpendicularly intersect the radial $\mathscr{E}$ field lines between anode and cathode. In this way a toroidal plasma discharge is created and confined over the target surface. The gun assembly is usually positioned centrally within the chamber creating a deposition geometry approximating that of a planar (ring) evaporation source. In this manner substrates on a planetary holder can be coated as uniformly as with e-gun sources. It was the invention of the 3-inch sputter gun in 1968 by Peter Clarke that spurred the development of commercial magnetron-sputtering processes.

5.3.3 Variant and Enhanced Magnetron Sputtering Processes

5.3.3.1 Pulsed-Power Magnetron Sputtering

It is not uncommon for metal targets to be *poisoned* or acquire a different surface composition from that of the bulk, i.e., $\theta_t = 1$. Reactive gases are usually responsible and often result in the formation of dielectrics, e.g., oxides, which tend to readily charge up. If the surface potential is negative, ion impingement may initially cause sputtering but since the secondary-electron emission coefficient is less than unity, a positive surface charge develops. Ion bombardment is discouraged and then ceases, which prevents sputtering. However, given sufficient charge buildup, local dielectric breakdown of the insulating films causes arcing over the target surface (unipolar arcs) as well as to neighboring hardware (bipolar arcs). As a result, the target erodes by cratering with the ejection of liquid droplets that create defects when incorporated into growing films.

In addition to particle splatters, arcing prevents the stable operation of DC power supplies, particularly during the reactive sputtering of dielectric

films. Eliminating this serious problem is simple in principle; discharge the positive surface charge! Pulsed-power sputtering technology was developed (Refs. 5, 21) to do this. Typically, square or sinusoidal voltage pulses are applied to the target at frequencies of several tens of kilohertz, where pulse magnitude, polarity, width, and off-time can all be varied. In the case of bipolar pulses, for example, asymmetric positive and negative voltage polarities are applied, e.g., a -400 V pulse for 80–90% of the voltage cycle followed by a $+100$ V pulse for 20–10% of the remaining time. During the positive portion of the cycle, electrons are drawn to the target in order to neutralize the positive charge buildup that leads to arcs. After discharging, the target is again negatively powered and sputtered until it is time to discharge it again, etc. Commercial pulsed-power supplies incorporate arc-suppression electronics which sense the initiation of arc currents and either rapidly reduce the power or apply a positive voltage in response.

Noteworthy industrial applications of pulse technology to the reactive magnetron sputtering of three of the most important coating materials, Al_2O_3, TiN, and SiO_2, have been described by Schiller *et al.* (Ref. 22). In all cases large targets (a meter or larger in size), powered at levels of ~ 100 kilowatts and pulse rates ranging from 1 to 100 kHz, with typical current densities of 1–10 mA/cm^2, are involved. Significantly, in the case of Al_2O_3, which is normally deposited at 1000°C by CVD methods, pulsed-magnetron reactive sputtering produced comparably transparent and hard crystalline (α-Al_2O_3) coatings at 750°C, not the usual γ-Al_2O_3. Similarly, high-speed steel tools maintained at 450°C have been coated with TiN having a structure that was denser, with a lower surface roughness than similar coatings deposited without target pulsing. The fact that 3.7 m-long silicon targets have been used in reactively coating architectural glass attests to the scale-up capabilities of pulsed-power sputtering.

5.3.3.2 Unbalanced Magnetron Sputtering

We have already noted the benefits of electron and plasma confinement. But too much confinement at the cathode surface may be undesirable if not enough plasma is available at the anode to activate reactive gases or promote ion bombardment for the desired modification of growing films. The problem can be overcome through selective strengthening of the magnetic field at the target ends so that more of the secondary electrons can escape their confinement as shown in the so-called "type-II" unbalanced magnetron configuration of Fig. 5-10 (bottom). In essence, so-called "magnetic bottle" confinement occurs enabling current densities at the substrate to reach 10 mA/cm^2, roughly an order of magnitude higher than with a balanced magnetron. Similarly unbalancing the magnetron can raise the

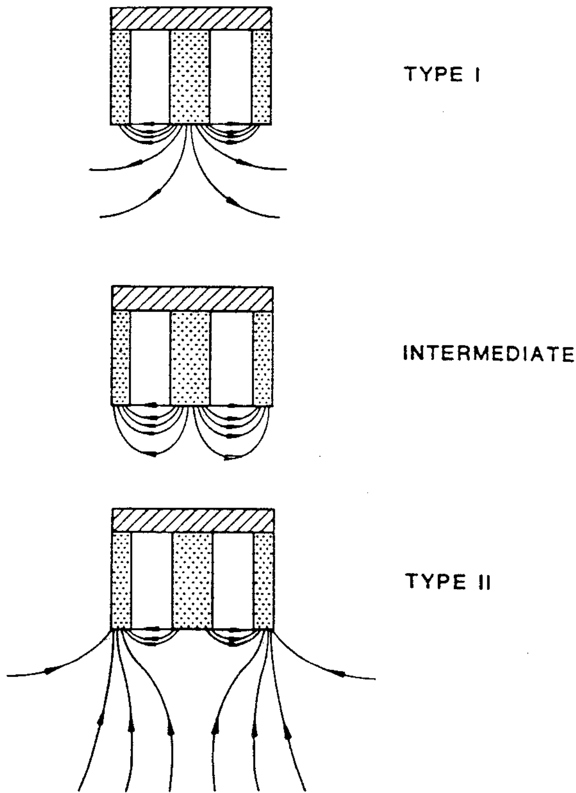


Figure 5-10 Planar magnetron configurations. (Top) Type-I (unbalanced). (Middle) Intermediate (balanced). (Bottom) Type-II (unbalanced). (From B. Window and N. Savvides, *J. Vac. Sci. Technol.* **A4**, 196, (1986). Reprinted with the permission of Dr. N. Savvides.)

plasma density at the substrate from the 10^6–$10^8\,cm^{-3}$ range to the 10^8–$10^{10}\,cm^{-3}$ range. Magnetrons can also be unbalanced through placement of stronger magnets in the middle relative to the outer end of the target ("type I"). Now the B field is expelled (Fig. 5-10, top) and lower electron and ion currents flow than for either of the former configurations. In practice, perfectly balanced magnetrons are difficult to achieve so that some ion bombardment of substrates usually occurs.

Unbalanced magnetron-sputtering processes and applications have been reviewed by Kelly and Arnell (Ref. 23) and Rohde (Ref. 24). The most significant commercial applications of this sputtering variant have been to the deposition of hard and wear-resistant coatings such as diamond-like

carbon, TiN, and assorted transition metal nitrides and carbides. Films for electronic (TiN, Ni/Cr), optical (indium–tin oxide, TiO_2), and corrosion-resistant (Al–Mg, Cd) purposes have also been deposited.

5.3.3.3 Dual Magnetrons

By independently powering two (or more) magnetrons of different composition simultaneously, the production of films with assorted material mixtures and graded compositions is possible. The process is conceptually similar to the idea of evaporation from two independently heated sources. Cosputtered mixed-oxide films of Al_2O_3/SiO_2, Al_2O_3/TiO_2, and ZrO_2/TiO_2, as well as similar mixed nitride films have been deposited this way (Ref. 25). One of the variants of pulsed-power reactive sputtering also employs two magnetron targets as shown in Fig. 5-11. But in this case bipolar pulses from a single power supply makes each target alternately behave as a cathode and anode, with an electrode discharge interval between each sputtering phase. In this way stable long-time reactive sput-

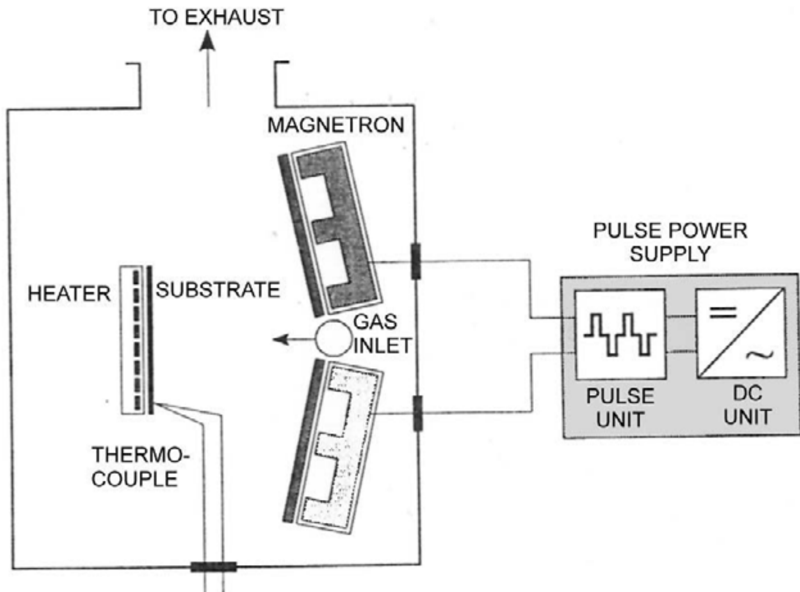


Figure 5-11 Arrangement for reactive pulsed-magnetron sputtering from dual Al targets. (From O. Zywitzki and G. Hoetzsch, International Conference on Metallurgical Coatings and Thin Films, San Diego, April 22–26, 1996). Reprinted with permission of the authors.

tering of dense Al_2O_3 films has been achieved without arcing. Closed-field unbalanced magnetron-sputtering methods employing two targets have been described (Ref. 23). Components include an N–S–N magnet array behind one target and an S–N–S magnet array behind the second target, and these are configured either parallel to the substrate or vertically opposed to each other sandwiching a rotating substrate in between.

One of the difficulties with DC reactive sputtering is the deposition of insulating films on the chamber walls, cathode, and anode as well as substrates. This essentially reduces the anode area, an effect known as the *disappearing anode*. Sputtering eventually terminates, and the electrodes must be cleaned to resume deposition. AC magnetron sputtering at 20–150 kHz in a dual-target configuration eliminates the problem because the targets clean on each half cycle (Ref. 5).

5.3.4 Magnetron Sputtering Concerns

5.3.4.1 Target Integrity

In planar magnetron sputtering, targets erode preferentially in the race-track region where the plasma is most intense. This leaves a kind of racetrack ditch or depression that progressively deepens, surrounded by a much larger target area that suffers less loss of metal. Such target erosion has a number of undesirable consequences that are discussed in this and the next section.

1. Low process efficiency is obvious since only 10–30% of the target surface area is utilized.
2. Inhomogeneous metal loss during reactive sputtering is such that the racetrack remains metallic while the surrounding target area becomes covered with insulator films, enhancing the probability of arcing.
3. The uniformity of deposited films, as much a key issue in sputtering as it is in evaporation, is adversely affected. First, the magnets are effectively closer to the eroded surfaces and the increasing magnetic field changes the operating conditions of the magnetron as a result. Second, as noted earlier (Section 4.5.2.4), the emission of sputtered atoms from a (small) planar cathode source is $\cos\phi$–like in character. But on the sloping surface of the target ditch the emission lobe is skewed at an angle relative to the cathode normal. This change in the directionality of sputtered material alters the film-thickness uniformity profile, making conformal coverage of steps and trenches more difficult.
4. Nonuniform plasma erosion tends to reduce target life. Initial processing of metal targets often leaves them in a state of spatially variable residual

stress that is in mechanical equilibrium when integrated over the total volume. But when metal is removed locally from the target during sputtering, the stresses relax and redistribute to achieve other equilibrium states. As target erosion continues, local tensile stresses may rise excessively to the point of fracture, prematurely terminating target life. Refractory metal-alloy and compound targets of limited ductility and complex shape are more susceptible to such failures than are pure metal targets of simple geometry.

5. Elastic distortion of the target may result in loss of contact to the water-cooled backing plate and a corresponding drop in cooling efficiency. Contributing to the problem are thermal stresses generated from ion-bombardment heating of the target.

A remedy for many of these interrelated problems is to maintain uniform target erosion during sputtering. This is practically accomplished by mechanically moving the array of permanent magnets over the back face of planar targets (Ref. 7). In such swept-field magnetrons, heart-shaped magnet arrays are eccentrically rotated to equalize the erosion pattern. Interestingly, the design of the magnet geometry and rotation may differ for different target metals because the angular sputter emission differs, e.g., Al and Ti.

5.3.4.2 Particulate Contamination

Plasmas can, surprisingly, be contaminant-laden environments. Because particulate incorporation in deposited sputtered films occurs with some frequency, the subjects of particle generation, transport, and removal have attracted considerable attention (Ref. 26). Particulates pose serious reliability concerns in microelectronic metallizations because of the potential for delamination, pinholes, and opening as well as shorting of interconnects. Similarly, particle incorporation in magnetic films used for storage disks can cause read–write errors, bad sectors, and total disk failure. A key question is, where do the particles come from? Conventional wisdom assumes that the primary sources of microcontamination are arcing, gas-phase nucleation (Section 1.7) of particles, and reactor wall and substrate films that have flaked off due to stress buildup. But elegant experiments involving *in situ* laser scattering of both targets and depositing films reveal that the mechanisms of particle formation during sputtering are more complex (Ref. 27).

The process starts at the target where the highly nonuniform plasma density, typical of magnetron sputtering, results in simultaneous material removal and redeposition in adjacent target regions. Redeposited metal films getter system gases, producing brittle oxide and nitride surface growths that often resemble acorns in appearance. These tend to charge up and fragment in the ensuing arcing. Target dust less than 50 nm in size may also

be generated by the explosive evolution of adsorbed water vapor in the first moments of ion bombardment. The particles get trapped and suspended in the plasma sheath where they grow by impact and sintering. Filaments, surface roughness, and nodules that develop in the weak plasma region between the racetrack and target edge tend to electrostatically trap the particles (Ref. 28). Because of particle impact the filaments grow and lengthen into the sheath, where like lightning rods they draw current, heat up, and glow. This stresses and weakens the negatively charged filaments to the point where they fracture and accelerate toward the substrate. Such a mechanism for particle formation was observed in carbon targets used to deposit diamond-like coatings on magnetic disks and during nonreactive sputtering of TiN, suggesting broad applicability to other plasma deposition and etching processes. Filament removal by properly scrubbing targets helps alleviate the problem. The use of pulsed power is effective in preventing arcing in the first place.

5.3.4.3 Film Uniformity: Step, Trench, and Via Coverage

As in the case in evaporation, the sputter deposition of uniformly thick blanket films on flat substrates is generally achieved without difficulty. It is when substrates contain steps and trenches that uniform film coverage of these features is considerably more challenging. This crucial need arises in the processing of integrated circuits where geometrically complex substrate topographies must be covered with thin films at different metallization levels (see Fig. 5-1). The many-faceted aspects of this important subject together with the role of sputter deposition techniques are subjects treated in a comprehensive monograph by R. A. Powell and S. M. Rossnagel (Ref. 7). Consider the problem of depositing sputtered metal into rectangular channels (trenches) as well as deep holes (vias) etched into a dielectric substrate. This usually first entails deposition of a thin diffusion-barrier film that conformally lines the trench, followed by subsequent metal filling. Because of the broad angular emission of sputtered atoms, the short target–substrate distance, and in-flight gas scattering, the impinging atom flux deposits in an isotropic fashion. When the aspect ratio (AR) of the hole (AR = hole depth/hole diameter) is high, early coverage of the corner and upper sidewalls physically shadows the lower regions from subsequent deposition. The result may be undesirable "bread-loafing" or overhanging films that bridge thin sidewalls (see Fig. 3-10b) and entrapment of keyhole voids.

Three strategies for addressing such trench coverage problems have been discussed by Rossnagel (Ref. 1). They all attempt to enhance the directionality of the impinging metal atoms. *Long-throw deposition* is the first approach. By increasing the target–substrate distance the atoms that move

laterally tend to be lost to the chamber walls, while those that are normally incident survive to deposit. For this method to be efficient, low pressures ($\sim 10^{-4}$ torr) are essential to minimize gas scattering. Long-throw deposition still suffers from the same line-of-sight geometrical constraints that govern the deposition of evaporated atoms, e.g., asymmetric coverage of sidewalls (see Section 3.3.3). Since deposition rates fall off as the square of the distance, compromises must be made with respect to the target–substrate distance. In this regard adequate coverage of features with an aspect ratio of 1 to 1.5 is possible for a 25-cm throw distance.

A second method pioneered by Rossnagel employs *collimation* to mechanically filter out those sputtered atoms having undesirable, nonnormal trajectories. The collimators are made of honeycomb sheet-metal arrays with hexagonal holes oriented perpendicular to both the target and substrate plane, and located between them. Although the transmission of the normal atom flux is enhanced, it comes at the price of much-reduced deposition rates, clogging of the collimator apertures, and a buildup of deposit on the top of the trench that must be polished off.

The third method to enhance trench coverage capitalizes on the deposition of metal ions through *ionized physical vapor deposition* (I-PVD). This is quite distinct from sputtering where neutrals are deposited. Once ions are produced they are steered toward the substrate by magnetic field lines oriented normal to it. The key issue of creating a rich density of metal ions is not as difficult as it sounds, because the first ionization potential ($\sim$5–7 eV) for many metals is less than that for argon (15.7 eV). In early systems, metal atoms were thermally evaporated and ionized when injected into Ar-based RF and microwave plasmas. Once evaporation began, the Ar flow was shut off and the metal ions sustained the plasma at pressures less than 1 mtorr. Importantly, directional deposition and trench lining was demonstrated. Later versions of I-PVD have employed magnetron sputtering targets as the metal source, surrounded by coils providing RF power such that a common plasma is shared. Figure 5-12a clearly demonstrates the deposition by I-PVD methods of highly conformal Ti/TiN linings for deep narrow trenches. The damascene process to produce these Al–Cu conductors will be discussed in Section 5.4.2.

5.4 PLASMA ETCHING

5.4.1 Introduction to Film Patterning

Until now the only film dimension considered has been thickness, a variable controlled by the growth or deposition process. However, the

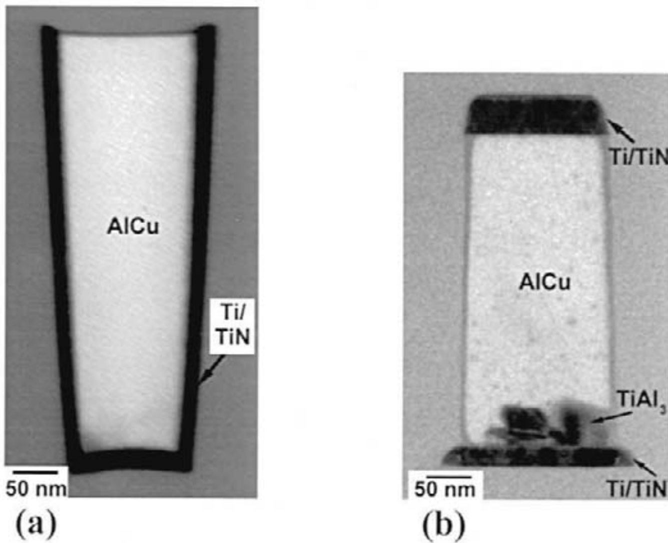


Figure 5-12 Al-Cu conducting interconnect channel produced by (a) hot sputter refill (damascene process); (b) metal reactive ion etch (standard process). In both cases a thin Ti/TiN barrier layer separates the AlCu from surrounding SiO_2. (Courtesy of L. Gignac, IBM T. J. Watson Research Center.)

fabrication of thin-film devices for all sorts of applications, not only microelectronics, generally requires that they be geometrically defined laterally or patterned in the film plane. The complexity of patterning processes depends on the nature of the film, the feature dimensions, and the spatial tolerances of the feature dimensions. For example, consider an evaporated metal film that must possess features 1 mm in size with a tolerance of ± 0.05 mm. The desired pattern could possibly be machined into a thin sheet-metal stencil or mechanical mask. Direct contact between this mask template and substrate ensures generation of the desired pattern in uncovered regions exposed to the evaporant flux. This method is obviously too crude to permit the patterning of features more than 1000 times smaller that are employed in integrated circuits.

Such demanding applications require lithographic techniques and the use of chemically reactive plasma etching, providing the rationale for inclusion of this subject matter here. Suppose it is desired to pattern an SiO_2 film that blankets a substrate. The lithographic process shown schematically in Fig. 5-13 consists of the following steps (Ref. 29):

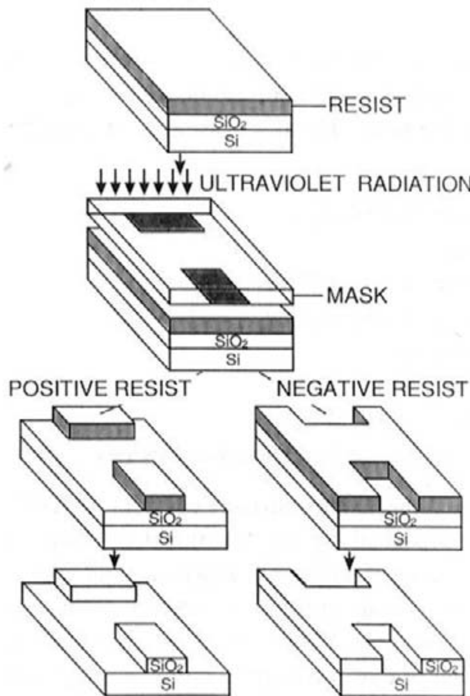


Figure 5-13 Schematic of the lithographic process for pattern transfer from mask to film. Both positive and negative resist behavior is illustrated.

1. *Generation of the mask.* The mask is essentially analagous to the negative in photography. It possesses the desired film geometry patterned in Cr or FeO thin films predeposited on a very flat quartz plate. Masks for integrated circuit use are generated employing computer-driven electron beams to precisely define regions that are either optically opaque or transparent. Other steps to produce the patterned mask film parallel those involved in subsequent pattern transfer to the involved film.

2. *Printing.* Printing of this negative mask requires the physical transfer of the pattern to the blanket film surface. This is accomplished by first spin coating a thin photoresist layer (~ 1 μm thick) on the film/substrate. As the name implies, photoresists are both sensitive to photons and resistant to chemical attack after exposure and development. Photoresists are complex photosensitive organic-compound mixtures, usually consisting of a resin, photosensitizer, and solvent. During exposure, light (usually UV) passes

through the mask and is imaged on the resist surface by appropriate exposure tools or printers. Either full-scale or reduced latent images can be produced in the photoresist layer. There are two types of photoresists whose behaviors are schematically distinguished in Fig. 5-13. The positive photoresist faithfully reproduces the (opaque) mask film pattern; in this case light exposure causes *scission* of polymerized chains, rendering the resist soluble in the developer. Alternatively, negative resists reproduce the transparent portion of the mask pattern because photoinduced *polymerization* leaves a chemically inert resist layer behind. For yet greater feature resolution, electron-beam and X-ray lithography techniques are required.

3. *Etching.* After exposure and development of the resist, it now becomes the mask that defines where the underlying film is to be removed. Liquid etchants attack and dissolve away the exposed film, leaving intact the film protected by resist. Equal rates of lateral and vertical material removal (isotropic etching) however, lead to loss of pattern definition due to undercutting and rounding of film features. This limits the resolution of wet etching to $\sim 3\ \mu$m feature sizes in $\sim 1\ \mu$m thick films. For submicron feature delineation in the patterning of ICs, dry or plasma etching is indispensable. In processes that resemble the inverse of film deposition, plasmas are employed to chemically erode or etch away the film surface.

4. *Resist removal.* The final step of resist removal can be carried out in either a liquid- or plasma-based stripping operation and leaves behind a high-fidelity thin-film copy of the mask pattern. As in etching, the choice of resist removal process depends on the application.

What has just been described is known as *subtractive* pattern transfer. If a blanket metal film is involved it can be patterned subtractively as described, or by *additive* pattern transfer to create the same final film geometry. In this case, however, a resist mask layer is applied to the substrate first and lithographically patterned. Then the film is deposited covering both the exposed substrate regions as well as the resist mask. Lastly, the resist is swelled and dissolved by a solvent, lifting off with it the unwanted film from the masked area of the pattern. Left behind is the desired patterned film.

5.4.2 Patterning and Fabricating a Metallization Level

As an example of patterning integrated-circuit features, let us return to the multilevel structure of Fig. 5-1. Each level consists of metal interconnections, trenches, and vias composed either of Al-base alloys or, more recently,

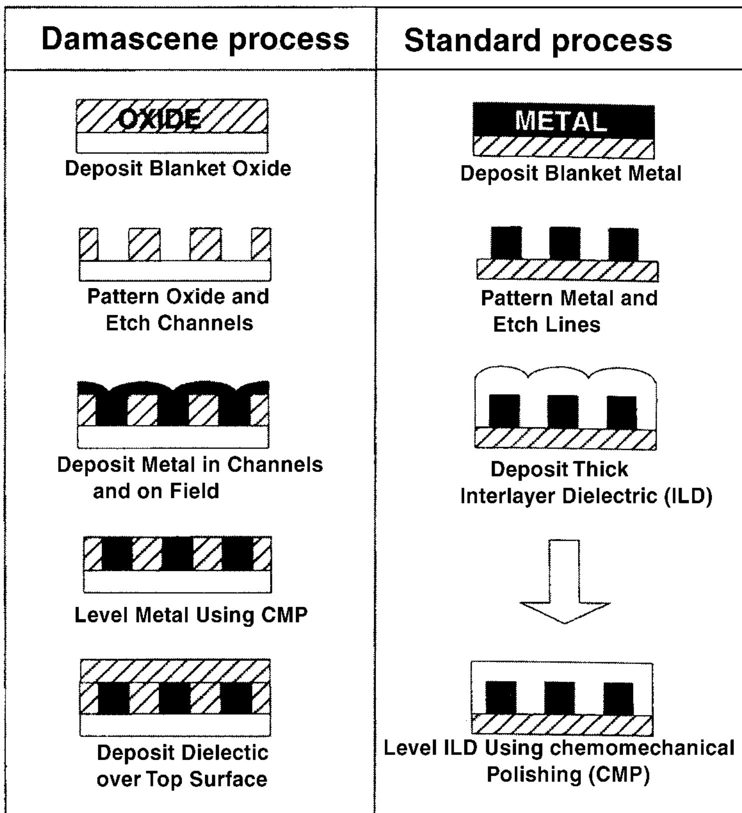


Figure 5-14 Comparison of damascene (left) and standard (right) metal interconnect wiring processes. (From Ref. 7.)

of Cu that must be completely surrounded by dielectric films, the intermetal filling, e.g., SiO_2. How is this structure, schematically shown at the bottom of Fig. 5-14, created? In particular, how is the alternating, narrow side-by-side metal–dielectric film array constructed, and how is the substrate surface planarized? Actually two routes have been used. In the standard process a blanket metal film is deposited and then patterned by subtractive plasma (reactive ion) etching. A dielecric layer is deposited by CVD methods and then actually mechanically polished until it is flat (e.g., Fig. 5-12b). The second technique, shown on the left, starts with a planar dielectric surface. Using conventional lithographic and reactive ion etching methods, the trenches are opened down to some buried contact pad, interconnection, or via. Metal liners (Ti/TiN) and metal are then deposited, filling these etched

cavities (e.g., Fig. 5-12a). Finally, in a chemical–mechanical polishing (CMP) technique the excess metal is removed in what is known as a *damascene* process. The state-of-the-art plasma deposition of the intermetal dielectric is further addressed in Section 6.7.3. Critical to successfully defining features is plasma etching, a subject we now address.

5.4.3 Plasma Etching Mechanisms

The etching or removal of atoms from film or substrate surfaces that are immersed in plasmas occurs by both physical and chemical means. Flamm (Refs. 30, 31) has claimed that all known plasma etching processes can be grouped into five categories (Fig. 5-15): sputter etching, chemical etching or gasification, accelerated ion-assisted etching, sidewall-protected ion-enhanced etching, and reactive-ion etching.

1. *Sputter etching* needs little introduction because this is what happens at cathodes during sputtering in low-pressure plasmas. Because their energies are high, ions striking a surface are sensitive to the magnitude of atomic bonding forces and structure, rather than the composition of different species which tend to sputter at similar rates. For this reason sputtering lacks discrimination or selectivity in the removal of surface atoms.

2. In *chemical etching* the gas-phase species reacts chemically with surface atoms producing a gasified or volatile product. The only purpose of the plasma is to make reactive etchant species. For example, the plasma feed gases F_2, NF_3, and CF_4 all generate F atoms, which do the etching. Chemical etching is the most selective of the mechanisms since it is sensitive to rather slight bonding-energy differences between atoms. However, etching occurs isotropically so that removal of atoms occurs at the same rate from both vertical and horizontal surfaces. It is therefore impossible to pattern very fine lines by this method.

3. Anisotropic etching due to *accelerated ion-assisted* processes removes such disadvantages by enhancing the vertical component of the impinging energetic ion flux. This in turn damages the surface by creating point defects, dislocations, and dangling bonds, all of which render the surface more reactive.

4. Etching via the related *inhibitor ion–enhanced chemistry* mechanism requires two conceptually different species, etchants and inhibitors. The latter form very thin protective films on vertical sidewalls and see very little ion bombardment. As a consequence the passivated sidewalls act as barriers, effectively enhancing anisotropic-etching processes.

5. The term *reactive-ion etching* has apparently been carelessly propagated in the literature to denote anisotropic etching. In low-density plasmas,

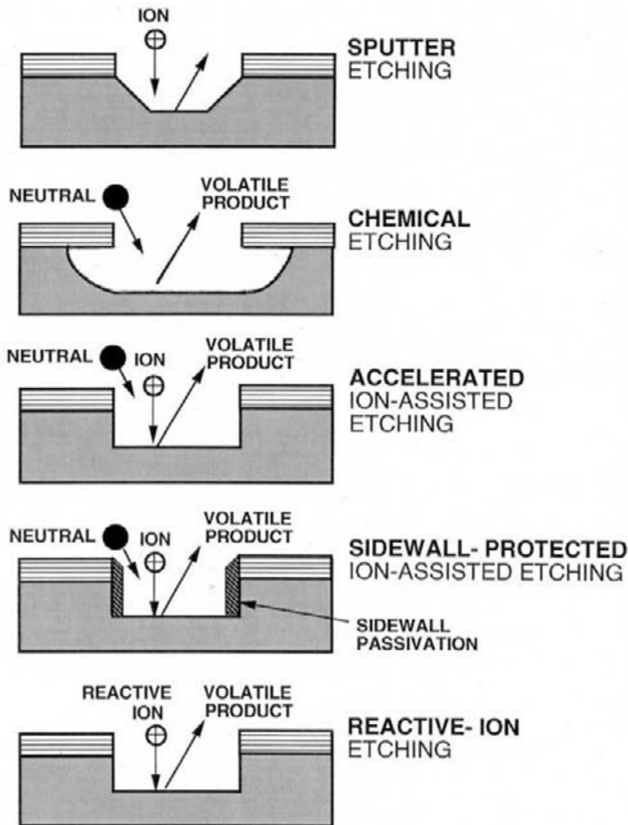


Figure 5-15 The five basic mechanisms of plasma etching. (From D. L. Flamm, *Plasma Processing of Semiconductors*, edited by P. F. Williams. Reprinted with the permission of Kluwer Academic Publishers.)

i.e., with current densities 0.01 to 1 mA/cm^2, there are simply too few impinging ions to achieve practical etch rates. However, in the more recent high-density plasma-etching systems (Section 5.4.5.4) bombarding ion fluxes, with current densities of 1 to 10 mA/cm^2, create a sufficient concentration of "hungry ions" to devour substrates. It is for such cases that the term reactive-ion etching is appropriate.

Changing the ion energies and pressures shifts the dominant material-removal processes. For example, physical sputter etching occurs at the lowest pressures ($\sim$1 mtorr) and highest energies (keV). Ion-assisted etching via the surface damage mechanism takes place at lower energies and somewhat higher pressures ($\sim$50 mtorr). In both cases surface etching tends

to be anisotropic. However, with chemical etching at elevated pressures of ~ 1 torr, energetic ion bombardment is precluded and the result is isotropic attack of films. Because the mass of many of the ionic species in practical plasma etching processes is large, their motion may not be in phase with the RF field. As a result, the ionic-displacement amplitude and energy are generally too low to cause sputtering.

Sometimes unique etching mechanisms can be capitalized upon to generate structures at the limit of etching resolution. As examples of historical interest, consider two fine-line metal nanostructures (Fig. 5-16) fabricated by contamination resist lithography. Here a finely focused electron beam writes a pattern in the form of a carbonaceous residue on blanket metal films supported by thin Si_3N_4 membranes. Subsequent ion-beam etching removes

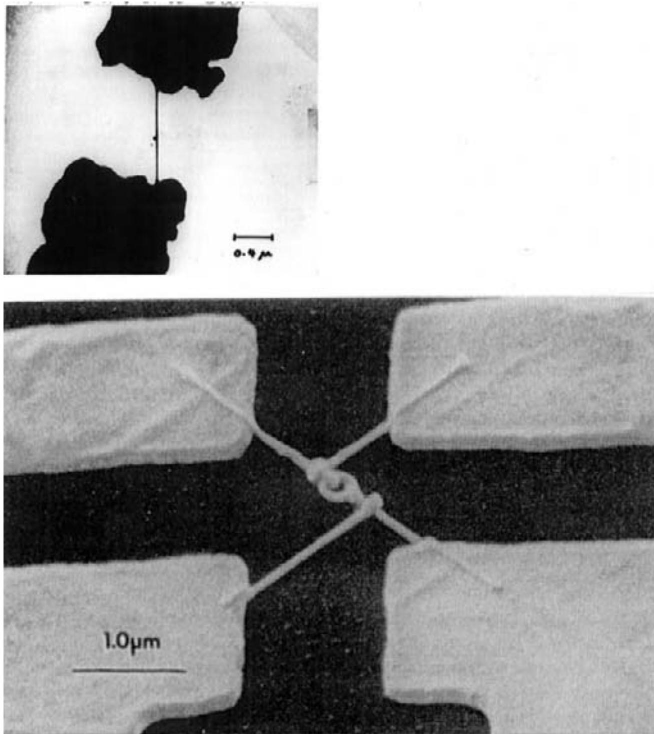


Figure 5-16 Scanning electron micrographs illustrating contamination resist lithography. (a) One of the earliest (1967) nanostructures (a very narrow Al line) ever fabricated. (Courtesy of R. B. Laibowitz, IBM T. J. Watson Research Center.) (b) Au superconducting device used to measure quantum magnetic effects. (From Ref. 32.)

metal everywhere except under the carbon contaminant layer. Metal lines less than 100 Å wide have been fabricated this way (Ref. 32).

Irrespective of the particular plasma-etching process, seven primary sequential steps appear to take place (Ref. 33). They are listed next, together with typical chemical reactions illustrative of the etching of silicon by chlorine.

1. Generation of the reactive chemical species (ions, radicals) in the flowing plasma gas (e.g., $e^- + Cl_2 \rightarrow Cl_2^+ + 2e^-$, $e^- + Cl_2 \rightarrow 2Cl + e^-$, and/or $e^- + Cl \rightarrow Cl^+ + 2e^-$)
2. Diffusion or drift of the reactive species through the plasma to the film surface
3. Adsorption (chemsorption) of these species on the film surface, e.g., Cl_2^+, Cl, or Cl^+ + Si (substrate) → Cl (adsorbed)
4. Surface diffusion of active ions and radicals
5. Chemical etching reaction with film atoms producing adsorbed by-products, e.g., $Si + nCl \rightarrow SiCl_n$ (adsorbed); $n \sim 4$
6. Desorption of the volatile reaction products, e.g., $SiCl_n$ (adsorbed) → $SiCl_n$ (desorbed gas)
7. Transport of the etching by-products to the plasma where they are removed by pumping

Simply listing each step gives no indication of its microscopic subtlety or complexity. For example, Cl atoms etch heavily *n*-doped Si some $\sim$20 times faster than undoped Si. The higher Fermi level in *n*-doped Si promotes charge transfer and a more ionic Cl–Si bond that enables Cl to access additional reactive sites. But in undoped Si, steric hindrance of adsorbed Cl atoms prevents etchant from reaching the surface, halting any further reaction.

5.4.4 Plasma Etching Parameters

The fact that there are a great number of experimental processing parameters (e.g., gas composition, reactor geometry, gas flow rates, power, frequency, temperature) that influence etching characteristics or parameters (e.g., etching rates, selectivity, anisotropy and loading effects) means that causal relationships between and among them will generally be difficult to draw. Despite this, much progress has been made in quantitatively simulating the plasma chemistry of film etching as well as deposition, a subject that is briefly reviewed in Section 6.7.4. Here we limit the discussion to defining and commenting upon some of the more important etching characteristics.

5.4.4.1 Etch Rate

The etch rate (R_e in units of Å/s, μm/min, etc.) is the basic measurable quantity that characterizes etching. First and foremost, as suggested by Table 5-1, etching processes are primarily dependent on plasma chemistry. In particular, halogen-containing gases are effective in etching many of the important film materials used in semiconductor technology. Since they are influenced by many processing variables, specific etch-rate values are not listed; however, R_e typically ranges from 0.01 to 1 μm/min. Temperature, a primary accelerator of chemical reactions, also strongly influences etching rates. A not unexpected Arrhenius-like behavior characterizes R_e in many situations. For example, the etch rates for Si and SiO_2 by fluorine of concentration C_F (atoms/cm^3) are given by (Ref. 34)

$$R_e(\text{Si}\langle 100\rangle) = 2.91 \times 10^{-12} T^{1/2} C_F \exp - \frac{0.108\,\text{eV}}{k_B T}\ \text{Å/min} \qquad (5\text{-}9a)$$

Table 5-1
Etching of Electronic Materials with Chlorine and Fluorine Atoms[a]

Materials etched	Source gas	Additive	Mechanism	Selective over
Si	Cl_2	None	Chemical	SiO_2
	CCl_4	O_2	Ion energetic	
	CF_4	O_2		
	$SiCl_4$	O_2		
	SF_6	O_2	Chemical	Resist
III–V semiconductors (GaAs base)	Cl_2, CCl_4, $SiCl_4$	None, O_2, O_2	Chemical/ crystallographic	SiO_2 resists
GaN[b]	Cl_2, H_2, Ar			
Al	Cl_2	$SiCl_4$, CCl_4, BCl_3	Ion inhibitor	SiO_2, Si_3N_4, some resists
{$TiSi_2$, $TaSi_2$, $MoSi_2$, WSi_2}	F_2, CF_4	None, O_2	Ion energetic	SiO_2, Si_3N_4, resists
{Ti, Ta, Mo, W, Nb}	F_2, CF_4	None, O_2	Ion energetic	SiO_2, resists
SiO_2/Si_3N_4	CF_4	O_2	Chemical	Resist

[a] Primarily from D. L. Flamm, in *Plasma Etching: An Introduction*, D. M. Manos and D. L. Flamm, eds. Academic Press, Boston, 1989.
[b] From O. Ambacher, *J. Phys. D: Appl. Phys.* **31**, 2653 (1998).

and

$$R_e(\mathrm{SiO_2}) = 6.14 \times 10^{-13} T^{1/2} C_F \exp - \frac{0.163 \text{ eV}}{k_B T} \text{ Å/min}, \qquad (5\text{-}9b)$$

where T is the substrate temperature.

5.4.4.2 Selectivity

The ratio of the etch rates between different materials (A and B) exposed to the same plasma is defined as the selectivity (S_e), i.e.,

$$S_e = R_e(\mathrm{A})/R_e(\mathrm{B}). \qquad (5\text{-}10)$$

Two kinds of selectivity can be distinguished, i.e., that between the substrate and pattern mask (for mask durability) and that between two material layers. Relative dry etch rates rarely differ by more than a factor of ~50 and are often much less. Therefore, it is important to optimize the value of S_e, and in this regard there are a few strategies. For example, materials that form volatile compounds etch much faster than those that form involatile products. Thus photoresists, which are largely composed of carbon and hydrogen, are etched or "ashed" in O_2 plasmas to form CO, CO_2, and H_2O. Selectivity may also be enhanced through a comparison of the thermodynamic stability of compounds. In the case of Al films on SiO_2, for example, exposure to a Cl_2 plasma results in the more rapid etching of Al.

Simply changing the plasma gas composition can significantly alter etching selectivity. For example, the SiO_2 etch rate exceeds that of poly-Si by only 25% in a pure CF_4 plasma. In an equimolar mixture of $H_2 + CF_4$, however, the etch rate of poly-Si drops almost to zero; the selectivity or ratio of etch rate of SiO_2 relative to poly-Si exceeds 45 (Ref. 35).

5.4.4.3 Anisotropy

The degree of etching anisotropy (A) may be defined as

$$A = 1 - [R_e(\mathrm{H})/R_e(\mathrm{V})] \qquad (5\text{-}11)$$

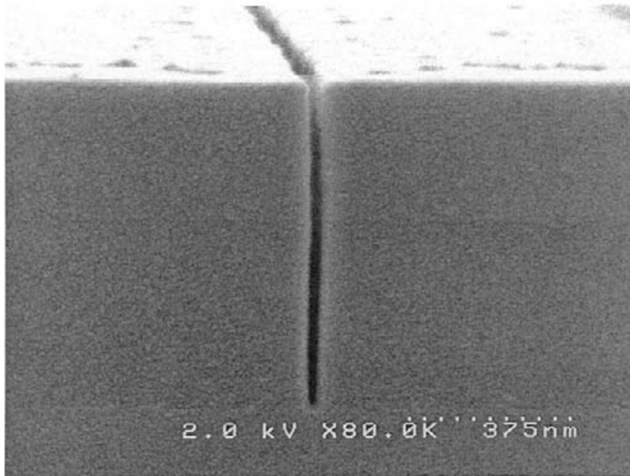


Figure 5-17 Scanning electron micrograph cross section of a 25-nm-wide 788 nm deep trench (aspect ratio = 31:1) etched into SiO_2 (TEOS). (From A. E. Braun, *Semiconductor International* **24**(2), 89, 2001. Courtesy of H. Huang, Trikon.)

where $R_e(\mathrm{H})$ and $R_e(\mathrm{V})$ are the horizontal and vertical etch rates. In isotropic etching $R_e(\mathrm{H}) = R_e(\mathrm{V})$ and $A = 0$, whereas for anisotropic etching A may actually closely approach unity. We have already dealt with etching anisotropy and noted some broad process parameters that affect it. In addition to ion bombardment of the surface, anisotropic etching is promoted by low substrate temperatures, favorable chemistry, and sidewall passivation. Crystallography of the surface being etched also affects the magnitude of A. Steep sidewall topography shown in the 31:1 aspect-ratio SiO_2 trench of Fig. 5-17 is the result of anisotropic material removal. For a Si/SiO_2 structure the ratio $R_e(100)/R_e(111)$ defines the etch rate anisotropy of Si while the ratio $R_e[\mathrm{Si}(100)]/R_e[\mathrm{SiO_2}]$ represents the selectivity.

5.4.4.4 Loading

The decrease in etch rate that is often observed as the area being etched increases is known as loading. It occurs because etch rates are gas concentration dependent, and larger areas simply cause greater etchant depletion. We may expect loading to occur when the active species etches rapidly, but has a long lifetime in the absence of etchable material.

5.4.5 Plasma-Etching Reactors

In the plasma etching of wafers a number of contrasting descriptor pairs have been used to classify reactors, such as batch vs single wafer, plasma vs ion etcher, and downstream vs plasma immersion. Among the common issues of concern are placement of wafers relative to the plasma, system pressure, electrode geometry, and generator frequency. Before describing the more common etching reactors in the context of these characteristics, it is instructive to first consider the operating frequency of these systems.

5.4.5.1 Operating Frequency

Suppose we wish to design a plasma etching reactor (Ref. 16). How do we account for reactor geometry, and what should the operating frequency of the discharge be? Too low a frequency may mean pulsating DC operation and plasma contamination due to electrode sputtering within the reactor. On the other hand, too high a frequency may result in inefficient power transfer because field reversals will occur more rapidly than plasma species can react. When dealing with AC etching discharges there are a broad range of characteristic frequencies and time constants to contend with. For example, there are the electron plasma (ω_e) and electron–neutral collision frequencies mentioned in Chapter 4. Other time-related quantities are the ion plasma frequency (ω_i) and plasma decay time (τ_p). For simplicity let us consider a high area-to-volume (e.g., pancake) reactor in which the ion density is n_i. In such a reactor the decay time is determined mainly by wall recombination of ions and electrons. If the radius of the cylindrical reactor is r and the spacing between electrodes is L, then an estimate of τ_p is obtained by dividing the charge in the system by the total current lost to the two electrodes and the cylindrical circumferential area, or

$$\tau_p = \frac{qn_i\pi r^2 L}{j_i(2\pi r^2 + 2\pi r L)}. \tag{5-12}$$

In this expression j_i, the ion current density at the plasma boundary, is taken as $j_i = qn_i v_i/4$ (Section 4.3.3), where $v_i = (k_B T_e/m_i)^{1/2}$, assuming the Bohm criterion (Ref. 6). The latter suggests that while electrons are repelled by a floating negative object, ions are accelerated toward it with an energy of at least $\frac{1}{2}k_B T_e$. If $r \gg L$,

$$\tau_p = \frac{qn_i L}{2j_i} \simeq L\left(\frac{m_i}{k_B T_e}\right)^{1/2}. \tag{5-13}$$

Typically, for $L = 5$ cm and $k_B T_e = 2$ eV, $\tau_p \simeq 20\,\mu$s, which is the half-cycle

time of a 25 kHz wave. Actual decay times for parallel-plate AC discharges are of the order of $\sim 100\,\mu$s, and the difference in these values is due to nonuniform plasma densities and currents.

The decay time essentially sets a lower limit to reactor operating frequencies. For electrons and argon ions, each present at a density of 10^{10} per cm^3, the electron and ion plasma frequencies (Eq. 4-18) are calculated to be $\omega_e = 0.9$ GHz and $\omega_i = 3.3$ MHz, respectively. As a consequence, three frequency bands are typically suggested for plasma etching as well as deposition processing. Whereas the low-frequency band ranges from 50 to 500 kHz, a high-frequency (radio) band spans the range from 10 to 50 MHz, and a still higher 1 to 50-GHz frequency band defines operation in the microwave regime. In each of these frequency bands we can be assured that reactions will be completed well before the plasma decays. Like the frequency of 13.56 MHz, 2.45 GHz has been reserved for industrial plasma-processing systems.

5.4.5.2 Barrel and Single Wafer Etchers

The first widely employed batch reactor for etching wafers was the barrel etcher. External electrodes are connected either capacitively or inductively to supply RF power to the plasma (Fig. 4-5). The wafers are stacked within the notches of quartz boats in no preferential orientation with respect to the electrodes. Etch rate and temperature uniformity are difficult to achieve in these reactors. However, they are cheap, have high capacity, and efficiently carry out noncritical operations such as photoresist stripping, surface cleaning, and isotropic etching.

In contrast to batch processing of wafers, an alternative approach is to etch them sequentially one at a time within cluster tools. To attain comparable wafer throughputs in such single-wafer etchers, higher etch rates are necessary. The more intense plasma used generates greater ion bombardment, sputtering, heating, and surface damage, together with less etching selectivity.

5.4.5.3 RF Planar and Hexode Reactors

The first of these etching reactors, due to Reinberg (Ref. 36), employs parallel-facing circular electrode plates and were initially used for chemical-vapor deposition of thin films. Gas introduced into the reactor flows radially over electrodes that are typically asymmetrically sized. Wafers placed on the smaller plate where the higher sheath potential develops (see Eq. 5-4) are exposed to enhanced ion bombardment and improved etching directionality. In the higher-pressure AC *plasma-etching* mode the wafer stage is

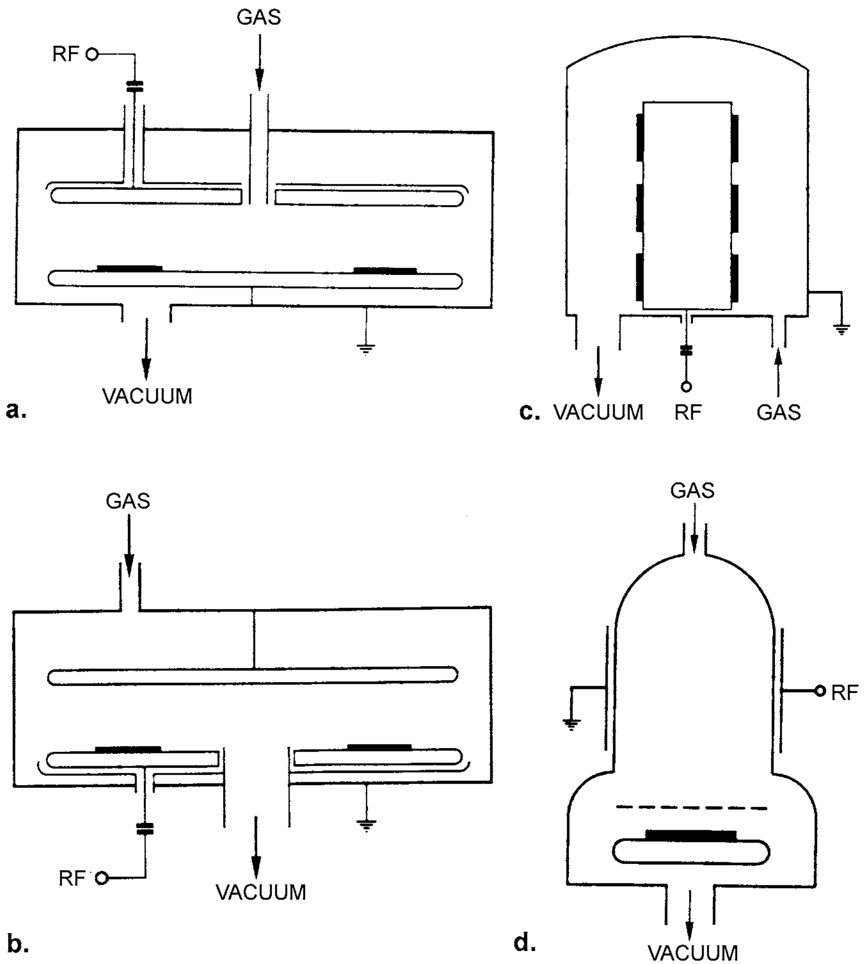


Figure 5-18 (a) Anode-coupled plasma etching reactor. Typical conditions: pressure = 30 Pa, power = 5 kW/m^2. (b) Cathode-coupled reactive-ion etching reactor. Typical conditions: pressure = 5 Pa, power = 5 kW/m^2. (c) Hexode plasma etcher. Typical conditions: pressure = 10 Pa, power = 5 kW/m^2. (d) Inductively-coupled remote RF plasma etching reactor. (From A. J. van Roosmalen, J. A. G. Baggerman and S. J. H. Brader, *Dry Etching for VLSI*. Reprinted with the permission of Plenum Publishing Corporation.)

grounded and the (anode-coupled) upper electrode is driven by RF power via an impedance matching network (Fig. 5-18a). In contrast the *reactive-ion etching* (RIE) mode operates at lower pressure with RF power supplied to the (cathode-coupled) wafer-loaded electrode (Fig. 5-18b). Reactive-ion etching combines two etching components, one due to isotropically reacting

chemical species (for selectivity) and the directed flux due to energetic (100–1000 eV) ions (for anisotropy). In both configurations plasma etching is primarily accomplished by neutral species and not ions.

A popular variant of the planar or diode reactor that enables a greater wafer throughput is the hexode plasma etcher shown schematically in Fig. 5-18c. Here an RF-powered hexagonal tower on which wafer substrates are mounted faces a grounded chamber-wall electrode in a coaxial configuration. Both the parallel plate and hexode reactors have been among the most widely used film-etching tools. Despite this fact, these capacitively coupled RF reactors have certain disadvantages, particularly in meeting the more stringent etching demands of shrinking microelectronic devices. First, substrates are immersed in a relatively harsh plasma/chemical environment and generally subjected to energetic ion bombardment. Because this may cause degradation by charging dielectrics or by heating of temperature-sensitive substrates, it is desirable to place them downstream in a charge-free region outside the discharge. Secondly, since gas scattering at higher pressures prevents etchant gases from penetrating the depths of trenches, low-pressure operation is essential. Finally, the ion flux and ion acceleration energy cannot be independently varied (Ref. 37). In summary, undesirable wafer damage, loss of linewidth dimensions, and generally low etch rates are some of the drawbacks of these reactors.

5.4.5.4 Microwave and High-Density Plasma Reactors

To address the shortcomings of the foregoing RF etching reactors, there has been much recent activity in the development and implementation of downstream and high-density plasma reactors. In downstream reactors (Fig. 5-18d) the etching gas is first fed into the discharge zone where neutral etchant radicals are formed. Because recombination of the charged species is generally more rapid than neutral gas-phase reactions, the longer-lived radicals then flow downstream to etch the substrates. In a popular version of these reactors, microwave power is employed. Similar, so-called ECR (electron cyclotron resonance) reactors are also used to deposit films, and their operation will be further discussed in Section 6.7.2. In contrast, other high-density plasma sources (helicon, helicon resonator), developed over the past decade or so, employ RF power that is *inductively* coupled, rather than capacitively coupled as in parallel-plate reactors. This means coupling to the plasma through a dielectric window in the former in contrast to electrodes in the latter. Magnetic fields are also applied in various ways to help confine the plasma. Like the ECR reactor, the plasma generation region is decoupled or downstream from the films being deposited (or etched). Further information on the design, operation, and performance of these inductively

powered high-density plasma reactors and their variants can be found in Refs. 37 and 38. These inductively coupled reactors typically operate at a 20-fold reduced pressure, a 2-fold higher power level, a 10-fold higher plasma density, and a 100-fold increase in ionization fraction relative to the capacitively coupled plate reactors; etch rates are thus correspondingly higher.

Typical electrode configurations, plasma excitation frequencies and charge characteristics, etching mechanisms, and operating conditions for all of the plasma etching reactors considered above have been summarized by Grill (Ref. 6).

5.4.6 TEMPERATURE RISE OF FILMS DURING PLASMA PROCESSING

One of the important common issues in both plasma etching and sputtering is the temperature rise in the film/substrate (Ref. 39). Plasma-etching species and sputtered atoms that impinge on surfaces are far more energetic, for example, than comparable atoms emanating from evaporant sources. During ionic impact, condensation, and reaction, the excess energy liberated must be dissipated by the substrate or otherwise it may heat excessively to the detriment of film quality. To address the question of substrate heating, we start with an equation for the heat power balance, namely,

$$\rho c d(dT/dt) = \bar{\mathscr{P}} - \mathscr{L}. \tag{5-14}$$

The term on the left is the net thermal energy per unit area per unit time (in typical units of W/cm^2) retained by a substrate whose density, heat capacity, effective thickness, and rate of temperature rise are given by ρ, c, d, and dT/dt, respectively.

The incident power flux $\bar{\mathscr{P}}$ has three important components:

1. Heat of condensation of atoms, ΔH_c (eV/atom)
2. Average kinetic energy of incident adatoms $\bar{E}_k$ (eV/atom)
3. Plasma heating from bombarding neutrals and electrons with energy E_p (eV/atom)

Table 5-2 contains values for these three energies during magnetron sputtering at 1 keV (Ref. 40). For a deposition rate $\dot{d}$ (Å/min),

$$\bar{\mathscr{P}} = \frac{2.67 \times 10^{-29}\dot{d}(\Delta H_c + \bar{E}_k + E_p)}{\Omega} \text{ W/cm}^2 \tag{5-15}$$

Table 5-2

Energies Associated with Magnetron Sputtering[a]

Metal	Heat of condensation (eV/atom)	Kinetic energy of sputtered atoms (eV/atom)	Plasma (eV/atom)	Estimated flux (eV/atom)	Measured flux (eV/atom)
Al	3.33	6	4	13	13
Ti	4.86	8	9	22	20
V	5.29	7	8	20	19
Cr	4.11	8	4	16	20
Fe	2.26	9	4	15	21
Ni	4.45	11	4	19	15
Cu	3.50	6	2	12	17
Zr	6.34	13	7	26	41
Nb	6.50	13	8	28	28
Mo	6.88	13	6	26	47
Rh	5.60	13	4	23	43
Cd	1.16	4	1	6	8
In	2.52	4	2	9	20
Hf	6.33	20	7	33	63
Ta	8.10	21	9	38	68
W	8.80	22	9	40	73
Au	3.92	13	2	19	23

[a] From Ref. 40.

where Ω is the film atomic volume in cm^3/atom. The remaining term, $\mathscr{L}$, represents the heat loss to the substrate holder by conduction or to cooler surfaces in the chamber by radiation. For the moment, let us neglect $\mathscr{L}$ and calculate the temperature rise of a thermally isolated substrate. Substituting Eq. 5-15 into Eq. 5-14 and integrating, we obtain

$$T(t) = \frac{2.67 \times 10^{-29} \dot{d}(\Delta H_c + \bar{E}_k + E_p)t}{\rho c d \Omega}. \tag{5-16}$$

Consider Al deposited at a rate of 10,000 Å/min on a Si wafer 0.050 cm thick. For Al, $(\Delta H_c + \bar{E}_k + E_p) = 13$ eV/atom and $\Omega = 16 \times 10^{-24}$ cm^3/atom, while for Si, $\rho = 2.3$ g/cm^3 and $c = 0.7$ J/g-C. In depositing a film 1 μm thick, $t = 60$ s and the temperature rise of the substrate is calculated to be 162°C. Higher deposition rates and substrates of smaller thermal mass will result in proportionately higher temperatures.

The temperature will not reach values predicted by Eq. 5-16 because of $\mathscr{L}$. For simplicity we only consider heat loss by radiation. If the front and rear substrate surfaces radiate to identical temperature sinks at T_0 with equal emissivity ε, then $\mathscr{L} = 2\sigma\varepsilon(T^4 - T_0^4)$, where σ is the Stefan–Boltzmann constant, 5.67×10^{-12} W/cm^2-K^4. Substitution in Eq. 5-14 and direct integration, after separation of variables, yields

$$t = \frac{1}{2\alpha^{3/2}\beta^{1/2}}\left[\tan^{-1}\sqrt{\frac{\beta}{\alpha}}T - \tan^{-1}\sqrt{\frac{\beta}{\alpha}}T_0 + \frac{1}{2}\ln\left(\frac{\sqrt{\alpha}+\sqrt{\beta}\,T}{\sqrt{\alpha}+\sqrt{\beta}\,T_0}\cdot\frac{\sqrt{\alpha}-\sqrt{\beta}\,T_0}{\sqrt{\alpha}-\sqrt{\beta}\,T}\right)\right] \quad (5\text{-}17)$$

where $\alpha = \sqrt{(2\sigma\varepsilon T_0^4 + \bar{\mathscr{P}})/\rho c d}$, and $\beta = \sqrt{2\sigma\varepsilon/\rho c d}$.

Equation 5-17 expresses the time it takes for a substrate to reach temperature T starting from T_0, assuming radiant cooling. For short times, Eq. 5-16 holds, whereas for longer times the temperature equilibrates to a radiation-limited value dependent on the incident power flux and substrate emissivity. Sputter deposition for most materials at the relatively high rate of 1 μm/min requires a typical substrate power flux of $\sim$250 mW/cm^2. The predicted rate of film heating is shown in Fig. 5-19. If substrate bias is also

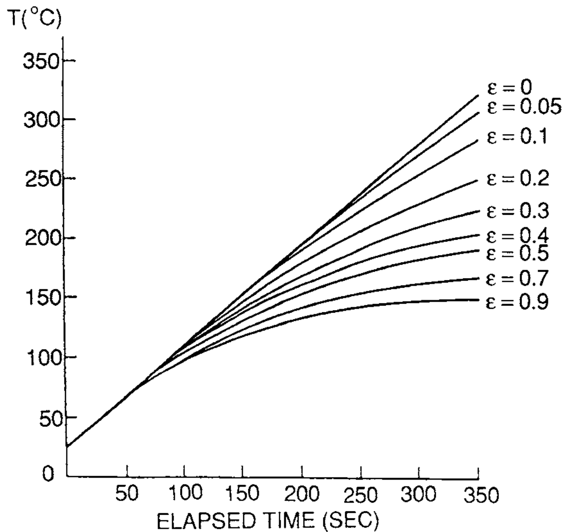


Figure 5-19 Temperature–time response for the film–substrate combination under the influence of a power flux of 250 mW/cm^2. Deposition rate is $\sim$1 μm/min. (Reprinted with permission from Cowan Publishing Co., from Ref. 39.)

applied, temperature increases can be quite substantial. In Al films, temperatures in excess of 200°C have been measured. This partially accounts for enhanced atom mobility and step coverage during ion bombardment. Heating effects are also important when sputtering metals onto heat-sensitive substrates. Various contributions to substrate heating have been measured and analyzed for both conventional as well as hollow cathode magnetron sputtering processes (Ref. 41). In the latter case the main contributions arise from recombination of charge carriers and convective heating by the hot process gas.

Finally, we note that at the same power level, plasma etching rates tend to be lower, by more than an order of magnitude, than film-deposition rates. This means that etching requires unusually high power levels, frequently in the range 1 to 2 W/cm^2. The combination of high power levels and long etching times causes substrates to reach high radiation-limited temperatures. In Al, for example, temperature increases well in excess of 300°C have been recorded during etching.

5.5 HYBRID AND MODIFIED PVD PROCESSES

This chapter concludes with a discussion of several PVD processes which are somewhat more complex than ones considered up to this point. As an example, laser-ablated carbon and magnetron-sputtered titanium have been combined to produce graded multilayer TiC coatings on ball-bearing races (Ref. 42). In the most general case, different admixtures of techniques are used to generate free atoms (e.g., evaporation, sputtering, PLD), alter their chemistry in the vapor phase (e.g., plasmas, reactive environments), and modify the way they deposit to form the film (e.g., ion beams, substrate bias). These composite processes demonstrate the diversity of hybridization and modification possible in producing films with unusual properties. And like the two or more components that form a materials composite, advantageous process features can be complementary and synergistic. Ion plating, reactive evaporation, and arc plasma deposition will be considered first. In these processes, which exist in many variant and overlapping versions, the material deposited often originates from a heated source that is immersed in a plasma atmosphere. Next ion-beam-assisted deposition and surface modification processing will be discussed. Here well-characterized ion beams bombard films deposited by either evaporation or sputtering. The chapter closes with a discussion of ionized-cluster beam deposition. This process is different from others considered in this chapter in that film formation occurs through condensation of collective groups of gas-phase atoms rather than individual atoms.

5.5.1 Ion Plating

Ion plating, developed by Mattox (Ref. 43), refers to evaporated film deposition processes in which the substrate is exposed to a flux of high energy ions capable of causing appreciable sputtering before and during film formation. The process has been reviewed by Ahmed (Ref. 44) and more recently by Mattox (Ref. 5). A schematic representation of a diode-type batch ion-plating system is shown in Fig. 5-20. Since it is a hybrid system, provision must be made to sustain the plasma, cause sputtering, and heat the vapor source. Prior to deposition, the substrate, negatively biased from 2 to 5 kV, is subjected to inert-gas ion bombardment at a pressure in the

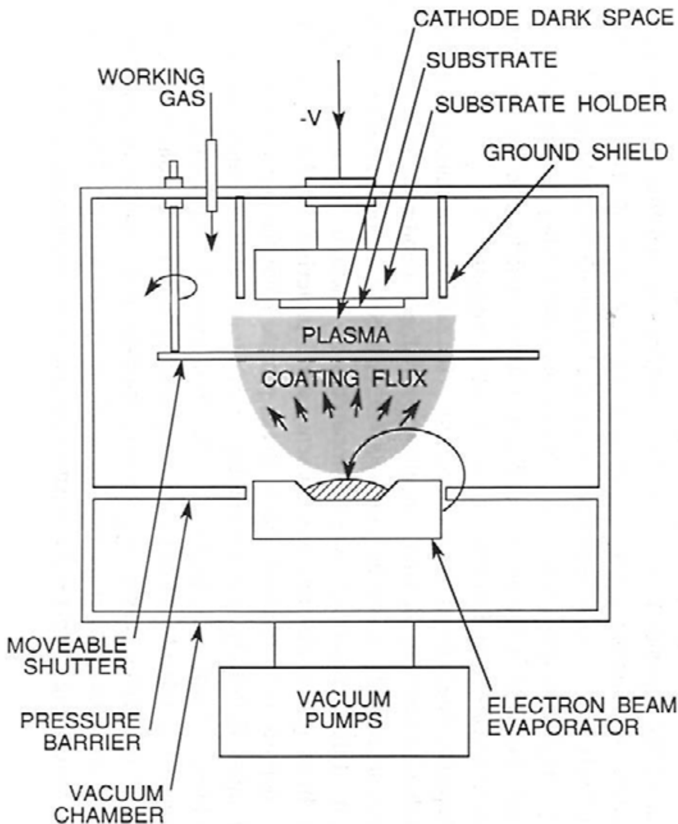


Figure 5-20 Ion plating system. (From Ref. 43.)

millitorr range for a time sufficient to sputter clean the surface and remove contaminants. Then the system pressure is reduced and source evaporation is begun without interrupting the sputtering. Clearly, the sputtering rate must be less than the deposition rate. Once the interface between film and substrate has formed, ion bombardment may or may not be continued. To circumvent the relatively high system pressures associated with glow discharges, high-vacuum ion-plating systems have also been constructed. They rely on directed ion beams targeted at the substrate. Such systems, which have been limited thus far to research applications, will be discussed further in Section 5.5.4.

Film deposition rates by ion plating (0.1–25 μm/min) fall between those for sputtering (0.01–2 μm/min) and evaporation (0.1–75 μm/min). Perhaps the chief advantage of ion plating is the ability to promote excellent adhesion between the film and substrate by the ion and particle bombardment mechanisms discussed in Section 4.6. In contrast to the relatively sharp film–substrate interfaces for evaporation and sputtering, ion plating results in graded and diffuse interfaces. A second important advantage is the high "throwing power," particularly when compared to vacuum evaporation. This results from gas scattering, entrainment and sputtering of the film, all of which enable deposition in recesses such as holes and on areas remote from the source–substrate line of sight. Coating of substrates having complicated shapes is thus achieved. Lastly, the quality of deposited films is frequently enhanced. The continual bombardment of the growing film by high-energy ions or neutral atoms and molecules serves to peen and compact it to near bulk densities. Sputtering of loosely adhering film material, increased surface diffusion, and reduced shadowing effects serve to suppress undesirable columnar growth.

A major use of ion plating has been to coat steel and other metals with very hard films to extend the life of cutting tools. Other important applications are for decorative, optical, corrosion- and wear-resistant, and high-temperature coatings. As an example of the last, considerable added value has been given to superalloy turbine blades by coating them with high-temperature refractory oxides (usually evaporated Y-stabilized ZrO_2). Between this outer coating and the turbine-blade substrate there is an intermediate ion-plated layer of MCrAlY ($M = Ni$, Co, Fe). For such purposes the required metals and alloys are electron-beam evaporated. To deposit compounds, metals such as Ti, Zr, and Si are similarly e-beam evaporated through an Ar plasma in the presence of reactive gases such as N_2, O_2, and CH_4. This variant of the process is known as reactive-ion plating (RIP) and coatings of nitrides (TiN, Si_3N_4), oxides (SiO_2, TiO_2), and carbides (TiC), have been deposited in this manner.

5.5.1.1 Coating Uniformity

One of the important issues in ion plating as well as other deposition processes carried out under soft-vacuum conditions (millitorr range) is coating uniformity. Since shaped objects are often coated, surfaces in the direct line of the source receive a greater evaporant or sputtered flux than surfaces that do not view the source. But these latter surfaces are also coated because of gas-phase scattering. Fancey and Matthews (Ref. 45) have obtained a quantitative measure of coating uniformity by comparing deposition on thin two-sided substrates. One surface (front side) facing the vapor source is located a distance L away, and the other surface (back side) faces away from the source. Importantly, an upward nonthermalized deposition flux impinges on the front side, while the downward flux on the back side is thermalized as a result of energy exchange during gas scattering (see Section 5.2.3). Based on collision-probability theory (Section 4.2.2), the nonthermalized flux can be represented by $\Psi \exp - (L/\lambda)$, where λ is the associated mean free path in the gas, and Ψ is a function of L and also accounts for dilution of the depositing flux with distance from the source. Therefore, the thermalized deposition flux is $\Psi[1 - \exp - (L/\lambda)]$.

Since the thermalized flux is independent of direction, the average component moving upward is $\frac{1}{2}\Psi[1 - \exp - (L/\lambda)]$. Therefore, the total vapor flux directed upward is $\Psi \exp - (L/\lambda) + \frac{1}{2}\Psi[1 - \exp - (L/\lambda)]$. Similarly, the average vapor flux moving downward is $\frac{1}{2}\Psi[1 - \exp - (L/\lambda)]$. Assuming that these are the only fluxes that influence the coating process, the ratio (R) of the front-to-back thickness is

$$R = \frac{1 + \exp - (L/\lambda)}{1 - \exp - (L/\lambda)} = \coth(L/2\lambda). \tag{5-18}$$

In Fig. 5-21 the coating process is visualized and the results of the model are plotted. The theory has been validated in RF ion plated and evaporated titanium films (Ref. 45) where predicted linear plots of L vs $- \ln\{(R - 1)/(R + 1)\}$ were obtained using measured values of L and R.

Although only thickness variations have been considered here, there is a related issue of possibly greater significance, namely the uniformity of film stoichiometry. Because the metal flux is concentrated near the source and the gas species is more uniformly dispersed, there is a greater tendency to produce metal-rich films on surfaces facing the melt or target than on back-side surfaces. Further away front- and back-side surfaces are more likely to have the same composition. Such effects have been observed during the reactive evaporation of TiN films (Ref. 46).

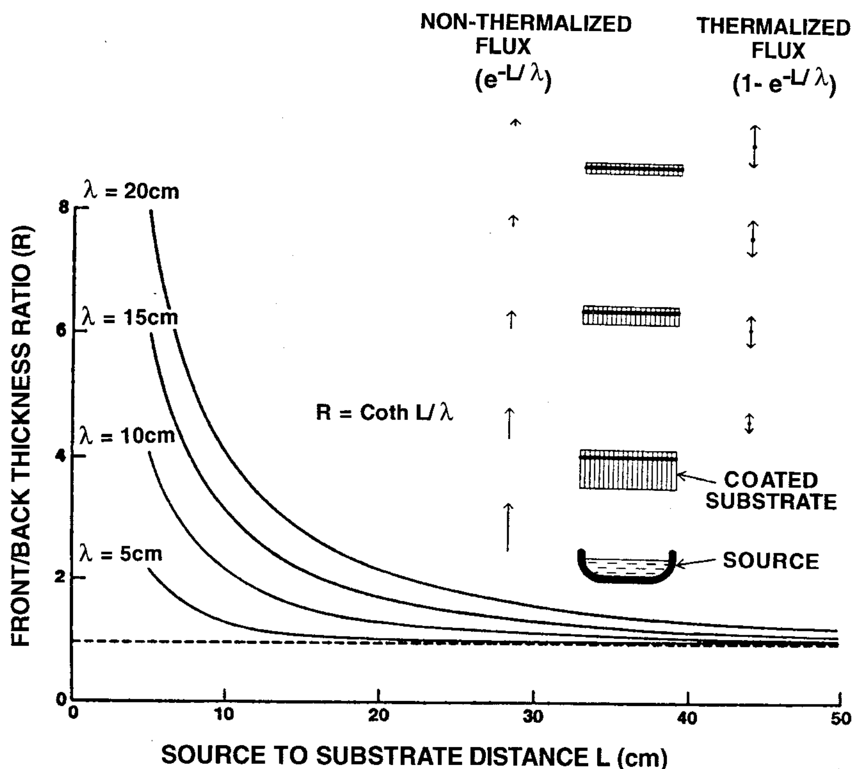


Figure 5-21 Plots of front/back film thickness ratio (R) vs source–substrate distance (L). Curves are for different values of λ. The coating thickness uniformity model is also physically depicted. (From K. S. Fancey and A. Matthews, *Advanced Surface Coatings*, edited by D. S. Rickerby and A. Matthews. Glasgow: Blackie/Chapman and Hall, 1991. Reprinted with the permission of the publishers and Dr. Alan Matthews.)

5.5.2 Reactive Evaporation Processes

In reactive evaporation the evaporant metal vapor flux passes through and reacts with gas introduced into the system (at $\sim 1–30 \times 10^{-3}$ torr) to produce compound deposits. The process has a history of evolution in which evaporation was first carried out without ionization of the reactive gas. In the activated reactive evaporation (ARE) processes developed by Bunshah and co-workers (Ref. 47), a plasma discharge is maintained directly within the reaction zone between the metal source and substrate. Both the metal vapor and reactive gases, such as O_2, N_2, CH_4, or C_2H_2, are therefore ionized, increasing their reactivity on the surface of the growing film or coating and promoting compound formation. One of the process configura-

tions is illustrated in Fig. 5-22, where the metal is melted by an electron beam. A thin plasma sheath develops on top of the molten pool. Low-energy secondary-electrons from this source are drawn upward into the reaction zone by a circular wire electrode placed above the melt biased to a positive DC potential (20–100 V), creating a plasma-filled region extending from the electron beam gun to near the substrate. The ARE process is endowed with considerable versatility since the substrates can be grounded, allowed to float electrically, or biased either positively or negatively. In the last variant ARE is quite similar to RIP. Other modifications of ARE include resistance-heated evaporant sources coupled with a low-voltage cathode (electron) emitter–anode assembly. Activation by DC and RF excitation has also been employed to sustain the plasma, and transverse magnetic fields have been applied to effectively extend lifetimes of plasma species.

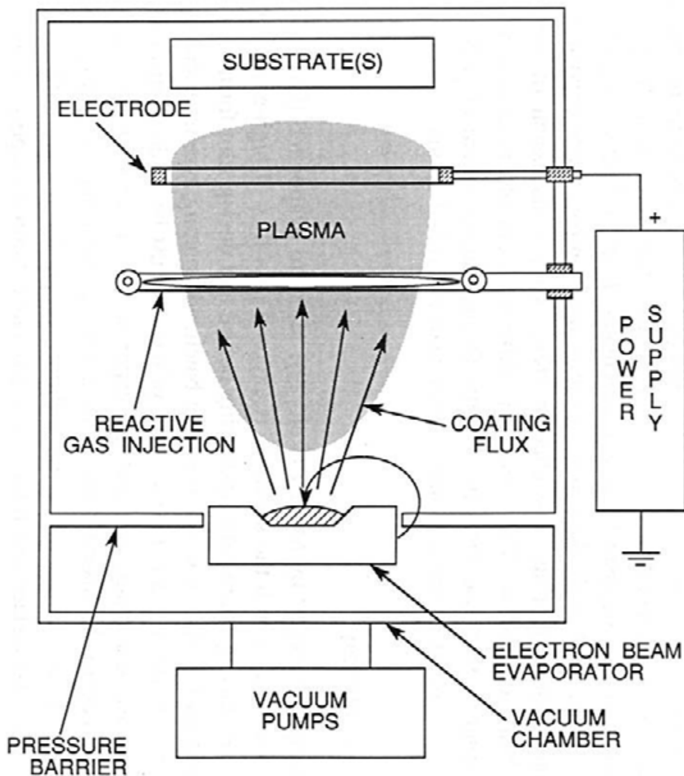


Figure 5-22 Activated reactive evaporation. (From Ref. 47.)

Before considering the variety of compounds produced by ARE, it is worthwhile to recall that thermodynamic and kinetic factors are involved in their formation. The high negative enthalpies of compound formation of oxides, nitrides, carbides, and borides indicate no thermodynamic obstacles to chemical reaction. The rate-controlling step in simple reactive evaporation is frequently the speed of the chemical reaction at the reaction interface. The actual physical location of the latter may be the substrate surface, the gas phase, the surface of the metal evaporant pool, or a combination of these. Plasma activation generally lowers the energy barrier for reaction by creating many excited chemical species. By eliminating impediments to reaction, deposition rates of a few thousand angstroms per minute are possible.

A partial list of compound films synthesized by ARE methods includes the oxides α-Al_2O_3, V_2O_3, TiO_2, and indium–tin oxide; the carbides TiC, ZrC, NbC, Ta_2C, W_2C, VC, and HfC; and the nitrides TiN, MoN, HfN, and cubic boron nitride. The extremely hard TiN, TiC, Al_2O_3, and HfN compounds have found extensive use as coatings for sintered-carbide cutting tools, high-speed drills, and gear cutters. In these applications ARE has been superceded by ion plating, reactive sputtering, arc plasma, and CVD methods. The fact that no volatile metal-bearing precursor is required as in CVD is an attractive advantage of ARE. Most significantly, these complex compound films are synthesized at relatively low temperatures, a unique feature of plasma-assisted deposition processes.

5.5.3 Arc Plasma Deposition

As noted in Section 4.2.3, arcs (Refs. 5, 48) are high-current (tens to hundreds of amps), low DC voltage (tens of volts) gas discharges. The discharge can occur in high vacuum but it should be realized that the term "vacuum arc" is a paradox, i.e., if a vacuum exists there can be no arc and vice versa. As in common welding, initiation of the vacuum arc requires two metal electrodes that first touch and then separate a small distance. A very luminous small cathode spot (10^{-8} to 10^{-4} m diameter) forms that passes extremely high current densities ($\sim 10^8$ to 10^{12} A/m^2). This causes erosion of the cathode by melting and vaporization as well as ejection of solid and molten particles. It is the vapor in the form of multivalent ions that sustains the vacuum arc. Furthermore, the spot rapidly jumps around the surface of the cathode in a manner dependent on the cathode composition, gas species and pressure, presence of magnetic fields, etc.; the spot velocity is typically ~ 100 m/s. During its motion it is also common for the arc to be extinguished. Meanwhile there is a much smaller current density at the anode

($\sim 10^5$ A/m^2) which, nevertheless, can cause it to melt and evaporate. The emitted ions are generally monovalent and the degree of ionization of the vaporized material is lower than in cathodic arcs.

Arc-deposition processes that have been developed capitalize on melting phenomena at either the anode or cathode, and interestingly, blend features of both plasma and evaporation environments.

5.5.3.1 Anodic Arcs

In anodic-arc sources the anode is molten and material that evaporates from it passes through the arc plasma to impinge on the substrate. Such arcs are categorized according to the source of the electrons, e.g., heated thermionic emitter, hollow cathode, or arc cathode. Through the use of magnetic fields the electron beam can be bent to prevent vapor impingement on the electron source. Except for immersion in a plasma, the evaporant geometry and crucible resemble those of the electron-beam gun evaporator (Fig. 3-13). Process advantages include safer, lower-voltage welding power supplies and beneficial substrate ion bombardment. Basically, anodic-arc deposition is a variant of ion plating.

5.5.3.2 Cathodic Arcs

When material vaporization primarily occurs by arc erosion of the cathode surface, we are dealing with cathodic-arc plasma deposition. A schematic of a cathodic-arc deposition system is shown in Fig. 5-23, where microscopic events at a cathode spot are also depicted. Particularly important in these systems is the arc ignition mechanism, usually a mechanical striker, and the means to confine the arc spot to the cathode surface. Arcs can be confined by boundary shields that extinguish the arc when it leaves the cathode, by magnetic fields, and by passive borders, e.g., rings made of boron nitride that encompass the cathode. To achieve uniform film deposition the arc is steered magnetically, which causes material to be eroded from the cathode in a series of flash evaporation events. Magnet configurations used are not unlike those employed in magnetron sputtering. From the standpoint of film deposition, the metal ions and neutrals emitted are normally the desired species, whereas microdroplets that manage to impinge on the substrate are a primary concern. They arise from ablation of molten or solid cathode particles due to thermal shock and hydrodynamic effects. Since they range in size from submicon dimensions up to several micrometers, their presence is as unwelcome here as it is in pulsed-laser deposition of films (Section 3.5.2.2). The majority of molten globules, also known as "macros," are fortunately emitted at low angles (0° to 30°) with respect to

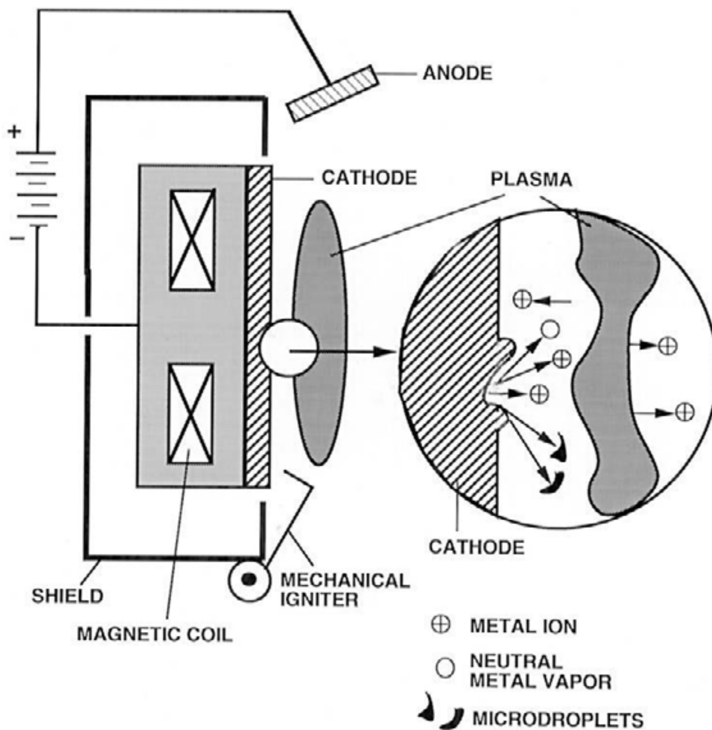


Figure 5-23 Cross section of cathodic-arc deposition system with a model of activity at a cathode spot. (From Ref. 48.)

the cathode surface. Therefore, low-angle shielding is effective in minimizing the number that reach the substrate. Other effective strategies for this purpose include reduction of arc current, adjustment of gas pressure in reactive processes, increasing the velocity of cathode-spot motion by magnetic methods, and employing curved-sector electromagnetic mass filters. In addition to low-voltage power supplies, there is the advantage of mounting cathodes in any orientation. Cathodic-arc deposition rates are greater than those for sputtering, but less than for e-gun evaporation.

5.5.3.3 Materials Deposited by Arc Methods

Cathodic arcs have been widely used in vacuum as well as reactive atmospheres to commercially deposit a large assortment of metal alloy and compound films (Refs. 5, 48). These include adherent coatings of refractory

metals (e.g., Zr, Mo, Ta), alloys (e.g., 304 stainless steel, Ti-6Al-4V, MCrAlY), nitrides (e.g., TiN, ZrN, CrN), ternary nitrides (e.g., TiAlN, TiZrN), carbides (e.g., TiC, WC), carbonitrides (e.g., TiC_xN_{1-x}, ZrC_xN_{1-x}), and oxides (e.g., TiO_2, ZrO_2). The major application of refractory nitrides, carbonitrides, and carbides has been to wear- and abrasion-resistant coatings in forming and cutting tools. Unlike CVD methods (Section 6.6.2.1), the coatings are deposited at lower temperatures which generally yield harder, finer grained deposits; furthermore, tool distortion is minimized. Interestingly, advantage is taken of the broad color range of several binary and ternary nitrides in decorative coatings, e.g., for watch bands and eyeglass frames. Additionally, amorphous and diamond-like carbon films have been deposited in cathodic arcs, and optical coatings have been produced using anodic arcs.

5.5.4 Ion-Beam-Assisted Deposition

5.5.4.1 Processes

In this section the focus is on processes that employ *directed* ion beams to modify growing films and surfaces. We have already noted on several occasions in the past two chapters that ion bombardment of films during sputter deposition is particularly effective in modifying film properties. However, process control in plasmas is somewhat haphazard because the direction, energy, and flux of the ions incident on the growing film cannot be easily regulated. Ion-beam-assisted processes were invented to provide independent control of the deposition parameters and, particularly, the characteristics of the ions bombarding the substrate. Importantly, the ion beams are not derived from elevated-pressure plasmas but rather stream from an ion source through a high-vacuum ambient.

Ion sources are generally employed in two distinct modes. The first uses an inert-gas ion beam to controllably sputter atoms from a target; here the function of the beam is to promote the supply of target atoms for deposition. In the second mode the beam, composed of either inert- (e.g., Ar^+) or reactive-gas ions (e.g., O_2^+), is directly aimed at the substrate; here the purpose of the beam is to modify the properties of the growing film by some combination of bombardment, reaction, or burial within it. A simple example of burial, presented earlier in Fig. 3-18, showed an ion gun used in conjunction with an evaporation source in a process known as ion-beam-assisted deposition (Ref. 49).

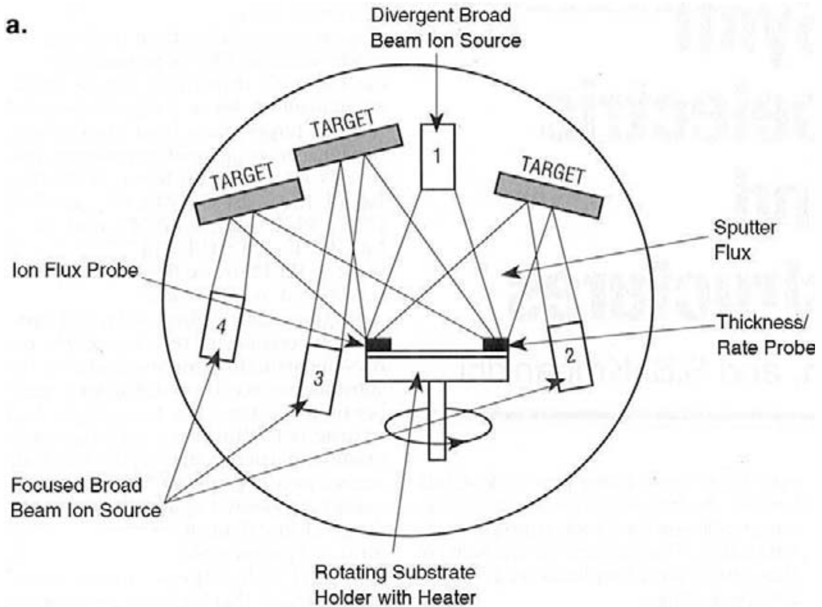
It is common, however, to incorporate both ion-beam functions in a single vacuum chamber. This necessitates at least two ion sources. For

example, in a simple dual ion-beam system, one ion source is used for sputtering while the second supplies ions to modify the depositing film. Separate film-thickness and ion-current monitors, fixed to the substrate holder, enable the incident beam fluxes to be independently controlled. When more stoichiometrically challenging films are involved, such as $PbZr_xTi_{1-x}O_3$ ferroelectrics (Ref. 50), four ion beams are used as shown in Fig. 5-24a. Focused inert-gas ion beams emerge from each of the three gun sources to sputter Pb, Zr, and Ti metal in parallel, while the fourth source floods the substrate with a divergent beam of oxygen ions to adjust the oxide chemistry. Alternatively, one focused inert-gas (Xe) ion beam has been used in conjunction with rotating targets to serially sputter the same metals (Fig. 5-24b). Anisotropic film properties can arise when the ion-beam angle of incidence is not normal to the substrate. Substrate rotation is, therefore, recommended if isotropy is desired.

It is not at all surprising that directed ion beams have also been used to etch films. After all, it has been the traditional way to thin specimens for examination in the transmission electron microscope. However, for other ion-beam-assisted etching operations, a chemically reactive gas is injected toward the substrate while an inert-gas ion beam simultaneously bombards the film. In this way chemical and physical etching effects are decoupled. Such systems, limited to low-pressure operation where etch rates are low, have yielded highly anisotropic features in GaAs films (Ref. 51). Relative to other dry-etching processes there are no plasma reactions.

5.5.4.2 Ion-Beam Sources

The heart of an ion-beam-assisted deposition system is the ion source. First used as ion thrusters for electric space propulsion, the three primary ion-source configurations are gridded, Hall-effect (gridless), and electron cyclotron resonance. Their design and operation have all been reviewed (Ref. 52), and we shall solely focus on the broad-beam Kaufman gridded-type source, the most widely used at present. In the configuration described, the cylindrical discharge-chamber axis lies horizontally. A hot filament or hollow cathode source, the emitter of energetic electrons, is located on this axis and surrounded coaxially by the anode. During operation, the working gas is fed into the discharge chamber where its atoms or molecules are ionized and raised to the desired potential. To efficiently confine and steer the ions down the chamber, an axial magnetic field is applied. These ions are then extracted and accelerated through matching apertures in a pair of fine-mesh grid screens. The resulting beamlets merge into a broad beam outside the discharge chamber where a neutralizer filament or cathode charge-neutralizes it, creating a beam with equal densities of ions and



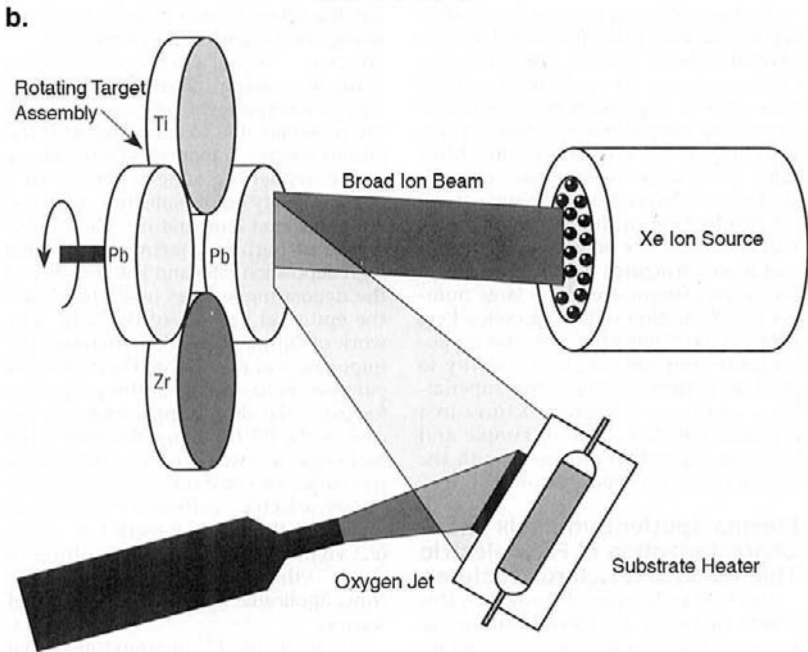


Figure 5-24 (a) Multiple ion-beam target system for depositing $PbZr_xTi_{1-x}O_3$ ferroelectric films. (b) Single ion beam, multitarget system to deposit similar films. (From C. Auciello, A. I. Kingon, and S. B. Krupanidhi, *MRS Bulletin* **21**(6). Reprinted with permission.)

electrons. Further details on the construction of gridded and other types of ion sources can be found in Ref. 53.

Sources with beam diameters as large as one meter have been constructed. Current densities of several mA/cm^2 are achieved. Note that 1 mA/cm^2 is equivalent to 6.25×10^{15} $ions/cm^2$ or several monolayers per second. The resulting beams have a low energy spread (typically 10 eV) and are well collimated with divergence angles of only a few degrees. Furthermore, the background pressure is quite low (10^{-4} torr) compared to typical sputtering or etching plasmas.

5.5.4.3 Film Materials Modified by Ion-Beam Processing

Examples of thin film property modification as a result of ion bombardment are given in Table 5-3 (Refs. 54, 55). Bombarding ion fluxes are generally smaller, but more energetic, than depositing atom fluxes. In general property enhancements tend to rise in step with increases in the ion/atom ratio. Additional materials deposited with ion-beam assist include multimetal dielectrics (Ref. 50), and very hard film materials, e.g., TiC, TiAlN, TiZrN (Ref. 56), carbon nitride (Ref. 57), and diamond (Ref. 58). The reader should appreciate the applicability to all classes of solids and to a broad spectrum of properties. Again we repeat that the main benefits from ion bombardment fall into the following categories (Ref. 59):

1. Enhancement of adatom surface mobility
2. Stimulation of the early stages of film formation, e.g., nucleation, growth, and coalescence
3. Development of preferred crystal orientation
4. Lower substrate temperatures for the onset of epitaxy
5. Crystallization of amorphous films and amorphization of crystalline films
6. Increased film/substrate adhesion
7. Modification of film stress
8. Stimulation of film-sorption effects and film surface reactivity

Some of these effects were elaborated upon in Section 4.6. Perhaps the most important application of ion bombardment is the enhancement of the density and index of refraction of assorted oxide films used for optical coatings.

5.5.5 IONIZED CLUSTER BEAM (ICB) DEPOSITION

The idea of employing energetic ionized clusters of atoms to deposit thin films is due to T. Takagi (Refs. 60, 61). In this novel technique, vapor-phase

Table 5-3
Property Modification by Ion Bombardment during Film Deposition[a]

Film material	Ion species	Property modified	Ion energy (keV)	Ion/atom arrival-rate ratio
Cu	Cu^+	Improved epitaxy	0.050–0.4	10^{-2}
Si	Si^+	Si/Si epitaxy	0.05–0.1	
Ge	Ge^+	Ge/Si epitaxy	0.05	
Ge	Ar^+	Stress, adhesion	0.065–3	2×10^{-4} to 10^{-1}
Nb	Ar^+	Stress	0.1–0.4	3×10^{-2}
Cr	Ar^+, Xe^+	Stress	3.4–11.5	8×10^{-3} to 4×10^{-2}
Cr	Ar^+	Stress	0.2–0.8	7×10^{-3} to 2×10^{-2}
Cu	N^+, Ar^+	Adhesion	50	10^{-2}
Ni on Fe	Ar^+	Hardness	10–20	~0.25
GdCoMo	Ar^+	Magnetic anisotropy	~0.001–0.15	~0.1
SiO_2	Ar^+	Step coverage	0.5	0.3
SiO_2	Ar^+	Step coverage	~0.001–0.08	4.0
AlN	N_2^+	Preferred orientation	0.3–0.5	0.96–1.5
BN	$(B\text{-}N\text{-}H)^+$	Cubic structure	0.2–1.0	~1.0
Si_3N_4	N_2^+		0.06, 0.1	2.1
TiN	N_2^+		1	0.01–0.03
ZrO_2, SiO_2, TiO_2	Ar^+, O_2^+	Refractive index, amorphous→crystal	0.6	2.5×10^{-2} to 10^{-1}
SiO_2, TiO_2	O_2^+	Refractive index	0.3	0.12
SiO_2, TiO_2	O_2^+	Optical transmission	0.03–0.5	0.05–0.25
Ta_2O_5	O_2^+		1.2	2.6

[a] From Refs. 54, 55.

aggregates or clusters, thought to contain a few hundred to a few thousand atoms, are created, ionized, and accelerated toward the substrate as depicted schematically in Fig. 5-25 (Ref. 62). As a result of impact with the substrate, the cluster breaks apart, releasing atoms to spread across the surface. Cluster production is, of course, the critical step and begins with evaporation from a crucible containing a small aperture or nozzle. The evaporant

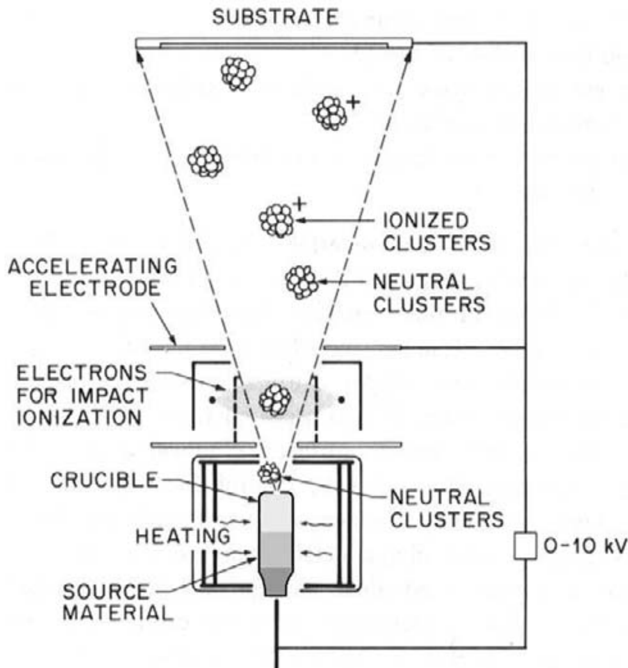


Figure 5-25 Schematic of an ICB deposition system. (Courtesy of W. L. Brown, AT&T Bell Laboratories. Reprinted with permission.)

vapor pressure is much higher (10^{-2}–10 torr) than in conventional vacuum evaporation. For cluster formation the nozzle diameter must exceed the mean free path of vapor atoms in the crucible. Viscous flow of atoms escaping the nozzle then results in an adiabatic supersonic expansion and the formation of stable cluster nuclei. Optimum expansion further requires that the ratio of the vapor pressure in the crucible to that in the vacuum chamber exceed 10^4 to 10^5.

The arrival of ionized clusters with the kinetic energy of the acceleration voltage (0–10 kV), and neutral clusters with the kinetic energy of the nozzle ejection velocity, modifies film nucleation and growth processes in many of the ways outlined in Section 5.5.4.3. As a result, strong adhesion to the substrate, elimination of the columnar-grain morphology, and epitaxial growth can be achieved at low temperatures. Moreover, the magnitude of these effects can be modified by altering the accelerating voltage and extent of electron impact ionization. Virtually all classes of film materials have been deposited by ICB and variant reactive versions, including pure metals,

alloys, intermetallic compounds, semiconductors, oxides, nitrides, carbides, halides, and organic compounds. These are deposited under clean vacuum evaporation conditions ($\sim 10^{-7}$ torr) and energetic ion bombardment of the substrate, two features that normally are mutually exclusive. Among additional ICB deposition effects that have been observed are high-yield lateral sputtering, shallow ion implantation, reactive formation of thin films, and smoothing or planarization of surfaces (Ref. 63). To visualize this last effect, we may imagine the cluster to be a bag of marbles that bursts on surface impact, scattering them in all directions along the surface. Materials such as diamond, soft metals (Au, Ag), nickel alloys, polysilicon, and high-temperature superconductors have been smoothed by ICB processing. For example, argon-gas clusters reduced the height variability of a nickel-alloy surface from 10 to 2 nm.

Despite the attractive features of ICB, the formation of clusters and their role in film formation have not been well understood. Experiment (Ref. 62) has indicated that the total number of atoms agglomerated in large metal clusters is actually very small (only 1 in 10^4) and that only a fraction of large clusters is ionized. The total energy brought to the film surface by ionized clusters is, therefore, quite small. Rather, it appears that individual atomic ions, which are present in much greater profusion than ionized clusters, are the dominant vehicle for transporting energy and momentum to the growing film. In this respect ICB deposition belongs to the class of process deriving benefits from the ion-beam-assisted film growth mechanisms that were discussed previously.

5.5.6 Plasma-Immersion Ion Implantation

The final process considered crosses the capabilities of plasmas and ion beams to modify surface properties, deposit thin films, or effect a combination of both (Ref. 64). The part to be treated by plasma-immersion ion implantation (PIII) is placed in a vacuum chamber where RF or microwave plasmas are created containing ions of the species to be implanted. Repetitive, high negative-voltage pulses ranging from 2 to 300 kV are then applied to the part, causing ions to accelerate through the sheath. Thus, PIII is very different from ion-implantation practice in microelectronics technology where mass-selected, high-energy, unidirectional ion beams scan over planar semiconductor surfaces that are not immersed in a plasma. Under the high negative bias in PIII, ions not only simultaneously bombard the part surface from all directions, but are typically implanted 100 nm or less beneath it. In nitrogen-, oxygen-, and carbon-containing gas plasmas, surface modification results from the formation of subsurface nitrides, oxides, and carbides,

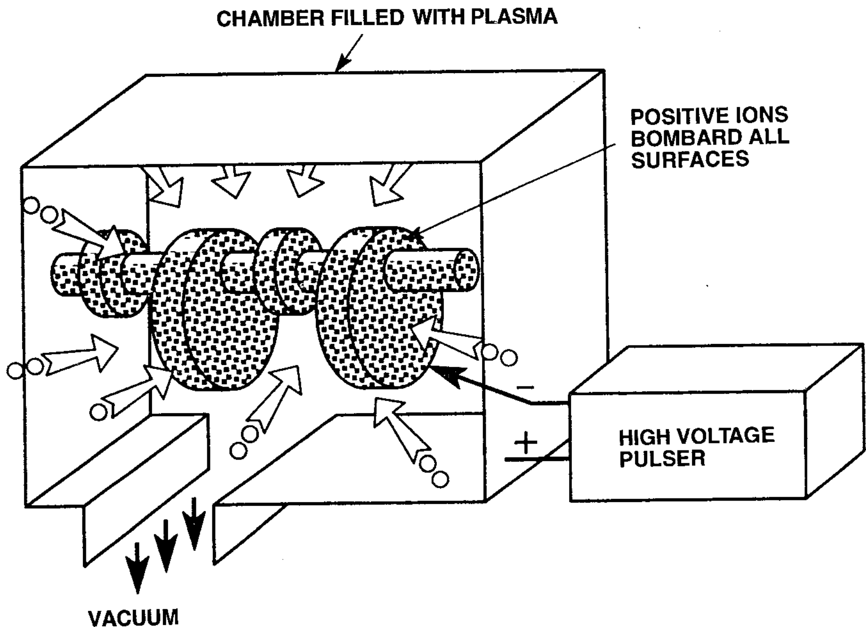


Figure 5-26 Illustration of the PIII process for automotive crankshafts. Because the plasma sheath surrounds the shaft, all of its surfaces are simultaneously ion bombarded without beam aiming or target manipulation. (From J. V. Mantese, I. G. Brown, N. W. Cheung, and G. A. Collins, *MRS Bulletin* **21**(8), 52 (1996). Reprinted with permission.)

respectively. These enhance hardness and wear and corrosion resistance by altering the surface chemistry (e.g., metastable compounds, precipitates) as well as structure (e.g., extended lattice defects, amorphous regions). A number of steels, as well as titanium and aluminum alloys, have been treated by PIII. For example, significant increases in wear resistance are produced in tool steels implanted by 45 keV N_2^+ ions at a dose of 1×10^{18} atoms/cm^2. For species that diffuse appreciably, e.g., N, postimplantation annealing can extend the case-hardened depth appreciably to many micrometers depending on temperature and time.

The PIII process is shown schematically in Fig. 5-26, where the surfaces of an automobile crankshaft are being exposed to ion bombardment in order to make it more wear resistant. Another promising application under consideration by the automobile industry is the conformal coating of cast Al–Si alloy auto pistons with a very hard diamond-like coating (DLC). This coating contains a large fraction of the sp^3 carbon-bonding orbitals that we associate with diamond (Section 6.8.5). But it is not normally adherent to untreated surfaces so the piston is first cleaned in an Ar plasma (step 1).

Then carbon is implanted by PIII methods to produce a bonding layer composed of both Al and Si carbides (step 2). After excess graphite is etched away by Ar ion bombardment (step 3), an adherent DLC film is deposited on the piston surface, thus completing the four-step process.

Finally, in what may have significant implications in microelectronics, PIII processing has already been commercially used to dope silicon layers less than 50 nm thick during the fabrication of CMOS devices (Ref. 65). In these devices, the very thin Si film of a Si-on-SiO_2 (SOI) structure (see Section 8.5.2.2) rather than bulk Si is employed. This means ion implantation at energies below 20 keV. Such low energies approach the low-end capabilities of traditional high-energy beamline ion-implantation systems that use electromagnets for ion-mass selection.

5.6 CONCLUSION

This chapter has built on foundations laid in the previous two chapters. In particular, the basic configuration of two electrically powered electrodes separated by a plasma is common to a number of film deposition and etching processes. This is usually true irrespective of whether processing occurs under DC, AC, or RF powering, or whether reactive gases are present or not. The simplest of the processes is sputter deposition of metal films in a planar diode configuration. Magnetron sputtering processes relying on magnetic confinement of the plasma are now universally employed because they yield films of all materials at relatively high deposition rates. Sputter deposition of nonmetallic films, e.g., oxides or nitrides, can be done either employing cathodes of these materials as in RF sputtering, or more commonly now by reactive sputtering using metal targets. In the latter, cathode arcing and low deposition rates are problems that have been overcome by pulsed-power magnetron sputtering. Film deposition is further complicated when the substrate topography contains trenches and vias that must be uniformly coated. In the case of metallizations, collimated sputtering and the use of ionized metal vapor are strategies that have led to better sidewall coverage.

Film etching and deposition may be viewed as related but inverse processes. However, material removal generally requires a more involved sequence of atomic motion and interaction than material deposition. Thus plasma etching often tends to be more challenging than sputtering. Not only must the complex chemistry of the plasma be optimized for etching selectivity, but material removal must often occur anisotropically on patterned films. In addition, halide-bearing etching and product gases pose a danger to health and the environment that is normally absent in sputtering

processes. The imperative for enhanced etching rates has led to the deployment of high-density plasma systems.

Hybrid film-deposition systems combining two or more of the following features, i.e., evaporation, sputtering, low voltage–high current arcs, inert-gas plasmas, chemically reactive plasmas, and ion beams, have been widely used. They include ion plating, activated evaporation, cathodic arc, and assorted ion-beam-assisted deposition processes. In these, ion bombardment of growing films is a common desideratum.

EXERCISES

1. The processes of electron impact ionization and secondary emission of electrons by ions control the current in a sputtering system acording to the Townsend equation (Eq. 4-2)

$$i = \frac{i_0 \exp \alpha d}{[1 - \gamma_e(\exp \alpha d - 1)]}$$

where all terms were defined previously.

(a) If the film deposition rate during sputtering is proportional to the product of i and S, calculate the proportionality constant for Cu in this system if the deposition rate is 200 Å/min for 0.5-keV Ar ions. Assume $\alpha = 0.1$ ion/cm, $\gamma_e = 0.08$ electron/ion, $d = 10$ cm, and $i_0 =$ 100 mA.

(b) What deposition rate can be expected for 1 keV Ar if $\alpha = 0.15$ ion/cm and $\gamma_e = 0.1$ electron/ion?

2. In a DC planar magnetron sputtering system at 1000 V, the cathode–anode spacing is 10 cm. What magnetic field should be applied to trap electrons within 0.5 cm of the target?

3. Provide good physical reasons for the following observations made during magnetron sputtering:

(a) The electrical resistivity of a 200-nm chromium film was 15 $\mu\Omega$-cm using a 99.95% pure Cr target from manufacturer A, but another target of the same metal purity from manufacturer B yielded 30 $\mu\Omega$-cm films.

(b) The pixel defect density in indium–tin oxide films used for liquid crystal displays was twice as high when the target was bonded to a copper backing plate than when it was bonded to a molybdenum backing plate.

(c) For targets of the same sintered density and operating power, there was less variation in the electrical resistivity of conducting films when the racetrack was wide rather than narrow.

(d) Increasing the target density reduced the nodule-defect density in the resultant films.

4. Estimate the frequency of target pulsing required for arc suppression during reactive sputtering given the following information: target current density = 1 mA/cm^2, dielectric constant of insulating film = 3.7, dielectric breakdown field = 30 kV/mm.

5. Sketch typical magnetic field distributions for three different dual magnetron-sputtering-target configurations shown below. In all cases each of the magnetrons is unbalanced. In the first case the substrate is parallel to coplanar targets; in the latter two cases a rotating substrate is sandwiched between the facing planar targets.

(a)		(b)	(c)
NSN	*SNS*	*SNS*	*NSN*
Substrate		(Rotating Substrate)	(Rotating Substrate)
		NSN	*NSN*

Comment on the deposition process in each case.

6. What are the advantages and disadvantages of having electrodes (a) inside or (b) outside plasma-etching reactors? Enumerate the advantages and disadvantages of having substrates inside (c) or a distance away (d) from the plasma-generation region during film deposition?

7. As a general rule of thumb, the ratio of the rate of etch-product formation to the flow rate of etch gas should be greater than 0.1 for uniform etching. Suppose a 7.5-cm diameter silicon wafer is etched at a rate of 500 Å/min in a CF_4 plasma.
   (a) How many atoms of Si are removed per minute?
   (b) What evolution rate of SiF_4 does this correspond to in standard cubic centimeters per minute (sccm)?
   (c) What minimum CF_4 flow rate should be maintained?

8. Estimate the maximum amplitude and energy of CF_3^+ ions in a 13.56-MHz discharge of CF_4 gas if the effective DC bias electric field is 100 kV/m.

9. Consider a Reinberg-type parallel plate diode reactor where an RF voltage $V_{RF} = V_0 \sin \omega t$ is applied between electrodes 1 and 2. Plasma sheaths develop at each electrode with potentials of V_1 and V_2. Derive expressions for V_1 and V_2 if the following electrical conditions hold:
   (a) The DC components of the sheath potential must cancel, i.e., $V_1(\text{DC}) = V_2(\text{DC})$.
   (b) V_1 and V_2 are 180° out of phase.
   (c) Kirchoff's law requires $V_{RF} = V_1 - V_2$.

10. Assume that plasma etching or gasification of one Si atom requires delivery of six atoms of Cl to the Si wafer surface. Below what Cl^- ion current density will the required process etch rate of 1000 Å Si per minute be unattainable?

11. (a) What concentration of fluorine is required to etch (Si ⟨100⟩) at a rate of 5000 Å/min at a temperature of 250°C?
    (b) By what percent must the temperature be raised to have the equivalent effect on etch rate that a 1% increase in fluorine concentration has?

12. At what sputter deposition rate of indium on a silicon wafer will the film melt within 1 minute? The melting point of In is 155°C.

13. A sheet steel (0.5 mm thick) enclosure of complex shape was ion plated with TiN. The coating was examined at two different locations. At 15 cm from the metal source the coating facing the source was 9 μm thick and consisted of a Ti_2N + TiN phase mixture; on the backside there was a 0.6-μm thick TiN deposit possessing a grain structure with preferential (200) crystal orientation. When a spot 25 cm away from the source was probed, the front and back coatings were found to be 1 μm and 0.5 μm thick, respectively, and both were (200) oriented.
    (a) Explain the coating thickness variations.
    (b) Provide reasons for the compositional differences in the coating closest to the source relative to the other coated regions.

14. Plasma are often used to clean surfaces. Two of the common surface contaminants are water and hydrocarbons (oil).
    (a) Consider a 0.1 μm thick water layer on a glass substrate that must be removed prior to the deposition of an optical coating. What plasma composition and experimental arrangement would you recommend to remove the water?
    (b) Steel sheet contains a 1 μm thick oil layer. Suggest a plasma-based cleaning process to degrease this metal.
    (c) Typical cleaning gases have an ionization energy of 5 eV per ion. Explain the electronic and chemical mechanisms involved in your process for cleaning steel.

15. A reactive metal element contains two isotopes. The heavier one is present in relatively small amounts but it is very valuable, especially when unoxidized. Design a process that would enable a separation of the isotopes from a piece of this metal. All you have at your disposal is a laboratory equipped with all of the PVD capabilities discussed in Chapters 3 and 5. You may use other commonly available laboratory components and pieces of equipment.

16. In a laboratory-scale cathodic arc plasma deposition process, 10 kW (50 volts, 200 amps) of electric power is delivered to a 10 cm × 10 cm metal alloy cathode. Assume a single cathode spot, 10 μm in diameter, raster scans across the cathode area at a speed of 100 m/s. How does this process compare with or differ from those of comparable laboratory-scale, pulsed laser deposition and electron-beam evaporation systems in terms of:
    (a) the magnitude of specific power and energy involved
    (b) evaporation or deposition rates
    (c) the uniformity of film deposition
    (d) film composition uniformity
    (e) applicability to deposition of different materials

REFERENCES

1. S. M. Rossnagel, *J. Vac. Sci. Technol.* **B16**, 2585 (1998).
2. P. F. Williams, ed., *Plasma Processing of Semiconductors*. Kluwer Academic, Dordrecht, The Netherlands, 1997.
3. J. L. Vossen and W. Kern, eds., *Thin Film Processes*. Academic Press, New York, 1978.
4. J. L. Vossen and W. Kern, eds., *Thin Film Processes II*. Academic Press, New York, 1991.
5. D. M. Mattox, *Handbook of Physical Vapor Deposition (PVD) Processing*. Noyes Publishing, Westwood, NJ, 1998.
6. A. Grill, *Cold Plasma in Materials Fabrication*, IEEE Press, New York, 1994.
7. R. A. Powell and S. M. Rossnagel, *PVD for Microelectronics: Sputter Deposition Applied to Semiconductor Manufacturing* (Vol. 26, Physics of Thin Films). Academic Press, San Diego, 1999.
8. M. H. Francombe and J. L. Vossen, eds., *Plasma Sources for Thin Film Deposition and Etching* (Vol. 18, Physics of Thin Films). Academic Press, San Diego, 1994.
9. D. M. Manos and D. L. Flamm, eds., *Plasma Etching*. Academic Press, Boston, 1989.
10. B. Heintz, *Vacuum and Thinfilm* **2**(10), 22 (1999).
11. G. C. Stutzin, K. Rozsa, and A. Gallagher, *J. Vac. Sci. Technol.* **A11**, 647 (1993).
12. J. J. Cuomo and R. J. Gambino, *J. Vac. Sci. Technol.* **14**, 152 (1977).
13. D. L. Smith, *Thin-Film Deposition*. McGraw-Hill, New York, 1995.
14. W. A. P. Claassen, W. G. J. N. Valkenburg, M. F. C. Willemsen, and W. M. v.d. Wijgert, *J. Electrochem. Soc.* **132**, 893 (1985).
15. M. Konuma, *Film Deposition by Plasma Techniques*. Springer Verlag, Berlin, 1992.
16. A. J. van Roosmalen, J. A. G. Baggerman, and S. J. H. Brader, *Dry Etching for VLSI*. Plenum, New York, 1991.
17. W. D. Westwood, in *Physics of Thin Films*, Vol. 14, H. M. Francombe and J. L. Vossen, eds. Academic Press, New York, 1989.
18. S. Berg, T. Nyberg, H. O. Blom, and C. Nender, in *Handbook of Thin Film Process Technology*, D. A. Glocker and S. I. Shah, eds. Institute of Physics Publishing, Bristol, 1998.
19. L. I. Maissel and M. H. Francombe, *An Introduction to Thin Films*. Gordon and Breach, New York, 1973.
20. Reported in *Vacuum and Thinfilm* **2**(1), 16 (1999).

21. S. Schiller, K. Goedicke, J. Reschke, V. Kirchoff, S. Schneider, and F. Milde, *Surf. Coatings Technol.* **61**, 331 (1993).
22. S. Schiller, V. Kirchoff, K. Goedicke, and N. Schiller, *Semicond. Int.* **S-11**, Dec. 1996.
23. P. J. Kelly and R. D. Arnell, *Vacuum* **56**, 159 (2000).
24. S. L. Rohde, in *Physics of Thin Films*, Vol. 18, M. H. Francombe and J. L. Vossen, eds. Academic Press, San Diego, 1994.
25. A. Belkind, E. Ezell, W. Gerristead, Z. Orban, P. Rafalco, D. Dow, J. Felts, and R. Laird, *J. Vac. Sci. Technol.* **A9**, 530 (1991).
26. See Workshop on Dusty Plasmas, *J. Vac. Sci. Technol.* **A14** (1995).
27. G. S. Selwyn, C. A. Weiss, F. Sequeda, and C. Huang, *J. Vac. Sci. Technol.* **A15**, 2023 (1997).
28. M. Dalvie, M. Surendra, G. S. Selwyn, and C. R. Guarnieri, *Plasma Sources Sci. Technol.* **3**, 442 (1994).
29. R. K. Watts in *VLSI Technology*, 2nd ed., S. M. Sze, ed. McGraw-Hill, New York, 1988.
30. D. L. Flamm and G. K. Herb, in *Plasma Etching*, D. M. Manos and D. L. Flamm, eds. Academic Press, Boston, 1989.
31. D. L. Flamm, in *Plasma Processing of Semiconductors*, P. F. Williams, ed. Kluwer, Dordrecht, The Netherlands, 1997.
32. R. B. Laibowitz, A. N. Broers, J. T. C. Yeh, and J. M. Viggiano, *Appl. Phys. Lett.* **35**, 891 (1979).
33. H. W. Lehmann, in *Thin Film Processes II*, J. L. Vossen and W. Kern, eds. Academic Press, Boston, 1991.
34. D. L. Flamm, V. M. Donnelly, and J. I. Mucha, *J. Appl. Phys.* **52**, 3633 (1981).
35. D. W. Hess, in *Microelectronic Materials and Processes*, R. A. Levy, ed. Kluwer Academic, Dordrecht, The Netherlands, 1989.
36. A. R. Reinberg, U.S. Patent, 3,757,733 (1975).
37. M. A. Lieberman and R. A. Gottscho, in *Physics of Thin Films*, Vol. 18, M. H. Francombe and J. L. Vossen, eds. Academic Press, San Diego, 1994.
38. O. A. Popov, *High Density Plasma Sources: Design, Physics and Performances*. Noyes Publications, Park Ridge, NJ, 1995.
39. L. T. Lamont, *Solid State Technol.* **22**(9), 107 (1979).
40. J. A. Thornton, *Thin Solid Films* **54**, 23 (1978).
41. T. Kalber and T. Jung, Int. Meeting of the Electrochem. Soc., Symposium on Metallized Plastics, VI, Paris, p. 1 (1997).
42. E. J. Lerner, *Laser Focus World* **35**(12), 79 (1999).
43. D. M. Mattox, *J. Vac. Sci. Technol.* **10**, 47 (1973).
44. N. A. G. Ahmed, *Ion Plating Technology*. John Wiley & Sons, Chichester, 1987.
45. K. S. Fancey and A. Matthews, in *Advanced Surface Coatings*, D. S. Rickerby and A. Matthews, eds. Blackie (Chapman and Hall), Glasgow, 1991.
46. K. S. Fancey and A. Matthews, *Surf. Coatings Technol.* **36**, 233 (1988).
47. R. F. Bunshah and C. Deshpandey, in *Physics of Thin Films*, Vol. 13, M. H. Francombe and J. L. Vossen, eds. Academic Press, New York, 1987.
48. P. C. Johnson, in *Thin Film Processes II*, J. L. Vossen and W. Kern, eds. Academic Press, Boston, 1991.
49. J. M. E. Harper and J. J. Cuomo, *J. Vac. Sci. Technol.* **21**(3), 737 (1982).
50. O. Auciello, A. I. Kingon, and S. B. Krupanidhi, *MRS Bull.* **21**(6), 25 (1996).
51. M. Geis, S. W. Pang, N. E. Efremow, G. A. Lincoln, G. D. Johnson, and W. D. Goodhue, in *Handbook of Ion Beam Processing Technology*, J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds. Noyes Publications, Park Ridge, NJ, 1989.
52. H. R. Kaufman and R. S. Robinson (pp. 8, 39), W. M. Holber (p. 21), in *Handbook of Ion Beam Processing Technology*, J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds. Noyes Publications, Park Ridge, NJ, 1989.

53. H. R. Kaufman and R. S. Robinson, *Operation of Broad-Beam Sources.* Commonwealth Scientific Corporation, Alexandria, VA, 1987.
54. J. M. E. Harper, J. J. Cuomo, R. J. Gambino, and H. R. Kaufman, in *Ion Beam Modification of Surfaces*, O. Auciello and R. J. Kelly, eds. Elsevier, Amsterdam, 1984.
55. J. K. Hirvonen, in *Materials and Processes for Surface and Interface Engineering*, Y. Pauleau, ed. Kluwer Academic, Dordrecht, The Netherlands, 1995.
56. J. Singh, *Adv. Mat. Processes*, p. 27, Dec. (1996).
57. M. Kohzaki, A. Matsumuro, T. Hayashi, M. Muramatsu, and K. Yamaguchi, *Thin Solid Films* **308–309**, 239 (1997).
58. M. Kitabatake and K. Wasa, in *Handbook of Ion Beam Processing Technology*, J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds. Noyes Publications, Park Ridge, NJ, 1989.
59. P. J. Martin and R. P. Netterfield, in *Handbook of Ion Beam Processing Technology*, J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds. Noyes Publications, Park Ridge, NJ, 1989.
60. T. Tagaki, in *Physics of Thin Films*, Vol. 13, M. H. Francombe and J. L. Vossen, eds. Academic Press, San Diego, 1987.
61. I. Yamada and T. Tagaki, in *Handbook of Ion Beam Processing Technology*, J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman, eds. Noyes Publications, Park Ridge, NJ, 1989.
62. W. L. Brown, M. F. Jarrold, R. L. McEachern, M. Sosnowski, G. Takaoka, H. Usui, and I. Yamada, *Nucl. Inst. Methods Phys. Res.* **B59/60**, 182 (1991).
63. W. Skinner, D. Fenner, and L. Allen, *Vacuum Solutions*, March/April, p. 29 (1999).
64. J. V. Mantese, I. G. Brown, N. W. Cheung, and G. A. Collins, *MRS Bull.* **21**(8), 52 (1996).
65. M. J. Current, S. W. Bedell, I. J. Malik, L. M. Feng, and F. J. Henley, *Solid State Technol.* **43**(9), 66 (2000).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 6

Chemical Vapor Deposition

6.1 INTRODUCTION

Chemical vapor deposition (CVD) is the process of chemically reacting a volatile compound of a material to be deposited, with other gases, to produce a nonvolatile solid that deposits atomistically on a suitably placed substrate. It differs from physical vapor deposition (PVD), which relies on material transfer from condensed-phase evaporant or sputter target sources. Because CVD processes do not require vacuum or unusual levels of electric power, they were practiced commercially prior to PVD. A century ago CVD methods were used to deposit a protective tungsten coating on carbon filaments in an attempt to extend the life of incandescent lamps (Ref. 1) Today, high-temperature CVD processes for producing thin films and coatings have found increasing applications in such diverse technologies as the fabrication of solid-state electronic devices, the manufacture of ball bearings and cutting tools, and the production of rocket engine and nuclear reactor components. In particular, the need for high-quality epitaxial (single crystal) films in both silicon and compound-semiconductor technology, coupled with the necessity to deposit associated insulating and passivating films, has served as a powerful driver spurring the development of CVD processing methods. A schematic view of the silicon MOS transistor structure in Fig. 6-1 (also see Fig. 5-1) indicates the extent to which CVD materials monopolize the films deposited above the original wafer. With the exception of the gate oxide and metals Al-Cu, Ti/TiN, and $TiSi_2$, all films are deposited by some variant of CVD processing. They include the (epitaxial) Si substrate, polysilicon, various SiO_2 films, low-temperature

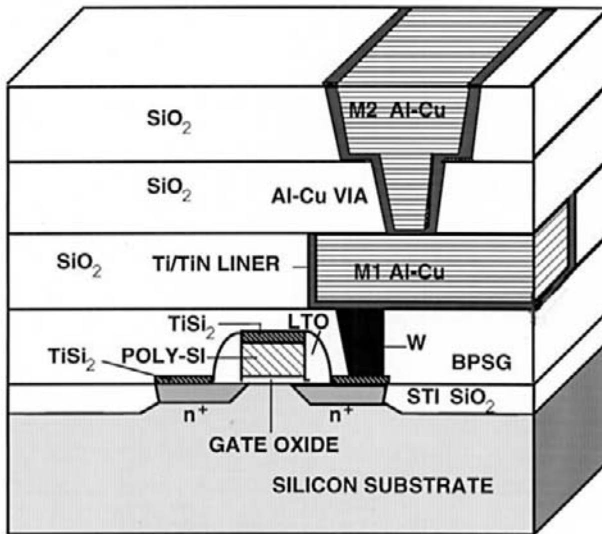


Figure 6-1 Schematic cross-sectional view of an MOS transistor structure with multilevel metallization scheme. Film materials deposited by CVD are indicated in the text. *Note:* LTO = low temperature oxide, BPSG = borophospho-silicate glass, STI = silicon trench insulator. After K. P. Rodbell, IBM, T. J. Watson Research Division.

oxide (LTO), borophosphosilicate glass (BPSG), and W plugs. Silicon nitride is another CVD material commonly used in these devices.

Among the reasons for the growing adoption of CVD methods is the ability to produce a large variety of films and coatings of metals, semiconductors, and inorganic as well as organic compounds in either a crystalline or vitreous form, possessing desirable properties. Furthermore, the ability to controllably create films of widely varying stoichiometry makes CVD unique among deposition techniques. Other advantages include the affordable cost of the equipment and operating expenses, the suitability for both batch and semicontinuous operation, and the compatibility with other processing steps. Because of this, many variants of CVD processing have been researched and developed, including atmospheric pressure (APCVD), low-pressure (LPCVD), plasma-enhanced (PECVD), and laser-enhanced (LECVD) chemical vapor deposition. Hybrid processes combining features of both physical and chemical vapor deposition have also emerged.

The fundamental sequential steps that occur in every CVD process are sketched in Fig. 6-2 and include:

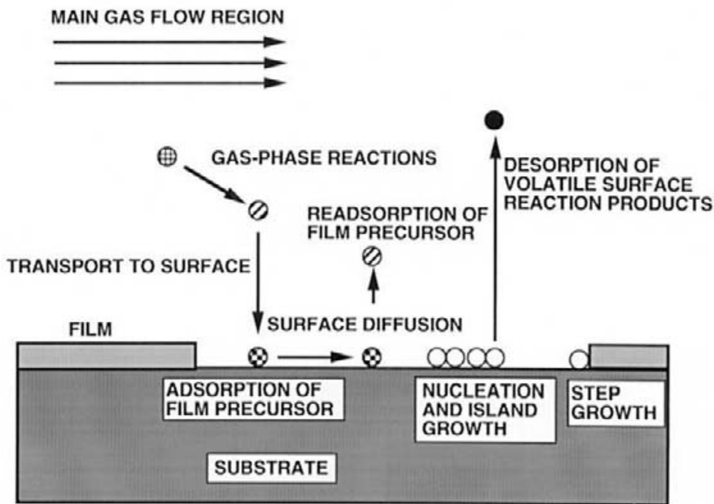


Figure 6-2 Sequence of gas transport and reaction processes contributing to CVD film growth. (From *Chemical Vapor Deposition*, edited by M. L. Hitchman and K. F. Jensen. Reprinted with the permission of Academic Press, Ltd., and Professor K. F. Jensen, MIT.)

1. Convective and diffusive transport of reactants from the gas inlets to the reaction zone
2. Chemical reactions in the gas phase to produce new reactive species and by-products
3. Transport of the initial reactants and their products to the substrate surface
4. Adsorption (chemical and physical) and diffusion of these species on the substrate surface
5. Heterogeneous reactions catalyzed by the surface leading to film formation
6. Desorption of the volatile by-products of surface reactions
7. Convective and diffusive transport of the reaction by-products away from the reaction zone

Figure 6-3 provides a perspective that integrates many of these steps, here subdivided into coordinates related to basic chemistry and physics, gas transport phenomena, and to the reactors that must deposit films efficiently. Accordingly, much of this chapter is devoted to exploring the scientific and engineering issues raised in this figure. Practical concerns of thermodynamics, gas transport, deposition rates, and film properties will be discussed in assorted thermal CVD processes. Continuing a thread running

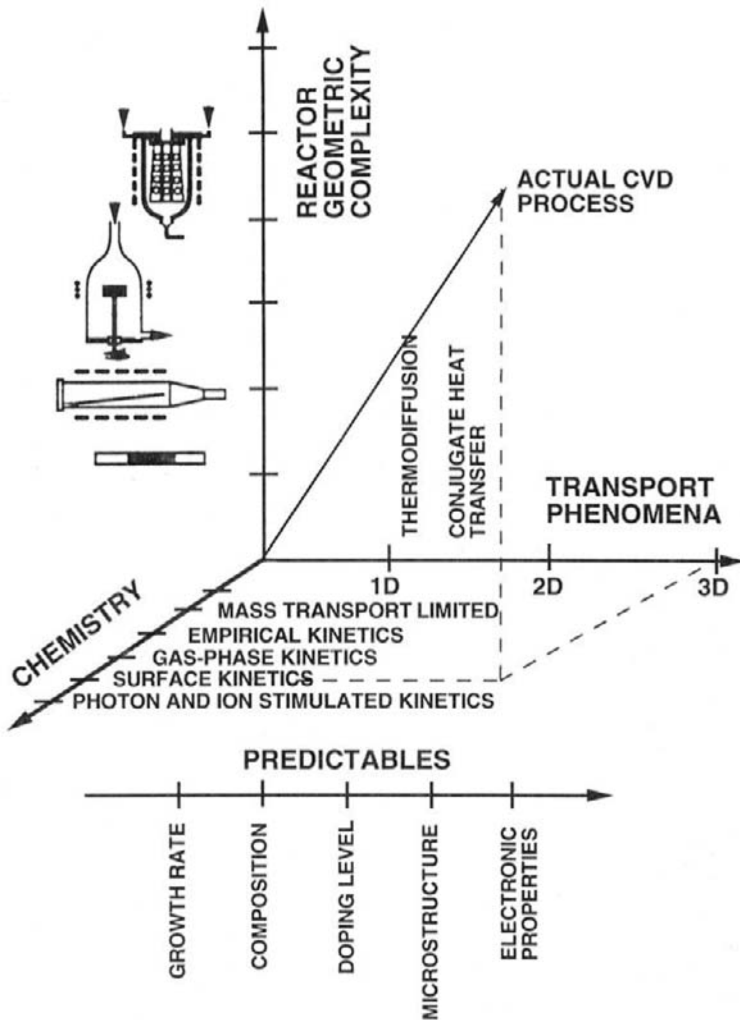


Figure 6-3 Schematic diagram of the chemical, transport, and geometrical complexities involved in modeling CVD processes. (From *Chemical Vapor Deposition*, edited by M. L. Hitchman and K. F. Jensen. Reprinted with the permission of Academic Press, Ltd., and Professor K. F. Jensen, MIT.)

through the two previous chapters, plasma-assisted CVD processes will also be treated.

This chapter largely focuses on CVD processes that yield amorphous or polycrystalline films used in a variety of technologies. An exception is the treatment of epitaxial Si, a material well suited to model the thermo-

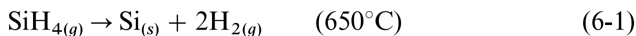
dynamics and kinetics of film formation. Recommended books (Refs. 2–6) and review articles (Refs. 1, 7–10) dealing with these aspects of CVD are listed in the references. The very important CVD methods used to grow epitaxial III–V compound-semiconductor films for electronic and optoelectronics applications will be deferred to Chapter 8, however. In this way, corollary issues related to film crystallinity, stress, defects, and electrical properties can be better appreciated. Additional references reviewing these applications of CVD will be given then.

6.2 REACTION TYPES

To view the scope of the subject broadly, it is useful to first briefly categorize the various types of chemical reactions that have been employed to deposit films and coatings. Corresponding examples are given for each by indicating the essential overall chemical equation and approximate reaction temperature. In these equations (*g*) and (*s*) refer to gas and solid, respectively.

6.2.1 PYROLYSIS

Pyrolysis involves the thermal decomposition of such gaseous species as hydrides, carbonyls and organometallic compounds on hot substrates. Commercially important examples include the high-temperature pyrolysis of silane to produce polycrystalline or amorphous silicon films, and the low-temperature decomposition of nickel carbonyl to deposit nickel films:



$$SiH_{4(g)} \rightarrow Si_{(s)} + 2H_{2(g)} \qquad (650^{\circ}C) \tag{6-1}$$

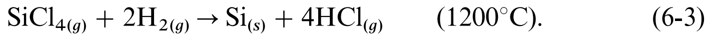
$$Ni(CO)_{4(g)} \rightarrow Ni_{(s)} + 4CO_{(g)} \qquad (180^{\circ}C). \tag{6-2}$$

Interestingly, the latter reaction is the basis of the Mond process, which has been employed for well over a century in the metallurgical refining of Ni.

6.2.2 REDUCTION

These reactions commonly employ hydrogen gas to effect the reduction of such gaseous species as halides, carbonyl halides, oxyhalides, or other oxygen-containing compounds. An important example is the reduction of $SiCl_4$ on single-crystal Si wafers to produce epitaxial Si films according to

the reaction



$$SiCl_{4(g)} + 2H_{2(g)} \rightarrow Si_{(s)} + 4HCl_{(g)} \qquad (1200^\circ C). \qquad (6\text{-}3)$$

Refractory metal films such as W and Mo have been deposited by reducing the corresponding hexafluorides, e.g.,

$$WF_{6(g)} + 3H_{2(g)} \rightarrow W_{(s)} + 6HF_{(g)} \qquad (300^\circ C) \qquad (6\text{-}4)$$

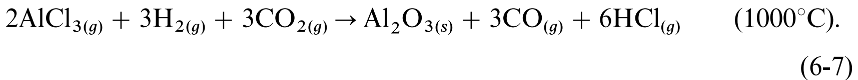
$$MoF_{6(g)} + 3H_{2(g)} \rightarrow Mo_{(s)} + 6HF_{(g)} \qquad (300^\circ C). \qquad (6\text{-}5)$$

Tungsten films deposited at low temperatures have served to interconnect levels of metallization in integrated circuits. Interestingly, WF_6 gas reacts directly with exposed silicon surfaces depositing thin W films while releasing the volatile SiF_4 by-product. In this way silicon contact holes can be *selectively* filled with tungsten while leaving neighboring insulator surfaces uncoated (see Section 6.8.6).

6.2.3 OXIDATION

Two examples of important oxidation reactions are:

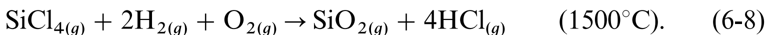
$$SiH_{4(g)} + O_{2(g)} \rightarrow SiO_{2(s)} + 2H_{2(g)} \qquad (450^\circ C) \qquad (6\text{-}6)$$



$$2AlCl_{3(g)} + 3H_{2(g)} + 3CO_{2(g)} \rightarrow Al_2O_{3(s)} + 3CO_{(g)} + 6HCl_{(g)} \qquad (1000^\circ C). \qquad (6\text{-}7)$$

The deposition of SiO_2 by Eq. 6-6 is carried out at a stage in the processing of integrated circuits where higher substrate temperatures cannot be tolerated. Hard alumina coatings that extend the life of cutting tools are produced by reaction 6-7.

In another process of technological significance, a very pure form of SiO_2 is produced by the oxidation reaction of Eq. 6-8:



$$SiCl_{4(g)} + 2H_{2(g)} + O_{2(g)} \rightarrow SiO_{2(g)} + 4HCl_{(g)} \qquad (1500^\circ C). \qquad (6\text{-}8)$$

The eventual application here is the production of optical fiber for communications purposes. Rather than a thin film, the SiO_2 forms a cotton-candy-like deposit consisting of soot particles less than 1000 Å in size. These are then consolidated by elevated-temperature sintering to produce a fully dense silica rod for subsequent drawing into fiber. Whether silica films deposit or soot forms is governed by process variables favorable to heterogeneous or homogeneous nucleation, respectively. Homogeneous soot formation is essentially the result of a high $SiCl_4$ concentration in the gas phase.

6.2.4 COMPOUND FORMATION

A variety of carbide, nitride, boride, etc., films and coatings can be readily produced by CVD techniques. What is required is that the compound elements exist in a volatile form and be sufficiently reactive in the gas phase. Examples of commercially important reactions for the deposition of hard, wear-resistant surface coatings include:

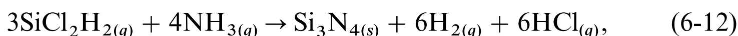
$$SiCl_{4(g)} + CH_{4(g)} \rightarrow SiC_{(s)} + 4HCl_{(g)} \qquad (1400°C) \qquad (6\text{-}9)$$

$$TiCl_{4(g)} + CH_{4(g)} \rightarrow TiC_{(s)} + 4HCl_{(g)} \qquad (1000°C) \qquad (6\text{-}10)$$

$$BF_{3(g)} + NH_{3(g)} \rightarrow BN_{(s)} + 3HF_{(g)} \qquad (1100°C). \qquad (6\text{-}11)$$

Films and coatings of compounds can generally be produced using a variety of precursor gases and reactions. For example, in the much-studied SiC system, layers were first produced in 1909 through reaction of $SiCl_4 + C_6H_6$ (Ref. 11). Subsequent reactant combinations over the years have included $SiCl_4 + C_3H_8$, $SiBr_4 + C_2H_4$, $SiCl_4 + C_6H_{14}$, $SiHCl_3 + CCl_4$, and $SiCl_4 + C_6H_5CH_3$, to name a few, as well as volatile organic compounds containing both silicon and carbon in the same molecule (e.g., CH_3SiCl_3, CH_3SiH_3, $(CH_3)_2SiCl_2$). Although the deposit is nominally SiC in all cases, resultant properties generally differ because of structural, compositional, and processing differences.

Impermeable insulating and passivating films of Si_3N_4 are required to hermetically seal integrated circuits. Although they can be deposited at 750°C by the reaction



$$3SiCl_2H_{2(g)} + 4NH_{3(g)} \rightarrow Si_3N_{4(s)} + 6H_{2(g)} + 6HCl_{(g)}, \qquad (6\text{-}12)$$

the necessity to deposit silicon-nitride films at lower temperatures has led to alternate processing involving the use of plasmas. Films can be deposited below 300°C with SiH_4 and NH_3 reactants, but considerable amounts of hydrogen are incorporated into the deposits.

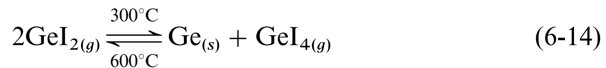
The very important and rapidly growing metalorganic CVD (MOCVD) processes, used to deposit assorted epitaxial compound-semiconductor films, also fit under the present category of reactions. As the name implies, volatile organic precursor-compounds such as trimethylgallium (TMGa or $(CH_3)_3Ga$), trimethylindium (TMIn), etc., are employed. They are reacted with group V hydrides to form the semiconductor compound, e.g.,

$$(CH_3)_3Ga_{(g)} + AsH_{3(g)} \rightarrow GaAs_{(s)} + 3CH_{4(g)}. \qquad (6\text{-}13)$$

Similar MOCVD reactions are exploited in producing a wide assortment of complex oxides and semiconductors, and these processes will be treated further in Sections 6.6.4 and 8.5.3.1, respectively.

6.2.5 DISPROPORTIONATION

Disproportionation reactions are possible when a nonvolatile metal can form volatile compounds having different degrees of stability depending on the temperature. This manifests itself in compounds, typically halides, where the metal exists in two valence states, e.g., GeI_4 and GeI_2, such that the lower-valent state is more stable at higher temperatures. As a result, the metal can be transported into the vapor phase by reacting it with its volatile, higher-valent halide to produce the more stable lower-valent halide. The latter disproportionates at lower temperatures to produce a deposit of metal while regenerating the higher-valent halide. This complex sequence can be simply described by the reversible reaction

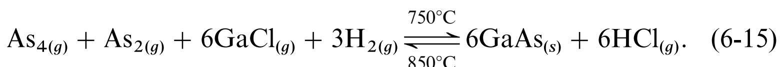


$$2GeI_{2(g)} \underset{600°C}{\overset{300°C}{\rightleftharpoons}} Ge_{(s)} + GeI_{4(g)} \tag{6-14}$$

and realized in systems where provision is made for mass transport between hot and cold ends. Elements which have lent themselves to this type of transport reaction include aluminum, boron, gallium, indium, silicon, titanium, zirconium, beryllium, and chromium. Single-crystal films of Si and Ge were grown by disproportionation reactions in the early days of CVD experimentation on semiconductors (Ref. 12) employing reactors like that shown in Fig. 6-4. The enormous progress made since then is evident.

6.2.6 REVERSIBLE TRANSFER

Chemical transfer or transport processes are characterized by a reversal in the reaction equilibrium at source and deposition regions maintained at different temperatures within a single reactor. An important example is the deposition of epitaxial GaAs films by the *chloride* process according to the reaction



$$As_{4(g)} + As_{2(g)} + 6GaCl_{(g)} + 3H_{2(g)} \underset{850°C}{\overset{750°C}{\rightleftharpoons}} 6GaAs_{(s)} + 6HCl_{(g)}. \tag{6-15}$$

Here $AsCl_3$ gas transports molten Ga, contained within the reactor, toward the substrates in the form of GaCl vapor. Subsequent reaction with As_4 causes deposition of GaAs at low temperatures. At elevated temperatures the reaction reverses and the film is etched. In the alternative *hydride* process, As is introduced in the form of AsH_3 (arsine) while HCl serves to

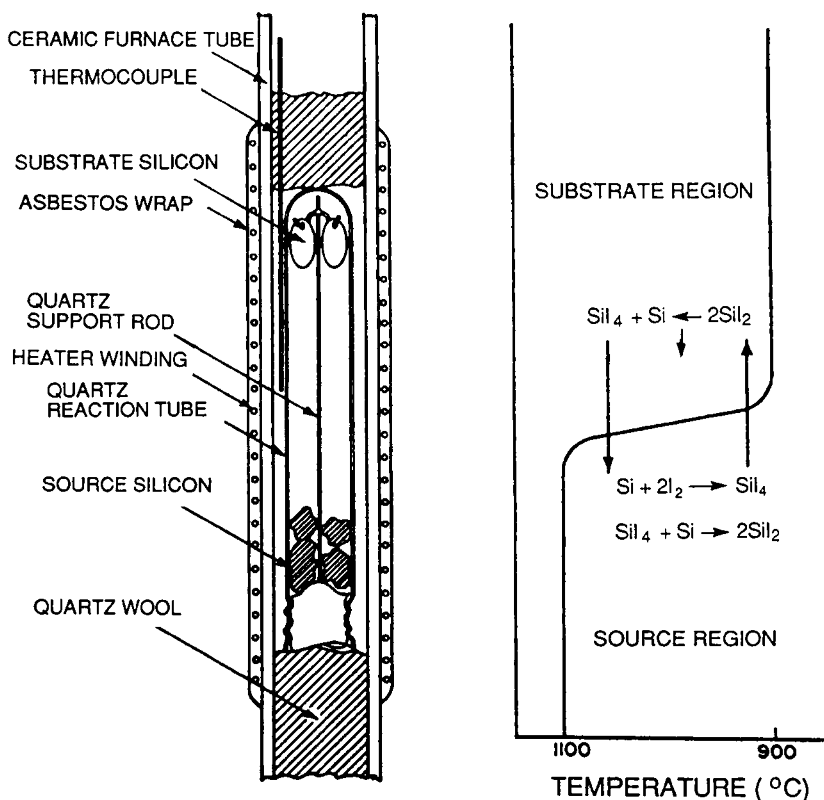


Figure 6-4 Early experimental reactor for epitaxial growth of Si films. (From Ref. 12, copyright © 1960 by the IBM Corp., reprinted with permission.)

transport Ga. Both processes essentially involve the same gas-phase reactions and are carried out in similar reactors. What is significant is that single-crystal, *binary* (primarily GaAs and InP but also GaP and InAs), *ternary*, e.g., (Ga, In)As and Ga(As, P), and *quaternary* epitaxial films containing controlled amounts of Ga, In, As, and P have been deposited by these processes. Combinations of gas mixtures, different substrates, and more complex reactors are required in these cases to achieve the desired film stoichiometries. These CVD processes have played an important role in fabricating the optoelectronic devices, e.g., lasers and detectors, that made long-distance optical communications possible. While they are still used, MOCVD processes are now supplanting these older CVD methods for depositing compound-semiconductor films.

Table 6-1
Thermal CVD Films and Coatings

Deposited material	Substrate	Input reactants	Deposition temperature (°C)	Crystallinity
Si	Single-crystal Si	$SiCl_2H_2$, $SiCl_3H$, or $SiCl_4 + H_2$	1050–1200	E
Si		$SiH_4 + H_2$	600–700	P
Ge	Single-crystal Ge	$GeCl_4$ or $GeH_4 + H_2$	600–900	E
GaAs	Single-crystal GaAs	$(CH_3)_3Ga + AsH_3$	650–750	E
InP	Single-crystal InP	$(CH_3)_3In + PH_3$	725	E
SiC	Single-crystal Si	$SiCl_4$, toluene, H_2	1100	P
AlN	Sapphire	$AlCl_3$, NH_3, H_2	1000	E
In_2O_3:Sn	Glass	In-chelate, $(C_4H_9)_2Sn(OOCH_3)_2$, H_2O, O_2, H_2	500	A
ZnS	GaAs, GaP	Zn, H_2S, H_2	825	E
CdS	GaAs, sapphire	Cd, H_2S, H_2	690	E
Al_2O_3	Si, cemented carbide	$Al(CH_3)_3 + O_2$, $AlCl_3$, CO_2, H_2	275–475 850–1100	A A
SiO_2	Si	$SiH_4 + O_2$, $SiCl_2H_2 + N_2O$	450	A
Si_3N_4	SiO_2	$SiCl_2H_2 + NH_3$	750	A
TiO_2	Quartz	$Ti(OC_2H_5)_4 + O_2$	450	A
TiC	Steel	$TiCl_4$, CH_4, H_2	1000	P
TiN	Steel	$TiCl_4$, N_2, H_2	1000	P
BN	Steel	BCl_3, NH_3, H_2	1000	P
TiB_2	Steel	$TiCl_4$, BCl_3, H_2	>800	P

Note: E = epitaxial; P = polycrystalline; A = amorphous.
Adapted from Refs. 1, 2, 3.

The previous examples are but a small sampling of the total number of film and coating deposition reactions which have been researched in the laboratory as well as developed for commercial applications. In Table 6-1 there is a brief listing of CVD processes for depositing metals, semiconductors, and assorted compounds. Only inorganic source gases are noted here. (Metalorganic precursors and processes used to deposit these and other film materials will be discussed later.) The particular entries are culled from the review articles given earlier where specific details on process variables can be found.

In carefully examining the foregoing categories of CVD reactions, the discerning reader will note the following two common features:

1. All of the chemical reactions can be written in the simplified generalized form

$$aA_{(g)} + bB_{(g)} \rightarrow cC_{(s)} + dD_{(g)} \tag{6-16}$$

where A, B ... refer to the chemical species and $a, b \ldots$ to the corresponding stoichiometric coefficients. A single solid and mixture of gaseous species categorizes each heterogeneous reaction.

2. Some reactions are reversible, and this suggests that standard concepts of chemical thermodynamics may prove fruitful in analyzing them. Other reactions occur far from thermodynamic equilibrium and are strongly driven toward decomposition and film deposition.

There is a further distinction between chemical vapor *deposition* and chemical vapor *transport* reactions that should be noted. In the former, one or more gaseous species enter the reactor from gas tanks or liquid bubbler sources maintained outside the system. The reactants then combine at the hot substrate to produce the solid film. In chemical vapor transport reactions, solid or liquid sources are contained within either closed or open reactors. In this case externally introduced carrier or reactant gases flow over the sources, lifting them into the vapor stream where they are transported along the reactor. Subsequently, deposition of solid from the gas phase occurs at the substrates. Both chemical vapor deposition and transport reactions are, however, described by the same type of chemical reaction. As far as thermodynamic analyses are concerned, no further distinction will be made between them, and the generic term CVD will be used for both. We now turn our attention to the subject of thermodynamics.

6.3 THERMODYNAMICS OF CVD

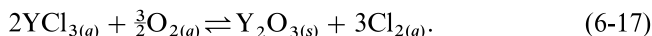
6.3.1 Reaction Feasibility

Thermodynamics addresses a number of important issues with respect to CVD. The question of whether a given chemical reaction is feasible or not is perhaps the most important of these. Once it is decided that a reaction is possible, thermodynamic calculation can frequently provide information on the partial pressures of the involved gaseous species and the direction of transport in the case of reversible reactions. Importantly, it provides an upper limit of what to expect under specified conditions. Thermodynamics does not, however, address questions related to the speed of the reaction and

resulting film growth rates. Indeed, processes which are thermodynamically possible frequently proceed at such low rates because of both vapor transport kinetics and vapor–solid reaction limitations that they are unfeasible in practice. Furthermore, the use of thermodynamics implies that chemical equilibrium has been attained. Although this may occur in a closed system, it is generally not the case in an open or flow reactor where gaseous reactants and products are continuously introduced and removed. In general, CVD may be viewed as an empirical science with thermodynamic guidelines.

Provided that the free-energy change, ΔG, can be approximated by the *standard* free-energy change, ΔG°, many simple consequences of thermodynamics with respect to CVD can be understood. For example, consider the selection of suitable chemical reactions in order to grow single-crystal films. In this case, it is essential that a single nucleus form as an oriented seed for subsequent growth. According to elementary nucleation theory, a small negative value of ΔG_v, the chemical free energy per unit volume, is required to foster a low nucleation rate of large critical-sized nuclei (Section 1.7). This, in turn, would require a ΔG° value close to zero. When this happens, large amounts of reactants and products are simultaneously present. If ΔG_v (i.e., ΔG°) were large and negative, however, the likelihood of a high rate of heterogeneous nucleation, or even homogeneous nucleation of solid particles within the gas phase, would be enhanced. The large driving force for chemical reaction tends to promote polycrystal formation in this case.

As an illustrative example it is worthwhile to follow the thought processes involved in the design of a CVD reaction to grow crystalline Y_2O_3 films. Following the treatment by Laudise (Ref. 13) consider the reaction

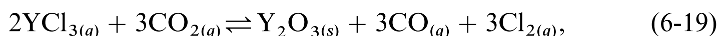


$$2YCl_{3(g)} + \tfrac{3}{2}O_{2(g)} \rightleftharpoons Y_2O_{3(s)} + 3Cl_{2(g)}. \quad (6\text{-}17)$$

At 1000 K, $\Delta G^\circ = -59.4$ kcal/mol, corresponding to an equilibrium constant (K_{eq}) given by $\log K_{eq} = +13$. The reaction is thus too far to the right for practical film growth. If the chloride is replaced by a bromide or iodide, the situation will be worse. YBr_3 and YI_3 are expected to be less stable than YCl_3, making ΔG° even more negative. The situation is improved by adding a gas-phase reaction with a positive value of ΔG°, e.g.,

$$CO_{2(g)} \rightleftharpoons CO_{(g)} + \tfrac{1}{2}O_{2(g)}; \qquad \Delta G^\circ = +46.7\ \text{kcal}. \quad (6\text{-}18)$$

Thus, the possible overall reaction is now



$$2YCl_{3(g)} + 3CO_{2(g)} \rightleftharpoons Y_2O_{3(s)} + 3CO_{(g)} + 3Cl_{2(g)}, \quad (6\text{-}19)$$

for which $\Delta G^\circ = -59.4 + 3(46.7) = +80.7$ kcal/mol. The equilibrium now falls too far to the left, but substituting YBr_3 and Br_2 for YCl_3 and Cl_2

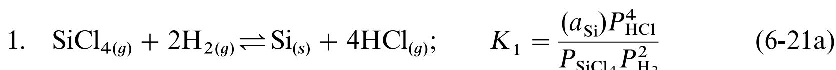
changes the sign of ΔG° once again. Thus for

$$2YBr_{3(g)} + 3CO_{2(g)} \rightleftharpoons Y_2O_{3(s)} + 3CO_{(g)} + 3Br_{2(g)}, \tag{6-20}$$

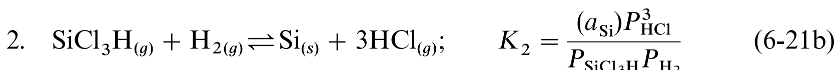
$\Delta G^\circ = -27$ kcal/mole. Although a value of ΔG° closer to zero would be more desirable, this reaction yields partial pressures of YBr_3 equal to 10^{-2} atm when the total pressure is 2 atm. Growth in other systems has occurred at such pressures. Pending availability of YBr_3 in readily volatile form and questions related to the operating temperature, reaction rates, and safety of gas handling, Eq. 6-20 appears to be a potential candidate for successful film growth. For analysis of chemical reactions good values of thermodynamic data are essential. Several sources of this information are listed among references E in Chapter 1.

6.3.2 Conditions of Equilibrium

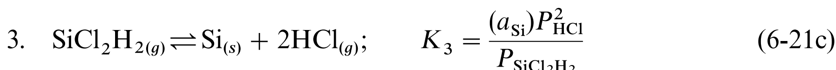
Thermodynamics can provide us with much more than a prediction of whether a reaction will proceed or not. Under certain circumstances, it can yield quantitative information on the operating intensive variables which characterize the equilibrium. The problem is to evaluate the partial pressures or concentrations of the involved species within the reactor given the reactant compositions and operating temperature. In practice, the calculation is frequently more complicated than initially envisioned because *in situ* mass-spectroscopic analysis of operating reactors has surprisingly revealed the presence of unexpected species which must be accounted for. For example, in the technologically important deposition of epitaxial Si films on silicon wafer substrates, no fewer than eight gaseous compounds have been identified during the reduction of chlorosilanes. They are part of the much-studied Si–Cl–H system, and the following example illustrates the method of calculation (Refs. 14, 15). The most abundant chemical species in this system are $SiCl_4$, $SiCl_3H$, $SiCl_2H_2$, $SiClH_3$, SiH_4, $SiCl_2$, HCl, and H_2. Partial pressures of these eight gaseous species are connected by the following six equations of chemical equilibrium.



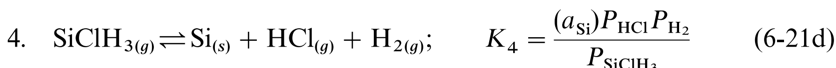
1. $$SiCl_{4(g)} + 2H_{2(g)} \rightleftharpoons Si_{(s)} + 4HCl_{(g)}; \qquad K_1 = \frac{(a_{Si})P^4_{HCl}}{P_{SiCl_4}P^2_{H_2}} \tag{6-21a}$$



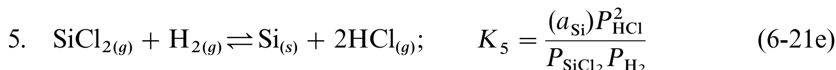
2. $$SiCl_3H_{(g)} + H_{2(g)} \rightleftharpoons Si_{(s)} + 3HCl_{(g)}; \qquad K_2 = \frac{(a_{Si})P^3_{HCl}}{P_{SiCl_3H}P_{H_2}} \tag{6-21b}$$



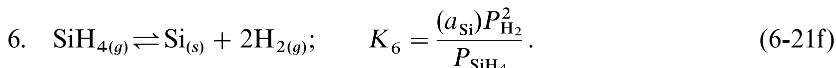
3. $$SiCl_2H_{2(g)} \rightleftharpoons Si_{(s)} + 2HCl_{(g)}; \qquad K_3 = \frac{(a_{Si})P^2_{HCl}}{P_{SiCl_2H_2}} \tag{6-21c}$$



$$4.\quad SiClH_{3(g)} \rightleftharpoons Si_{(s)} + HCl_{(g)} + H_{2(g)}; \qquad K_4 = \frac{(a_{Si})P_{HCl}P_{H_2}}{P_{SiClH_3}} \tag{6-21d}$$



$$5.\quad SiCl_{2(g)} + H_{2(g)} \rightleftharpoons Si_{(s)} + 2HCl_{(g)}; \qquad K_5 = \frac{(a_{Si})P_{HCl}^2}{P_{SiCl_2}P_{H_2}} \tag{6-21e}$$



$$6.\quad SiH_{4(g)} \rightleftharpoons Si_{(s)} + 2H_{2(g)}; \qquad K_6 = \frac{(a_{Si})P_{H_2}^2}{P_{SiH_4}}. \tag{6-21f}$$

Throughout the activity of solid Si, a_{Si} will be taken to be unity.

To solve for the eight partial pressures, two more equations relating these unknowns are required. The first specifies that the total pressure in the reactor, equal to the sum of the individual partial pressures, is fixed, e.g., at 1 atm. Therefore,

$$P_{SiCl_4} + P_{SiCl_3H} + P_{SiCl_2H_2} + P_{SiClH_3} + P_{SiH_4(g)} + P_{SiCl_2(g)} + P_{HCl(g)} + P_{H_2(g)} = 1. \tag{6-22}$$

The final equation involves the Cl/H molar ratio, which may be taken to be fixed if there is no net change in the concentrations of Cl and H. Therefore,

$$\frac{Cl}{H} = \frac{4P_{SiCl_4} + 3P_{SiCl_3H} + 2P_{SiCl_2H_2} + 2P_{SiCl_2} + P_{SiClH_3} + P_{HCl}}{2P_{H_2} + P_{SiCl_3H} + 2P_{SiCl_2H_2} + 3P_{SiClH_3} + P_{HCl} + 4P_{SiH_4}}. \tag{6-23}$$

The numerator represents the total amount of Cl in the system and is equal to the sum of the Cl contributed by each species. For example, the mass of Cl in $SiCl_4$ is given by $m_{Cl} = 4M_{Cl}(m_{SiCl_4}/M_{SiCl_4})$ where m and M refer to the mass and molecular weight, respectively. But, by the perfect gas law,

$$(m_{SiCl_4}/M_{SiCl_4}) = P_{SiCl_4}V/RT.$$

Therefore, the number of moles of $Cl = m_{Cl}/M_{Cl}$ or $4P_{SiCl_4}V/RT$, and similarly for all other terms in the numerator and denominator. The common factor V/RT, involving the volume V and the temperature T of the reactor, cancels out and all that is left is given by Eq. 6-23.

There are now eight independent equations relating the eight unknown partial pressures which can be determined, at least in principle. First, however, the individual equilibrium constants K_i must be specified, and this requires a slight excursion requiring additional thermodynamic calculation. K_i is fixed by specifying T and ΔG°. A convenient summary of thermodynamic data in the Si–Cl–H system is given in Fig. 6-5 where the free energies of compound formation are plotted versus temperature in an Ellingham-type diagram. Each line represents the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, from which ΔH°, ΔS°, and ΔG° can be calculated at any temperature for the particular compound in question. For example, consider the

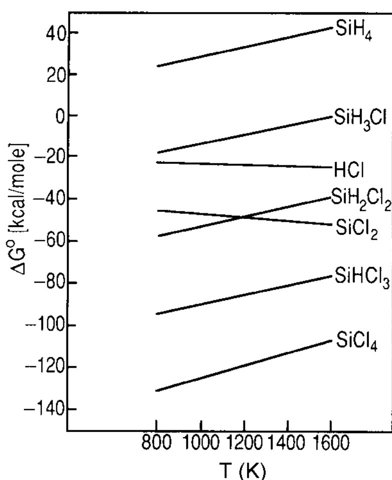
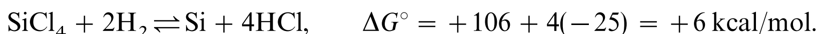


Figure 6-5 Free energies of formation of important gaseous species in the Si–Cl–H system in the temperature range 800–1600 K. (Reprinted with permission from the publisher, The Electrochemical Society, from Ref. 14.)

formation reactions for $SiCl_4$ and HCl at 1500 K. From Fig. 6-5,

$$Si + 2Cl_2 \rightleftharpoons SiCl_4, \qquad \Delta G^\circ = -106\ \text{kcal/mol}$$

$$\tfrac{1}{2}H_2 + (1/2)Cl_2 \rightleftharpoons HCl, \qquad \Delta G^\circ = -25\ \text{kcal/mol}$$



$$SiCl_4 + 2H_2 \rightleftharpoons Si + 4HCl, \qquad \Delta G^\circ = +106 + 4(-25) = +6\ \text{kcal/mol.}$$

Therefore $K_1 = \exp -6000/(1.99)1500 = 0.13$, and similarly for other values of K.

The results of the calculation are shown in Fig. 6-6 for the case of a molar ratio of [Cl]/[H] = 0.01, which is typical of conditions used for epitaxial deposition of Si. Through application of an equation similar to 6-23, the molar ratio of [Si]/[Cl] was obtained and is schematically plotted in the same figure. A reactor operating temperature in the vicinity of 1400 K is suggested, because as a result of film deposition the Si content in the gas phase is minimized. Such temperatures are employed in practice. Analogous calculations have also been made for the case where [Cl]/[H] = 0.1, which is typical of conditions favoring deposition of polycrystalline Si. At equivalent temperatures the [Si]/[Cl] ratios are somewhat higher than obtained for epitaxial deposition, reflecting the greater Si gas concentration operative during polycrystal growth. In both cases hydrogen is by far the most abundant species in the gas phase.

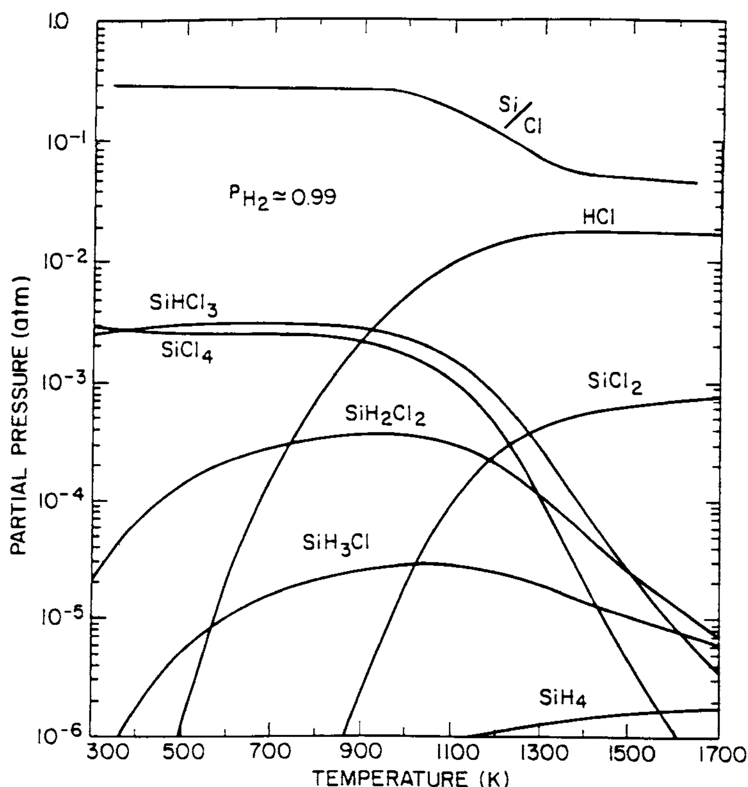


Figure 6-6 Equilibrium compositions of the $SiCl_4/H_2$ gas phase as a function of reactor temperature. Total pressure = 1 atm, Cl/H = 0.01. (Reprinted with permission from the publisher, The Electrochemical Society, from Ref. 14.)

Now that a very laborious, traditional thermodynamic analysis of this multicomponent system has been presented, the reader will be relieved to know that there are now far simpler ways to compute the same results. In fact, by using the Outokumpu HSC Chemisty, computer-based thermodynamics database and computation software (Outokumpu Research Oy, P.O. Box 60, Fin-28101 Pori, Finland), the results of Fig. 6-6 can be reproduced in a matter of minutes. As another example, by means of this or other similar software, the equilibrium partial pressures of gas species participating in the reduction of CH_4 by H_2 were determined as a function of temperature. The results are shown in Fig. 6-7 and have been used as a guide in the synthesis of diamond films (Ref. 16).

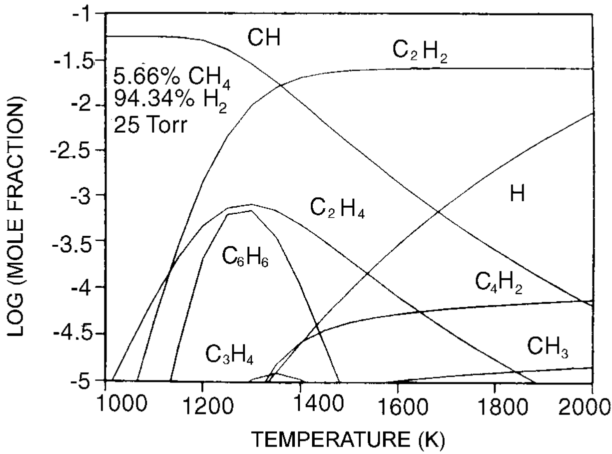


Figure 6-7 Methane/hydrogen equilibrium compositions. Total pressure = 25 torr, $CH_4/H_2 = 0.06$. (From S. O. Hay, W. C. Roman, and M. B. Colket, *J. Mater. Res.* **5**, 2387 (1990). Reprinted with the permission of the Materials Research Society and the authors.)

6.4 GAS TRANSPORT

6.4.1 INTRODUCTION

Gas transport is the process by which volatile species flow from one part of a reactor to another. It is important to understand gas transport phenomena in CVD systems for the following reasons:

1. The deposited film or coating thickness uniformity depends on the delivery of equal amounts of reactants to all substrate surfaces.
2. Rapid deposition growth rates are dependent on optimizing the flow of reactants through the system and to substrates.
3. More efficient utilization of generally expensive process gases can be achieved as a result.
4. Computer modeling of CVD processes will be more accurate enabling improved reactor design and better predictive capability with regard to performance.

At the outset it is important to distinguish between diffusion and bulk flow processes in gases. Diffusion involves the motion of individual atomic or molecular species, whereas in bulk transport processes such as viscous flow or convection, parts of the gas move as a whole. Different driving forces and resulting transport equations define and characterize these two broad

types of gas flow. Each of these will now be discussed briefly as a prelude to considering the flow combinations that take place in actual CVD reactors.

6.4.2 VISCOUS FLOW

The viscous flow regime is operative when gas transport occurs at pressures of roughly 0.01 atm and above in reactors of typical size. This is the pressure range characterisic of most CVD systems. At typical flow velocities of tens of centimeters per second, the reactant gases exhibit what is known as laminar or streamline flow. The theory of fluid mechanics provides a picture of what occurs under such circumstances. We shall consider the simplest of flow problems, i.e., that parallel to a flat plate. As shown in Fig. 6-8, the flow velocity has a uniform value v_0, but only prior to impinging on the leading edge of the plate. However, as flow progresses, velocity gradients must form because the gas clings to the plate. Far away the velocity is still uniform but drops rapidly to zero at the plate surface, creating a boundary layer. The latter grows with distance along the plate and has a thickness $\delta(x)$ given by $\delta(x) = 5x/(\mathrm{Re}_x)^{1/2}$, where Re_x is the Reynolds number defined as $\mathrm{Re}_x = v_0 \rho x/\eta$. The quantities η and ρ are the

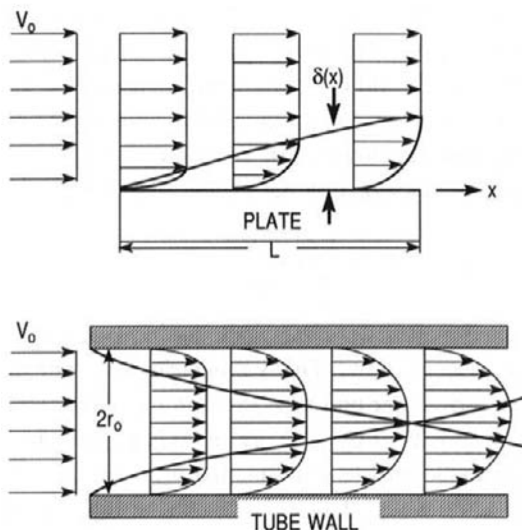


Figure 6-8 Laminar gas flow patterns. (Top) Flow across flat plate. (Bottom) Flow through circular pipe.

gas viscosity and density, respectively. More will be said about η, but here it should be noted that the viscosity essentially establishes the frictional viscous forces which decelerate the gas at the plate surface.

The average boundary layer thickness over the whole plate is

$$\bar{\delta} = 1/L \int_0^L \delta(x)\,dx = \frac{10}{3} L \sqrt{\frac{\eta}{\rho v_0 L}} = \frac{10}{3} \frac{L}{\sqrt{\mathrm{Re}_L}} \qquad (6\text{-}24)$$

where Re_L is defined as $\mathrm{Re}_L = v_0 \rho L/\eta$. Because both gaseous reactants and products must pass through the boundary layer separating the laminar stream and film deposit, low values of δ are desirable in enhancing mass-transport rates. This can be practically achieved by increasing the gas flow rate (v_0), which raises the value of Re. Typical values of Re in CVD reactors range up to a few hundred. If, however, Re exceeds approximately 2100, a transition from laminar to turbulent flow occurs. The resulting erratic gas eddies and swirls are not conducive to uniform, defect-free film growth and are to be avoided.

It is instructive to now consider gas flow through a tube of circular cross section. The initial uniform axial-flow velocity is altered after the gas enters the tube. Boundary layers develop at the walls and grow with distance along the tube as shown in Fig. 6-8. The Reynolds number is now given by $\sim 2v_0 \rho r_0/\eta$, where r_0 is the tube radius. Beyond a certain critical entry length, $L_e \simeq 0.07 r_0 \mathrm{Re}$, the flow is fully developed and the velocity profile no longer changes. At this point, the boundary layers around the tube circumference have merged and the whole cross section consists of "boundary layer." The resulting axial flow is now described by the Hagen–Poiseuille relation

$$\dot{V} = \frac{\pi r_0^4}{8\eta} \frac{\Delta P}{\Delta x} \qquad \mathrm{cm^3/s} \qquad (6\text{-}25)$$

where $\dot{V}$ is the volumetric flow rate and $\Delta P/\Delta x$ is the pressure gradient driving force for viscous flow. Defined as the volume of gas which moves per unit time through the cross section, the volumetric flow rate is related to the average gas velocity $\bar{v}$ by $\dot{V} = \pi r_0^2 \bar{v}$. Within the tube the gas velocity, $v(r)$, assumes a parabolic profile as a function of the radial distance r from the center given by $v(r) = v_{\max}(1 - r^2/r_0^2)$, where $v_{\max}$ is the maximum gas velocity. The gas flux J is given by the product of the concentration of the species in question and the velocity with which it moves:

$$J_i = C_i \bar{v}_i. \qquad (6\text{-}26)$$

Upon substitution of $C_i = P_i/RT$ from the perfect gas law, and $\bar{v}_i = \dot{V}/\pi r_0^2$,

$$J_i = \frac{P_i r_0^2 \Delta P_i}{RT 8\eta \Delta x}. \tag{6-27}$$

Provided the molar flux of any gaseous species in a chemical reaction is known, and equilibrium conditions prevail, the fluxes of other species can be determined from the stoichiometric coefficients.

Viscous flow is characterized by the coefficient of viscosity η. Kinetic theory of gases predicts that η varies with temperature as $T^{1/2}$ but is independent of pressure. Experimental data bear out the lack of a pressure dependence at least to several atmospheres, but indicate that η varies as T^n with n having values between 0.6 and 1.0. Gas viscosities measured in poise (P) typically range between 0.01 centipoise (cP) at 0°C and 0.1 cP at 1000°C. (1 P = 0.1 Pa-s = 1 dyne-s/cm^2.)

6.4.3 Diffusion in Gases

The phenomenon of diffusion applies to mass transport in gases as well as condensed phases. In the case of two different gases that are initially separated and then allowed to mix, each will interdiffuse and increase the entropy of the system. Elementary kinetic theory of gases predicts that the diffusivity in gases depends on pressure and temperature as $D \sim T^{3/2}/P$. It is therefore usual to represent D in gases by

$$D = D_0 \frac{P_0}{P}\left(\frac{T}{T_0}\right)^n \tag{6-28}$$

where n is experimentally found to be approximately $\sim$1.8. D_0, the value of D measured at standard temperature, T (273 K), and pressure, P (1 atm), is dependent on the particular gas combination in question. Typical D_0 values at temperatures of interest span the range 0.1–10 cm^2/s. Thus, diffusivity values in gases are many orders of magnitude higher than even the largest D values in solids. If the gas composition is reasonably dilute so that the perfect gas law applies for a given species i, $C_i = P_i/RT$ and Eq. 1-22 can be equivalently expressed by

$$J_i = -\frac{D}{RT}\frac{dP_i}{dx}. \tag{6-29}$$

This formula can be applied to the diffusion of gas through the stagnant boundary layer of thickness δ adjacent to the substrate. The flux is then

given by

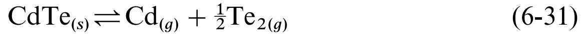
$$J_i = -\frac{D(P_g - P_s)}{\delta RT}, \tag{6-30}$$

where P_g is the vapor pressure in the bulk gas and P_s is the vapor pressure at the substrate surface.

Since D varies inversely with pressure, gas mass-transfer rates can be enhanced by reducing the pressure in the reactor. Advantage of this fact is taken in low-pressure CVD (LPCVD) systems, which are now extensively employed in semiconductor processing. Their operation will be discussed later in Section 6.6.3.

6.4.3.1 Close-Spaced Vapor Transport

As an example that integrates both thermodynamics and diffusion in a CVD process, consider the deposition of CdTe films by close-spaced vapor transport (CSVT) (Ref. 17). In this process mass is transferred from a solid CdTe source at temperature T_1 located a very short distance l (typically 1 mm) from the substrate maintained at T_2 ($T_1 > T_2$). The present objective is to establish conditions necessary to derive an expression for the film growth rate. It is assumed that chemical equilibrium prevails at the respective temperatures. The basic reaction is



$$\mathrm{CdTe}_{(s)} \rightleftharpoons \mathrm{Cd}_{(g)} + \tfrac{1}{2}\mathrm{Te}_{2(g)} \tag{6-31}$$

for which $\Delta G^\circ = 68.64 - 44.94 \times 10^{-3}T$ kcal/mol. Therefore, the two equations

$$P_{\mathrm{Cd}}(T_1)P_{\mathrm{Te}_2}^{1/2}(T_1) = \exp - \frac{\Delta G^\circ(T_1)}{RT_1} = K(T_1) \tag{6-32a}$$

and

$$P_{\mathrm{Cd}}(T_2)P_{\mathrm{Te}_2}^{1/2}(T_2) = \exp - \frac{\Delta G^\circ(T_2)}{RT_2} = K(T_2) \tag{6-32b}$$

express the equilibria at source and substrate. If the concentrations of the gas phase species vary linearly with distance, the individual mass fluxes (in units of mol/cm^2-s) are expressed by

$$J_{\mathrm{Cd}} = \frac{D_{\mathrm{Cd}}}{l}\left\{\frac{P_{\mathrm{Cd}}(T_1)}{RT_1} - \frac{P_{\mathrm{Cd}}(T_2)}{RT_2}\right\} \tag{6-33a}$$

$$J_{\mathrm{Te}_2} = \frac{D_{\mathrm{Te}_2}}{l}\left\{\frac{P_{\mathrm{Te}_2}(T_1)}{RT_1} - \frac{P_{\mathrm{Te}_2}(T_2)}{RT_2}\right\}. \tag{6-33a}$$

Note the use of the perfect gas law and the neglect of the temperature dependence of D. Maintenance of stoichiometry requires that

$$J_{Cd} = 2J_{Te_2}, \qquad (6\text{-}34)$$

the factor of 2 arising because Te_2 is a dimer. The film growth rate is obtained by the relation

$$\dot{G}(\mu m/min) = (60 \times 10^4) J_{Cd} M_{CdTe}/\rho \qquad (6\text{-}35)$$

where M and ρ are the molar mass and density of CdTe, respectively. If T_1 exceeds T_2 by approximately 100°C, then $P_i(T_1) \gg P_i(T_2)$ where i refers to both Cd and Te_2. By neglecting the T_2 terms in Eqs. 6-33a and 6-33b,

$$\frac{P_{Cd}(T_1)}{P_{Te_2}(T_1)} = \frac{2D_{Te_2}}{D_{Cd}} = 1.1. \qquad (6\text{-}36)$$

The value of 1.1 is derived from kinetic theory of gases, which suggests that $D_{Cd} = 1.85 D_{Te_2}$ in H_2, or Ar ambients. Equations 6-32, 6-33, 6-34, and 6-36 enable all the partial pressures to be determined. By knowing the value of either D_{Cd} or D_{Te_2}, $\dot{G}$ may then be determined.

The technique of CSVT for semiconductor film growth has been reviewed by Perrier and Philippe (Ref. 18) and more recently applied to epitaxial SiC film growth (Ref. 19). Growth conditions in this study included source temperatures over 1900°C, 1–50 torr, a temperature gradient of 3.5°C/mm, and a source–substrate spacing of 1.5 mm.

6.4.4 CONVECTION

Convection is a bulk gas-flow process that can be distinguished from both diffusion and viscous flow. Whereas gas diffusion involves the statistical motion of atoms and molecules driven by concentration gradients, convection arises from the response to gravitational, centrifugal, electric, and magnetic forces. It is manifested in CVD reactors when there are vertical gas-density or temperature gradients. An important example occurs in cold-wall reactors such as depicted in Fig. 6-15 where heated susceptors are surrounded above as well as on the sides by the cooler walls. Cooler, more dense gases then lie above hotter, less dense gases. The resultant convective instability causes an overturning of the gas by buoyancy effects. Subsequently, a complex coupling of mass and heat transfer serves to reduce both density and temperature gradients in the system. Another example of convective flow occurs in two-temperature-zone vertical reactors. In the disproportionation process considered previously in Fig. 6-4, it is immaterial

whether the hotter zone is physically located above or below the cooler zone insofar as thermodynamics is concerned. But efficient gas flow considerations mandate the placement of the cooler region on top to enhance gas circulation.

It is important to note that film growth is limited by viscous, diffusive, and convective mass transport fluxes which, in turn, are driven by gas pressure gradients. In open reactors the metered gas (volumetric) flow rates establish these pressure gradients. In closed reactors the latter arise because of imposed temperature differences which locally alter equilibrium partial pressures.

6.4.5 Modeling Complex Gas Transport Effects

When gas transport is physically more complex because of combined flows in three dimensions, the fundamental equations of fluid dynamics become the starting point of the analysis. Although there are standard treatments of the mathematical theory of transport phenomena, the texts by Lee (Ref. 20) and Middelman and Hochberg (Ref. 21) are recommended because of their chemical-engineering approach to CVD processing. Without delving into a detailed formal analysis it is instructive to note some of the features of the involved formidable partial differential equations. Three basic equations describing the conservation of mass, momentum, and energy provide the foundation of the subject. They are simply enumerated with brief verbal descriptions:

1. *Continuity.* The conservation of mass requires that the net rate of mass accumulation in a region be equal to the difference between the rate of mass flow in and out.
2. *Navier–Stokes.* Momentum conservation requires that the net rate of momentum accumulation in a region be equal to the difference between the rate of momentum in and out, plus the sum of forces acting on the system.
3. *Energy.* The rate of accumulation of internal and kinetic energy in a region is equal to the net rate of internal and kinetic energy in by convection, plus the net rate of heat flow by conduction, minus the rate of work done by the fluid.

In multicomponent fluid systems there are additional equations that describe concentration changes due to thermal and mass transport, as well as chemical reactions involving each of the species.

In principle, all of the above coupled equations together with the boundary conditions are sufficient to describe gas transport in reactors.

However, in practice exact solutions are rare and only exist in the simplest cases, e.g., the isothermal tubular reactor of Sections 6.4.2. For this reason numerical methods based on finite-element analysis are often used to analyze gas flow in reactors of complex geometry where steep temperature gradients exist. Before starting, however, it is often found to be very useful to transform the equations into dimensionless forms by scaling the variables and properties that describe gas behavior relative to reference values. As a result, certain dimensionless numbers that may be interpreted as a ratio of the magnitude of two physical phenomena emerge in the equations.

These as well as other useful dimensionless groups are defined in Table 6-2 together with their interpretations and typical magnitudes in atmospheric pressure (APCVD) and low-pressure (LPCVD) reactors (Refs. 22, 23). Although only Damkohler numbers Da_g and Da_s specifically refer to chemical reaction rates among species within the gas or at the substrate surface, the other dimensionless groups describe gas transport in both multicomponent and homogeneous systems. In particular, the Peclet, Grashof/Rayleigh, Reynolds, Damkohler, and Knudsen numbers are important descriptors of CVD reactor processes. For example, when the Peclet number is large, species transfer is mainly due to convection, and downstream reaction products and impurities will not diffuse back to the reaction zone; for small Peclet numbers, transport due to diffusion dominates the flow. Large Grashof/Rayleigh numbers favor natural convection which tends to disturb the flow field. When the surface Damkohler number is large the film deposition process will be diffusion limited (see Section 6.5.3), but when the reverse is true surface-reaction limited deposition occurs. If the gas-phase Damkohler number is large, chemical reactions will proceed rapidly to establish thermodynamic equilibrium within the gas; but for small Damkohler numbers the residence time in the reactor is too short for gas-phase reactions to occur and the system will be far from equilibrium.

The following computer simulation of gas flow in a cylindrically symmetrical cold-wall, single-wafer reactor due to Kleijn (Ref. 23) provides an instructive application. This LPCVD reactor, containing radially impinging gas jets, was used to develop a process for depositing tungsten on 20-cm diameter silicon wafers through reduction of WF_6 by H_2. Subsequent figures depicted in Figs. 6-9a, 6-9b, and 6-9c illustrate the interplay of the dimensionless groups Re and Gr, at fixed Pr ($Pr = 0.7$) and Ga ($Ga = 0.8$) numbers. The resultant gas flow is expressed in terms of streamlines (Ψ). Curves of Ψ ranging from 0 to 1 represent the paths followed by “particles” of fluid, while gradients normal to the stream flow lines are measures of the gas velocity. Corresponding gas isotherms are also depicted. From these figures it is clear that increasing Gr at constant

Table 6-2
Dimensionless Parameter Groups in CVD

Name	Definition	Physical interpretation	Typical magnitude	
			APCVD	LPCVD
Knudsen	$\mathrm{Kn} = \lambda/L$	Ratio of gas mean free path to characteristic length	10^{-6}–10^{-5}	10^{-3}–10^{-2}
Prandtl	$\mathrm{Pr} = C_p\eta/K$	Ratio of momentum diffusivity to thermal diffusivity	~0.7	~0.7
Schmidt	$\mathrm{Sc} = C_p\eta/D$	Ratio of momentum diffusivity to mass diffusivity	1–10	1–10
Reynolds	$\mathrm{Re} = \rho vL/\eta$	Ratio of inertia forces to viscous forces	10^{-2}–10^{2}	10^{-2}–10^{2}
Peclet (mass)	$\mathrm{Pe_m} = \mathrm{ReSc}$	Ratio of convective mass flux to diffusive mass flux	10^{-1}–10^{3}	10^{-1}–10^{3}
Grashof (thermal)	$\mathrm{Gr_t} = \dfrac{g\rho^2 L^3 \Delta T}{\eta^2 T_r}$	Ratio of buoyancy force to viscous force	10^{2}–10^{7}	0–10
Rayleigh	$\mathrm{Ra} = \mathrm{GrPr}$	Ratio of buoyancy force to viscous force	10^{2}–10^{7}	0–10
Damkohler (gas phase)	$\mathrm{Da_g} = \dfrac{\dot{R}_g L}{C_{in} v}$	Ratio of chemical reaction rate to bulk flow rate	10^{-3}–10^{3}	10^{-3}–10^{3}
Damkohler (surface)	$\mathrm{Da_s} = \dfrac{\dot{R}_s L}{C_{in} D}$	Ratio of chemical reaction rate to diffusion rate	10^{-3}–10^{3}	10^{-3}–10^{3}
Arrhenius	$\mathrm{Arrh} = \dfrac{E}{RT_r}$	Ratio of activation energy to potential energy	0–100	0–100
Gay–Lussac	$\mathrm{Ga} = \Delta T/T_r$	Ratio of temperature difference to reference temperature	1–1.3	0.6–1

Note: L = reactor dimension (m), C_p = specific heat (J/kg-K), η = viscosity (Pa-s), K = thermal diffusivity (m^2/s), g = gravitational constant, $\Delta T = T_{hot} - T_{cold}$, T_r = reference temperature, $\dot{R}_g$ = gas reaction rate (mol/m^3-s), $\dot{R}_s$ = surface reaction rate (mol/m^2-s), C_{in} = input gas concentration (mol/m^3).
Adapted from Refs. 22, 23.

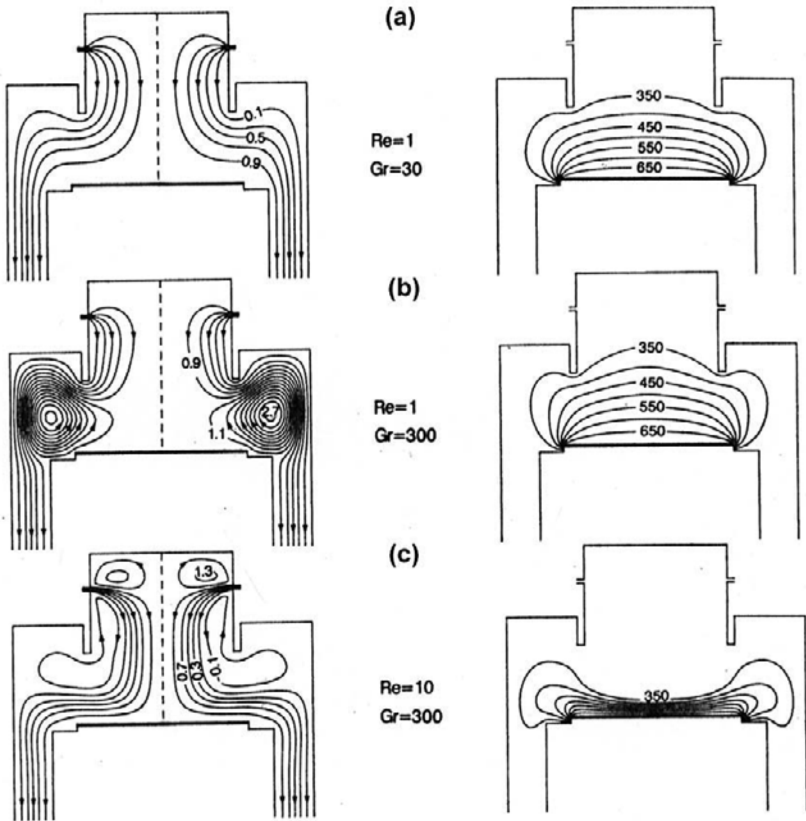


Figure 6-9 Gas streamlines (left) over the range 0–1, and corresponding isotherms in K (right) for different combinations of Reynolds and Grashof numbers. The reactor profile is outlined. (a) Re = 1, Gr = 30; (b) Re = 1, Gr = 300; (c) Re = 10, Gr = 300. (From C. R. Kleijn, "Transport Phenomena in Chemical Vapor Deposition Reactors." Ph.D. Thesis, Technical University of Delft, 1991. Reprinted with permission.)

Re causes buoyancy-induced recirculation of gas near the edge of the heated susceptor, which then rises above it and finally falls at the cold reactor wall. However, the temperature profiles are relatively unaffected. Increasing Re suppresses the buoyancy-induced gas recirculation, produces inertial gas rolls, and creates a relatively thin thermal boundary layer above the wafer (horizontal). Clearly such simulations have the potential for suggesting reactor conditions that minimize undesirable buoyancy and inertial rolls and promote uniform flow across the wafer. Streamlining the reactor is effective in the latter regard.

6.5 FILM GROWTH KINETICS

The growth kinetics of CVD films is dependent on a number of factors associated with transport of reactants to the substrate surface and the subsequent heterogeneous surface reactions that lead to film formation. Intimate microscopic details of these reactions are often unknown and, therefore, the growth kinetics is frequently modeled in macroscopic terms. This is a simpler approach since it makes no atomistic assumptions, yet is capable of predicting deposit growth rates and uniformity.

6.5.1 Axial Growth-Rate Uniformity

In this and the next section we treat the analysis of film growth on substrates placed in a horizontal reactor. We first consider the film thickness deposition rate on substrates lying parallel to the reactor axis, as a function of distance (Ref. 24). For the reactor configuration shown in Fig. 6-10a, the following is assumed:

1. The whole system is at constant temperature.
2. Along the reactor axis the gas has a constant velocity component.
3. The reactor extends a large distance in the z direction so that the problem reduces to one of two dimensions.

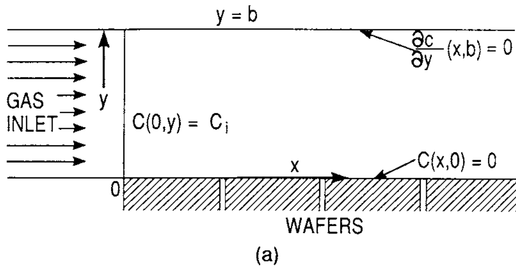
Furthermore, the flow is simply treated by assuming the mass flux $\mathbf{J}$ vector at any point to be composed of two terms involving the gas phase concentration of the depositing species, $C(x, y)$,

$$\mathbf{J}(x, y) = C(x, y)\bar{v} - D\nabla C(x, y). \tag{6-37}$$

Whereas the first represents a bulk (plug) flow where the source gas moves as a whole with drift velocity $\bar{v}$, the second term is due to diffusion of individual gas molecules, with diffusivity D, along concentration gradients. By taking the difference of the mass flux into and out of an elemental volume and equating it to the mass accumulation we obtain

$$\frac{\partial C(x, t)}{\partial t} = D\left\{\frac{\partial^2 C(x, y)}{\partial x^2} + \frac{\partial^2 C(x, y)}{\partial y^2}\right\} + \bar{v}\,\frac{\partial C(x, y)}{\partial x}. \tag{6-38}$$

Only the steady-state solutions are of concern so that $\partial C(x, t)/\partial t = 0$. The resulting equation is then subject to three conditions noted in Fig. 6-10a. Boundary value problems of this kind are readily solved by the well-known



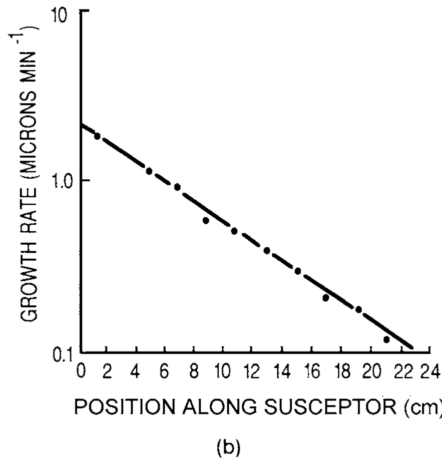


Figure 6-10 (a) Horizontal reactor geometry with conditions for flow. (b) Variation of growth rate with position along susceptor. Reactor conditions: $\bar{v} = 7.5$ cm/s, $b = 1.4$ cm, $T = 1200°C$, and $C_i = 3.1 \times 10^{-5}$ g/cm^3. (From P. C. Rundle, *Int. J. Electron.* **24**, 405, © 1968 Taylor and Francis, Ltd.)

techniques of partial differential equations involving separation of variables or Laplace-transform methods. The solution is

$$C(x, y) = \frac{4C_i}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin\left[\frac{(2n+1)\pi y}{2b}\right]$$

$$\times \exp\{\bar{v}/D - [\bar{v}^2/D^2 + (2n+1)^2\pi^2/b^2]^{1/2}\}x. \qquad (6\text{-}39)$$

To obtain a more readily usable form of the solution, it is assumed as a first approximation that $\bar{v}b > D\pi$. Except for short distances into the reactor or small values of x, only the first term in the series need be retained. These

simplifications yield

$$C(x, y) = \frac{4C_i}{\pi} \sin\left(\frac{\pi y}{2b}\right) \exp - \left(\frac{\pi^2 D x}{4\bar{v} b^2}\right), \tag{6-40}$$

a solution whose concentration dependence is sinusoidal in the y direction and decays exponentially along x. The flux of source gas to the substrate surface is defined by

$$J(x) = D \frac{\partial C(x, y)}{\partial y}\bigg|_{y=0} \qquad (\text{g/cm}^2\text{-s}). \tag{6-41}$$

What we are after is the resultant deposit growth rate, $\dot{G}(x)$. Aside from simple material constants, this quantity is directly proportional to $J(x)$ by $\dot{G}(x) = MJ(x)/\rho M_s$, where M and M_s are the molecular weights of the depositing species (Si) and source gas, respectively, and ρ is the film density. A combination of equations yields

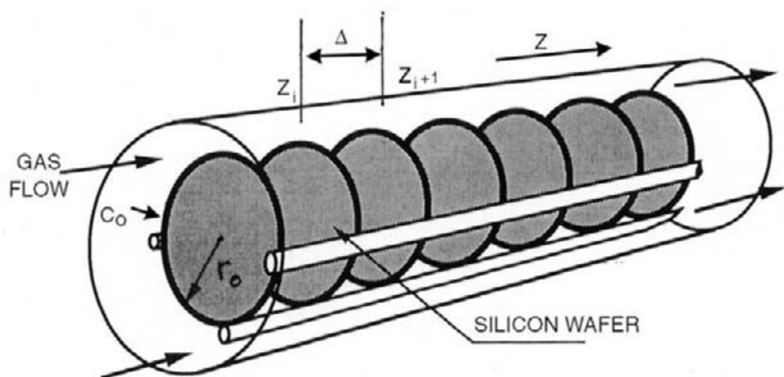
$$\dot{G}(x) = \frac{2C_i MD}{b\rho M_s} \exp - \frac{\pi^2 D x}{4\bar{v} b^2} \qquad (\text{cm/s}). \tag{6-42}$$

Values for D, $\bar{v}$, and C_i are strictly those pertaining to the mean temperature of the reactor, T. An exponential decay in the growth rate with distance along the reactor is predicted. This is not too surprising since the input gases are progressively depleted of reactants. The implicit boundary condition requiring that $C = 0$ at $x = \infty$ accounts for this loss. Despite the extreme simplicity of the assumptions, the model provides rather good agreement with experimental data on the variation of Si film growth rate with reactor distance as indicated in Fig. 6-10b. Although the specific reaction considered is the hydrogen reduction of chlorosilane, the results can be broadly applied to other CVD processes as well.

From the standpoint of reactor performance, high growth rates and uniform deposition are the two most important concerns, assuming film quality is not comprised. The equation for $\dot{G}(x)$ serves to provide design guidelines, but these are not always simple to implement. Earlier it was noted that the boundary layer increased in the axial direction. To compensate for this effect and improve growth uniformity, δ must be correspondingly reduced. Tilting the susceptor as shown in Fig. 6-15 is an effective remedy because the velocity of gas flow in the tapered space above the substrates is forced to increase with distance. Commercial horizontal and barrel reactors are designed with tilted susceptors so that enhanced transport across the stagnant layer compensates for reactant depletion. Continuously increasing the temperature downstream within the reactor is another way to enhance deposition uniformity.

6.5.2 Radial Growth-Rate Uniformity on Wafers

In LPCVD reactors, silicon wafers are often stacked vertically along the tube axis and gases flow around and between them as shown schematically in Fig. 6-11a. The objective is to deposit polycrystalline and amorphous silicon films. In such cases it is not obvious that film thickness uniformity can be maintained across the wafer surface. Hitchman and Jensen (Ref. 25) have treated this problem whose analysis is more difficult than the one just



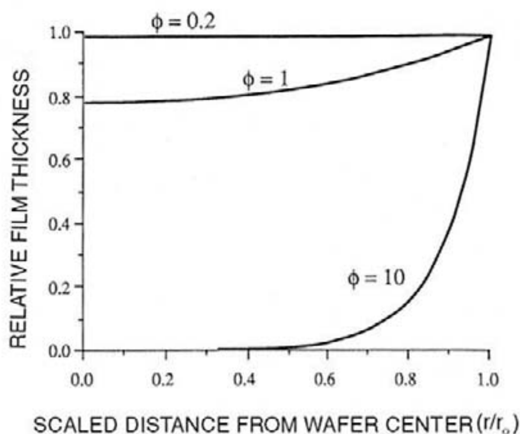


Figure 6-11 (a) Schematic representation of hot-wall, multiple wafer LPCVD reactor geometry with gas flow boundary conditions. (b) Film thickness variation as a function of the scaled radial distance along the wafer for different values of ϕ. (From *Chemical Vapor Deposition*, edited by M. L. Hitchman and K. F. Jensen. Reprinted with the permission of Academic Press, Ltd., and Professor K. F. Jensen, MIT.)

considered. The reason is due to the more complex geometries of the wafer and their vertical stacking. Different gas flow and reaction in the interwafer spacing is thus established relative to the annular region between the tube walls and the periphery of the wafers. In the annulus, viscous flow or more simply plug flow may be assumed as previously. However, we are primarily interested in the interwafer region, where we shall assume the deposited film thickness depends solely on radial diffusional transport driven by compositional variations in the gas phase. Therefore, the steady-state diffusion equation in cylindrical coordinates for the circular wafers is

$$D\left\{\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r\partial C(r,z)}{\partial r}\right)+\frac{\partial^2 C(r,z)}{\partial z^2}\right\}=0, \tag{6-43}$$

where r is the radial distance from the wafer center outward, and z is in the axial direction. The four boundary conditions are:

1. $\partial C/\partial r = 0$ at $r = 0$ (because flow is axiosymmetric across the wafer).
2. $C = C_0$ at $r = r_0$, where r_0 is the wafer radius. (At r_0, the gas concentration due to annular flow is joined to that of diffusion at the wafer edge.)
3. $D\partial C/\partial z = kC$ at $z = z_i$. At the back surface (z_i) of wafer i, the diffusional flux is consumed by a first-order chemical reaction whose rate constant is k.
4. $-D\partial C/\partial z = kC$ at $z = z_i + \Delta$, the front surface of wafer $i + 1$.

For a small wafer spacing (Δ) relative to r_0 (typically $\Delta/r_0 < 0.05$), we may replace differentials by averages so that $\partial^2 C/\partial z^2$ may be replaced by $\simeq (2/D\Delta)kC$, leaving

$$D\left[\frac{1}{r}\frac{d}{dr}\left(\frac{rdC(r)}{dr}\right)\right]+\frac{2kC(r)}{\Delta}=0. \tag{6-44}$$

The corresponding deposit growth rate on the wafer is again expressed by $\dot{G} = MJ/\rho M_s$, where now $J = kC(r)$. Exact $C(r)$ vs r solutions of Eq. 6-44 subject to boundary conditions 1 and 2 are in the form of incomplete Bessel functions of zero order, I_0. A comparison of the film thickness (d) as a function of r relative to that at the wafer edge is instructive and given by

$$\frac{d(r/r_0)}{d(r=r_0)}=\frac{C(r)}{C_0}=\frac{I_0(\sqrt{2r_0^2k/\Delta D}\,[r/r_0])}{I_0(\sqrt{2r_0^2k/\Delta D})}. \tag{6-45}$$

Defining the dimensionless quantity $\phi = \sqrt{2r_0^2k/\Delta D}$, when $\phi^2 \ll 1$ diffusion is rapid compared to surface reaction and the depositing species is uniformly distributed in the interwafer space; such conditions yields a uniform deposit. On the other hand when $\phi^2 \gg 1$, diffusion cannot keep up

with the surface reaction and the precursor gas is depleted at the wafer center, resulting in a thinner deposit; a "bull's-eye" effect now characterizes the nonuniform film thickness. These results are graphically illustrated in Fig. 6-11b.

6.5.3 Temperature Dependence

In order to understand the effect of temperature on film growth it is instructive to reproduce the classic treatment of film growth by Grove (Ref. 26). The essentials of this simple model are noted in Fig. 6-12 where the environment in the vicinity of the gas/growing film interface is shown. A drop in reactant concentration occurs from C_g in the bulk gas to C_s at the interface. The corresponding mass flux is given by

$$J_{gs} = h_g(C_g - C_s) \tag{6-46}$$

where h_g is the gas-phase mass-transfer coefficient to be defined later. The flux consumed by the reaction taking place at the surface of the growing film is approximated by

$$J_s = k_s C_s \tag{6-47}$$

where first-order kinetics are assumed, and k_s is the rate constant for surface

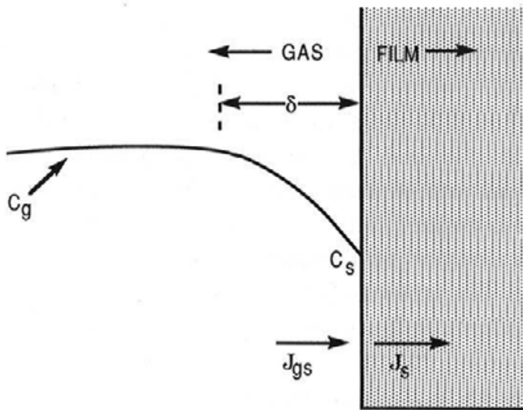


Figure 6-12 Model of CVD growth process. Gas flows normal to plane of paper. (From Ref. 26. Reprinted with permission from John Wiley & Sons, copyright © 1967.)

reaction. In the steady state $J_{gs} = J_s$ and therefore,

$$C_s = \frac{C_g}{1 + k_s/h_g}. \tag{6-48}$$

This formula predicts that the surface concentration drops to zero if $k_s \gg h_g$, a condition referred to as *mass-transfer* control. In this case slow gas transport through the boundary layer limits the otherwise rapid surface reaction. Conversely, *surface-reaction* control dominates when $h_g \gg k_s$, in which case C_s approaches C_g. Here the surface reaction is sluggish even though sufficient reactant gas is available. The film growth rate, $\dot{G}$, is given by $\dot{G} = J_s/N_0$, where N_0 is the atomic density or number of atoms incorporated into the film per unit volume. Therefore,

$$\dot{G} = \frac{k_s h_g C_g}{(k_s + h_g)N_0}. \tag{6-49}$$

The temperature dependence of $\dot{G}$ hinges on the properties of k_s and h_g. A Boltzmann-like behavior dominates the temperature dependence of the surface chemistry, that is, $k_s \sim \exp - (E/RT)$ where E is the characteristic activation energy involved. Comparison of Eqs. 6-30 and 6-46 reveals that h_g is related to D/δ. Since D_g varies as T^2 at most, and δ is weakly dependent on T, h_g is relatively insensitive to variations in temperature. At low temperatures, film growth is surface-reaction controlled, that is, $\dot{G} = k_s C_g/N_0$, whereas at high temperatures the gas mass-transfer or diffusion-controlled regime is dominant where $\dot{G} = h_g C_g/N_0$. The predicted behavior is borne out by growth rate data for epitaxial Si as shown in the Arrhenius plots of Fig. 6-13. Actual film growth processes are carried out in the gas diffusion-controlled region where the temperature response is relatively flat. At lower temperatures the same activation energy of about 1.5 eV is obtained irrespective of chlorosilane used. Migration of Si adatoms is interpreted to be the fundamental rate-limiting step in this temperature regime.

6.5.4 INFLUENCE OF THERMODYNAMICS

The previous discussion implies that all reaction rates increase with temperature. Although this is generally true, it is sometimes observed that higher reactor temperatures lead to lower film growth rates in certain systems. This apparent paradox can be explained by considering the reversibility of chemical reactions (Ref. (27). In Chapter 1 the net rate for a forward exothermic reaction (and reverse endothermic reaction) was given by Eq. 1-36 and modeled in Fig. 1-19. Recall that exothermic reactions mean

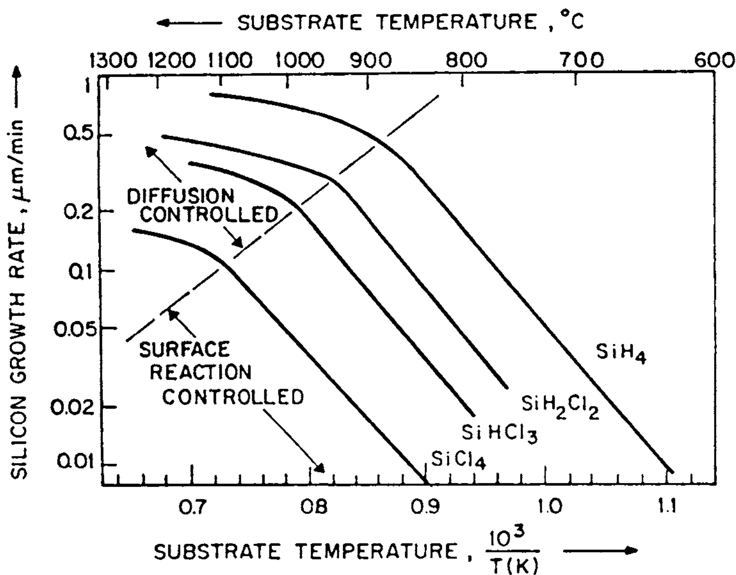
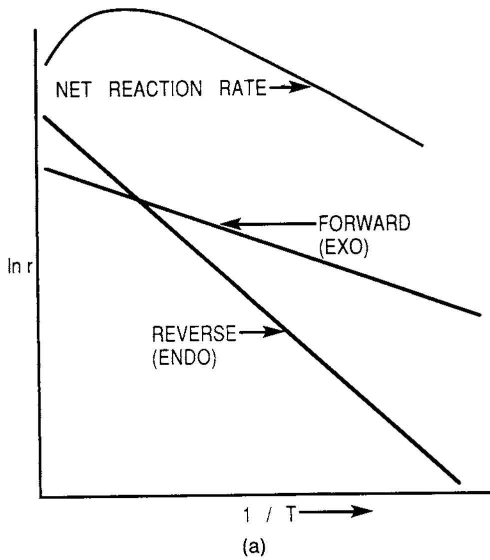


Figure 6-13 Deposition rate of Si from four different precursor gases as a function of temperature. (From W. Kern, in *Microelectronic Materials and Processes*, ed. R. A. Levy, reprinted with permission of Kluwer Academic Publishers, 1989.)

that the sign of ΔH° is negative, that is, the reactants have more energy than the products. For endothermic reactions, ΔH is positive. The individual forward and reverse reaction components are now shown in Fig. 6-14a on a common plot. Clearly, the activation energy barrier (or slope) for the reverse reaction exceeds that for the forward reaction. The *net* reaction rate or difference between the individual rates is also indicated. Interestingly, it reaches a maximum and then drops with temperature. A practical manifestation of this is etching—the reverse of deposition—in the high temperature range.

In Fig. 6-14b the alternate case is considered, that is, a forward endothermic deposition reaction and a reverse exothermic reaction. Here the net reaction rate increases monotonically with temperature and film growth rates will always increase with temperature. It is left for the reader to show that ΔH° for the reduction of $SiCl_4$ by H_2 (Eq. 6-3) is endothermic as written, for example, $\Delta H^\circ \cong +60\,\text{kcal/mol}$. There is actually a thermodynamic driving force which tends to transport Si from the cooler regions, the walls, to the hottest part of the reactor. This is where inductively heated



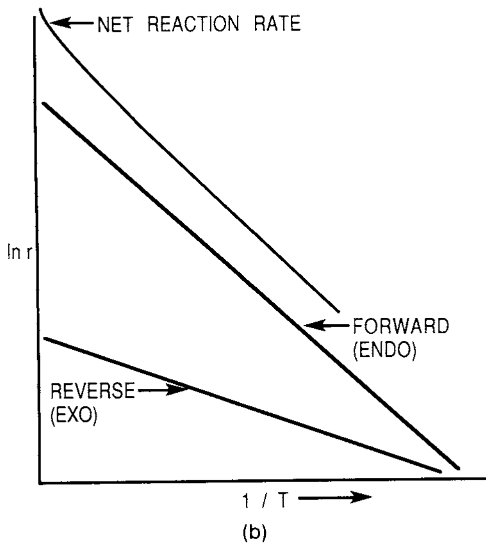


Figure 6-14 Chemical reaction energetics. (a) Activation energy for forward exothermic reaction is less than for reverse endothermic reaction. (b) Activation energy for forward endothermic reaction is greater than for reverse exothermic reaction. (After Ref. 27.)

substrates are placed and where film growth rates are highest. Epitaxial Si is most efficiently deposited in so-called *cold-wall* reactors for this reason.

The opposite is true when $\Delta H < 0$. Reversible reactions (Eq. 6-15) for the deposition of GaAs are exothermic and *hot-wall* reactors are employed in this case to prevent deposition on the walls. These results are a direct consequence of the well-known van't Hoff equation

$$\frac{d \ln K_{\mathrm{eq}}}{dT} = \frac{\Delta H^{\circ}}{RT^2}. \tag{6-50}$$

6.6 THERMAL CVD PROCESSES

6.6.1 SCOPE

The remainder of the chapter builds on the fundamentals of gas flow and reactions in CVD processes by describing their practical exploitation in producing films and coatings of interest. Most CVD processes can be conveniently subdivided into two categories. Thermal CVD processes employ heat energy to activate the required gas and gas-solid phase reactions. In contrast, the plasma-enhanced CVD processes treated in Section 6.7 derive benefit from plasma activation of the involved chemical species. As we shall see, however, even with similar input gases used to deposit the same nominal materials, films produced by thermal and plasma CVD processes can differ widely with respect to film structure, composition, and properties.

The great variety of materials deposited by thermal CVD methods has inspired the design and construction of an equally large number of processes and systems. Thermal processes have been broadly differentiated by such descriptors as low and high temperature, atmospheric and low pressure, cold and hot wall, and closed and open. Within a specific category of process, the variations in design and operating variables frequently make it difficult to compare performance of individual systems or reactors, even when depositing the same material. Regardless of process type, however, the associated equipment must have the following capabilities:

1. Deliver and meter the reactant and diluent gases into the reactor
2. Supply heat to the substrates so that reaction and deposition can proceed efficiently
3. Remove the by-product and depleted gases

In this section we will discuss some of the more widely practiced thermal CVD processes as well as some less common CVD methods; notably absent here is a treatment of CVD techniques for compound-semiconductor

epitaxy, a subject extensively covered in Chapter 8. Since more information is published on CVD applications and equipment for microelectronic device fabrication, the subsequent discussion will reflect this reality.

6.6.2 Atmospheric-Pressure CVD

6.6.2.1 High-Temperature Atmospheric-Pressure Systems

Although there is a continual imperative to reduce semiconductor processing temperatures, the growth of high-quality epitaxial thin films is often achieved by high-temperature, atmospheric-pressure CVD (APCVD) methods. This is true of Si as well as compound semiconductors. The reactors employed can be broadly divided into hot-wall and cold-wall types. Three popular cold-wall reactor configurations are depicted in Fig. 6-15. Of note in both the horizontal and barrel reactors are the tilted susceptors. As

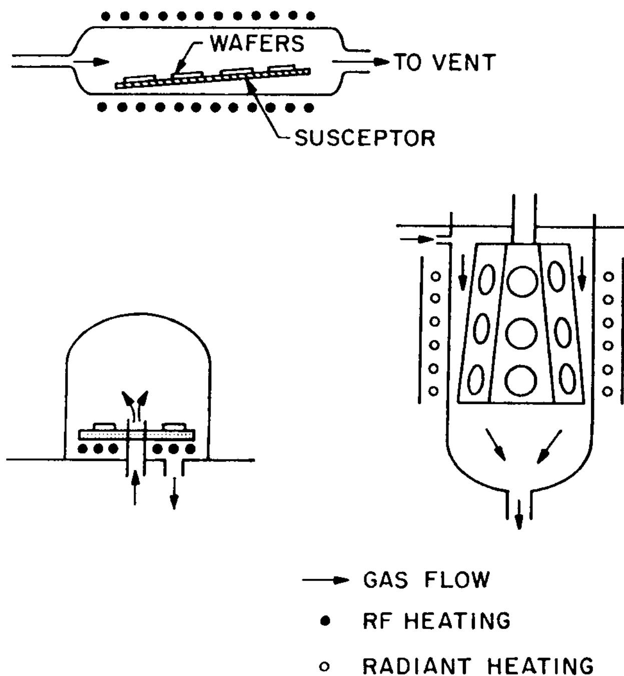


Figure 6-15 Schematic diagrams of reactors employed in epitaxial Si deposition. (Top) Horizontal; (lower left) pancake; (lower right) barrel. (From S. M. Sze, *Semiconductor Devices: Physics and Technology*, copyright © 1985, John Wiley & Sons. Reprinted with permission.)

previously discussed, this feature compensates for reactant depletion, which results in progressively thinner deposits downstream. In contrast to the other types, the wafer substrates lie horizontal in the pancake reactor. Incoming reactant gases flow radially over the substrates where they partially mix with the product gases.

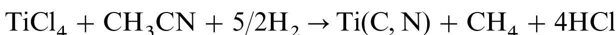
Barrel reactors are commonly utilized for the deposition of epitaxial Si films where the reactions listed in Eqs. 6-21a–f occur. Silicon wafers are placed in good thermal contact with SiC-coated graphite susceptors which can be inductively heated while the nonconductive chamber walls are simultaneously air or water cooled. These reactors typically operate with H_2 flow rates of several hundred standard liters per minute, and 1 vol% $SiCl_4$. Silicon film growth rates of 0.2 to 3 μm per minute are attained under these conditions. Substantial radiant heat loss from the susceptor surface and consumption of large quantities of gas, 60% of which is exhausted without reacting at the substrate, limit the efficiency of these reactors.

High-temperature APCVD systems are also extensively employed in the deposition of hard metallurgical coatings (Refs. 5, 28). One of the important advantages of CVD methods is the ability to batch coat large numbers of small tools at one time. For this purpose commercial hot-wall reactor systems capable of individual or sequential TiC, TiN, and Al_2O_3 film depositions have been constructed as schematically indicated in Fig. 6-16a. Perhaps billions of cemented tungsten-carbide cutting tools have now been coated since the mid-1960s in such reactors. As many as 10–20,000 WC tool inserts ($\sim$1 cm square) can be simultaneously given a $\sim$6 μm thick coating during a several-hour treatment cycle at temperatures within the range 700–1100°C. Since hard coatings are deposited on varied substrates, such as carbides, high-speed tool steels, and ball-bearing steels, a wide range of substrate temperatures is required, necessitating different deposition techniques and conditions. For example, TiN can be deposited by CVD over a broad range of temperatures using the same $TiCl_4$ precursor. Some typical reactions leading to the attractive, lustrous golden TiN deposits are:

1. High temperature, $1200°C > T > 850°C$:

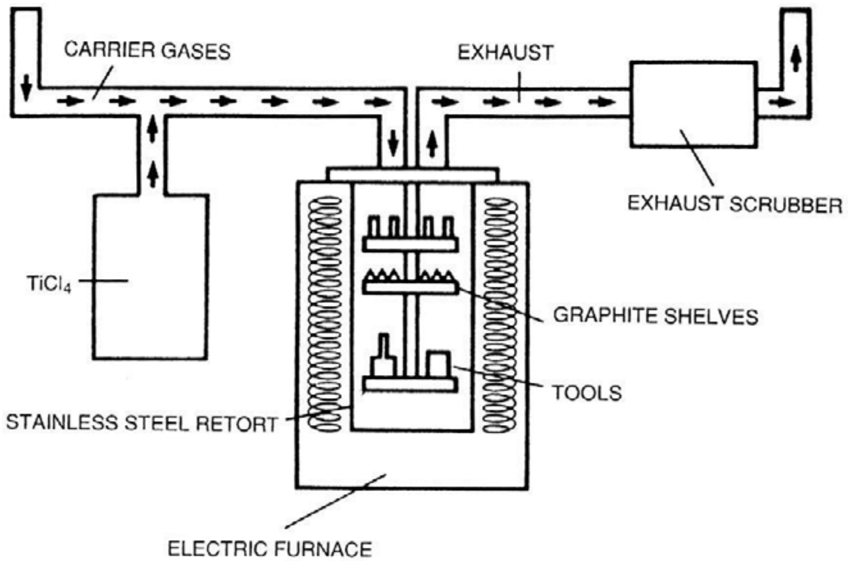
$$2TiCl_4 + N_2 + 4H_2 \rightarrow 2TiN + 8HCl$$

2. Moderate temperature, $850°C > T > 700°C$:



$$TiCl_4 + CH_3CN + 5/2H_2 \rightarrow Ti(C, N) + CH_4 + 4HCl$$

Deposition at temperatures as low as 300°C is possible, but this requires operation in a plasma discharge. Analogous reactions involving methane result in TiC deposits. The microstructure of such a multilayer CVD coating is shown in Fig. 6-16b.



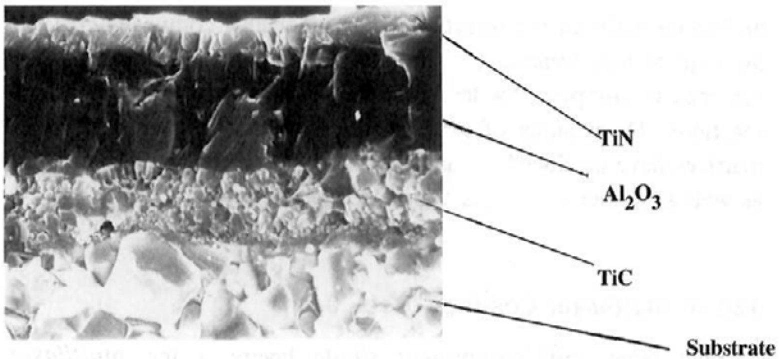
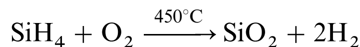


Figure 6-16 Schematic view of a commercial CVD reactor for deposition of TiC, TiN, and Al_2O_3 on carbide cutting tools. (Courtesy of A. Gates, Multi-Arc Scientific Coatings Inc.) (b) SEM image of CVD multilayer coating for cutting tool inserts. Carbide substrate/TiC/Al_2O_3/TiN (3500 ×). Courtesy of S. Wertheimer, ISCAR Ltd.

6.6.2.2 Low-Temperature Atmospheric Pressure Processes

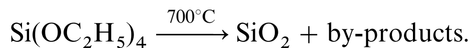
It has already been noted (Section 6.1) that in the fabrication of both memory and logic IC chips, assorted dielectric thin films, such as SiO_2, BPSG glasses, and silicon nitride, are deposited in order to insulate, passivate, mask, or hermetically seal various parts of the underlying circuitry. Several types of atmospheric-pressure, low-temperature reactors have been devised over the years in order to deposit these films. They include resistance-heated rotary reactors of radial configuration and reactors featuring a close-spaced gas-nozzle geometry. In the latter, gases impinge on wafers translated past the nozzles on a metal conveyer belt. As we shall see subsequently, atmospheric CVD methods have been supplanted in many cases by low pressure CVD and plasma CVD processes; the former enable greater wafer throughputs while the latter promote film deposition at lower temperature.

The widely deployed SiO_2 films have been deposited via the oxidation of silane



$$SiH_4 + O_2 \xrightarrow{450^\circ C} SiO_2 + 2H_2$$

and the decomposition of tetraethyloxysilicate or TEOS ($Si(OC_2H_5)_4$),



$$Si(OC_2H_5)_4 \xrightarrow{700^\circ C} SiO_2 + \text{by-products.}$$

Instead of pyrophoric silane, the latter reaction employs an inert liquid precursor. Another advantage is the improved step and trench coverage with TEOS. This is due to the fact that the very reactive SiH_4 decomposes immediately as it hits the surface. On the other hand the more complex TEOS molecules decompose slowly and diffuse further along the surface or down a trench before leaving SiO_2 behind (Ref. 6). TEOS is widely used to produce the intermetal-dielectric filling in multilevel IC structures. Although APCVD (ozone/TEOS) and low-pressure CVD have been employed for this purpose (Ref. 29), the low-pressure, high-density plasma-assisted (O_2-TEOS) CVD processes discussed in Section 6.7.3.1 appear to be the wave of the future (Ref. 30).

For the premetal dielectric layer that fills the space between the transistors and the first level of metallization (see Fig. 6-1), films of BPSG are traditionally employed (Ref. 31). This important layer has two functions. It electrically isolates transistors from the metal conductors, and protects them from being exposed to harmful Na^+ and K^+ ions, which degrade device performance. Containing about 2–6 wt% each of B and P, such films are

deposited by thermal CVD methods at 500–700°C and annealed at 800–1000°C. When SiH_4 is cooxidized with PH_3, films of phosphosilicate glass (PSG) form, but reaction with $PH_3 + B_2H_6$ mixtures generates BPSG. The role of phosphorus is to getter or trap the mobile ions, while boron "softens" the layer facilitating viscous flow and planarization during reflow annealing.

6.6.3 Low-Pressure CVD (LPCVD)

One of the significant developments in semiconductor device processing was the introduction of CVD reactors that operate at lower than atmospheric pressure. Historically they were first employed to deposit polysilicon films for gate electrodes. However, assorted dielectric films such as SiO_2, silicon oxynitrides, and silica glass, in addition to tungsten, other refractory metals, and silicide films, have also been deposited by LPCVD methods. If APCVD was capable of producing some of these films, what accounted for the adoption of LPCVD processing with its more complex reactors? The chief reason is the cost savings achieved by packing wafers more densely than is possible in APCVD. In practice, large batches of wafers, a hundred or more, can be processed at a time. This coupled with generally high deposition rates, improved film-thickness uniformity, better step coverage, greater control over stoichiometry and contamination problems, lower particle density, and fewer pinhole defects has given LPCVD important advantages in the processing of dielectric films.

Depending on how much below atmospheric pressure the CVD reactor operates, three different process regimes have been distinguished. From 100 torr to 1 torr we speak of reduced pressure CVD (RPCVD), while LPCVD is commonly practiced in the range from 1 to 10 mtorr. Ultrahigh vacuum CVD at $\sim 10^{-7}$ torr, a subject discussed in Section 8.6.3.2, has enabled the realization of low-temperature silicon epitaxy. To compensate for low pressures of operation relative to 760 torr in APCVD reactors, the input reactant gas concentration must be correspondingly enriched to maintain comparable deposition rates. Actually, when normalized to the same reactant partial pressure, LPCVD film growth rates may indeed exceed those for conventional atmospheric CVD. Low gas pressures primarily enhance the mass flux of gaseous reactants and products through the boundary layer between the laminar gas stream and substrates. This effect on mass transport can be semiquantitatively estimated. According to Eq. 6-30 the mass flux of the gaseous species is directly proportional to D/δ. Since the diffusivity varies inversely with pressure, D is roughly 1000 times higher in the case of LPCVD. This more than offsets the increase in δ, which is inversely

proportional to the square root of the Reynolds number. In an LPCVD reactor the gas-flow velocity is generally a factor of 10–100 times higher, the gas density a factor of 1000 lower, and the viscosity unchanged relative to APCVD. Therefore, Re is a factor of 10 to 100 times lower and δ about 3 to 10 times larger. Because the change in D dominates that of δ, a mass transport enhancement of well over an order of magnitude can be expected for LPCVD.

Now consider wafers in an APCVD reactor. Efficient film growth is limited by mass transport through the boundary layer of the flowing gases, and to overcome this barrier wafers are laid parallel to the gas stream. But in LPCVD reactors, gas molecules have a larger mean free path and greatly enhanced diffusivity. Gas will readily impinge on the wafers and deposition is no longer mass-transfer limited but rather reaction-rate controlled. This means that wafers can be stacked closer together and parallel to each other (Section 6.5.2), resulting in high wafer throughputs.

Commercial LPCVD systems commonly employ resistively heated, horizontal hot-wall reactors, as shown schematically in Fig. 6-11a. Big mechanical pumps as well as blower booster pumps are required to accommodate the large gas load at low operating pressures. One significant difference between atmospheric and LPCVD systems concerns the nature of deposition on reactor walls. Dense, adherent deposits accumulate on the hot walls of LPCVD reactors whereas thinner, less adherent films form on the cooler walls of atmospheric reactors. In the latter case, particulate detachment and incorporation in films is a problem especially on horizontally placed wafers. It is less of a problem for LPCVD reactors where vertical stacking is employed.

6.6.4 Metalorganic CVD (MOCVD) Processes

Originally developed by Manasevit and co-workers (Ref. 32) in the late 1960s to grow epitaxial compound semiconductor films MOCVD (also known as organometal, OMCVD) is now widely used for that purpose. But MOCVD has also assumed additional importance in the deposition of assorted dielectrics and metals as well. Unlike other CVD process variants that are differentiated physically on the basis of pressure, or use of plasmas or photons (see Section 6.6.5), MOCVD is distinguished by the chemical nature of the precursor gases. The great advantage of metalorganics is their generally high volatility at moderately low temperatures. Since all constituents are in the vapor phase, precise electronic control of gas flow rates and partial pressures is possible without dealing with troublesome liquid or solid

sources in the reactor. This, combined with pyrolysis reactions which are relatively insensitive to temperature, allows for efficient and reproducible deposition. Carbon contamination of films is a disadvantage, however. MOCVD reactions are usually carried out at low pressures, and sometimes with plasma assist. The reactors employed, which are discussed more fully in Section 8.5.3.1, are carefully designed to optimize the flow and efficient use of generally expensive precursors.

It is useful to divide the discussion of MOCVD along two major lines of applications. The more important ones deal with epitaxial compound-semiconductor deposition, a subject that will be treated in Chapter 8. On the other hand applications considered here are to the deposition of oxide and metal films. In general, semiconductor deposition needs are more stringent with regard to reactor design and performance, as well as film quality.

At the top of Table 3-4 we have already noted a number of electroceramic films that have been prepared by pulsed-laser deposition. Because they include important applications in microelectronics, electrooptics, nonlinear optics, and information storage and display, there has been an imperative to prepare these materials by other methods including MOCVD. Assorted ferroelectric (Pb, Ba, Zr, Sr)TiO_3 films, in particular, have been the subject of much attention (Refs. 33–35). By virtue of hysteresis effects in electric fields and two remanent polarization states, these titanates possess a natural memory. This, coupled with their very large dielectric constants, makes them obvious choices as capacitor dielectrics for high-density DRAM applications. In producing these films, precursors are typically heated to $\sim$200°C while deposition occurs anywhere from $\sim$300 to 1000°C. Amorphous as well as oriented microcrystalline films often deposit on silicon wafers; epitaxial films of $BaTiO_3$ have been grown on $LaAlO_3$ (Ref. 33).

Although the production of technologically significant surface areas of electroceramic films by laser ablation is not yet practical, MOCVD technology employing metal alkoxides or β-diketonate precursors offers a flexible approach for depositing oxides (Ref. 36). MOCVD possesses the potential for large-area deposition, good composition control and film uniformity, high film-deposition rates and densities, and conformal step coverage. Success strongly hinges on the availability of suitable precursors; in this regard, commercial scale-up of MOCVD requires the following: (1) batch precursor amounts must rise from a few grams to a few kilograms, (2) production yields must rise and quality be consistent, (3) the environmental impact must be minimal, and (4) safety must be assured. Typical MOCVD precursors for electroceramic metal-oxide film deposition are listed in Table 6-3.

Before leaving this subject, it is worth noting a very versatile and inexpensive thin-film preparation method that employs metalorganic pre-

Table 6-3
MOCVD Precursors for Assorted Metals and Electroceramic Metal Oxides

Metals[a]	Alkoxides	β-Diketonates[c]	Alkyls
Ag		Ag(acac)	
Al			$AlMe_3$, $AlEt_3$
Au		Me_2Au(fhac)	
Cu	$Cu(OBu)_4$	$Cu(hfac)_2$, $Cu(acac)_2$	
Pt		$Pt(acac)_2$	$C_5H_5Pt(Me)_3$
Metal oxides[b]			
TiO_2	$Ti(OR)_4$ [b]		
ZrO_2	$Zr(OR)_4$	$Zr(acac)_4$, $Zr(thd)_4$	
Ta_2O_5, Nb_2O_5	$Ta(OEt)_5$ [c], $Nb(OEt)_5$		
(Ba, Sr)TiO_3	$Ti(OR)_4$, $Ti(OPr)_2(thd)_2$	$Ba(thd)_2$, $Ba(hfac)_2$, $Sr(thd)_2$	
Pb(Zr, Ti)O_3, (Pb, La)(Zr, Ti)O_3	$Zr(OR)_4$, $Ti(OR)_4$, $Ti(OR)_2(thd)_2$	$Pb(thd)_2$, $Pb(fod)_2$, $Zr(thd)_4$, $La(thd)_3$	$PbEt_4$, (neopentoxy)$PbEt_3$
Pb(Mg)NbO_3	$Nb(OEt)_5$	$Pb(thd)_2$, $Mg(thd)_2$, $Nb(thd)_4$	
(Ni, Zn)Fe_2O_4		$Ni(thd)_2$, $Ni(acac)_2$, $Zn(thd)_2$, $Zn(acac)_2$, $Fe(thd)_3$, $Fe(acac)_3$	
$YBa_2Cu_3O_{7-x}$		$Y(thd)_3$, $Ba(thd)_2$, $Ba(hfac)_2$, $Cu(thd)_2$, $Cu(hfac)_2$	

[a] From Ref. 36.
[b] From Ref. 40.
[c] Abbreviations for β-diketonate ligands: acac: 2,4-pentanedionate; thd: 2,2,6,6-tetramethyl-3,5-heptanedionate; hfac: 1,1,1,5,5,5-hexafluoropentane-2,4-dionate; fod: 1,1,1,2,2,3,3-heptafluoro-7,7-dimethyloctane-4,6-dionate.
$R = (C_mH_{2m+1})$, Me = methyl, Et = ethyl, Pr = propyl, Bu = butyl.

cursors in a novel way (Refs. 37, 38). Instead of being injected into a reactor as a gas, the precursors are directly applied to the substrate by spin coating, spraying, or dipping. Then a high-temperature treatment decomposes the compound, leaving the desired thin film behind. By this metalorganic decomposition (MOD) process, assorted oxide films such as $SrTiO_3$ and $YBa_2Cu_3O_7$ have been prepared. It is claimed that epitaxial films are sometimes obtained at very high decomposition temperatures.

6.6.5 Laser-Enhanced CVD Deposition

There is considerable interest in depositing thin films at selected locations on unpatterned or broad-area substrates. This can be achieved by focusing beams of photons, electrons, or ions that are scanned or made to pass through mask openings. Where these beams hit substrates immersed in a CVD environment, reactions are triggered, resulting in *writing* or initiation of deposition. Laser- or, more generally, photoassisted chemical processing involves the use of monochromatic light to enhance and control film deposition.

Two mechanisms are involved during laser-enhanced chemical-vapor deposition (LECVD) and these are illustrated in Fig. 6-17. In the *pyrolytic* mechanism the laser heats the substrate to decompose gases above it and enhance rates of chemical reactions there. Many different lasers are suitable for pyrolytic deposition. However, substrates that melt above the gas decomposition temperatures are required. *Photolytic* processes, on the other hand, depend on direct dissociation of molecules by energetic photons.

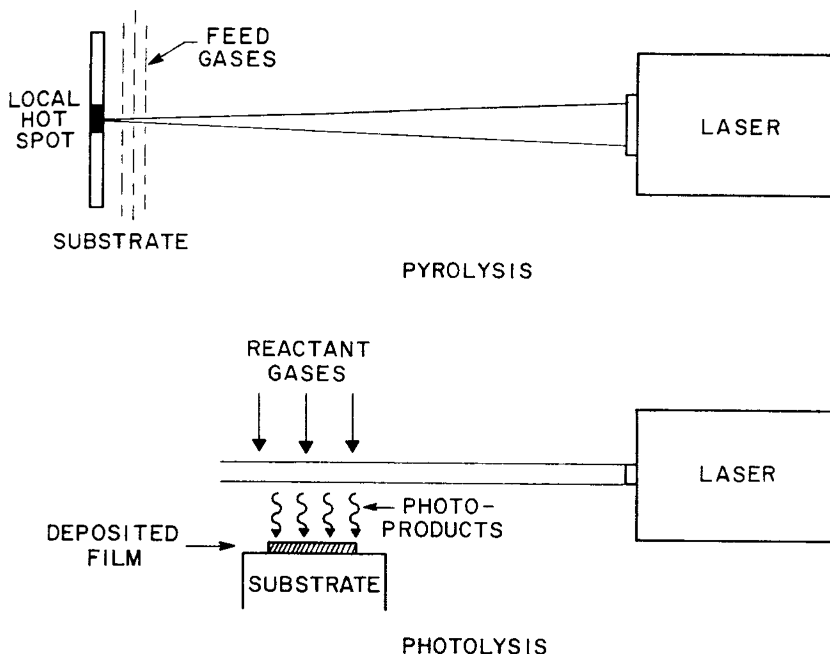


Figure 6-17 (a) Pyrolytic and (b) photolytic laser-induced chemical-vapor deposition of films (From *Chemical Vapor Deposition*, edited by M. L. Hitchman and K. F. Jensen. Reprinted with the permission of Academic Press, Ltd., and Professor K. F. Jensen, MIT.)

Organic metal dialkyl and trialkyl precursors are often used. Ultraviolet light sources are essential because many useful parent molecules (e.g., SiH_4, Si_2H_6, Si_3H_8, N_2O) require absorption of photons with wavelengths of less than 220 nm to initiate dissociation reactions. Practical lasers for this purpose are essentially the same as those employed for pulsed-laser deposition (Section 3.5.2.2) but operate at typical powers of a few tens of milliwatts. Such power levels are too low to enable high deposition rates over large areas, but are sufficient to initiate deposits selectively. It should be noted that there is frequently an admixture of pyrolytic and photolytic deposition processes occurring simultaneously with deep UV sources. Alternately, pyrolytic deposition is accompanied by some photodissociation of loosely bound complexes if the light source is near the UV.

Photolytically deposited metals are usually contaminated with carbon and display poor conductivity. For this reason direct writing of materials is usually accomplished by pyrolytic processes. Metal films so deposited have been used to repair open circuit defects in integrated circuits as well as "clear" defects in lithographic masks (Ref. 39). A number of metals such as Al, Au, Cr, Cu, Ni, Ta, Pt, and W have been deposited by LECVD using assorted volatile metal carbonyls, hydrides, and metalorganic compounds (see Table 6-3). Results for several metals indicate what may be expected of LECVD.

1. Aluminum films with electrical resistivities 1.2 to 3 times that of bulk Al have been deposited at $\sim 100°C$ from dimethylaluminum hydride and trimethylamine alane. Vertical film-growth rates up to 20 μm-s^{-1}, while scanning at speeds of 600 μm-s^{-1}, have been achieved.
2. Copper films of $>95\%$ purity, with resistivities of 2–2.2 $\mu\Omega$-cm, have been deposited at less than 200°C.
3. Gold films deposited at growth rates of 3 μm-s^{-1}, with resistivities 1.4–10 times that of bulk Au, have been derived from gold diketonate sources.

Additional information has been published on the CVD of metals, including deposition temperatures (Ref. 40) and selective deposition (Ref. 41).

Research on the photo-assisted deposition of elemental semiconductors (Si, Ge, diamond), compound semiconductors (GaAs, InP, GaP, HgTe, HgCdTe), dielectrics (SiO_2, TiO_2, silicon nitride) has been reviewed by McCrary and Donnelly (Ref. 42). As with metals, LECVD processing enables selective deposition of these materials at temperatures hundreds of degrees lower than possible by other methods. The exciting possibilities of low-temperature processing are offset by generally inferior film properties.

6.7 PLASMA-ENHANCED CVD PROCESSES

6.7.1 OVERVIEW

In PECVD processes, glow-discharge plasmas are sustained within chambers where simultaneous vapor-phase chemical reactions and film deposition occur. Just as the demands of semiconductor technology provided the driving force for advances in sputtering and plasma etching, PECVD processing arose from similar imperatives. A major early commercial application of PECVD was the low-temperature deposition of silicon nitride films for the passivation and encapsulation of completely fabricated microelectronic devices. At this final stage IC chips cannot be heated much above 300°C, a far lower temperature than that used in traditional thermal CVD processes for silicon-nitride deposition.

In the majority of PECVD processing activity, glow-discharge plasmas are excited by an RF field. The reason is that most films deposited by this method are dielectrics and DC discharges are not feasible. Generally, the RF frequencies employed range from about 100 kHz to 40 MHz. In the reduced gas pressure environment, typically sustained between 50 mtorr and 5 torr, electron and positive ion densities number between 10^9 and $10^{11}/cm^3$, and average electron energies range from 1 to 10 eV. This energetic discharge is sufficient to decompose gas molecules into a variety of component species, i.e., ions, atoms and molecules in ground and excited states, molecular fragments, free radicals, etc. The net effect of the interactions among these reactive entities is to cause chemical reactions to occur at much lower temperatures than in thermal CVD reactors not benefiting from plasma activation. Therefore, previously unfeasible high-temperature reactions can be made to occur on temperature-sensitive substrates.

To illustrate that thermodynamic limitations are overcome by plasmas consider the formation of TiC by the reaction of Eq. 6-10, for which the free energy is given by ΔG° (kcal/mol) $= 72 - 0.059T$. The temperature that divides thermodynamically possible from impossible reaction occurs when $\Delta G^\circ = 0$, or 1218 K. Below this temperature $\Delta G^\circ > 0$ and no TiC ought to form, a conclusion in rough accord with the typical ~1000°C temperature for film deposition. However, in the presence of a plasma, the energy barrier for reaction is reduced and deposition occurs at 700 K (Ref. 43).

A sampling of assorted materials that have been deposited by PECVD methods is included in Table 6-4 together with typical operating temperatures and source gas compositions (Refs. 43, 44). Included are elements such as carbon in the form of diamond and diamond-like films, metals, oxides, nitrides, and borides. Although semiconductor films have been prepared by PECVD, their stoichiometry and crystalline quality are simply insufficient for device purposes.

Table 6-4
PECVD Films, Source Gases, and Deposition Temperatures

Film	Source gases	Deposition temperature (°C)
Elemental		
Al	$AlCl_3$–H_2	100–250
a-B	BCl_3–H_2	400
a-C	C_nH_m–H_2/Ar	25–250
a-Si	SiH_4–H_2	300
c-Si	SiH_4–H_2	400
Oxides		
Al_2O_3	$AlCl_3$–O_2	100–400
SiO_2	$SiCl_4$–O_2	100–400
TiO_2	$TiCl_4$–O_2	100–500
Nitrides		
AlN	$AlCl_3$–N_2	<1000
BN	B_2H_6–NH_3	300–700
	BCl_3–NH_3/Ar	300–700
Si_3N_4	SiH_4–NH_3–N_2	25–500
TiN	$TiCl_4$–N_2–H_2	100–500
Carbides		
B_4C	B_2H_6–CH_4	400
BCN	B_2H_6–CH_4–N_2	~ 25
	$C_8H_{18}BN$	250
SiC	SiH_4–C_nH_m	140–600
TiC	$TiCl_4$–CH_4–H_2	400–900
Borides		
TiB_2	$TiCl_4$–BCl_3–H_2	480–650

From Refs. 43 and 44.

6.7.2 PECVD Reactors

Plasma CVD processes are distinguished by the source of excitation, e.g., RF, microwave, and type of coupling, e.g., capacitive and inductive. Conforming to these descriptors are several PECVD reactor configurations. Tube or tunnel reactors are the simplest and consist of a coil wrapped

around a tube (Fig. 4-5). They are inductively coupled, but if external electrode plates are used instead, they can be capacitively coupled. Gases passing through the tube become ionized and react with each other, while solid films deposit on suitably placed substrates. In both configurations a symmetric potential develops on the walls of the reactor. High-wall potentials are avoided to minimize sputtering of wall atoms and their incorporation into growing films.

The parallel-plate, plasma-deposition reactor of the Reinberg type (Section 5.4.5.3) shown in Fig. 6-18 has been a very widely used configuration for PECVD. Reactant gases first flow along the axis of the chamber and then radially outward across rotating substrates which rest on one plate of an RF-coupled capacitor. This diode configuration enables reasonably uniform and controllable film deposition to occur. Since the process is carried out at low pressures, advantage is taken of enhanced mass transport; a typical deposition rate for silicon nitride is 300 Å/min at power levels of 500 W.

The reactors just described are used in what are known as conventional or *direct* PECVD processes. In these, the reactant and product gases, diluents, plasma, and substrates are immersed in and interact within the same space. Such direct processes are to be distinguished from the indirect,

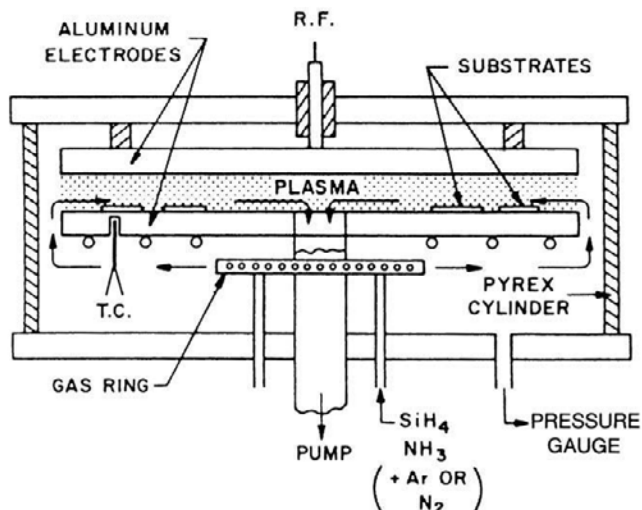


Figure 6-18 Reinberg-type cylindrical radial-flow plasma reactor for the deposition of silicon-nitride films. [From M. J. Rand, *J. Vac. Sci. Technol.* **16**(2), 420 (1979). Reprinted with permission.]

downstream, or *remote* processes in which substrates lie outside the plasma-generation zone. The major advantage of such remote reactors is the reduction in the number of possible deposition and etching reaction pathways (Ref. 45). Eliminating plasma-activated species of by-product gases from being incorporated in films enables better control of oxygen, nitrogen, and hydrogen stoichiometric ratios in silicon oxides, nitrides, and oxynitrides. Relative to direct plasma processing, concentrations of bonded hydrogen (see Section 9.6.3) are significantly lower in assorted amorphous dielectrics and semiconductors; this accounts for the use of remote reactors in depositing amorphous thin-film silicon solar cells and transistors.

Downstream, high-density plasma (HDP) reactors have also been increasingly used in integrated-circuit film-deposition and etching processes. An important example of an HDP reactor is the vertical microwave-frequency version shown in Fig. 6-19. In this reactor, also known as an electron cyclotron resonance (ECR) reactor, an applied static magnetic field (B) confines the plasma. During operation, microwave energy is coupled to the resonant frequency of the plasma electrons. The condition for absorption requires that the microwave frequency ω_m (commonly 2.45 GHz) be equal to qB/m_e, a formula previously defined in connection with magnetic confinement of plasmas (Section 4.3.4.2). Physically, plasma electrons under-

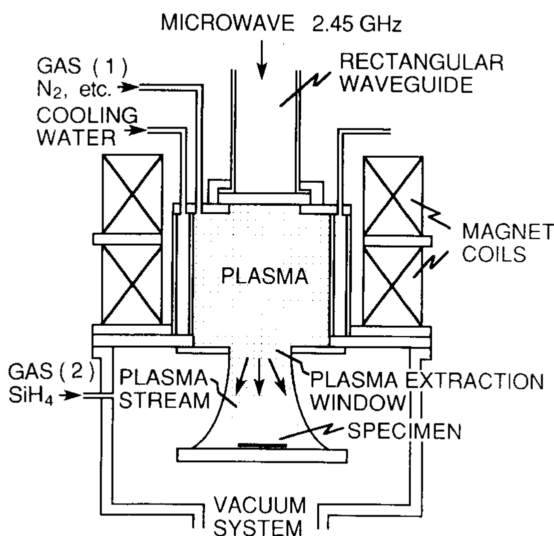


Figure 6-19 ECR plasma deposition reactor. (From S. Matuso, in *Handbook of Thin Film Deposition Processes and Techniques*, ed. K. K. Schuegraf. Noyes, Park Ridge, NJ, 1988. Reprinted with permission from Noyes Publications.)

go one circular orbit during a single period of the incident microwave. Whereas RF plasmas contain a charge density of $\sim 10^{10}$ cm^{-3} in a 10^{-2} to 1 torr environment, the ECR discharge is denser and easily generated at pressures of 10^{-5} to 10^{-3} torr. At such pressures gas diffusion is more rapid relative to charged-particle recombination. Therefore, the degree of ionization is about 100–1000 times higher in ECR relative to RF plasmas. This, coupled with low-pressure operation, low plasma sheath potentials, and absence of source contamination (no electrodes!), has made ECR plasmas attractive for both film deposition and etching (see Section 5.4.5.4). Furthermore, collimated ion beams of controlled energy can be also extracted through biased grids employing magnetic fields. In this way the benefits of controlled ion bombardment are realized. Materials such as SiO_2, Al_2O_3, SiN, Ta_2O_5, and diamond films have been deposited in ECR reactors. Production of high-quality films at low substrate temperatures is a significant advantage of microwave plasma processing.

Other inductively coupled high-density plasma sources (helicon resonator), developed primarily for etching and briefly described in Section 5.4.5.4 have also been used for film deposition (Ref. 46). In general, the types and properties of films produced parallel those deposited in ECR reactors.

6.7.3 PECVD PROCESSES

6.7.3.1 Microelectronics Applications

In recent years the need for new materials coupled with lower processing-temperature requirements in different but related applications, e.g., ULSI (ultralarge-scale integrated) circuits, solar cells, and flat-panel displays, have made PECVD film deposition and plasma-etching methods indispensable. This is illustrated in a recent issue of the *IBM Journal of Research and Development* (Ref. 47) devoted to plasma processing. Among the involved materials are silicon-based semiconductors, insulators such as amorphous silicon and boron or phosphorus-doped SiO_2, silicon nitride, and silicon oxynitride.

Plasma-processing methods of these materials can be best appreciated by reviewing a couple of cutting-edge applications in ULSI technology (Ref. 48). The first deals with the intermetal dielectric in logic integrated circuit chips. Previously (in Sections 5.4.2 and 5.3.4.3), methods of patterning and sputtering metal interconnections at various metallization levels were described. Here we comment on the intermetal SiO_2 dielectric which must fill, without containing internal voids, the gaps between the interconnections. In this regard a comparison between different oxide deposition

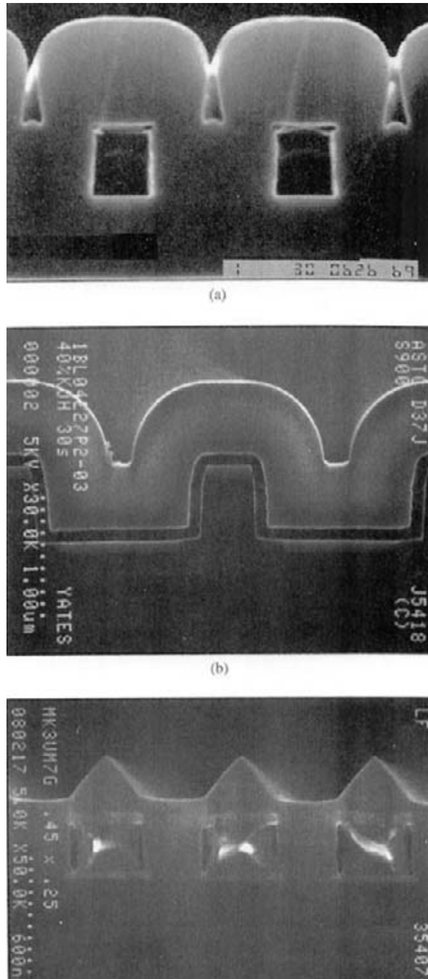


Figure 6-20 Scanning electron microscope cross-sectional images of oxide films grown by plasma and thermal CVD means. (a) TEOS-based PECVD; (b) TEOS-based atmospheric-pressure thermal CVD; (c) silane-based high-density plasma CVD. (From Ref. 48. Copyright © 1999 by IBM Corp. Reprinted by permission.)

methods is instructive. Conventional PECVD dielectric films derived from a SiH_4 source have a pronounced breadloaf appearance when deposited over an interconnect-stepped substrate. There is less breadloaf cusping (Fig. 6-20a) if TEOS is used because of the higher surface-mobility of the reactants. Nevertheless, a void still forms if the trench gap is too small

because the oxide is not completely conformal, i.e., the deposit thickness is greater at the top than bottom of the trench. In contrast, conformal coverage is typical for ozone-TEOS deposition by thermal atmospheric as well as subatmospheric CVD (Fig. 6-20b). To eliminate breadloafing during gap filling by PECVD, some combination of sequential deposition and etching steps is required, e.g., PECVD-argon sputter etch-PECVD. But, if RF bias is applied to the substrate within high-density plasma reactors, significant ion bombardment and sputter etching occurs simultaneously with deposition. A triangular instead of breadloaf oxide structure (Fig. 6-20c) emerges that is more readily planarized by subsequent chemical–mechanical polishing.

The deposition/sputter-rate ratio (D/S) is an important measure of the gap filling capability of this one-step process. Low D/S ratios facilitate filling of high-aspect-ratio trench structures, but if it is too low, corners of the features to be filled tend to be sputtered off (i.e., corner clipping). On the other hand, high D/S ratios lead to seams and voids. Typical deposition conditions for SiO_2 gap filling of a 0.25-μm wide, 0.62-μm deep structure in a HDPCVD system include the following: RF power, 2 to 4 kW; gases, $SiH_4/Ar/O_2$; pressure, <5 mtorr; $D/S = 3.2/1$; deposition temperature, 250–350°C; deposition rate, 100–400 nm/min.

A second set of PECVD applications in ULSI-DRAM cells for memory chips is schematically depicted in Fig. 6-21. Aside from the heart of the DRAM cell consisting of the interconnected field-effect transistor/deep-trench capacitor, the metal, oxide, glass and nitride films that are deposited parallel those for logic chips.

6.7.3.2 Plasma Modification of Metal Surfaces

In the plasma nitriding and carburizing processes treated in this section, nitride and carbide films are *not* deposited. Instead, the atoms of nitrogen and carbon that deposit on metal surfaces modify them by diffusing into the underlying matrix. Except for the plasma assist and low pressures, these processes are very much like traditional nitriding and carburizing, which in essence are high-temperature, atmospheric-pressure CVD treatments. Plasma (or ion) nitriding of steels, one of the earliest commercial uses of plasmas, is generally accomplished in DC glow discharges generated at a potential difference of 300–1000 V between the workpiece cathode and the chamber walls (Ref. 44). Since the former may have complex three-dimensional shapes, it is important to conformally surround them with a "glow seam." These workpieces are maintained below 600°C and subjected to N_2/H_2 mixtures at pressures between $\sim$0.5 and 5 torr. Under these conditions a very hard nitrogen-rich layer some 2–10 μm thick, consisting

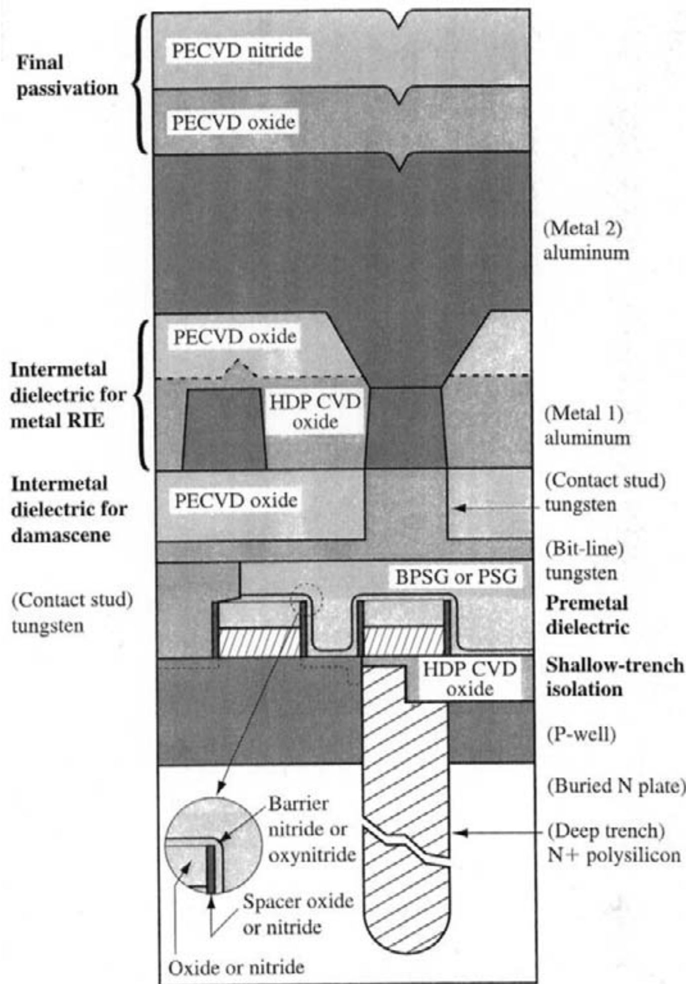


Figure 6-21 Schematic drawing of a three-level DRAM cell illustrating actual and potential (bold font) plasma CVD applications. (From Ref. 48. Copyright © 1999 by IBM Corp. Reprinted with permission.)

of assorted Fe_xN ($x = 2$–4) compounds, forms within a diffusion zone a few hundred microns deep. By adding CH_4 and CO_2 to the gas mix, carbonitriding of surfaces can be achieved.

The process of plasma carburizing is carried out in an Ar/H_2 atmosphere containing CH_4 or C_3H_8 reactants at pressures between ~ 3 and 20 torr at temperatures of $\sim 1000°C$. Among the advantages claimed for such plasma

treatment are enhanced mass transfer of carbon and prevention of surface and intergranular oxidation, a feature that confers greater resistance to mechanical fatigue damage. Despite this, plasma carburizing of steel has not displaced conventional heat-treating to the same extent that its low-temperature plasma-nitriding counterpart has. Plasma modification of other metals, e.g., nitriding of titanium alloys, and by different processes, e.g., plasma boriding of steel, are additional variations on the theme. By extrapolation to high-energy pulsed-plasmas, there is the PIII process discussed in Section 5.5.6. In common, these processes all rely on the undersized atoms of N, C, and B. These rapidly diffuse interstitially or can be readily implanted into metal matrices, and once incorporated they readily react to form generally hard metal nitride, carbide, and boride compounds.

It is not difficult to estimate the depth profiles in these processes. If there are negligible ion-implantation effects, ions simply impinge on the solid surface and diffuse in. We may assume the surface ion current-density (j) is essentially equivalent to a diffusion flux (J), i.e., $J = j/q$, where q is the charge per ion and J has units of ions/cm^2-s. A boundary-value problem somewhat different from the two presented in Section 1.6.1, but like the one in Section 3.5.2.3, is then suggested. As before, we assume that there are no subsurface atoms (ions) initially, i.e., $C(x, 0) = 0$, and that $C(x = \infty, t) = 0$. However, at the surface of our one-dimensional semiinfinite solid, $D(dC/dx)$ $(x = 0, t > 0) = j/q$, where D is the effective diffusion coefficient.

Under these conditions the solution to Eq. 1-24 is

$$C(x, t) = \frac{2j}{qD}\{(Dt/\pi)^{1/2} \exp(-x^2/4Dt) - \tfrac{x}{2}\,\mathrm{erfc}[x/2(Dt)^{1/2}]\}. \quad (6\text{-}51)$$

The surface concentration,

$$C(x = 0, t > 0) = \frac{2j}{qD}\{(Dt/\pi)^{1/2}\},$$

is linearly dependent on j. As $C(x = 0)$ rises parabolically in time, the slope $(dC(x)/dx)_{x=0}$ remains invariant.

6.7.4 Modeling Chemical Reactions in Plasmas

Since the complexity of plasmas is so great and our ability to treat chemical interactions in them so limited, computer simulation often offers the only salvation for predicting the kinetics of film deposition and etching processes in actual reactors. Much progress has been made in this regard

starting from first principles as reviews (Refs. 49, 50) of this complicated subject indicate. It is therefore worth considering typical steps involved in modeling such plasma processes. One formulation conceptually breaks the overall simulation into three tasks: (1) the *electron kinetics*, (2) the *plasma chemistry*, and (3) the *surface reactions*, each with separate models.

1. The electron kinetics submodel assumes or calculates the electric (and magnetic) field and generates electron-impact cross sections and rate constants for use in the plasma-chemistry model. This phase involves modeling of sheath characteristics and calculation of electron densities that vary spatially and with time (for AC discharges), as a result of drift, diffusion, and assorted inelastic ionization, dissociation, attachment, etc., collisions. Values of σ_e for various electron–particle collisions are the goal and much progress has been made in this regard.
2. The plasma-chemistry submodel provides a self-consistent accounting of the concentration of neutrals, radicals, and charged species as a function of space and time while conserving their mass, energy, and charge. Each reactant and product specie is treated separately and its motion due to diffusion, drift, and convective gas transport is folded together with information (from submodel 1) on collision-induced chemical reactions. Chemistries that have been modeled include those for hydrocarbon (e.g., CH_4) and silane-based deposition systems as well as those for fluorocarbon (e.g., CF_4), chlorine, and SF_6-based etching processes.
3. The surface reaction submodel attempts to describe net sticking rates of gas-phase species, the kinetics of adsorption, and interaction of bombarding ions and radicals with the first few layers of the substrate. Objectives include an understanding of the physical and chemical mechanisms of film deposition and etching where a particular focus is anisotropic etching. In general, surface interactions are the least well understood of all plasma-related reactions.

Each of the submodels is intimately coupled to the other submodels. Thus the output of submodel 1 in the form of electron impact rates serves as the input for the mass conservation equations of the plasma-chemistry submodel. Similarly, computational feedforward as well as feedback paths exist between the plasma-chemistry and surface-reaction submodels. When possible, additional experimental data in the form of discharge breakdown fields, spectral information, and light output from added actinometer gases (whose optical emission-rate coefficients are known) assist the computational process.

Modeling of the type described above for SiO_2 film deposition in high-density (microwave) plasma reactors has been reviewed by Meeks *et*

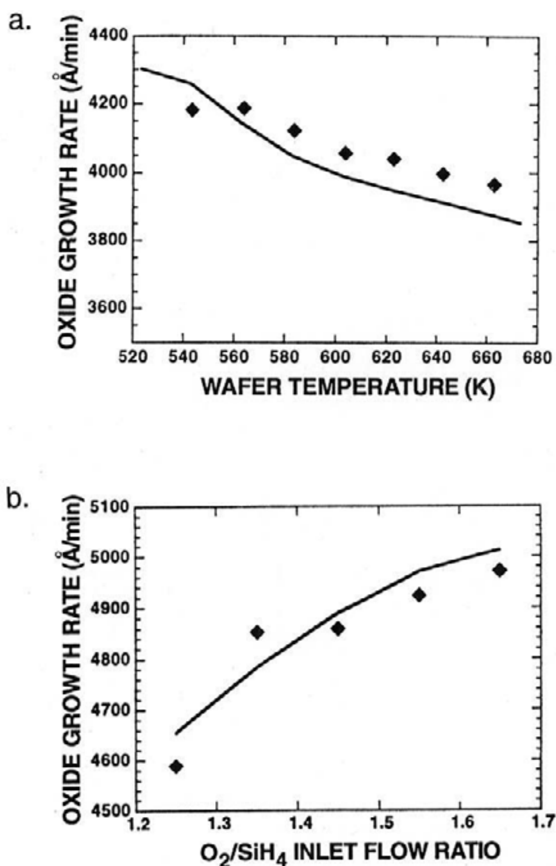


Figure 6-22 Comparison between experiment and theory in the growth of SiO_2 in a high-density (microwave) plasma reactor. (a) Oxide growth rate vs temperature. —— model; ♦ data from D. R. Denison, LAM Research Corp. (b). Oxide growth vs O_2/SiH_4 flow ratio. —— model; ♦ data from C. Apblett, Sandia National Labs. (From Ref. 51, Fig. 7. Reprinted with the permission of Dr. Ellen Meeks.)

al. (Ref. 51). Rate coefficients were calculated and tabulated for some 167 gas-phase reactions involving electron and particle collisions within simulated Ar, O_2, and SiH_4 plasmas. The results were reported in the form $k = AT^{B} \exp -(C/T)$, where A, B, and C are constants. An additional 96 reaction-rate parameters were also tabulated for surface reactions in the same simulated plasmas. Comparisons of theoretically derived oxide growth rates with measured rates, as a function of substrate temperature and gas composition, are shown in Fig. 6-22.

6.8 SOME CVD MATERIALS ISSUES

6.8.1 INTRODUCTION

In this section we consider several unconventional CVD materials, explore their composition, microstructure, and properties, and discuss processes to deposit them. As specific examples we shall treat silicon and silicon-nitride films, present the extraordinary properties of diamond films and other ultrahard materials, and consider novel processes to selectively deposit films. Our discussion will naturally invite comparisons with competitive materials or processes, and further expose the diversity of CVD methods. Inasmuch as it is the key to understanding properties, a brief discussion of film structure is a good place to start.

6.8.2 STRUCTURE OF THERMAL CVD SILICON FILMS

Actual film and coating structural morphologies that develop are the result of complex nucleation and growth processes. Since the kinetic details are similar in all film formation and growth processes irrespective of whether deposition is chemical or physical in nature, they will be treated within a common framework in the next three chapters. Perhaps the two most important variables affecting growth morphologies are vapor supersaturation and substrate temperature. The former influences the film nucleation rate while the latter affects the growth rate. In concert they influence whether epitaxial films, platelets, whiskers, dendrites, coarse-grained polycrystals, fine-grained polycrystals, amorphous deposits, gas-phase powder, or some combination of these form. Thus, single-crystal film growth is favored by low gas supersaturation and high substrate temperatures, whereas amorphous film formation is promoted at the opposite extremes of high supersaturations and low temperatures.

An example of the way substrate temperature affects the structure of deposited Si films (Ref. 52) is shown in Fig. 6-23. Decomposition of silane at temperatures of about 600°C and below yields amorphous films with no detectable structure. Polysilicon films deposited from 600 to 650°C have a columnar structure with grain sizes ranging from 0.03 to 0.3 μm, and possess a [110] preferred orientation. Larger Si crystallites form at higher temperatures, and eventually single-crystal or epitaxial film growth can be achieved at 1200°C.

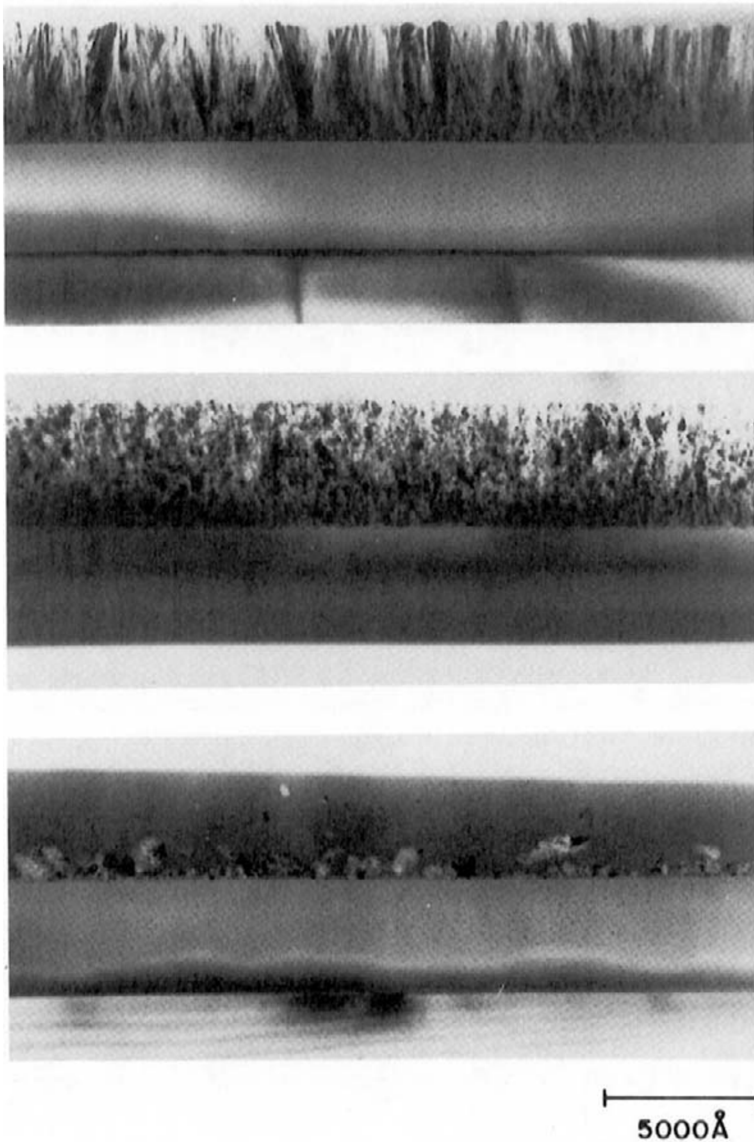


Figure 6-23 Morphology of poly-Si deposited from SiH_4 on an SiO_2 substrate. (Top) Columnar grains deposited above 650°C. (Middle) Fine-grained poly-Si. (Bottom) Partly amorphous structure deposited at 625°C. Deposition at temperatures below 600°C produces structureless amorphous films. (From Ref. 52. Courtesy of R. B. Marcus, reprinted with permission from John Wiley & Sons.)

6.8.3 Deposition and Structure of Amorphous Silicon Films

Amorphous silicon (a-Si) and hydrogenated versions of it, e.g., a-Si:H, are the basic materials discussed in this section. They have been deposited by both PVD and CVD techniques. Of the elements, silicon and others in column IV of the periodic table are among the easiest to prepare in amorphous form. Further, they do not require particularly low substrate temperatures (T_S) for their preparation. When normalized to the melting point (T_M), the ratio of T_S/T_M for amorphous CVD materials is higher by at least a factor of ~ 5 relative to what is required to amorphize typical PVD metal alloys. What accounts for the particular ease in amorphizing Si by CVD is the steric hindrance provided by molecular groups. Such structural obstruction leads to the high viscosity conducive to glass formation in bulk hydrocarbon polymers and oxide melts. In the case of a-Si:H films derived from SiH_4–H_2 mixtures, for example, the assorted radical and molecular fragments, e.g., SiH_3, SiH_2, Si_2H_6, are the source of the steric hindrance. A continuum of Si–H structures links the gas phase of $SiH_4 + H_2$ to the deposit surface. Contained within this transition region are assorted fragmented molecular species, which in turn connect to the film growth zone and eventually to the stable a-Si:H network beneath the film surface. Atom incorporation reactions are complex and not known with certainty (Refs. 53, 54); they resemble network polymerization of assorted Si_nH_{2n} molecular fragments and include dissociation and hydrogen desorption as well. The net result is the incorporation of as much as 35 at.% hydrogen (Ref. 55) into a silicon deposit disordered by bond stretching and distortion.

6.8.4 Amorphous Silicon Nitride

Silicon nitride is normally prepared by reacting silane with ammonia in an argon plasma, but a N_2/SiH_4 discharge can also be used. During plasma deposition, as much as 30 at.% hydrogen can be incorporated which apparently forms bonds to both Si and N. It is in this sense that silicon nitride is often described as a ternary solid-solution alloy, SiNH. Amorphous silicon nitride should be distinguished from the stoichiometric compound, Si_3N_4, formed by reacting silane and ammonia at 900°C in an atmospheric CVD reactor. It is instructive to further compare the physical and chemical property differences in three types of silicon nitride (Refs. 56, 57), and this is done in Table 6-5. While Si_3N_4 is denser and more resistant to chemical attack and has higher resistivity and dielectric breakdown strength, SiNH tends to provide better step coverage. An upper temperature

Table 6-5
Physical and Chemical Properties of Silicon Nitride Films from $SiH_4 + NH_3$

Property	Si_3N_4 1 atm CVD (900°C)	Si_3N_4(H) LPCVD (750°C)	SiNH PECVD (300°C)
Density (g/cm^3)	2.8–3.1	2.9–3.1	2.5–2.8
Refractive index	2.0–2.1	2.01	2.0–2.1
Dielectric constant	6–7	6–7	6–9
Dielectric breakdown field (V/cm)	10^7	10^7	6×10^6
Bulk resistivity (Ω-cm)	10^{15}–10^{17}	10^{16}	10^{15}
Stress at 23°C (GPa)	0.7–1.2(T)	0.6(T)	0.3–1.1(C)
Color transmitted	None		Yellow
H_2O permeability	Zero		Low–none
Thermal stability	Excellent		Variable >400°C
Si/N ratio	0.75	0.75	0.8–1.0
Etch rate, 49% HF (23°C)	80 Å/min		1500–3000 Å/min
Na^+ penetration	<100 Å		<100 Å
Step coverage	Fair		Conformal

Note: T = tensile; C = compressive.
Adapted from Refs. 56 and 57.
Stress data from S. M. Hu, *J. Appl. Phys.* **70**, R60 (1991).

limit that can be tolerated in some completed integrated circuits is ~300°C because interconnection metallizations begin to react with the surrounding insulation beyond this point. Since SiHN can be deposited at such temperatures and is an effective barrier against moisture penetration, it has been used to encapsulate IC chips.

Further aspects of the structure of amorphous elemental and alloy films, including silicon nitride, will be deferred to Section 9.6.

6.8.5 Ultrahard Thin Films

In this section we discuss films based on the three hardest substances known, diamond, boron nitride, and carbon nitride, and consider their deposition, properties, and uses. In common, they are all primarily produced by CVD methods. Diamond is by far the most important of these materials and is therefore given the greatest attention.

6.8.5.1 Introduction to Diamond and Carbons

Derived from the Greek *αδαμασ* (adamas), which means unconquerable, diamond is the most extraordinary of materials. Diamond is indeed an invincible material. In addition to being the most costly on a unit weight basis, and capable of unmatched beauty when polished, diamond has a number of other remarkable properties. It is the hardest substance known ($H_k > 8000$ kg/mm^2) and has a higher modulus of elasticity (1050 GPa) than any other material. When free of impurities, it has one of the highest electrical resistivities ($\sim 10^{16}$ Ω-cm). It also combines a very high thermal conductivity (2000 W/m-K), which is some 4 to 5 times that of Cu and Ag, with a low thermal expansion coefficient (1.2×10^{-6} K^{-1}) to yield a high resistance to thermal shock. Other less extraordinary but nevertheless interesting properties include resistance to chemical attack, high breakdown voltage, large bandgap energy, and transparency to visible, infrared, and microwave radiation. Diamond's remarkable attributes, the first three in particular, spurred one of the most exciting and competitive quests in the history of materials science — the synthesis of bulk diamond. Success was achieved with the 1954 GE Corp. process utilizing extremely high pressures and temperatures in the diamond-stable region of the carbon phase diagram (Fig. 1-12).

Isolated C atoms have distinct 2s and 2p atomic orbitals. When these atoms condense to form diamond, electronic admixtures occur resulting in four equal hybridized sp^3 molecular orbitals. Each C atom is covalently attached to four other atoms in tetragonal bonds 1.54 Å long, creating the well-known diamond cubic structure (Fig. 1-2c). Graphite, on the other hand, has a layered structure. The C atoms are arranged hexagonally with strong trigonal bonds (sp^2) and have an interatomic spacing of 1.42 Å in the basal plane. A fourth electron in the outer shell forms weak van der Waals bonds between planes which account for such properties as good electrical conductivity, lubricity, lower density, softness, and a grayish-black color.

In addition, C exists in a variety of metastable and amorphous forms which have been characterized as degenerate or imperfect graphitic structures. Here layer planes are disoriented with respect to the common axis and overlap each other irregularly. Beyond the short-range graphitic structure, the matrix consists of amorphous C. A complex picture now emerges of the manifestations of C ranging from amorphous to crystalline forms in a continuum of structural admixtures. Similarly, the proportions of sp^2–sp^3 (and even sp^1) bonding are variable, causing the different forms to have dramatically different properties. Not surprisingly, this broad spectrum of metastable carbons have been realized in thin-film deposits. What now complicates matters further is that the many techniques to produce carbon

films use hydrocarbon gases. Hydrogen is, therefore, inevitably incorporated, and this adds to the complexity of the deposit structure, morphology, and properties. Given the structural and chemical diversity of carbon films, an understandable confusion has arisen with regard to the description of these materials. Labels such as hard carbon, amorphous carbon (a-C), hydrogenated amorphous carbon (a-C:H), and diamond-like carbon (DLC), as well as diamond, have all been used in the literature. The ensuing discussion will treat the deposition processes and properties of these films with the hope of clarifying some of their distinguishing features.

6.8.5.2 Film Deposition Processes

Attempts to produce diamond from low-pressure vapors date back at least to 1911 (Ref. 58), and by the mid 1970s Derjaguin and Fedeseev (Ref. 59) had apparently grown epitaxial diamond films and whiskers during the pyrolysis of various hydrocarbon–hydrogen gas mixtures. After a decade of relative quiet, an explosive worldwide interest in the synthesis of diamond films and their properties erupted which persists to the present day.

6.8.5.2.1 Diamond

Thin diamond films involve metastable synthesis in the low-pressure graphite region of the phase diagram. The possibility of synthesizing diamond in this region is based on the small free-energy difference, i.e., 500 cal/mol, between diamond and graphite under ambient conditions (Ref. 60). Therefore, a finite probability exists that both phases can nucleate and grow simultaneously, especially under conditions where kinetic factors dominate, i.e., high energy or supersaturation. In particular, the key is to prevent graphite from forming or to remove it preferentially, leaving diamond behind. The way this is done practically is to generate a supersaturation or superequilibrium of atomic H. The latter has been most commonly produced utilizing 0.2–2% CH_4–H_2 mixtures in DC, RF, and microwave plasmas, or in thermal CVD reactors containing hot filaments. Under these conditions the atomic H that is generated fosters diamond growth by inhibiting graphite formation, by dissolving it if it does form, by stabilizing sp^3 bonding, or by promoting some combination of these factors. In general, hydrocarbon, e.g., CH_4, C_2H_2, decomposition at substrate temperatures of 800–900°C in the presence of atomic H is conducive to diamond growth on nondiamond substrates. Paradoxically, the copious amounts of atomic H result in very little hydrogen incorporation in the deposit. The beautiful SEM images (Fig. 6-24) of small faceted diamond crystallites produced in the manner just described have captured the imagination of the world.

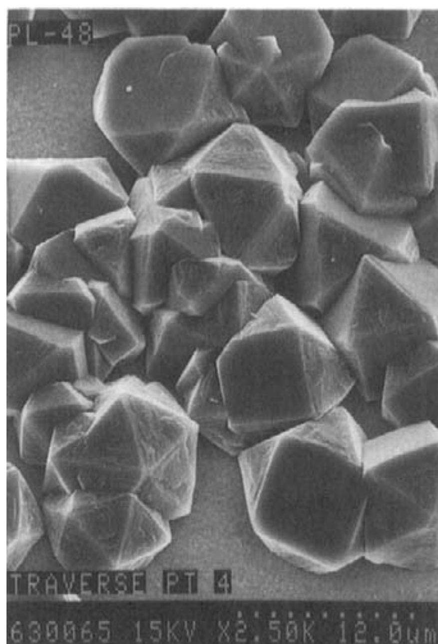


Figure 6-24 Diamond crystals grown by CVD employing combined microwave and heated filament methods. (Courtesy of T. R. Anthony, GE Corporate Research and Development.)

In retrospect, low-pressure diamond synthesis has proven to be not nearly as prohibitively difficult as originally imagined. For example, diamond deposits readily at rates of $\sim 15\,\mu$m/h on abraded molybdenum substrates heated by the feather portion of an oxyacetylene torch flame (Ref. 61). High-quality crystals ($>10\,\mu$m in size) have been produced at 500–750°C with an O_2/C_2H_2 ratio of 0.85–0.98.

An important consideration in commercial deposition processes employing cathodic arcs or CVD methods is the growth rate, which typically ranges up to 1000 μm/h (1 mm/h). This should be compared to even higher rates for commercial production of diamond abrasive grain.

6.8.5.2.2 Diamondlike Carbon

Amorphous carbons containing hydrogen, are identified as a-C:H materials and possess diamondlike properties. Films are formed when hydrocarbons impact relatively low-temperature substrates with energies in the range of a few hundred eV. Plasma CVD techniques employing RF and DC glow discharges in assorted hydrocarbon gas mixtures commonly produce

a-C:H deposits (Ref. 62). Substrate temperatures below 300°C are required to prevent graphitization and film softening. The energetic molecular ions disintegrate upon hitting the surface and this explains why the resulting film properties are insensitive to the particular hydrocarbon employed. It is thought that the incident ions undergo rapid neutralization and the carbon atoms are inserted into C–H bonds to form acetylenic and olefinic polymer-like structures, e.g., $C + R{-}CH_3 \rightarrow R{-}CH{=}CH_2$, where R is the remainder of the hydrocarbon chain. The resultant diamondlike films, therefore, contain variable amounts of hydrogen with H/C ratios ranging anywhere from $\sim$0.2 to $\sim$0.8 or more. Containing an admixture of sp^3, sp^2, and sp^1 bonding, they may be thought of as glassy hydrocarbon ceramics and can be even harder than SiC.

6.8.5.2.3 Amorphous Carbon (a-C)

These diamond-like films are prepared at low temperatures in the absence of hydrocarbons by ion-beam-assisted or sputter-deposition techniques. Both essentially involve deposition of carbon under the bombardment of energetic ions. Simple thermal evaporation of carbon will, however, yield highly conductive, soft films whose properties are far removed from the hard, very resistive, high-energy bandgap diamond-like materials. Ion impact energy, therefore, appears to be critical in establishing the structure of the deposit. More diamond-like properties are produced at lower energies; microcrystalline diamond ceases to form when the ion energy exceeds $\sim$100 eV, in which case the amorphous structure prevails.

6.8.5.3 Properties and Applications

The properties of CVD synthesized diamond, a-C, and a-C:H film materials (Refs. 60, 63) are compared with those of bulk diamond and graphite in Table 6-6. Basic differences in structure and properties of diamond and diamond-like films ultimately stem from the sp^3/sp^2 bond concentration ratios. Considerable bond admixtures occur in both the a-C:H and a-C films and much experimental effort has been expended in determining bonding proportions. Techniques such as Raman spectroscopy, nuclear magnetic resonance, and X-ray photoelectron spectroscopy (XPS) are used to characterize films and bolster claims for the presence of the elusive diamond crystals. Although there is a great deal of scatter in many of the film properties due to differing deposition conditions, it is clear that the films are extremely hard, chemically inert, and highly insulating.

Diamond and diamond-like film materials are already commercially exploited in a number of technologies. For example: high thermal conductivity is the reason that diamond substrates are used for semiconductor-laser

Table 6-6
Properties of Carbon Materials

	Thin films			Bulk	
Property	CVD diamond	a-C	a-C:H	Diamond	Graphite
Crystal structure	Cubic $a_0 = 3.561$ Å	Amorphous mixed sp^2–sp^3 bonds	Amorphous mixed sp^2–sp^3 bonds	Cubic $a_0 = 3.567$ Å	Hexagonal $a = 2.47$ Å
Form	Faceted crystals	Smooth to rough	Smooth	Faceted crystals	—
Hardness H_v	3000–12,000	1200–3000	900–3000	7000–10,000	—
Density g/cm^3	2.8–3.5	1.6–2.2	1.2–2.6	3.51	2.26
Refractive index	—	1.5–3.1	1.6–3.1	2.42	2.15 1.81
Electrical resistivity (Ω-cm)	$>10^{13}$	$>10^{10}$	10^6–10^{14}	$>10^{16}$	0.4 0.20
Thermal conductivity (W/m-K)	1100	—	—	2000	150
Chemical stability	Inert (inorganic acids)	Inert (inorganic acids)	Inert (inorganic acids and solvents)	Inert (inorganic acids)	Inert (inorganic acids)
Hydrogen content (H/C)	—	—	0.25–1	—	—
Growth rate (μm/h)	∼1	2	5	1000 (synthetic)	—

From Refs. 60 and 63.

and microwave-device heatsinks; the ultrahardness of diamond coatings is exploited in extending the life of cutting tools and surfaces that wear; diamond's extreme stiffness has been used on resonator diaphragms of tweeter loudspeakers to extend the frequency response to 60 kHz. For many applications crystalline diamond is not essential. Thus, DLC films are employed to protect magnetic information storage media. When such disks are started or stopped, the sliding contact that occurs beween the disk and slider can cause mechanical damage. To minimizae this while retaining high recording density, ultrasmooth DLC films less than 50 nm thick are used to coat hard disks and magnetic heads.

6.8.5.4 Boron Nitride and Carbon Nitride

Bulk cubic boron nitride (c-BN) is also a remarkable material. It has a hardness (70 GPa) second only to that of diamond (100 GPa), and its thermal conductivity, at half the diamond value, is still very high. However, diamond tools oxidize at low temperatures and cannot be used to machine Fe, Co, and Ni because of carbide formation. But c-BN is resistant to oxidation and is widely used to machine cast irons and steels. As with diamond-film synthesis, it is critical to form sp^3-hybridized c-BN during deposition to attain superior properties. Thermal CVD methods below 1250°C, using B_2H_6, BF_3, and $B_3N_3H_6$ to provide the boron and N_2 and NH_3 the nitrogen, solely produce the noncubic, sp^2-hybridized version of boron nitride that is amorphous or poorly crystallized. The hot-filament CVD so widely used to make diamond films does not appear to be successful for c-BN deposition. Plasma-assisted CVD methods for depositing c-BN have been reviewed by Konyashin *et al.* (Ref. 64). Typically, the same gases have been reacted together in a variety of RF, DC, ECR, laser enhanced, microwave, pulsed plasma, and inductively coupled plasma CVD reactors.

To explain the often contradictory effects of deposition variables on c-BN content in the films, these authors distinguish between *physical* and *chemical* routes taken in plasma CVD processes. The physical route generally means high-energy bombardment of biased substrates by Ar^+ and N^+ ions. Compressively stressed amorphous and hexagonal-BN layers deposit that are poorly adherent. Such structures are also seen during physical vapor deposition of BN films, where higher c-BN contents, albeit with an sp^2-hybridized BN interlayer, generally deposit. In the chemical route, films deposit in much the same was as diamond, i.e., a hydrogen-rich plasma with high concentrations of atomic hydrogen. The latter etches away the sp^2 bonds and converts them to sp^3 bonds, enabling c-BN crystals to grow directly from the seeded substrate. Schematically depicted in Fig. 6-25 is a

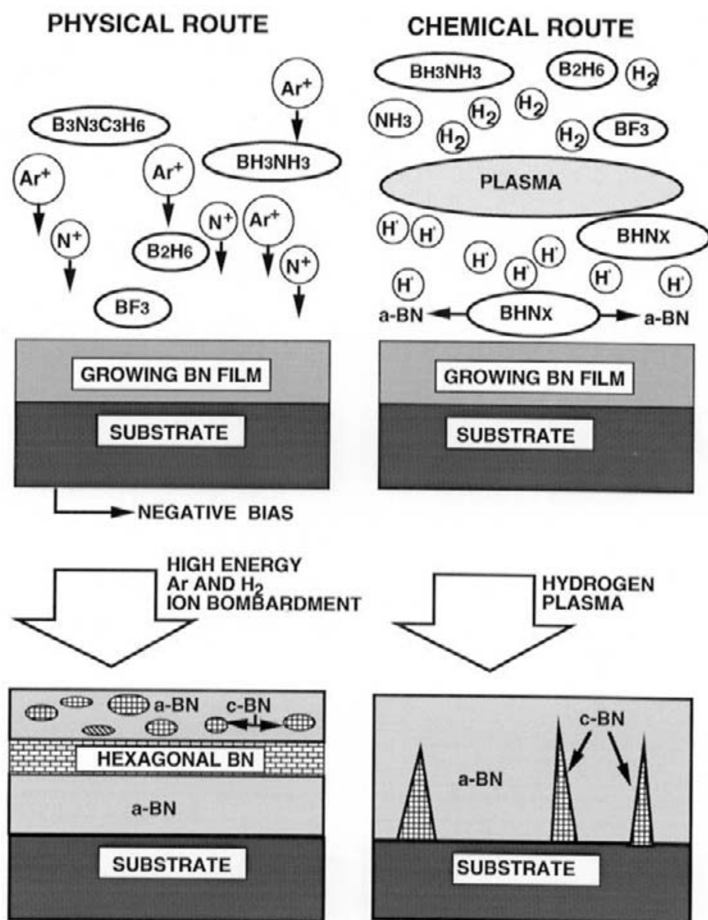


Figure 6-25 Illustration of physical (left) and chemical (right) routes in the production of boron nitride films. In the physical route Ar and N ion bombardment of a negatively biased substrate is emphasized, whereas PECVD in a hydrogen plasma is featured in the chemical route. Note the difference in c-BN growth morphology in each case. (From Ref. 64. Reprinted with the permission of Professor F. Aldinger.)

model for this reaction as well as the distinction between the two routes. Relative to bulk c-BN, the properties of deposited boron nitride films have clearly not advanced to the state enjoyed by diamond films relative to bulk diamond.

Deposition of carbon nitride films has many of the earmarks of the earlier quest for the synthesis of diamond films. The excitement began when Liu

and Cohen (Ref. 65) theoretically predicted that the bulk modulus of the hypothetical compound β-C_3N_4, which is structurally similar to β-Si_3N_4, should be 427 GPa; this compares with 440 GPa for diamond. Further, they suggested this phase, like diamond, has a high atomic density coupled with short strong bonds and could be metastable under ambient conditions. Badding (Ref. 66) has claimed that there is no convincing evidence for the successful synthesis of either crystalline or amorphous carbon nitrides containing primarily sp^3-bonded carbon. Nevertheless, a hardness of 6500 kg/mm^2 (64 GPa) has been reported in films physically deposited by ion-beam assisted methods (Ref. 67). Although PVD methods have been primarily used to deposit assorted carbon nitrides, CVD processes to do this have also been reported (Ref. 67).

6.8.6 Selective Deposition of Films

Of all the attractive features of CVD, perhaps none is as intriguing as the ability to selectively deposit thin films in desired regions and nowhere else. Although CVD is usually thought of as a large-area or blanket deposition technique, there are ways of promoting deposition on selected areas. Different techniques and mechanisms are involved depending whether the substrate is unpatterned or patterned. Selectivity on unpatterned or broad-area substrates by means of lasers has already been discussed in Section 6.6.5.

It is a much greater challenge to selectively deposit films on a patterned rather than blanket substrate. The former is the case in microelectronics where the geometric complexity and tiny dimensions involved preclude beam writing. Instead, selectivity is based on differences in interfacial chemical reactions taking place on different materials (Refs. 68, 69). Ideally, film nucleation must be respectively simultaneously stimulated and inhibited on the desired and undesired surfaces. An outstanding example of CVD selective growth is shown in Fig. 6-26, where tungsten is seen to deposit on exposed Si and not on the neighboring SiO_2. Although tungsten can be deposited by thermal dissociation of WF_6, reduction by hydrogen makes it possible to produce W films at a lower temperature of 300°C according to the chemical reaction of Eq. 6-4. But deposition of W occurs indiscriminately on all surfaces because the source (WF_6) and reducing agent (H_2) are gases. However, if the reducing agent is a solid, e.g., patterned elemental Si, then reaction occurs selectively on the exposed Si according to the reaction

$$2WF_6 + 3Si \rightarrow 2W + 3SiF_4. \quad (6\text{-}52)$$

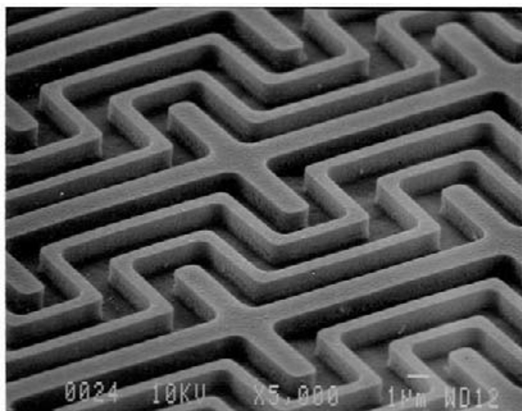


Figure 6-26 Scanning electron micrograph of an array of tungsten crosses and meandering lines selectively deposited over silicon through reduction of WF_6. Between the deposited tungsten is the uncoated SiO_2 substrate. (Courtesy of K. P. Rodbell, IBM, T. J. Watson Research Center.)

The Si in neighboring SiO_2 cannot act as a reducing agent because it is already fully oxidized. As long as it is exposed to WF_6, elemental Si continues to be sacrificed and selective growth proceeds. But after a while a thin film of W will fully cover the Si and the reaction ceases; growth is thus self-limiting. To grow thicker W films an excess concentration of H_2 gas must be now added to reduce a diluted WF_6 source gas. This results in a low supersaturation of W in the gas phase, which makes deposition on the previous selectively deposited tungsten more likely. Higher W supersaturation levels would more likely cause nucleation and growth everywhere including SiO_2, thus destroying selectivity in the process.

A similar mechanism for selectivity is predicted to be operative for deposition of W on Al, Ti, $TiSi_2$, and $CrSi_2$ simply based on thermodynamic considerations; these substrates yield driving forces for WF_6 reduction as high as that for Si (Ref. 68). For example, with a concentration ratio of $H_2/WF_6 = 40$, at a total pressure of 0.75 torr, the free energies for reactions with the above four metals (e.g., $WF_6 + Al \rightarrow W + AlF_3$) are, respectively, -27.7, -17.2, -18.5, and -17.8 kJ/mol at 500 K. These values compare with -18.4 kJ/mol for Si. The review by Hampden-Smith and Kodas (Ref. 41) lists mechanisms for selective deposition of metal films and the metals that are amenable to such processes; for example, volatile transition-metal organic-precursors (e.g., containing Cu) can selectively deposit on specific metals (e.g., Pt) relative to the nongrowth SiO_2 surface, in a replacement type reaction.

In Section 8.7.3 we shall continue the discussion of selective deposition as it applies to epitaxial films.

6.9 SAFETY

The safe handling of gases employed in CVD systems is a concern of paramount importance. Safety is part of the larger concern for chemical hazards in the semiconductor industry. This subject has been reviewed (Ref. 70) with respect to risk assessment, redundant control technology, hardware issues, and government regulation. In the case of CVD reactant or product gases we are typically dealing with toxic, flammable, pyrophoric, or corrosive substances. Because they frequently possess a combination of these attributes, they present particular health hazards to humans. Exposure of reactor hardware and associated gas-handling equipment to corrosive environments also causes significant maintenance problems and losses due to downtime. Table 6-7 contains a list of gases commonly employed in CVD processes together with some of their characteristics. A simple entry in the table does not accurately reflect the nature of the gas in practice. Silane, for example, more so than other gases employed in the semiconductor industry, has an ominous and unpredictable nature. It is stable but pyrophoric so that it ignites on contact with air. If it accumulates in a stagnant air space, however, the resulting mixture may explode upon ignition. In simulation tests of leaks, high flow rates of silane have resulted in violent explosions. For this reason, silane cylinders are stored outside buildings in concrete bunkers. The safety problems are magnified in low-pressure processing where concentrated gases are used. For example, in the deposition of polysilicon pure silane is used during LPCVD, whereas only a few percent silane is employed in atmospheric-pressure CVD processing. Often cited for their safety, many MOCVD precursors are pyrophoric.

Corrosive attack of gas-handling equipment (e.g., valves, regulators, piping) occurs in virtually all CVD systems. The problems are particularly acute in LPCVD processing because of the damage to mechanical pumping systems. Since many reactors operate at high temperatures, the effluent gases are very hot and capable of further downstream reactions in the pumping hardware. Furthermore, the exhaust stream generally contains corrosive species such as acids, oxidizers, and unreacted halogenated gases, in addition to large quantities of abrasive particulates. In semiconductor processing, for example, silica and silicon nitride particles are most common. All of these products are ingested by the mechanical pumps and the chamber walls become coated with precipitates or particu-

Table 6-7
Hazardous Gases Employed in CVD

Gas	Corrosive	Flammable	Pyrophoric	Toxic	Bodily hazard
Ammonia (NH_3)	x			x	Eye and respiratory irritation
Arsine (AsH_3)		x		x	Anemia, kidney damage, death
Boron trichloride (BCl_3)	x				
Boron trifluoride (BF_3)	x				
Chlorine (Cl_2)	x			x	Eye and respiratory irritation
Diborane (B_2H_6)		x	x	x	Respiratory irritation
Dichlorosilane (SiH_2Cl_2)	x	x			
Germane (GeH_4)		x		x	
Hydrogen chloride (HCl)	x				
Hydrogen fluoride (HF)	x				Severe burns
Hydrogen (H_2)		x			
Phosphine (PH_3)		x	x	x	Respiratory irritation, death
Phosphorus pentachloride (PCl_5)	x				
Silane (SiH_4)		x	x	x	
Silicon tetrachloride ($SiCl_4$)	x				
Stibine (SbH_3)		x		x	

late crusts. Oils used are degraded through polymerization and incorporation of solids. This hampers the lubrication of moving parts and hardware, and they tend to corrode and wear out more readily. All of this is a small price to pay, however, for the wonderful array of film materials that CVD has made possible.

6.10 CONCLUSION

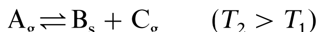
Instead of the physical transfer of atoms from a condensed evaporation source or sputtering target to the substrate, chemical vapor deposition primarily relies on gas phase and gas–solid chemical reactions to produce thin films. Because they are subject to thermodynamic and kinetic limitations and constrained by the flow of gaseous reactants and products, CVD processes are generally more complex than those involving PVD. However, guidelines adopted from the disciplines of physical chemistry, heat and mass transport, and fluid dynamics have provided a useful basis for analyzing and modeling these processes. Just as there are liquid and solid solutions, we may think of chemically reactive gas solutions in CVD, where an insoluble solid thin film can be made to controllably precipitate on a substrate under suitable temperatures and pressures. This chapter has been concerned with the strategies used to ensure that this gas-phase precipitation occurs at appreciable rates, to yield amorphous, polycrystalline, and epitaxial films with desired film compositions, morphologies, and structures.

An impressive number of different CVD materials (metals, elemental and compound semiconductors, oxides, nitrides, carbides, diamond, etc.) are of scientific and technical interest for a variety of electronically, optically, mechanically, and environmentally functional purposes. These materials and applications needs have been fulfilled through a variety of CVD processes involving different reactor designs and operating conditions. On the one hand there are thermal CVD processes that are conducted at both high and low temperatures, maintained at atmospheric as well as reduced pressures. Alternatively there are low-pressure, plasma-enhanced CVD processes that have been developed to deposit films at lower temperatures and on thermally sensitive substrates. In both categories the number of suitable gaseous precursors for CVD processes has been expanded through the synthesis of volatile metalorganic compounds. Developed particularly for the deposition of epitaxial semiconductor films, MOCVD processes are now increasingly used for the deposition of multicomponent oxide films as well. Included among the many desirable features of CVD processes is the unique capability of selective deposition.

In concluding this chapter our discussion of film deposition processes has ended for the moment. However, we shall return to other CVD processes in Chapter 8 and consider the demanding application of the epitaxial deposition of compound semiconductors.

EXERCISES

1. Consider the generic reversible CVD reaction



$$A_g \rightleftharpoons B_s + C_g \qquad (T_2 > T_1)$$

at 1 atm pressure ($P_A + P_C = 1$), where the free energy of the reaction is $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. Through consideration of the equilibria at T_1 and T_2:
   (a) Derive an expression for $\Delta P_A = \Delta P_A(T_1) - \Delta P_A(T_2)$ as a function of T, ΔH°, and ΔS°.
   (b) Plot ΔP_A as a function of ΔH°.
   (c) Comment on the gas transport direction and magnitude as a function of the sign and value of ΔH°.

2. It is desired to deposit epitaxial silicon films from a $SiHCl_3$ precursor gas in an atmospheric pressure CVD reactor.
   (a) Write a chemical reaction that would accomplish this.
   (b) Is this deposition reaction endothermic or exothermic? Why?
   (c) Based on thermodynamics, i.e., that $\Delta G^\circ \simeq 0$, what deposition temperature would you recommend?
   (d) What deposition temperature would you suggest based on kinetics considerations?

3. Plot $\ln \dfrac{P^4_{HCl}}{P_{SiCl_4} P^2_{H_2}}$ vs $1/TK$ for the temperature range 800 to 1500 K, using the results of Fig. 6-6.
   (a) What is the physical significance of the slope of this Arrhenius plot?
   (b) Calculate ΔH° for the reaction given by Eq. 6-21a, using data in Fig. 6-5.

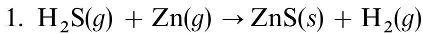
4. In growing epitaxial Ge films by the disproportionation reaction of Eq. 6-14, the following thermodynamic data apply:

$I_2(g) = 2I(g)$ $\quad \Delta G^\circ = -38.4T$ cal/mol
$Ge(s) + I_2(g) = GeI_2(g)$ $\quad \Delta G^\circ = -1990 - 11.2T$ cal/mol
$Ge(s) + GeI_4(g) = 2GeI_2(g)$ $\quad \Delta G^\circ = 36{,}300 - 57.5T$ cal/mol

   (a) What is ΔG° for the reaction $Ge(s) + 2I_2(g) = GeI_4(g)$?
   (b) Suggest a reactor design. Which region is hotter? Which is cooler?
   (c) Roughly estimate the operating temperature of the reactor.
   (d) Suggest how you would change the reactor conditions to deposit polycrystalline films.

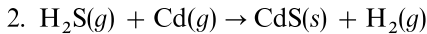
5. Assume you are involved in a project to deposit ZnS and CdS films for infrared optical coatings. Thermodynamic data reveal

1. $H_2S(g) + Zn(g) \rightarrow ZnS(s) + H_2(g)$



$$\Delta G^\circ = -76{,}400 + 82.1T - 5.9T \ln T \text{ (cal/mol)}$$

2. $H_2S(g) + Cd(g) \rightarrow CdS(s) + H_2(g)$



$$\Delta G^\circ = -50{,}000 + 85.2T - 6.64T \ln T \text{ (cal/mol)}$$

(a) Are these reactions endothermic or exothermic?

(b) In practice, reactions 1 and 2 are carried out at 680°C and 600°C, respectively. From the vapor pressures of Zn and Cd at these temperatures, estimate the P_{H_2}/P_{H_2S} ratio for each reaction, assuming equilibrium conditions.

(c) Recommend a reactor design to grow either ZnS or CdS, including a method for introducing reactants and heating substrates.

6. The disproportionation reaction

$$Si + SiCl_4 = 2SiCl_2 \ (\Delta G^\circ = 83{,}000 + 3.64T \log T - 89.4T \text{ (cal/mol)})$$

is carried out in a closed tubular atmospheric pressure reactor whose diameter is 15 cm. Deposition of Si occurs on a substrate maintained at 750°C and located 25 cm away from the source, which is heated to 900°C. Assuming thermodynamic equilibrium prevails at source and substrate, calculate the flux of $SiCl_2$ transported to the substrate if the gas viscosity is 0.08 cP.

7. Consider the deposition of silicon carbide utilizing two independent source gases CH_4 and $SiCl_4$ according to Eq. 6-9. According to the Gibbs phase rule, what thermodynamic variables and compositions would have to be specified before the system is determinate?

8. (a) In integrated circuits, films of BPSG are deposited by thermal rather than plasma CVD methods. Provide at least one reason why.
   (b) Interlevel dielectrics, i.e., SiO_2, are deposited by plasma CVD methods. Why?
   (c) The ultrathin gate oxide (SiO_2) is thermally grown by oxidation of silicon rather than deposited by a thermal CVD or PECVD process. Why?
   (d) Tungsten films are deposited selectively using WF_6. Attempts to employ a tungsten carbonyl precursor for this purpose proved unsuccessful. Why?

9. Polysilicon deposits at a rate of 30 Å/min at 540°C. What deposition rate can be expected at 625°C if the activation energy for film deposition is 1.65 eV?

10. Tetrachlorosilane diluted to 0.5 mol% in H_2 gas flows through a 12-cm-diameter, tubular, atmospheric CVD reactor at a velocity of 20 cm/s. Within the reactor is a flat pallet bearing Si wafers resting

horizontally. If the viscosity of the gas is 0.03 cP at 1200°C:

(a) What is the Reynolds number for the flow?

(b) Estimate the boundary layer thickness at a point 5 cm down the pallet.

(c) If epitaxial Si films deposit at a rate of 1 μm/min, estimate the diffusivity of Si through the boundary layer.

11. Find the stoichiometric formula for the following films:

(a) PECVD silicon nitride containing 20 at.% H with a Si/N ratio of 1.2.

(b) LPCVD silicon nitride containing 6 at.% H with a Si/N ratio of 0.8.

(c) LPCVD SiO_2 with a density of $2.2\,g/cm^3$, containing 3×10^{21} H atoms/cm^3.

12. Consider a long tubular CVD reactor in which one-dimensional steady-state diffusion and convection processes occur together with a homogeneous first-order chemical reaction. Assume the concentration $C(x)$ of a given species satisfies the ordinary differential equation

$$D\frac{d^2C}{dx^2} - v\frac{dC}{dx} - kC = 0,$$

where k is the chemical rate constant and x is the distance along the reactor.

(a) If the boundary conditions are $C(x = 0) = 1$ and $C(x = 1\,m) = 0$, derive an expression for $C(x)$.

(b) If $C(x = 0) = 1$ and $dC/dx\ (x = 1\,m) = 0$, derive an expression for $C(x)$.

(c) Calculate expressions for the concentration profiles if $D = 1000$ cm^2/s, $v = 100$ cm/s, and $k = 1\,s^{-1}$. *Hint:* A solution to the differential equation is $\exp \alpha x$, where α is a constant.

13. An ion nitriding system operates at a nitrogen pressure of 1 torr. Nitriding was done at 500°C, and the diffusivity of N in steel is given by

$$D = 0.003 \exp - \frac{18.2\ \text{kcal/mol}}{RT}\ \text{cm}^2/\text{s}.$$

After a 3-h exposure of a tool steel to the discharge, the nitrogen content rose from zero to 0.5% at a distance 10 μm beneath its surface.

(a) What was the discharge current density?

(b) Approximately what fraction of the nitrogen was ionized?

14. Select any film material (e.g., semiconductor, oxide, nitride, carbide, metal alloy) that has been deposited or grown by both PVD and CVD methods. In a report, compare the resultant structure, stoichiometries, and properties. The *Journal of Vacuum Science and Technology* and *Thin Solid Films* are good references for such information.

15. As manager of a coating facility with complete PVD and CVD capabilities, recommend the tooling and procedures necessary to coat the following items:
    (a) The internal cooling passages and external surfaces of cast superalloy turbine blades used in jet engines must be coated with 5 μm of ZrO_2 to thermally protect them.
    (b) Small spherical sapphire (Al_2O_3) microlenses ($\sim$1 mm diameter) used in fiber optic communication devices such as optical transmitters and receivers must be uniformly coated with 2240 Å of SiO_2 for antireflection purposes.
    (c) Kilometer-long lengths of silica optical fiber, 125 μm in diameter, must be coated with 200 Å of tungsten for hermetic sealing purposes.
16. A gas mixture flows along a tubular reactor of cross-sectional area A with velocity v. Downstream along the x direction, a first-order chemical reaction with rate constant k occurs on substrates of area A_s, thus continuously depleting the gas.
    (a) Show that the gas concentration declines exponentially according to $\sim C_0 \exp -(kA_s ax/vA_0)$, where C_0 is the initial gas concentration and a is a constant.
    (b) How does the gas concentration change with time?
    (c) Show that the film growth rate remains roughly constant if the temperature increases with distance along the reactor.
17. List several ways to suppress unstable gas flows that adversely affect films deposited in CVD reactors.

REFERENCES

1. E. M. Sherwood and J. M. Blocher, *J. Metals* **17**, 594 (1965).
2. S. Sivaram, *Chemical Vapor Deposition — Thermal and Plasma Deposition of Electronic Materials*. Van Nostrand Reinhold, New York, 1995.
3. M. J. Hampden-Smith and T. T. Kodas, eds., *The Chemistry of Metal CVD*. VCH, Weinheim, 1994.
4. M. L. Hitchman and K. F. Jensen, eds., *Chemical Vapor Deposition, Principles and Applications*. Academic Press, London, 1993.
5. H. O. Pierson, *Handbook of Chemical Vapor Deposition, Principles, Technology and Applications*. Noyes, Park Ridge, NJ, 1992.
6. A. Sherman, *Chemical Vapor Deposition for Microelectronics*. Noyes, Park Ridge, NJ, 1987.
7. Several review articles are devoted to plasma processing in *IBM J. Res. Develop.* **43**(1/2) (1999).
8. Several review articles on thermal and plasma-assisted CVD appear in *Thin Film Processes II*, eds. J. L. Vossen and W. Kern. Academic Press, New York, 1991.
9. W. Kern, in *Microelectronic Materials and Processes*, ed. R. A. Levy. Kluwer Academic, Dordrecht, The Netherlands, 1989.

10. W. Kern and V. S. Ban, in *Thin Film Processes*, eds. J. L. Vossen and W. Kern. Academic Press, New York, 1978.
11. J. Schlichting, *Powder Metal Int.* **12**(3), 141 (1980).
12. E. S. Wajda, B. W. Kippenhan, and W. H. White, *IBM J. Res. Develop.* **7**, 288 (1960).
13. R. A. Laudise, *The Growth of Single Crystals.* Prentice Hall, Englewood Cliffs, NJ, 1970).
14. E. Sirtl, I. P. Hunt, and D. H. Sawyer, *J. Electrochem. Soc.* **121**, 919 (1974).
15. V. S. Ban and S. L. Gilbert, *J. Electrochem. Soc.* **122**(10), 1382 (1975).
16. S. O. Hay, W. C. Roman, and M. B. Colket, *J. Mater. Res.* **5**, 2387 (1990).
17. T. C. Anthony, A. L. Fahrenbruch, and R. H. Bube, *J. Vac. Sci. Technol.* **A2**(3), 1296 (1984).
18. G. Perrier and R. Philippe, *J. Mater. Res.* **3**(5), 1031 (1988).
19. S. Nishino, K. Matsumoto, T. Yoshida, Y. Chen, and S. K. Lilov, *Mat. Sci. Eng.* **B61-62**, 121 (1999).
20. H. H. Lee, *Fundamentals of Microelectronics Processing.* McGraw-Hill, New York, 1990.
21. S. Middleman and A. K. Hochberg, *Process Engineering Analysis in Semiconductor Device Fabrication.* McGraw-Hill, New York, 1993.
22. K. F. Jensen and W. Kern, in *Thin Film Process II*, eds. J. L. Vossen and W. Kern. Academic Press, Boston, 1991.
23. C. Kleijn, "Transport Phenomena in Chemical Vapor Deposition Reactors." Ph.D. thesis, Technical University of Delft, The Netherlands (1991).
24. P. C. Rundle, *Int. J. Electron.* **24**, 405 (1968).
25. M. L. Hitchman and K. F. Jensen, in *Chemical Vapor Deposition, Principles and Applications*, eds. M. L. Hitchman and K. F. Jensen. Academic Press, London, 1993.
26. A. S. Grove, *Physics and Technology of Semiconductor Devices.* John Wiley & Sons, New York, 1967.
27. W. S. Ruska, *Microelectronic Processing.* McGraw-Hill, New York, 1987.
28. D. S. Rickerby and A. Matthews, eds., *Advanced Surface Coatings.* Blackie (Chapman and Hall), Glasgow, 1991.
29. J. Li, J. P. McVittie, J. Ferziger, and K. C. Saraswat, *J. Vac. Sci. Technol.* **B13**, 1867 (1995).
30. S. E. Lassig and J. D. Tucker, *Microelectron. J.* **26**(8), xi (1995).
31. S. Nag, R. Ramamurthy, W. J. Lei, C. Montell, and M. Hickey, *Solid State Technol.* **41**(9), 69 (1998).
32. H. M. Manasevit and W. I. Simpson, *Appl. Phys. Lett.* **12**, 172 (1975).
33. B. W. Wessels, *Ann. Rev. Mater. Sci.* **23**, 525 (1995).
34. V. R. Palkar, S. C. Purandare, and R. Pinto, *J. Phys. D: Phys.* **32**, R1 (1999).
35. D. E. Kotecki *et al.*, *IBM J. Res. Develop.* **43**(3), 367 (1999).
36. A. C. Jones, *Chem. Vap. Deposition* **4**(5), 169 (1998).
37. G. Braunstein and G. R. Paz-Pujalt, *Thin Solid Films* **216**, 1 (1992).
38. B. A. Tuttle and R. W. Schwartz, *MRS Bull.* **21**(6), 49 (1996).
39. T. H. Baum and P. B. Comita, *Thin Solid Films* **218**, 84 (1992).
40. M. J. Hampden-Smith and T. T. Kodas, *Chem. Vap. Deposition* **1**, 8 (1995).
41. M. J. Hampden-Smith and T. T. Kodas, *Chem. Vap. Deposition* **1**, 39 (1995).
42. V. R. McCrary and V. M. Donnelly, in *Chemical Vapor Deposition, Principles and Applications*, eds. M. L. Hitchman and K. F. Jensen. Academic Press, San Diego, 1993.
43. S. Veprek, *Thin Solid Films* **130**, 135 (1985).
44. K.-T. Rie, E. Menthe, A. Matthews, K. Legg, and J. Chin, *MRS Bull.* **21**(8), 46 (1996).
45. G. Lucovsky, D. V. Tsu, R. A. Rudder, and R. J. Markunas, in *Thin Film Processes II*, eds. J. L. Vossen and W. Kern. Academic Press, New York, 1991.
46. S. V. Nguyen, *IBM J. Res. Develop.* **43**(1/2), 109 (1999).
47. Y. Kuo, ed., *IBM J. Res. Develop.* **43**(1/2) (1999).

48. D. R. Cote, S. V. Nguyen, A. K. Stamper, D. S. Armbrust, D. Tobben, R. A. Conti, and G. Y. Lee, *IBM J. Res. Develop.* **43**(1/2), 5 (1999).
49. L. E. Klein and M. J. Kushner, *Crit. Rev. Solid State Mater. Sci.* **16**, 1 (1989).
50. S. Hamaguchi, *IBM J. Res. Develop.* **43**(1/2), 199 (1999).
51. E. Meeks, R. S. Larson, P. Ho, C. Apblett, S. M. Han, E. Edelberg, and E. S. Aydil, *J. Vac. Sci. Technol.* **A16**(2), 544 (1998).
52. R. B. Marcus and T. T. Sheng, *Transmission Electron Microscopy of Silicon VLSI Circuits and Structures*. John Wiley & Sons, New York, 1983.
53. D. L. Smith, *Thin-Film Deposition*. McGraw-Hill, New York, 1995.
54. J. Perrin, in *Plasma Processing of Semiconductors*, ed. P. F. Williams. Kluwer, Dordrecht, The Netherlands, 1997.
55. D. W. Hess and D. B. Graves, in *Chemical Vapor Deposition, Principles and Applications*, eds. M. L. Hitchman and K. F. Jensen. Academic Press, San Diego, 1993.
56. A. C. Adams, in *VLSI Technology*, 2nd ed., ed. S. M. Sze. McGraw-Hill, New York, 1988.
57. J. R. Hollahan and S. R. Rosler, in *Thin Film Processes*, eds. J. L. Vossen and W. Kern. Academic Press, New York, 1978.
58. W. von Bolton, *Z. Electrochem.* **17**, 971 (1911).
59. P. D. Bridgeman, *Sci. Am.* **233**, 102 (1975).
60. R. C. DeVries, *Ann. Rev. Mater. Sci.* **7**, 161 (1987).
61. K. V. Ravi, C. A. Koch, H. S. Hu, and A. Joshi, *J. Mater. Res.* **5**, 2356 (1990).
62. A. Grill, *IBM J. Res. Develop.* **43**, 147 (1999).
63. H. C. Tsai and D. B. Bogy, *J. Vac. Sci. Technol.* **A5**, 3287 (1987).
64. I. Konyashin, J. Bill, and F. Aldinger, *Chem. Vap. Deposition* **3**(5), 239 (1997).
65. A. Y. Liu and M. L. Cohen, *Science* **245**, 841 (1989).
66. J. V. Badding, *Adv. Mater.* **9**(11), 877 (1997).
67. T. Inoue, S. Ohshio, H. Saitoh, and K. Kamata, *Appl. Phys. Lett.* **67**, 353 (1995).
68. J.-O. Carlsson, *Crit. Rev. Solid State Mater. Sci.* **16**(3), 161 (1990).
69. J-O. Carlsson and U. Jansson, *Prog. Solid State Chem.* **22**, 237 (1993).
70. C. F. Chelton, M. Glowatz, and J. A. Masovsky, *IEEE Trans. Education* **34**(3), 269 (1991).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 7

Substrate Surfaces and Thin-Film Nucleation

7.1 INTRODUCTION

In keeping with the theme and subtitle, the preceding half of the book was broadly devoted to the *deposition* of thin films. This chapter is the first of three that directly deal with the multifaceted concerns of the second theme — film *structure*. These three chapters encompass the progression of events that starts with the condensation of isolated atomic clusters on a bare substrate, continues with film thickening due to additional deposition, and ends with fully developed epitaxial single-crystal, polycrystalline, or amorphous films and coatings. Included will be issues related to the physical and crystallographic nature of substrates, film–substrate interfacial interactions, and the evolution of the film grain structure and morphology. The overall objective of these three chapters is to reconcile film structure with prior processing on the one hand, and serve as a springboard for understanding film properties on the other. In this sense these chapters are the filling sandwiched between prior and subsequent chapters.

Interest in thin-film formation dates at least to the 1920s. In the course of research at the Cavendish Laboratories in England on evaporated thin films, the concept of formation of nuclei which grew and coalesced to form the film was advanced (Ref. 1). During the earliest stages of film formation a sufficient number of vapor atoms or molecules condense and establish a permanent residence on the substrate. Many such film birth events occur in this so-called nucleation stage. Numerous investigations employing assorted

techniques, e.g., transmission and scanning electron microscopy, Auger-electron spectroscopy, and now scanning probe microscopies, have focused on the earliest stages of film formation. These studies typically reveal that the demarcation between the end of nucleation and the onset of nucleus growth is not sharp. The sequence of film nucleation and growth events can be well appreciated in the transmission-electron microscopy images of Fig. 7-1. Soon after exposure of the substrate to the incident vapor, a uniform distribution of small but highly mobile clusters or islands is observed. In this stage the prior nuclei incorporate impinging atoms and subcritical clusters and grow in size while the island density rapidly saturates. The next stage involves merging of the islands by a coalescence phenomenon which, in some cases, is liquid-like, especially at high substrate temperatures. Coalescence decreases the island density resulting in local denuding of the substrate where further nucleation can then occur. Crystallographic facets and orientations are frequently preserved on islands and at interfaces between initially disoriented, coalesced particles. Continued coalescence results in the development of a connected network with unfilled channels in between. With further deposition, the channels fill in and shrink leaving isolated voids behind. Finally, even the voids fill in completely and the film is said to be continuous. This sequence of events occurs during the early stages of deposition, typically accounting for the first few hundred angstroms of film thickness (Ref. 2).

Many observations of subsequent film formation have pointed to three basic growth modes: (1) *island* (or Volmer–Weber), (2) *layer* (or Frank–Van der Merwe), and (3) *Stranski–Krastanov*, which are illustrated schematically in Fig. 7-2. Island growth occurs when the smallest stable clusters nucleate on the substrate and grow in three dimensions to form islands. This happens when atoms or molecules in the deposit are more strongly bound to each other than to the substrate. Metal and semiconductor films deposited on oxide substrates initially form such islands. The opposite characteristics are displayed during layer growth. Here the extension of the smallest stable nucleus occurs overwhelmingly in two dimensions, resulting in the formation of planar sheets. In this growth mode the atoms are more strongly bound to the substrate than to each other. The first complete monolayer is then covered with a somewhat less tightly bound second layer. Providing the decrease in bonding energy is continuous toward the bulk-crystal value, the layer growth mode is sustained. Single-crystal epitaxial growth of semiconductor films, a subject treated in Chapter 8, is the most important example of this growth mode. The layer plus island or Stranski–Krastanov (S-K) growth mechanism is an intermediate combination of the preceding two modes. In this case after forming one or more monolayers, subsequent layer growth becomes unfavorable and islands form. Film growth by the

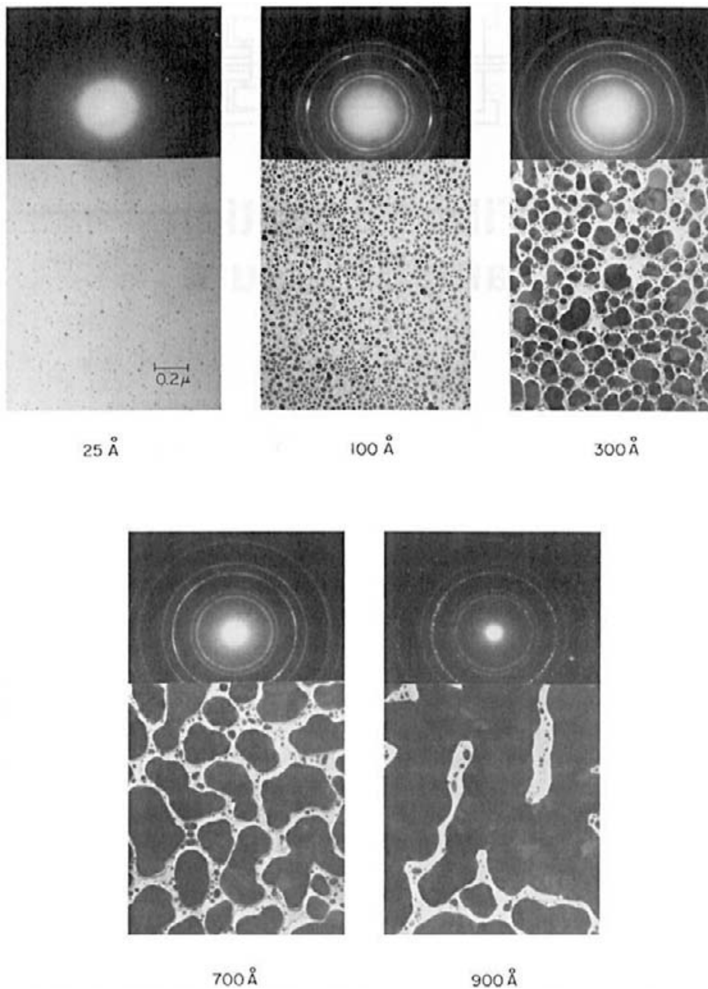


Figure 7-1 Transmission electron microscope images of nucleation, growth, and coalescence of Ag films on (111) NaCl substrates. Corresponding diffraction patterns are shown. (From Ref. 2, courtesy of R. W. Vook.)

S-K mode is fairly common and has been observed in metal–metal and metal–semiconductor systems.

This chapter traces the odyssey of nucleation events that starts with the clustering of isolated atoms on a substrate, follows with the incorporation of additionally deposited atoms, and ends after the first several layers of film

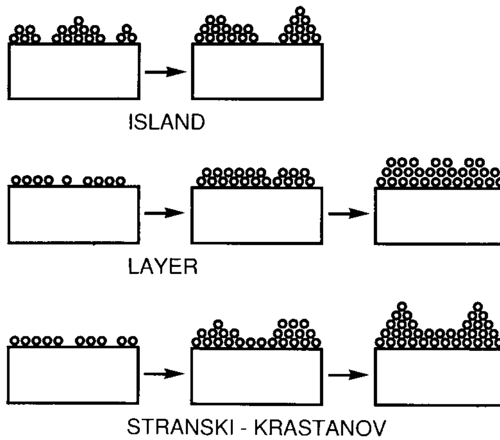


Figure 7-2 Basic modes of thin-film growth.

have formed. In this regime the nature of the substrate surface often plays an important role. We therefore start by discussing the nature of solid surfaces and see what happens when atoms impinge on them.

7.2 AN ATOMIC VIEW OF SUBSTRATE SURFACES

7.2.1 INTRODUCTION

Since they envelop and confine all solids including thin films and substrates, surfaces provide a key to understanding atomic behavior immediately beneath them as well as just above them. For many engineering purposes, these few atom layers collectively and seamlessly merge into what we call the "surface." Real surfaces are usually contaminated with adsorbed gases and assorted compound layers. However, in this chapter surfaces will generally represent an atomically sharp interfacial demarcation between condensed-phase and gas-phase atoms.

There are many important reasons for studying the nature of clean surfaces. First of all, substrate surfaces usually provide the template for subsequent deposition and growth of thin films. In addition, surfaces also serve as the vehicle for imparting interfacial structural, mechanical, and physical properties to films. Secondly, assorted surface-analytical techniques are widely employed to reveal the chemical composition and structure of films. Through stimulation by photons, electrons, and ion beams, the electronic and physical structures of surface and near-surface atoms are

altered in ways that can be detected, thus serving to fingerprint their identity and characterize their behavior. Thin films also confront the environment and interact chemically, mechanically, electrically, etc., with the ambient through their surfaces. And, as film thicknesses continue to shrink, which is the trend in many applications, a larger fraction of atoms will reside at the top film surface and bottom film–substrate interface; these surface atoms will then play increasingly greater roles in influencing the behavior of the atoms in between.

In this section we shall start by treating both the electronic and physical structures of subsurface atom layers of solids, and in particular, substrates. Then we will consider geometric placement of different atoms in the surface plane itself, our main interest being the first atomic layer of the film that deposits above the substrate surface plane. All in all perhaps only 15 Å of material at most is involved. Amazing as it seems, a profile through this vanishingly small thickness reveals that a minimum of two chemically different species interact in a continuum of bonding arrangements, and that at least three crystallographic structures can be differentiated. It is a testament to the great advances in surface science experimentation and theory that have enabled such distinctions to be made. An excellent review of the progress in this area to date is detailed in *Surface Science: The First Thirty Years* (Ref. 3). Importantly, through the electronic and optoelectronic devices it has made possible, surface science is intimately connected to the story of a "... technological revolution which is changing the world as we know it, transforming a planet into a global village."

7.2.2 Electronic Nature of Surfaces

7.2.2.1 Metals

Solid-state physics has long considered the quantum nature of bulk crystalline solids. By now we are all familiar with the notion that the periodic electrostatic potential in lattices is responsible for electron energy bands and gaps, as well as the classification of solids into metals, semiconductors, and insulators. Although solutions to the Schrodinger wave equation for the many-body periodic potential are rarely possible, useful approximations exist. For example, assorted theoretical models, e.g., free electron, nearly free electron, tight binding, and pseudopotential, have yielded wave functions and energies for electron states in metals (Ref. 4). The electronic structure is commonly depicted by the band diagram shown in Fig. 1-9a, a representation applicable to bulk regions of the metal. At the surface, however, the assumed uniform positive potential due to ion cores is abruptly terminated. As a consequence, the mobile electrons within the bulk

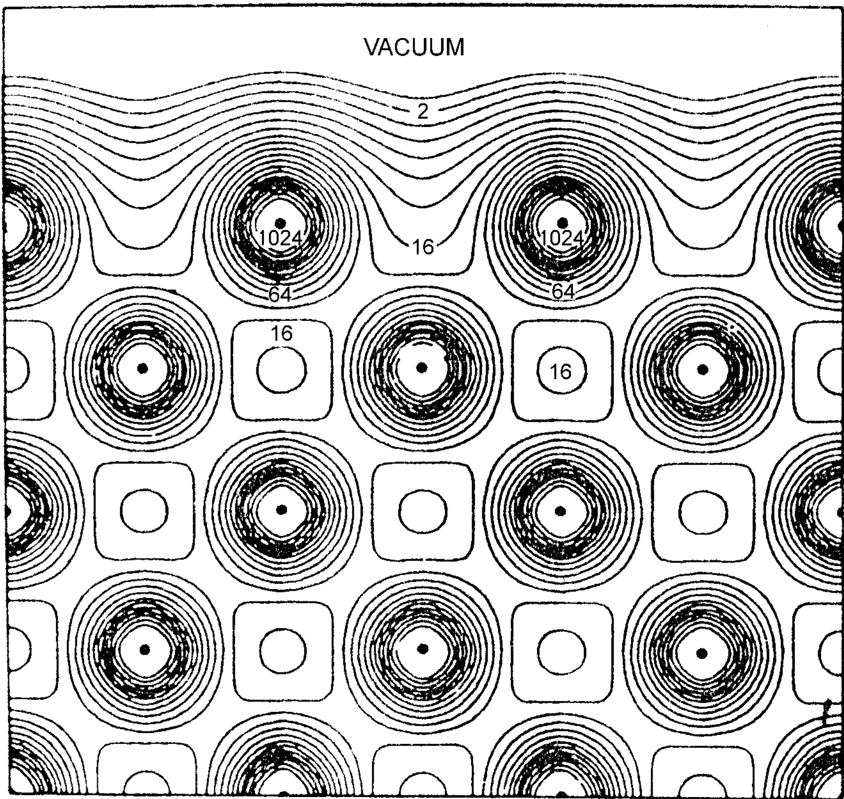


Figure 7-3 Electron charge density contours around Ni (001) atoms near the surface. (After F. J. Arlinghaus, J. G. Gay, and J. R. Smith, *Phys. Rev.* **B21**, 2055, (1980). Reprinted with the permission of John R. Smith.)

attempt to screen out the now unbalanced attractive potential at the surface. This results in an increased electron density at the surface which dies away rapidly in an oscillatory manner within the interior (Ref. 5). In solids with low electron densities the charge oscillations die away more slowly. Some of these features are shown in Fig. 7-3, where electron charge–density contours are plotted for nickel. At the surface the charge density is smeared out but it becomes periodic as the interior ion cores are reached. This smoothing of the surface charge is the compromise electrons strike when they lower both their potential and kinetic energies by respectively penetrating into the interior and by spreading out toward the surface.

Surface or Tamm states were first predicted based on Schrodinger equation solutions to electrons in terminated periodic lattices. From the

standpoint of quantum mechanics, the loss of three-dimensional lattice periodicity at the surface allows for quantum states that lie within the (bulk) energy gap (Ref. 6). Localized surface states are characterized by waves that travel parallel to the surface but not into the solid. When electron or hole charge occupies these states, electrostatic fields penetrate the interior and produce a variation of potential that is manifested as band bending. Adsorbed impurities, emergent dislocations, and surface relaxation (or reconstruction; see Section 7.2.4) also create surface states. But in all cases the extent of band bending in metals is generally negligible because there are sufficient numbers of mobile electrons to screen out or compensate for the surface-state charge; this is generally not true in semiconductors.

The work function (ϕ), an important surface property, is a consequence of the foregoing model. We may view ϕ as related to the difference in electron potential between that associated with the surface-charge relative to the average value within the bulk lattice. Thus $q\phi$, where q is the electronic charge, is the energy barrier electrons must surmount when leaving the solid. Measured values for ϕ in most elements cluster around 3.5 ± 1.5 V. Contributing to ϕ is a charged double layer generated by the electron asymmetry at the surface. Alteration of this double layer due to adsorbed contaminants, alloying, or the crystallographic face exposed typically changes the magnitude of ϕ by a few tenths of a volt.

7.2.2.2 Semiconductors

There are significant structural as well as electronic differences between semiconductor and metal surfaces. This fact is immediately apparent if we consider the interface between a (111), III–V compound semiconductor surface and vacuum shown in Fig. 7-4a. Forgetting distinctions between III– and V-type atoms for the moment, this ball-and-stick model is applicable to elemental semiconductors such as (111) silicon as well. The severed covalent bonds at the surface mean the presence of uncompensated charge and an electric field. However, dangling bonds from neighboring surface atoms bridge to complete the bonding requirements. In the process surface atoms shift position and the electrostatic energy in the vacuum is reduced. As we shall soon see (Sections 7.2.3.1 and 7.2.4.1), the surface reconstructs with important structural and property implications.

The same thing happens at III–V semiconductor surfaces but with interesting added twists. As the figure indicates, the surface consists entirely of Group V atoms (or light B atoms, e.g., As), but it could just as easily consist of Group III atoms (or dark A atoms, e.g., Ga). These unique and distinct surfaces arise because cuts can only be made between planes AA and BB. The reasons for such cuts are: (a) only one bond between atoms needs

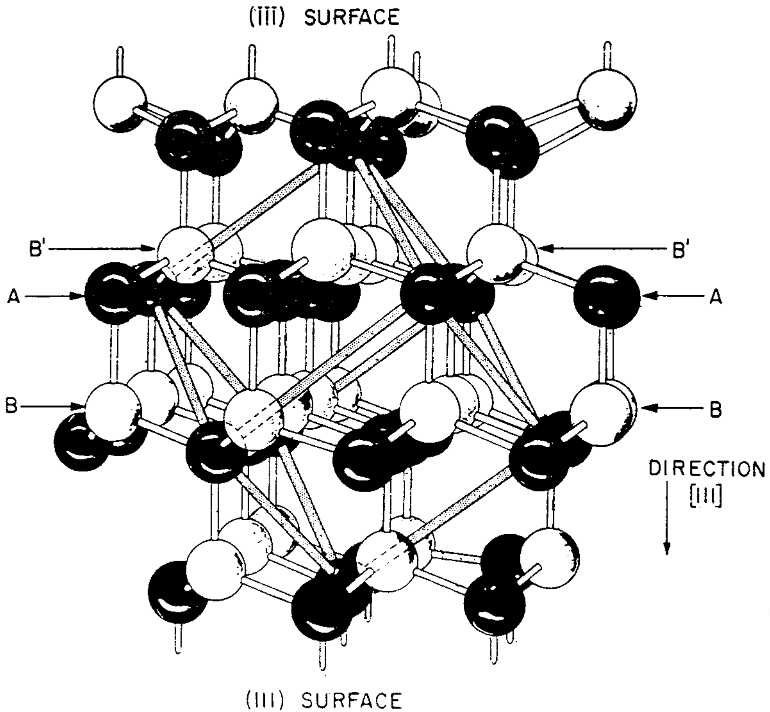


Figure 7-4a Model of (111) surfaces in the zinc blende crystal structure of a III–V compound semiconductor. (From H. C. Gatos, *Surf. Sci.* **299/230**, 1, (1994). Reprinted with the permission of Professor Harry C. Gatos, MIT.)

to be disrupted and (b) the surface atoms remain triply bonded to the lattice. Other cuts such as those between AA and B′B′ are not possible because three bonds must be severed and resultant surface atoms are singly bonded to the lattice. For the BB, so-called (111) surface, we note that only three of five valence electrons in each arsenic atom are attached or bonded, while on an AA (111) surface all three of boron's valence electrons are bonded. As a consequence, the As surface is more electronically active than the Ga surface, and can, for example, be smoothed and polished more readily by etching. In contrast, Ga surfaces etch slowly and tend to be rougher. Such differences lead to varied epitaxial phenomena when films are deposited on these surfaces.

Unlike metals, two bands — a lower valence as well as an upper conduction band — play important roles in semiconductors and devices based on them. Another difference is the band bending that occurs at semiconductor surfaces that contact other materials such as insulators, as shown in

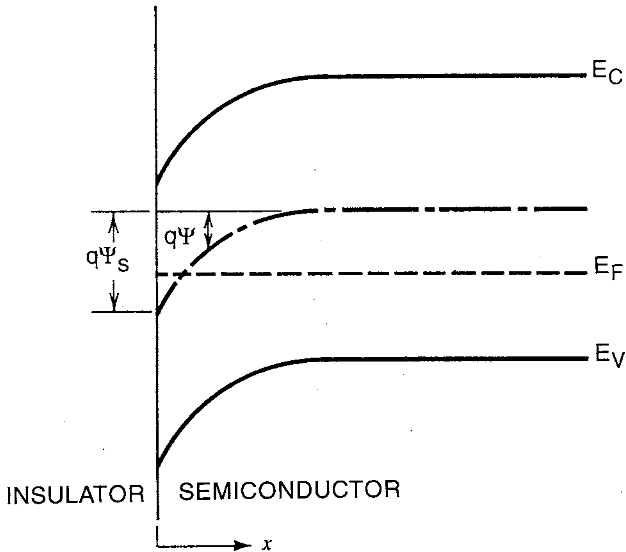


Figure 7-4b Energy band diagram at the surface of a p-type semiconductor. The potential ψ is defined to be zero in the bulk but assumes a nonzero value at the surface that depends on the phase contacting the semiconductor, e.g., vacuum, SiO_2, or other semiconductor. (After S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. John Wiley & Sons, New York, 1981.)

Fig. 7-4b. Band bending or the absence of a flat band condition means that a nonzero electrical potential ψ_s exists at the surface. What generates ψ_s are the severed covalent bonds at the semiconductor surface which are essentially localized surface-states that trap electrons or holes. Together with ionized donors or acceptors they create a spatially varying charge distribution near the surface. The situation parallels that of the plasma-sheath region (Section 4.3.5). A solution to the Poisson–Boltzmann equation in one dimension essentially yields a form of ψ similar to that for the electrostatic potential in a plasma (e.g., Eq. 4-17, also problem 5 in Chapter 4). Because there are relatively few carriers to shield surface charge, band bending is much more pronounced in semiconductors than in metals, e.g., the electrostatic potential typically penetrates $\sim 1\,\mu m$ below the surface. A space charge region is said to exist where the bands are bent. Within it the semiconductor exhibits dielectric behavior and supports the internal electric-field, i.e., $\mathscr{E} = -d\psi/dx$, that makes possible rectifying action at a p–n junction.

For semiconductor surfaces to perform electrical functions in devices they are often contacted by either insulating, dielectric, or conducting thin films.

In silicon, for example, these films are invariably either grown or deposited SiO_2, and deposited metals (M) that serve as electrical contacts. Because $Si{-}SiO_2$ and M–Si structures are of such great technical importance, these interfaces are perhaps the most studied in all of physical science. As a result there is a broad accessible literature on interfacial $Si{-}SiO_2$ charge and barriers to electron flow across M–Si contacts; for this reason these topics will not be discussed here.

7.2.3 SURFACE STRUCTURES

7.2.3.1 Introduction

As we have already seen, minimization of the overall electronic and bonding energies in the topmost atom layers causes the surface crystallography to differ from that of the bulklike interior as suggested by the schematic cross-sectional views of Fig. 7-5. If the surface structure is the

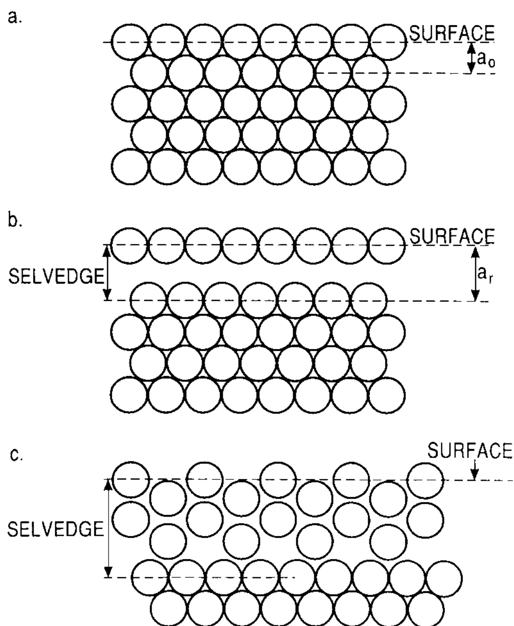


Figure 7-5 Schematic cross-sectional views of close-packed atomic positions at a solid surface. (a) Bulk exposed plane. (b) Atomic relaxation outward. (c) Reconstruction of outer layers. (From Ref. 5, © Oxford University Press, by permission.)

predictable extension of the underlying lattice, we have the case shown in Fig. 7-5a. However, the loss of periodicity in the vertical direction alters surface electronic properties and leaves energetic broken bonds. It is then more likely that the structure shown in Fig. 7-5b will prevail. The absence of bonding forces to underlying atoms results in new equilibrium positions which deviate from those in the bulk lattice. A disturbed surface layer known as the “selvedge” may then be imagined. Within this layer the atoms relax in such a way as to preserve the symmetry of the bulk lattice parallel to the surface, but not normal to it. A more extreme structural disturbance is depicted in Fig. 7-5c, where surface atoms rearrange into a structure with a symmetry that differs from that of the bulk solid. This phenomenon is known as *reconstruction* and can alter many surface structure-sensitive properties, e.g., chemical, electrical, optical, and sorption behavior.

Whereas atomic positioning in the selvedge cross-section is not amenable to simple analysis, the crystallography in the surface plane is, and we now proceed to discuss it. In particular, we are interested in such surfaces covered by a periodic array of adsorbed atoms (adatoms).

7.2.3.2 Crystallography of Substrate Meshes and Adatom Arrays

Just as there are 14 Bravais lattices in three dimensions, so there are just five unit meshes or nets, corresponding to a two-dimensional surface as shown in Fig. 7-6. Points representing atoms may be arranged to outline

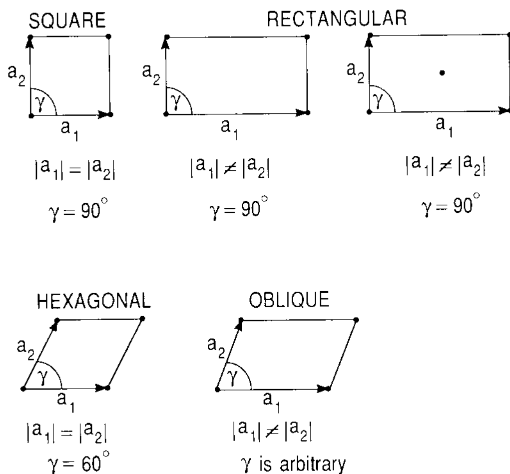


Figure 7-6 The five diperiodic surface nets.

(1) squares, (2) rectangles, (3) centered rectangles, (4) rhombi, and (5) parallelograms with included angles of 60° and 120°. Miller-type indices are used to denote atom coordinates, directions, and distances between lines within the surface.

Consider a mesh of substrate atoms that is covered with adatoms arrayed in a monolayer pattern. The methodology for identifying the structural geometry employs simple vector concepts (Refs. 5, 7). Vectors describing the adsorbate positions (**b**) are simply related to those of the substrate lattice (**a**). Possessing components b_i and a_i, with $i = 1, 2$ denoting the x and y directions, we may generally write

$$b_1 = M_{11}a_1 + M_{12}a_2 \quad \text{(7-1a)}$$

$$b_2 = M_{21}a_1 + M_{22}a_2. \quad \text{(7-1b)}$$

Terms M_{ij} denote the components of the transformation matrix between adatom and substrate coordinates. Let us apply these equations to Fig. 7-7a where a square monolayer pattern of adatoms denoted by ◯ (shaded circles) covers a net of substrate atoms represented by • (dots). This combination might correspond to the early growth of an epitaxial layer on the (100) plane of a BCC or FCC crystalline substrate; for the latter, however, a larger face-centered cell must be chosen. In the case shown, $b_1 = 2a_1$ and $b_2 = 2a_2$.

Thus

$$M = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}$$

and we speak of a $P(2 \times 2)$ overlayer structure, where P indicates a primitive unit cell effectively containing only one atom. If, for example, this structure physically corresponds to a deposit of sulfur atoms on (100) Fe, the notation would be Fe(100) $P(2 \times 2)$-S. In a similar vein, suppose there is an additional adatom positioned centrally as shown in Fig. 7-7b. Intuition would suggest that such a structure be identified as $C(2 \times 2)$ where C denotes the centered lattice. The matrix notation for the dashed primitive cell in this case is

$$M = \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix}.$$

Because the overlayer primitive cell dimensions are $\sqrt{2}$ times larger and rotated relative to substrate unit cell, an alternative notation is $(\sqrt{2} \times \sqrt{2})\,R\,45°$, where R stands for rotation.

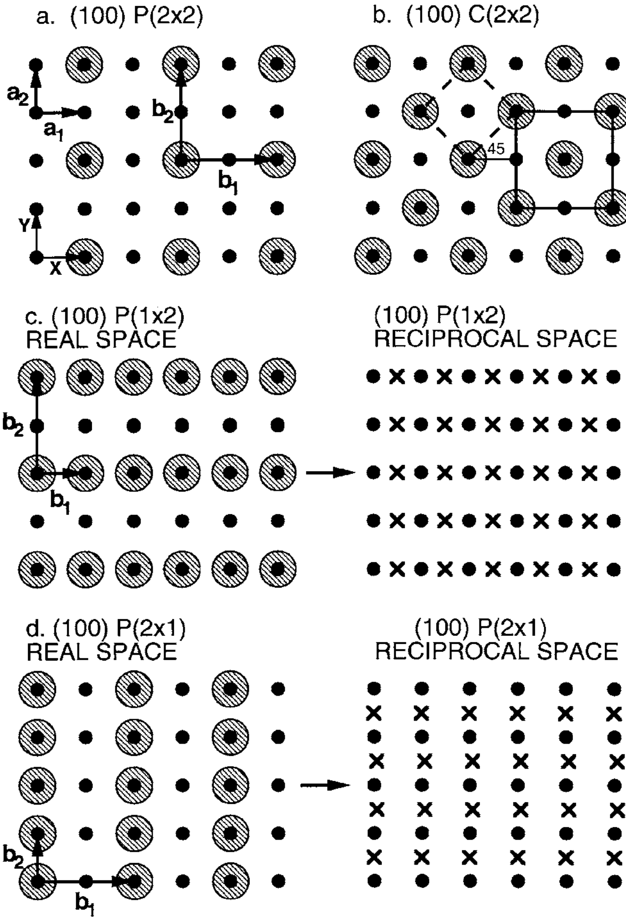


Figure 7-7 Atomic positions of adatoms (◯) relative to substrate atoms • arrayed on a (100) simple cubic plane. (a) $P(2 \times 2)$ overlayer structure. (b) $C(2 \times 2)$ centered lattice overlayer or $(\sqrt{2} \times \sqrt{2})\mathrm{R}\,45^\circ$ structure. (c) $P(1 \times 2)$ overlayer structure. (d) $P(2 \times 1)$ overlayer structure. Corresponding reciprocal lattices for (c) and (d) are shown (x refers to adatoms).

Consider now the repetitive rectangular adatom cell pattern in Fig. 7-7c where $b_1 = a_1$ and $b_2 = 2a_2$. Therefore,

$$M = \begin{vmatrix} 1 & 0 \\ 0 & 2 \end{vmatrix}$$

and the overlayer is denoted by $P(1 \times 2)$. Similarly, Fig. 7-7d shows an overlayer of the same geometry but rotated 90° with respect to case c; the notation would be $P(2 \times 1)$.

It should be noted that there is a certain ambiguity in identifying overlayers. We have only considered cases where the adatoms fall directly above substrate atoms. But adatoms may also lie in pockets between substrate-atom sites. This can happen, for example, by shifting each adatom by $\frac{1}{2}a_1$ and $\frac{1}{2}a_2$. But vector operations are independent of the choice of origin and such overlayers would be identified by exactly the same notations as before, e.g., $P(2 \times 2)$, $P(1 \times 2)$.

7.2.3.3 The Reciprocal Lattice

It is the geometry of the reciprocal lattice, not that of the real lattice, which appears as the physical image in diffraction patterns. As the name implies, the reciprocal lattice has dimensions and features which are "inverse" to those of the real lattice. In rectangular coordinates long dimensions in real space appear rotated 90° and are shortened in inverse proportion within the reciprocal space. The relation between b_1 and b_2 and the reciprocal vector components b_1^* and b_2^* of the unit mesh in the reciprocal lattice is expressed by

$$b_i \cdot b_j^* = 2\pi\delta_{ij} \tag{7-2a}$$

and similarly,

$$a_i \cdot a_j^* = 2\pi\delta_{ij} \tag{7-2b}$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$.

Reciprocal lattices corresponding to the real space structures of Figs. 7-7c and 7-7d are sketched intuitively in the same figure. The scale is arbitrary, but the lattice periodicity and symmetry are preserved. When two adatom orientations, $P(2 \times 1)$ and $P(1 \times 2)$ are admixed in roughly equal proportions, as might happen during deposition, the reciprocal lattices from each simply superimpose. It is left as an exercise for the reader to sketch the resultant pattern.

Two classic techniques for analyzing the structure of crystalline surfaces and epitaxial films in particular, are low-energy electron diffraction (LEED) and reflection high-energy electron diffraction (RHEED). Both are described in the next chapter in connection with epitaxial films and require the unfamiliar visualization in reciprocal space. Leaving less to the imagination is scanning tunneling microscopy (STM), an important new technique that provides direct images of adatoms on substrate surfaces. Many of our

present notions of surface crystallography have emerged through LEED and STM methods. In the next two sections as well as one at the chapter's end (Section 7.5.4) we shall see how STM has enhanced our understanding of semiconductor surfaces.

7.2.4 Reconstructed Silicon Surfaces

7.2.4.1 Crystallography of (001) Si

On a clean (001) Si surface, atoms are surprisingly not arranged according to the much reproduced bulk diamond-cubic lattice structure where the nearest-neighbor atom separation is 0.385 nm. Instead, the isotropic face-centered square array of Si atoms assumes an anisotropic reconstructed structure of lower symmetry (Ref. 8) possessing two characteristic directions. These lie in the surface plane either along or perpendicular to dimer (paired atom) rows in orthogonal $\langle 110 \rangle$ orientations. The parallel dimer rows aligned at 135° with respect to the page horizontal, are easily visible in the STM image of Fig. 7-8a.

More will be said about STM methods in Chapter 10. Suffice it to say that in this technique a metal probe tip sharpened to atomic dimensions is brought to within a few angstroms of the surface and scanned over its periodically bumpy atomic terrain, much like a stylus over a record. The current that flows between probe and surface reflects the topography of individual atoms on surfaces as well as the geometry of terraces and ledges or steps on the substrate surface.

Schematically shown in both side and top views of Figs. 7-8b and 7-8c is the (2×1) reconstructed surface structure on (001) Si. In these figures the position of silicon atoms, differentiated by circle size, is shown at three successive levels. Within dimers the Si atoms would like to merge even closer together but are restrained from doing so by surrounding atoms. Therefore, a tensile stress develops in the dimer-bond direction; similarly along the dimer row a smaller tensile or perhaps compressive stress is expected. Because of stress anisotropy the reconstructed surface of (001) Si decomposes into two alternating (2×1) terrace-like domains with monatomic steps at the boundaries. Dimer rows rotate 90° as successive terraces are accessed giving rise to two types of monatomic steps. The so-called S_A step is parallel to upper-terrace dimer rows, while S_B steps are perpendicular to these same rows. Because it is difficult to disrupt the linearity of dimer rows, the energy to form a kink in an S_A step is higher than for an S_B step. Thus S_A steps tend to appear smooth while S_B steps are rough. What happens when Si atoms deposit on such a terraced substrate is discussed in Section 7.5.4.1.

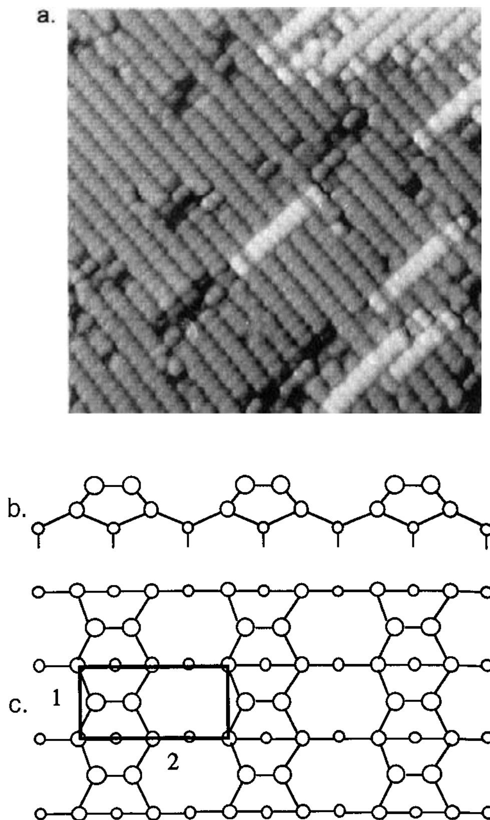


Figure 7-8 (a) Scanning tunneling microscope image of parallel dimer rows. (From Ref. 35.) (b) Schematic side view of the reconstructed (001) Si surface having a (2 × 1) structure. (c) Top view of reconstructed (2 × 1) Si surface. The intensified unit cell measures 0.385 nm × 0.770 nm. (From Ref. 8.) (Reprinted with the permission of Max G. Lagally.)

7.2.4.2 Crystallography of (111) Si

Reconstruction of (111) Si surfaces into (7 × 7) structures was first observed in LEED patterns (Ref. 9) similar to those shown in Fig. 8-33a. Disbelieved at first, these beautiful patterns were attributed to contaminated surfaces. It was some 25 years later that the now ubiquitous STM image of (111) Si shown in Fig. 7-9 independently confirmed the existence of the (7 × 7) reconstruction. Unfolding the surface and subsurface atomic positions that give rise to such a complex crystallographic structure is a formidable task. A minimum of 49 atom positions are involved and since

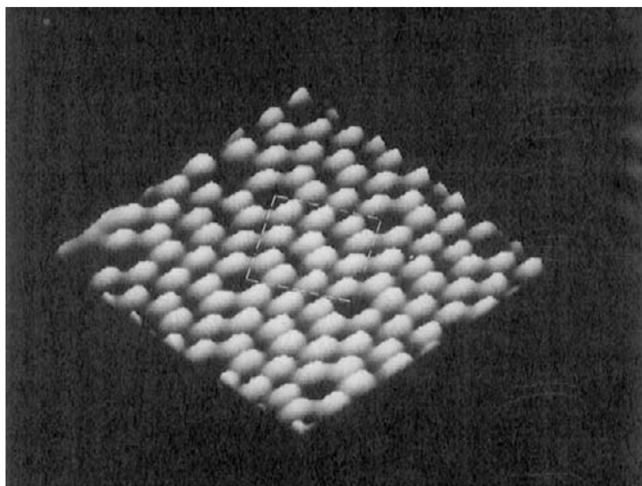


Figure 7-9 Scanning tunneling microscope image of the reconstructed (7×7) surface of (111) Si. (Courtesy of Y. Kuk, AT&T Bell Laboratories.)

the reconstruction occurs over several layers many more atom coordinates (perhaps hundreds) must be specified. Detailed crystallographic models for (7×7) Si have been proposed. This subject is clearly beyond our scope and therefore, the interested reader is referred to Refs. 10 and 11.

We shall return to the reconstructed (001) and (111) Si surfaces at the end of the chapter where the early stages of film nucleation on them are considered.

7.2.5 Adsorption Reactions on Solid Surfaces

Let us now take the pristine substrate surfaces we have been considering and expose them to gas particles. This is what occurs, for example, in thin-film deposition processes or during the operation of catalysts (Ref. 12). The first thing that happens is surface adsorption — the process in which impinging atoms and molecules enter and interact within the transition region between the gas phase and substrate surface (Ref. 13). Two kinds of adsorption processes, namely physical (*physisorption*) and chemical (*chemisorption*), can be distinguished depending on the strength of the atomic interactions. If the particle (molecule) is stretched or bent but retains its identity, and van der Waals forces bond it to the surface, then we speak of physisorption. However, chemisorption is involved when the particle

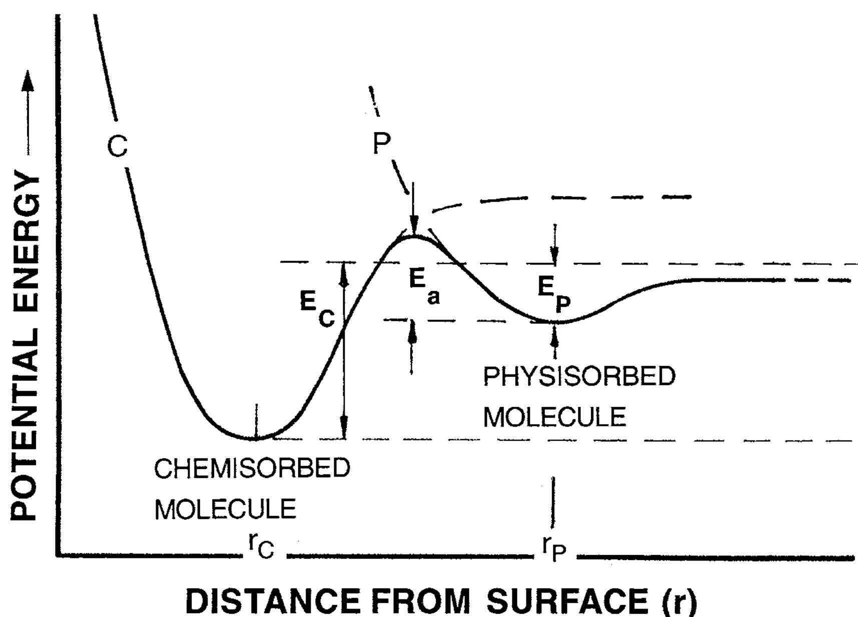


Figure 7-10 Model of physisorption and chemisorption processes in terms of the interaction potential between adsorbate and surface vs distance.

changes its identity through ionic or covalent bonding with substrate atoms. The two sorptions can be quantitatively distinguished on the basis of adsorption heats or energies, E_p and E_c, for physisorption and chemisorption, respectively. Typically $E_p \sim 0.25$ eV while $E_c \sim 1$–10 eV.

Individual physi- and chemisorption processes can be visualized in the respective curves P and C of Fig. 7-10, representing the potential energies of interaction as a function of distance r (see Fig. 1-8). It is evident that physisorbed particles achieve equilibrium further from the surface (i.e., at $r = r_p$) than chemisorbed species (at $r = r_c$). In addition to E_p and E_c we must consider the dissociation energy of molecules, E_{dis}. If E_{dis} is less than E_c, dissociative chemisorption is thermodynamically possible. When both adsorption processes occur the curves merge and assume a roller-coaster shape with an effective energy barrier E_a that governs the rate of the transition from the physisorbed to chemisorbed states. In the subsequent treatment of film nucleation a composite energy of desorption (E_{des}) that integrates over all desorption states will be used. Similarly, an energy of adsorption (E_{ads}) related to either E_p or E_c, or some combination of these, can be defined.

7.2.5.1 Adsorption Effects

To apply these ideas to deposition processes, let us consider a vapor phase containing potential adatoms at a partial pressure P. During adatom condensation on the substrate, the net time (t) rate of surface coverage (θ) is given by

$$d\theta/dt = k_{\text{ads}}P(1-\theta) - k_{\text{des}}\theta, \tag{7-3}$$

where θ assumes values between 0 (no deposit) and 1 (monolayer coverage). The first term reflects the rate of adatom coverage at the available fraction of unoccupied surface sites; this process is proportional to the adsorption rate constant k_{ads}, which in turn varies as $k_{\text{ads}} \sim \exp-(E_{\text{ads}}/k_{\text{B}}T)$. Likewise, adatom desorption is directly proportional to the product of the rate constant $k_{\text{des}} \sim \exp-(E_{\text{des}}/k_{\text{B}}T)$ and the coverage. Integration yields

$$\theta = \frac{KP}{1+KP}\{1-\exp[-k_{\text{des}}(1+KP)t]\} \tag{7-4}$$

where the equilibrium constant $K = k_{\text{ads}}/k_{\text{des}}$. At very long times, the equilibrium coverage is given by $\theta = KP/(1+KP)$, a relation known as the Langmuir isotherm. When $KP \gg 1$ the surface coverage is unity.

7.2.5.2 Statistical Theory Surface Roughening

Given an adatom population, values of θ have theoretical implications for the morphology of the surface deposit. To see why, we note that substrate sites are either occupied by adatoms or unoccupied. Furthermore, adatoms lower their *internal energy* by bonding to the substrate as well as each other, and raise the configurational *entropy* through random occupation of substrate sites. We wish to predict the equilibrium state of a system governed by the eternal thermodynamic tug (Eq. 1-5) between changes in enthalpy, or equivalently, internal energy (U) in this case, and entropy (S). Taking the Gibbs free energy of the adatoms as $G = U - TS$, well-established concepts of statistical mechanics provide U and S in terms of θ, yielding

$$G/n_{\text{s}} = 4\theta w(1-\theta) - k_{\text{B}}T[\theta\ln\theta + (1-\theta)\ln(1-\theta)] \tag{7-5}$$

where n_{s} is the number of surface sites. Equations of this form have long been used in many contexts within metallurgy and solid-state physics, e.g., theories of binary alloying and magnetism (Ref. 14). More relevant to our case, Jackson *et al.* (Ref. 15) showed that this very equation described smooth vs rough growth of crystals from melts. There was thus a precedent for Tsao (Ref. 16), who more recently applied Eq. 7-5 to the issue of smooth

vs rough growth, but in vapor deposited layers. Only the first term (U) on the right will be derived, since the expression for S is given in many places (Ref. 14). It is assumed that when an adatom attaches to the surface one downward vertical bond is annihilated during reaction with a substrate atom, one dangles upward, and the remaining four bonds dangle laterally. For an individual bond energy of w, the adatom has a total energy of $4w$. If the adatom now associates with other like adatoms, the energy decreases by $4\theta w$ because there are 4 reaction sites and a probability θ that each will be occupied. The energy per adatom is $4w - 4\theta w = 4w(1 - \theta)$, and therefore the change in internal energy per atom site is the product of the latter and the coverage, or $4w\theta(1 - \theta)$.

Thermodynamic equilibrium is achieved when free energy (G/n_s) is minimized or $d(G/n_s)/d\theta = 0$. Upon differentiation,

$$4w(1 - 2\theta) + k_B T \ln\left(\frac{\theta}{1 - \theta}\right) = 0.$$

When $k_B T/4w$ is plotted vs θ a kind of phase diagram emerges with a symmetric miscibility gap. Thus at low temperatures for $\theta = 0.5$, separation into two phase regions, a near-zero coverage ($\theta \simeq 0$) layer and a near-unity coverage ($\theta \simeq 1$) layer, occurs; in either case the surface appears to be nearly atomically flat. However, a roughened surface emerges at elevated temperatures reflecting the dominance of entropy effects. At the roughening temperature (T_r), a transition occurs between the phase-separated smooth adlayers and the microscopically rough adlayer. For a value of $\theta = 0.5$, T_r is calculated to be $2w/k_B$; physically, w is estimated to be $\sim\frac{1}{3}\Delta H_S$ where ΔH_S is the sublimation energy (typically 2–5 eV). This leads to unreasonably large values of T_r, indicating the rough qualitative nature of the theory.

We now shift direction from this focus on atomistic details of crystalline surfaces, and in the next few sections treat a complementary macroscopic view of layers deposited on substrates. Such an approach is often more useful pedagogically in describing film formation and growth effects. When it comes to epitaxial films (in Chapter 8), however, a microscopic view is again more appropriate.

7.3 THERMODYNAMIC ASPECTS OF NUCLEATION

7.3.1 SCOPE

Paradoxically, nucleation events which may only involve very few atoms can be profitably described in terms of classical thermodynamics, a subject

that deals with the continuum behavior of many atoms. Among the questions addressed in this section are:

1. Under what conditions are nuclei of thin-film deposits thermodynamically stable?
2. What role do surface energies of films and substrates play in nucleation events?
3. Under given deposition conditions (i.e., vapor supersaturation), how large are the nuclei predicted to be, and what is the energy barrier that must be surmounted in their formation?
4. How are island, planar, and S-K growth modes rationalized thermodynamically?
5. What influence do film deposition rate and temperature have on nucleation energetics and nucleus size?

However, issues related to atomic-scale nucleation events, the rates at which they occur, and the numbers of stable nuclei produced per unit time are not easily amenable to thermodynamics and will be treated in subsequent sections. Since surface energies of materials play a key role in these discussions, this subject is a good place to start.

7.3.2 Surface Energies

In the well-known capillarity or droplet theory of homogeneous nucleation, a solid nucleates from a prior unstable liquid or vapor phase by respectively establishing a solid–liquid (s-l) or solid–vapor (s-v) interface. Associated with these interfaces or surfaces are corresponding interfacial energies. Atoms at free surfaces are more energetic than atoms within the underlying bulk because they make fewer bonds with surrounding atoms and are thus less constrained. The difference in interatomic energy of atoms at these two locations is the origin of surface energy. Alternatively, there is a thermodynamic driving force to reduce the number of necessarily cut, dangling bonds at the surface through rebonding between atoms. We may then equivalently view the surface energy γ (J/m^2) in terms of the energy reduction per unit area.

In a recommended discussion of surface energies of solids by Tu *et al.* (Ref. 17) a number of issues relevant to surfaces are raised. With respect to material classes and systematics, metals have higher surface energies than oxides, alkali halides, sulfides, and organic substances. For virtually all elements in the liquid and solid states the surface energies roughly span the range 0.2 to 3 J/m^2 with 1 J/m^2 ($=1000$ erg/cm^2) being typical. The higher values are associated with the transition metals while the lower

values are indicative of alkali and divalent metals, metalloids, and inert gases. For organic and aqueous solutions γ is often less than 100 mJ/m^2. Furthermore, the surface energy drops slightly with temperature T; typically, $d\gamma/dT \sim -0.05\,\mathrm{mJ/m^2\text{-}^\circ C}$.

Surface energies are often interchangeably called surface tensions. These terms are not strictly identical although units of the latter, i.e., force/length, are equivalent to energy/area. The basic definition of surface energy is related to the reversible work (dW) on a material when its surface area (A) is increased, or $dW = \gamma\, dA$. Surfaces can alter their energies through change in either A or γ. The latter possibility implies altering atomic arrangements or reconstructing the surface and may be thought of as the work in mechanically stretching or straining the surface. For solids, surface stresses are thus involved, but these vanish in liquids because they cannot support strains. It is the shortening of the surface bond length plus the effective conversion of bulk to surface bonds that essentially causes liquid mercury to ball up; in the process A is reduced. We shall make no further distinction between surface energy and surface tension.

7.3.3 Capillarity Theory of Heterogeneous Nucleation

We have already considered a simple model for *homogeneous* nucleation of a condensed phase from a prior supersaturated vapor (see Section 1.7). Homogeneous nucleation, although rare in other contexts, occurs with some frequency in CVD reactors where too highly supersaturated gases rain "snow" down on substrates. The process is easy to model since it occurs without benefit of the complex heterogeneous sites that exist on an accommodating substrate surface. In a similar spirit we shall extend this capillarity theory to the case of *heterogeneous* nucleation of a condensed film on a substrate. Capillarity theory (Refs. 17–19) possesses the mixed virtue of yielding a conceptually simple qualitative model of film nucleation which is, however, quantitatively inaccurate. The lack of detailed atomistic assumptions gives the theory an attractive broad generality with the power of creating useful connections among such variables as substrate temperature, deposition rate, and critical film nucleus size.

We start by assuming that film forming atoms or molecules in the vapor phase impinge on the substrate creating nuclei of mean dimension r. The free-energy change accompanying the formation of such an aggregate is given by

$$\Delta G = a_3 r^3 \Delta G_{\mathrm{V}} + a_1 r^2 \gamma_{\mathrm{fv}} + a_2 r^2 \gamma_{\mathrm{fs}} - a_2 r^2 \gamma_{\mathrm{sv}}. \tag{7-6}$$

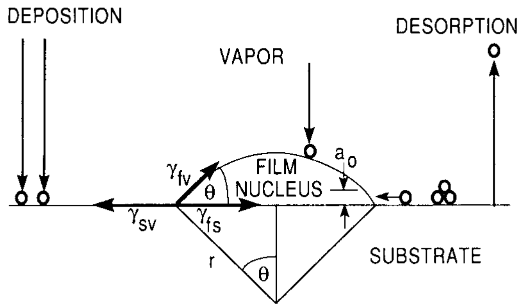


Figure 7-11 Schematic of basic atomistic nucleation processes on substrate surface during vapor deposition.

As in homogeneous nucleation the chemical free-energy change per unit volume, ΔG_V, drives the condensation reaction. There are several interfacial tensions, γ, to contend with now and these are identified by the subscripts f, s, and v representing film, substrate, and vapor, respectively. A pair of subscripts refers to the interface between the indicated phases. For the spherical cap-shaped solid nucleus shown in Fig. 7-11 the curved surface area ($a_1 r^2$), the projected circular area on the substrate ($a_2 r^2$), and the volume ($a_3 r^3$) are involved, and the respective geometric constants are $a_1 = 2\pi(1 - \cos\theta)$, $a_2 = \pi \sin^2\theta$, and $a_3 = \frac{\pi}{3}(2 - 3\cos\theta + \cos^3\theta)$.

The mechanical equilibrium among the horizontal components of the interfacial tensions or forces surrounding the nucleus yields Young's equation,

$$\gamma_{sv} = \gamma_{fs} + \gamma_{fv}\cos\theta \qquad \text{or} \qquad \cos\theta = (\gamma_{sv} - \gamma_{fs})/\gamma_{fv}. \tag{7-7}$$

Importantly, the contact or wetting angle θ, which figures prominently in the formulas for r^* and ΔG^* developed later, depends solely on the surface properties of the involved materials.

Returning now to Eq. 7-6, we note that any time a new interface appears there is an increase in surface free energy and hence the positive sign for the first two surface terms. Similarly the loss of the circular substrate–vapor interface under the cap implies a reduction in system energy and a negative contribution to ΔG. Thermodynamic equilibrium is achieved when $d\Delta G/dr = 0$, which occurs at a critical nucleus size $r = r^*$ given by

$$r^* = \frac{-2(a_1\gamma_{fv} + a_2\gamma_{fs} - a_2\gamma_{sv})}{3a_3\Delta G_V}. \tag{7-8}$$

Correspondingly, ΔG evaluated at $r = r^*$ is

$$\Delta G^* = \frac{4(a_1\gamma_{fv} + a_2\gamma_{fs} - a_2\gamma_{sv})^3}{27a_3^2\Delta G_V^2}. \tag{7-9}$$

Both r^* and ΔG^* scale in the manner shown in Fig. 1-20. An aggregate smaller in size than r^* disappears by shrinking, lowering ΔG in the process. Critical nuclei grow to supercritical dimensions by further addition of atoms, a process which lowers ΔG still more. In heterogeneous nucleation the accommodating substrate catalyzes vapor condensation by lowering the energy barrier ΔG^* through a reduction of the contact angle. After substitution of the geometric constants, it is easily shown that ΔG^* is essentially a product of two factors, i.e.,

$$\Delta G^* = \frac{16\pi(\gamma_{fv})^3}{3(\Delta G_V)^2}\left\{\frac{2 - 3\cos\theta + \cos^3\theta}{4}\right\}. \tag{7-10}$$

The first is the value for ΔG^* derived for homogeneous nucleation. It is modified by the bracketed term, a wetting factor dependent on θ which has the value of zero for $\theta = 0$ and unity for $\theta = 180°$. When the film wets the substrate, there is no barrier to nucleation. At the other extreme of dewetting, ΔG^* is maximum and equal to that for homogeneous nucleation.

The preceding formalism provides a generalized framework for inclusion of other energy contributions. If, for example, the film nucleus is elastically strained throughout because of the bonding mismatch between film and substrate, then a term $a_3r^3\Delta G_S$, where ΔG_S is the strain free-energy change per unit volume, would be appropriate. In the calculation for ΔG^*, the denominator of Eq. 7-9 would then be altered to $27a_3^2\ (\Delta G_V + \Delta G_S)^2$. Because the sign of ΔG_V is negative while ΔG_S is positive, the overall energy barrier to nucleation increases in such a case. If, however, deposition occurred on an initially strained substrate, i.e., one with emergent cleavage steps or screw dislocations (see Fig. 8-10), then stress relieval during nucleation would be manifested by a reduction of ΔG^*. Substrate charge and impurities would similarly influence ΔG^* by affecting terms related to either surface and volume electrostatic, chemical, etc., energies.

The most important thing to remember about ΔG^* is its strong influence on the density (N^*) of stable nuclei that can be expected to survive. Based on our previous experience (Eq. 1-3) of associating the probable concentration of an entity with its characteristic energy through a Boltzmann factor, it is appropriate to take

$$N^* = n_s \exp - \Delta G^*/k_B T, \tag{7-11}$$

where n_s is the *total* nucleation site density. We shall return to this equation subsequently.

7.3.4 FILM GROWTH MODES

Young's equation (Eq. 7-7) provides a way to distinguish and better understand the three modes of film growth displayed in Fig. 7-1. For island growth, $\theta > 0$, and therefore,

$$\gamma_{sv} < \gamma_{fs} + \gamma_{fv}. \tag{7-12}$$

If γ_{fs} is neglected, this relation suggests that island growth occurs when the surface tension of the film exceeds that of the substrate. This is why deposited metals tend to cluster or ball up on ceramic or semiconductor substrates.

In the case of layer growth the deposit "wets" the substrate and $\theta \simeq 0$. Therefore,

$$\gamma_{sv} \geqslant \gamma_{fs} + \gamma_{fv}. \tag{7-13}$$

A special case of this condition is ideal homo- or "auto-epitaxy." Because the interface between film and substrate essentially vanishes, $\gamma_{fs} = 0$. High-quality epitaxial films require the avoidance of any disruption to layer growth. Superlattices composed of alternating A and B epitaxial films ordered in a stack represent a particular challenge in this regard. Sequential depositions mean A deposited on a B film substrate followed by B deposited on an A film substrate. This stacking asymmetry is often not a problem when A and B are heteroepitaxial compound semiconductors; high-performance lasers have been fabricated in this way. Although roughly equal surface energies are involved here, this is not usually the case in metal–metal or metal–semiconductor superlattices. In general, materials with low surface energy will wet substrates with a higher surface energy. Thus Benjamin Franklin observed that a little oil smoothed the surface roughness of a large water pond. In this case the very thin oil film that formed dramatically modified the surface properties of water.

Lastly, for S-K growth, initially at least,

$$\gamma_{sv} > \gamma_{fs} + \gamma_{fv}. \tag{7-14}$$

In this case the strain energy per unit area of the film overgrowth is large with respect to γ_{fv} permitting nuclei to form above the initial layers. The transition from two- to three-dimensional growth which typically occurs after 5–6 monolayers is not completely understood; any factor which disturbs the monotonic decrease in binding energy characteristic of layer growth may be the cause. For example, because of film–substrate lattice mismatch, strain energy accumulates in the growing film. When released, the high energy at the deposit intermediate-layer interface may trigger island formation.

7.3.4.1 Morphological Stability of Strained Layers

To address strain (free) energy (G_S) semiquantitatively we note that elasticity theory (Eq. 1-45) suggests $G_S = \frac{1}{2}Y\varepsilon^2$ where Y is the elastic modulus of a layer and ε is its strain. A useful measure of ε at the interface between a film and substrate is the lattice mismatch strain or simply misfit, f, defined as $\text{f} = [a_0(\text{s}) - a_0(\text{f})]/a_0(\text{f})$, where a_0 is the lattice parameter and f and s refer to the film and substrate, respectively. As we shall see, the quantity f assumes an important role in epitaxy (Chapter 8).

Rather than view the breakdown of epitaxy due to dislocation generation as in Section 8.3.1, let us consider the critical film thickness beyond which a planar film surface roughens due to the growth of islands. This creates a morphological instability that can be treated in terms of macroscopic heterogeneous nucleation theory. Following Wessels (Ref. 20), the net free energy change for the nucleation of a hemispherical island (not spherical cap) on top of a growing, strained epitaxial layer of thickness h is given by

$$\Delta G = \frac{2\pi r^3}{3}\Delta G_V + \pi r^2\gamma + \Delta G_S. \tag{7-15}$$

In the process the island is assumed to be incoherent (or relaxed) while the epilayer partially relaxes. By reference to Eq. 7-6, the volume free energy of the island and interfacial and strain energy interaction between the island and epitaxial layer are, respectively, ΔG_V, γ, and ΔG_S. The term ΔG_S is interpreted as the difference in epilayer strain energy per unit area after the island nucleates ($\frac{1}{2}Y\varepsilon^2 h$) relative to that in the epilayer prior to island nucleation ($\frac{1}{2}Y\text{f}^2 h$). Therefore, $\Delta G_S = \frac{1}{2}(\varepsilon^2 - \text{f}^2)YhA$, where ε is the mean misfit in the strained film and A $(=\pi r^2)$ is the area affected by strain. As before, the critical nucleus radius (r^*) is determined by the condition that $d\Delta G/dr = 0$. Therefore,

$$r^* = -\frac{[\gamma - \frac{1}{2}(\varepsilon^2 - \text{f}^2)Yh]}{\Delta G_V} \tag{7-16}$$

and in the limit that $r^* = 0$, the critical film thickness (h^*) for the onset of the rough island morphology (e.g., S-K growth) is

$$h^* = \frac{2\gamma}{(\varepsilon^2 - \text{f}^2)Y}. \tag{7-17}$$

This relation implies that h^* varies approximately as f^{-2}. For InGaAs on GaAs, Wessels suggests that $h^*(\text{cm}) \times \text{f}^2 = 1.8 \times 10^{-10}$ holds after substituting appropriate constants. When $h^*\text{f}^2 > 1.8 \times 10^{-10}$ there is a transition from 2-D layer to 3-D island film growth.

Now we may better understand Fig. 7-12, which distinguishes the regimes of influence for the three film growth modes in terms of surface energy ratio

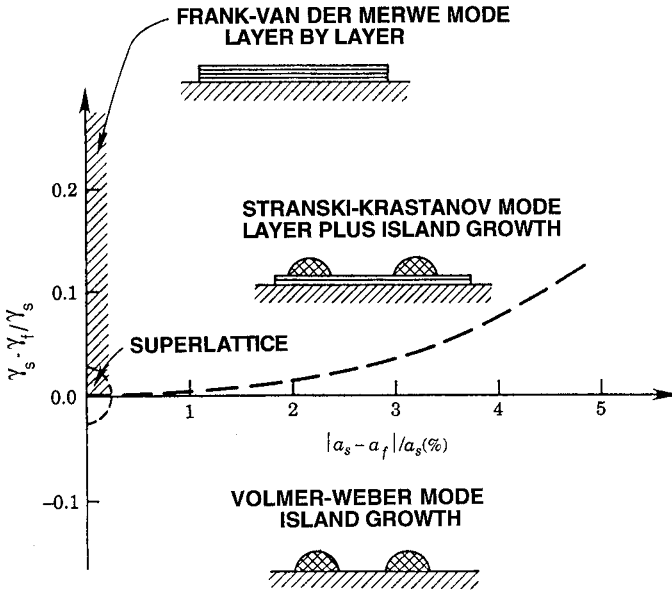


Figure 7-12 Stability regions of the three film growth modes in coordinates of surface energy difference between film and substrate (vertical) and lattice misfit (horizontal). (From K. N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science for Electrical Engineers and Materials Scientists*, Macmillan, 1992. Originally published in *IBM J. Res. Develop.* Reprinted with the permission of Professor K. N. Tu.)

W ($W = (\gamma_s - \gamma_f)/\gamma_s$) and $\mathbf{f}$. As noted earlier, when $\gamma_f > \gamma_s$, i.e., $W < 0$, island growth dominates. But its range of dominance expands when there is additional misfit present. Layer growth is possible only when $W > 0$; surprisingly, however, this growth mechanism can tolerate a small amount of misfit, a fact that makes strained-layer epitaxy feasible. In between island and layer film morphologies, and competing with them, is the S-K growth mode.

7.3.5 Nucleation Dependence on Substrate Temperature and Deposition Rate

Substrate temperature and deposition rate $\dot{R}$ (atoms/cm^2-s) are among the chief variables affecting deposition processes. Calculating their effect on r^* and ΔG^* is both instructive and simple to do. First we assume that $\dot{R}$ is proportional to P_V in Eq. 1-38 so that

$$\Delta G_V = -\frac{k_B T}{\Omega} \ln\left(\frac{\dot{R}}{\dot{R}_e}\right) \tag{7-18}$$

where $\dot{R}_e$ is the equilibrium evaporation rate from the film nucleus at the substrate temperature. Assuming an inert substrate, i.e., $\gamma_{fv} = \gamma_{fs}$, direct differentiation of Eq. 7-8 yields (Ref. 18)

$$\left(\frac{\partial r^*}{\partial T}\right)_{\dot{R}} = \frac{2}{3}\left[\frac{\gamma_{fv}(\partial \Delta G_V/\partial T) - (a_1 + a_2)\Delta G_V\,\partial\gamma_{fv}/\partial T}{a_3(\Delta G_V)^2}\right]. \tag{7-19}$$

Assuming typical values, $\gamma_{fv} = 1\ \mathrm{J/m^2}$ and $\partial\gamma_{fv}/\partial T = -0.05\ \mathrm{mJ/m^2\text{-}K}$. An estimate for $\partial\Delta G_V/\partial T$ is the entropy change for vaporization, which is roughly $8 \times 10^6\ \mathrm{J/m^3\text{-}K}$ for many metals. As long as $|\Delta G_V| \ll 1.6 \times 10^{10}\ \mathrm{J/m^3}$, direct substitution shows that

$$(\partial r^*/\partial T)_{\dot{R}} > 0. \tag{7-20}$$

Similarly, by the same assumptions and arguments

$$(\partial \Delta G^*/\partial T)_{\dot{R}} > 0. \tag{7-21}$$

It is also simple to show that

$$(\partial r^*/\partial \dot{R})_T < 0. \tag{7-22}$$

Direct chain-rule differentiation using Eqs. 7-8 and 7-18 yields

$$\left(\frac{\partial r^*}{\partial \dot{R}}\right)_T = \left(\frac{\partial r^*}{\partial \Delta G_V}\right)\left(\frac{\partial \Delta G_V}{\partial \dot{R}}\right) = \left(-\frac{r^*}{\Delta G_V}\right)\left(-\frac{k_B T}{\Omega \dot{R}}\right). \tag{7-23}$$

Since ΔG_V is negative, the overall sign is negative. Similarly, it is easily shown that

$$(\partial \Delta G^*/\partial \dot{R})_T < 0. \tag{7-24}$$

The preceding four inequalities have interesting implications and summarize a number of common effects observed during film deposition. From Eq. 7-20 we note that a higher substrate temperature leads to an increase in the size of the critical nucleus. Also, a discontinuous island structure is predicted to persist to a higher average coverage than at low substrate temperatures. The second inequality (Eq. 7-21) suggests that a nucleation barrier may exist at high substrate temperatures whereas at lower temperatures it is reduced in magnitude. Also, because of the exponential dependence of N^* on ΔG^*, the number of supercritical nuclei decreases rapidly with temperature. Thus, a continuous film will take longer to develop at higher substrate temperatures. From Eq. 7-22 it is clear that increasing the deposition rate results in smaller islands. Because ΔG^* is also reduced, nuclei form at a higher rate suggesting that a continuous film is produced at lower average film thickness. If for the moment we associate a large r^* and ΔG^* with being conducive to large crystallites or even monocrystal

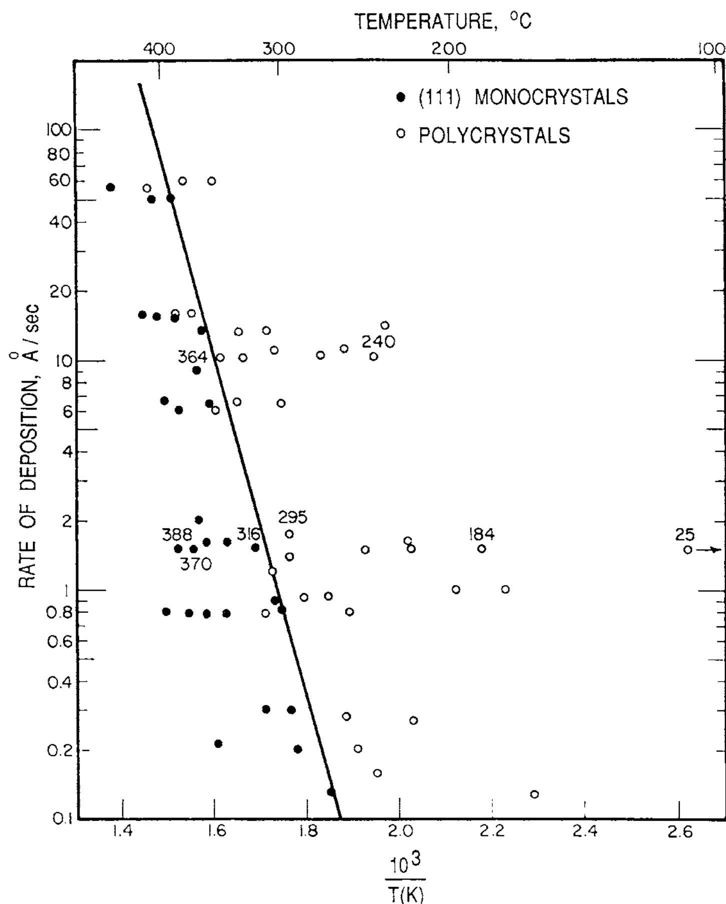


Figure 7-13 Dependence of microstructure on deposition rate and substrate temperature for Cu films on (111) NaCl. Some substrate temperatures are noted. (Reprinted with permission from the Metals Society. From Ref. 2, courtesy of R. W. Vook.)

formation, such conditions prevail at high substrate temperatures and low deposition rates. Alternately low substrate temperatures and high deposition rates yield polycrystalline deposits. These simple guidelines summarize much practical deposition experience in a nutshell.

A plot of these structural zones as a function of deposition rate and substrate temperature is shown in Fig. 7-13 for Cu films deposited on a (111) NaCl substrate. Maps of this sort have been graphed for other film–substrate combinations. In semiconductor systems a regime of

amorphous film growth is additionally observed when $\dot{R}$ is large and T is low.

Despite the qualitative utility of capillarity theory, it is far from being quantitatively correct. For example, let us estimate the value of r^* for a typical metal film at 300 K. To simplify matters the homogeneous nucleation formula, $r^* = -2\gamma/\Delta G_V$, will be used where $\Delta G_V = -k_B T/\Omega \ln P_V/P_S$. Assuming $\Omega = 20 \times 10^{-30}/\text{m}^3$, $\gamma = 1\ \text{J/m}^2$, $P_V = 10^{-3}$ torr, and $P_S = 10^{-10}$ torr, then $r^* = -2(1.0)/-\{[1.38 \times 10^{-23} \times 300/20 \times 10^{-30}]\ln[10^{-3}/10^{-10}]\}$ or 0.6 nm.

A nucleus of this size contains just a few atoms, and it is therefore doubtful that continuum concepts such as surface tension and nucleus radius have much significance. It is more realistic to interpret vapor condensation phenomena in terms of a heterogeneous nucleation theory based on atomistic modeling. This is one of the topics considered in the next section.

7.4 KINETIC PROCESSES IN NUCLEATION AND GROWTH

7.4.1 SCOPE

Among the questions addressed in this section are:

1. What is the nucleation rate ($\dot{N}$) and how does it depend on time, temperature, deposition rate, and the nature of the film and substrate?
2. Once formed, what are the possible mechanisms for subsequent growth and coalescence of nuclei?
3. What is the time dependence for the growth and coalescence of nuclei?

In particular, distinctions between the macroscopic (e.g., capillarity) and microscopic (atomistic) views of these processes will be stressed. For the most part theoretical modeling is emphasized.

7.4.2 NUCLEATION RATE

The nucleation rate is a convenient synthesis of terms that describes how many nuclei of critical size form on a substrate per unit area, per unit time. Nuclei can grow through direct impingement of gas-phase atoms, but this is unlikely in the earliest stages of film formation when nuclei are spaced far apart. Rather, the rate at which critical nuclei grow depends on the rate at which adsorbed monomers (adatoms) attach to it. In the model of Fig. 7-11

energetic vapor atoms which impinge on the substrate may immediately desorb but usually they remain on the surface for a length of time τ_S given by

$$\tau_S = \frac{1}{v}\exp\left(\frac{E_{des}}{k_B T}\right). \tag{7-25}$$

The vibrational frequency of an adatom on the surface is v (typically $10^{13}\,s^{-1}$) while E_{des} is the energy required to desorb it back into the vapor. Adatoms, which have not yet thermally accommodated to the substrate, execute random diffusive jumps and may, in the course of their migration, form pairs with other adatoms or attach to larger atomic clusters or nuclei. When this happens it is unlikely that these atoms will return to the vapor phase. This is particularly true at substrate heterogeneities such as cleavage steps or ledges where the binding energy of adatoms is greater relative to a planar surface. The proportionately large numbers of atom bonds available at these accommodating sites leads to higher E_{des} values. Thus a significantly higher density of nuclei is usually observed near cleavage steps and other substrate imperfections. The presence of impurities similarly alters E_{des} in a complex manner dependent on the type and distribution of atoms or molecules involved.

We now exploit some of these microscopic notions by noting that the nucleation rate $\dot{N}$ is essentially proportional to the product of three terms, namely,

$$\dot{N} = N^* A^* \omega \text{ (nuclei/cm}^2\text{-s)}. \tag{7-26}$$

N^* is the equilibrium concentration (per cm^2) of stable nuclei (Eq. 7-11) and ω is the rate at which atoms impinge (per cm^2-s) onto the nuclei of critical area A^* (cm^2). Of the total nucleation site density n_s, a certain number will be occupied by adatoms. The surface density of these adatoms, n_a, is given by the product of the vapor impingement rate (Eq. 2-8) and adatom lifetime or

$$n_a = \frac{\tau_s P N_A}{\sqrt{2\pi MRT}}. \tag{7-27}$$

Surrounding the cap-shaped nucleus of Fig. 7-11 are adatoms poised to attach to the circumferential belt whose area is

$$A^* = 2\pi r^* a_0 \sin\theta. \tag{7-28}$$

Quantities r^*, a_0, and θ were defined previously.

Lastly, the impingement rate onto area A^* requires adatom diffusive jumps on the substrate with a frequency given by $v \exp - (E_S/k_B T)$, where E_S is the activation energy for surface diffusion. The overall impingement

flux is the product of the jump frequency and n_a or

$$\omega = \frac{\tau_s P N_A \nu \exp - (E_S/k_B T)}{(2\pi MRT)^{1/2}} \qquad (\text{cm}^{-2}\,\text{s}^{-1}). \tag{7-29}$$

There is no dearth of adatoms which can diffuse to and be captured by the existing nuclei. During their residence time adatoms are capable of diffusing a mean distance X from the site of incidence given by

$$X = (2D_S\tau_S)^{1/2}. \tag{7-30}$$

The surface diffusion coefficient, D_S, is essentially (see Eq. 1-31)

$$D_S = \tfrac{1}{2}a_0^2\nu \exp - (E_S/k_B T) \tag{7-31}$$

and therefore,

$$X = a_0 \exp\left(\frac{E_{des} - E_S}{2k_B T}\right). \tag{7-32}$$

Large values of E_{des} coupled with small values of E_S serve to extend the nucleus capture radius.

Upon substitution of Eqs. 7-11, 7-27, 7-28, and 7-29 in Eq. 7-26, we finally obtain

$$\dot{N} = 2\pi r^* a_0 \sin\theta \frac{PN_A}{(2\pi MRT)^{1/2}} n_s \exp\left(\frac{E_{des} - E_S - \Delta G^*}{k_B T}\right). \tag{7-33}$$

The nucleation rate is a very strong function of the nucleation energetics which are largely contained within the term ΔG^*. It is left to the reader to develop the steep dependence of $\dot{N}$ on the vapor supersaturation ratio. As noted earlier, a high nucleation rate encourages a fine-grained or even amorphous structure while a coarse-grained deposit develops from a low value of $\dot{N}$.

7.4.3 Atomistic Models of the Nucleation Rate

Atomistic theories of nucleation describe the role of individual atoms and associations of small numbers of atoms during the earliest stages of film formation. An important advance in the atomistic approach to nucleation was the theory proposed by Walton *et al.* (Ref. 21) which treated clusters as macromolecules and applied concepts of statistical mechanics in describing them. They introduced the critical dissociation energy, E_i, defined to be that required to disintegrate a critical cluster containing i atoms into i separate adatoms. The critical concentration of clusters per unit area of size i, N_i, is

then given by

$$N_i^*/n_0 = (N_1/n_0)^{i^*} \exp(E_i^*/k_B T), \tag{7-34}$$

which expresses the chemical equilibrium between clusters and monomers. In this equation, E_i^* may be viewed as the negative of a cluster formation energy, n_0 is the total density of adsorption sites, and N_1 is the monomer density. The latter, by analogy to Eq. 7-27, is given by $N_1 = \dot{R}\tau_s$. Hence

$$N_1 = \dot{R}\nu^{-1} \exp \frac{E_{des}}{k_B T}. \tag{7-35}$$

Lastly, the critical monomer supply rate is essentially given by the vapor impingement rate and the area over which adatoms are capable of diffusing before desorbing. Therefore, through Eq. 7-32,

$$\dot{R}X^2 = \dot{R}a_0^2 \exp \frac{E_{des} - E_S}{k_B T}. \tag{7-36}$$

By combining Eqs. 7-34 to 7-36, the critical nucleation rate (cm^{-2}-s^{-1}) emerges as

$$\dot{N}_{i^*} = \dot{R}a_0^2 n_0 \left(\frac{\dot{R}}{n_0 \nu}\right)^{i^*} \exp\left(\frac{(i^* + 1)E_{des} - E_S + E_{i^*}}{k_B T}\right). \tag{7-37}$$

This expression for the nucleation rate of small clusters has been central to subsequent theoretical treatments and in variant forms has also been used to test much experimental data. Unlike our previous expression for $\dot{N}$ (Eq. 7-33), to which it should be compared, it has the advantage of expressing the nucleation rate in terms of measurable parameters rather than macroscopic quantities such as ΔG^*, γ, or θ; these quantities characteristic of capillarity theory are not known with certainty, nor are they easily estimated. Now, however, the uncertainties are in i^* and E_i^*.

One of the important applications of this theory has been to the subject of epitaxy, where the crystallographic geometry of stable clusters has been related to different conditions of supersaturation and substrate temperature. In Fig. 7-14 experimental data for Ag on (100) NaCl are shown together with the atomistic model of evolution to stable clusters upon attachment of a single adatom. At high supersaturations and low temperatures the nucleation rate is frequently observed to depend on the square of the deposition rate, i.e., $\dot{N}^* \sim \dot{R}^2$, indicating that $i^* = 1$. This means that a single adatom is in effect the critical nucleus. At higher temperatures two or three atom nuclei are critical and the stable clusters now assume the planar atomic pattern suggestive of (111) or (100) packing, respectively. Epitaxial films would then evolve over macroscopic dimensions provided the original

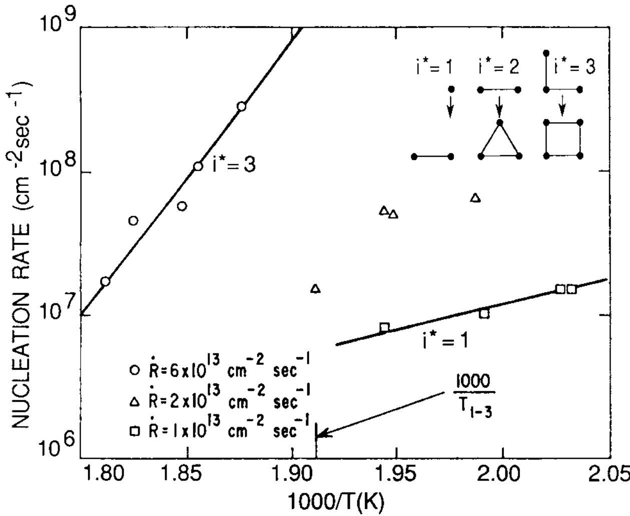


Figure 7-14 Nucleation rate of Ag on (100) NaCl as a function of temperature. Data for three different deposition rates are plotted. Also shown are smallest stable epitaxial clusters corresponding to critical nuclei. (From Ref. 21.)

nucleus orientation was preserved with subsequent deposition and cluster impingement.

A thermally activated nucleation rate whose activation energy is dependent on the size of the critical nucleus is predicted by Eq. 7-37. This suggests the existence of critical temperatures where the nucleus size and orientation may undergo change. For example, the temperature $T_{1\to2}$ at which there is a transition from a one- to a two-atom nucleus is given by equating the respective nucleation rates $\dot{N}_{i^*=1} = \dot{N}_{i^*=2}$. After substitution and some algebra,

$$T_{1\to2} = -\frac{E_{\text{des}} + E_2}{k_B \ln(\dot{R}/n_0 \nu)} \tag{7-38a}$$

or

$$\dot{R} = n_0 \nu \exp\left(-\frac{E_{\text{des}} + E_2}{k_B T_{1\to2}}\right). \tag{7-38b}$$

Use can be made of Eq. 7-38b in analyzing the data shown in Fig. 7-13. The transition of polycrystalline Cu to single-crystal Cu films can be thought of as the onset of (111) epitaxy ($i^* = 2$) from atomic nuclei ($i^* = 1$).

An Arrhenius plot of the line of demarcation between mono- and polycrystal deposits yields an activation energy of $E_{\text{des}} + E_2$ equal to 1.48 eV. At a deposition rate of 8.5×10^{14} atoms/cm^2-s or 1 Å/s, and $n_0 \nu = 6.9 \times 10^{27}$ atoms/cm^2-s, the epitaxial transition temperature is calculated to be 577 K. Equations similar to Eq. 7-38 can be derived for transitions between $i^* = 1$ and $i^* = 3$, or $i^* = 2$ and $i^* = 3$, etc.

It is instructive to compare these estimates of epitaxial growth temperatures with ones based on surface-diffusion considerations (Ref. 22). For the case of smooth layer-by-layer growth on high-symmetry crystal surfaces, adatoms must reach growth ledges by diffusional hopping before other islands nucleate. Atoms must, therefore, migrate a distance of some 100–1000 atoms, the typical terrace width on well-oriented surfaces during the monolayer growth time of ~0.1–1 s. By Eq. 7-30 a surface diffusivity greater than $\sim 10^{-8}$ cm^2/s is, therefore, required irrespective of the material deposited. This means a different critical epitaxial-growth temperature (T_E) depending on whether a semiconductor, metal, or alkali halide films is involved. In the Arrhenius plot of Fig. 7-15, D_S values are graphed vs T_M/T where T_M is the melting point. Similar relationships for lattice, grain-boundary, and dislocation diffusivities will be introduced in Chapter 11. At the critical value of $D_S = 10^{-8}$ cm^2/s, Fig. 7-15 indicates that $T_E \sim 0.5T_M$, $\sim 0.3T_M$, and $0.1T_M$ for layer growth on group IV semiconductors, metals, and alkali halides, respectively. These temperatures agree qualitatively with experimental observations.

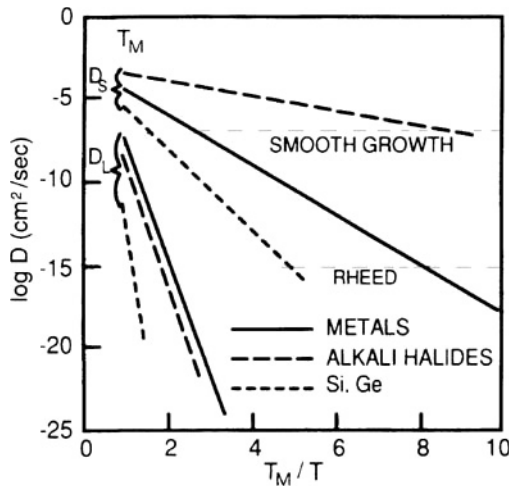


Figure 7-15 Lattice (D_L) and surface diffusivities (D_S) as a function of T_M/T for metals, semiconductors, and alkali halides. (From Ref. 22.)

7.4.4 Kinetic Models of Nucleation

Microscopic approaches to the modeling of nucleation processes have stressed the kinetic behavior of atoms and clusters containing a small number of atoms. Rate equations similar to those describing the kinetics of chemical reactions are used to express the time-dependent change of cluster densities in terms of the processes which occur on the substrate surface. Because these theories and models are complex mathematically as well as physically, the discussion will stress the results without resorting to extensive development of derivations. It is appropriate to start with the fate of the mobile monomers. If coalescence is neglected, then

$$\frac{dN_1}{dt} = \dot{R} - \frac{N_1}{\tau_s} - K_1 N_1^2 - N_1 \sum_{i=2}^{\infty} K_i N_i. \tag{7-39}$$

The equation states that the time rate of change of the monomer density is given by their incidence rate, minus their desorption rate, minus the rate at which two monomers combine to form a dimer. This latter term follows second-order kinetics with a rate constant K_1. The last term represents the loss in monomer population due to their capture by larger clusters containing two or more atoms. There are similar equations describing the population of dimer, trimer, etc., clusters as they interact with monomers. The general form of the rate equation for clusters of size i is

$$\frac{dN_i}{dt} = K_{i-1} N_1 N_{i-1} - K_i N_1 N_i \tag{7-40}$$

where the first term on the right expresses their increase by attachment of monomers to smaller $i - 1$ sized clusters, and the second term their decrease when they react with monomers to produce larger $i + 1$ sized clusters. There are i of these coupled rate equations to contend with, each one of which is dependent on direct impingement from the vapor as well as desorption through their link back to Eq. 7-35. Inclusion of diffusion terms, i.e., $d^2 N_i/dx^2$, enables change in cluster shape to be accounted for. When cluster mobility and coalescence are also taken into account, a fairly complete chronology of nucleation events emerges, but at the expense of greatly added complexity.

Transient as well as steady-state (i.e., where $dN_i/dt = 0$) solutions have been obtained for the foregoing rate equations for a variety of physical situations and for arbitrary values of i. They typically predict that $N(t)$ increases with time, eventually saturating at the value N_S. In practice the stable nucleus density is observed to increase approximately linearly with deposition time, then saturate at values ranging from $\sim 10^9$ to $10^{12}\ \mathrm{cm}^{-2}$

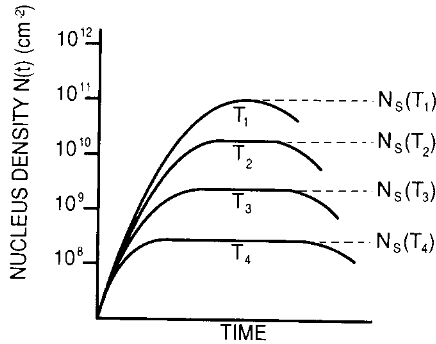


Figure 7-16 Schematic dependence of $N(t)$ with time and substrate temperature. $T_1 > T_2 > T_3 > T_4$. (From Ref. 19.)

depending on deposition rate and substrate temperature (Ref. 19). With further deposition $N(t)$ decreases as a result of coalescence. Furthermore, N_S is larger the lower the substrate temperature, as shown schematically in Fig. 7-16.

Venables (Ref. 23) has neatly summarized nucleation behavior for cases where i assumes any integer value. In general, the stable cluster density is given by

$$N_S = An_0(\dot{R}/n_0 v)^p \exp(E/k_B T) \tag{7-41}$$

where A is a calculable dimensionless constant dependent on the substrate coverage. As a result of such generalized equations, experimental data for N_S have been tested as a function of the substrate temperature and deposition rate. The parameters p and E have respectively yielded values for i^* and the energies of desorption, diffusion, and cluster binding. Values for E_{des} and E_S have been extracted for a number of systems, and these are entered in Table 7-1. Because $E_{des} > E_S$, adatom motion is largely confined to the substrate plane rather than normal to it. Although the major application of the kinetic model has been to island growth, the theory is also capable of describing S-K growth.

7.4.5 Cluster Coalescence and Depletion

As previously noted, the density of stable nuclei increases with time up to some maximum level before decreasing because of coalescence phenomena. Growth and coalescence of nuclei are generally characterized by the

Table 7-1

Nucleation Desorption and Surface Diffusion Energies[a]

Deposit	Substrate	E_{des} (eV)	E_S (eV)	Experimental method
Ag	NaCl (100)	0.41	0.19	TEM
	W (110)	2.2	0.15	SEM
	Ag (111)	(2.23)	0.097	STM
	Pt (111)	(2.94)	0.157	
	Si (111)	2.25	0.40	STM
	Ge (111)	2.65	0.35	STM
Pt	Pt (111)		0.26	STM, FIM[c]
	Ni (100)		0.69	FIM
Fe	Fe (100)		0.45	STM
Ni	Ni (100)		0.47	STM
	Pt (100)		0.45	FIM
Cu	Cu (100)		0.40	STM
Si[b]	Si (100)		0.65	STM

[a] After J. A. Venables, *Physica A* **239**, 35 (1997).
[b] From Y.-W. Mo, J. Kleiner, M. B. Webb, and M. G. Lagally, *Surf. Sci.* **268**, 275 (1992).
[c] FIM = Field ion microscopy.

following features:

1. A decrease in the total projected area of nuclei on the substrate occurs.
2. There is an increase in the height of the surviving clusters.
3. Nuclei with well-defined crystallographic facets sometimes become rounded.
4. The composite island generally reassumes a crystallographic shape with time.
5. When two islands of very different orientation coalesce, the final compound cluster assumes the crystallographic orientation of the larger island.
6. The coalescence process frequently appears to be liquid-like with islands merging and undergoing shape changes after the fashion of liquid droplet motion. This is especially true at elevated temperatures.
7. Prior to impact and union, clusters have been observed to migrate over the substrate surface in a process described as cluster-mobility coalescence.

Several mass transport mechanisms have been proposed to account for these coalescence phenomena, and they are discussed in turn.

7.4.5.1 Ostwald Ripening

Prior to coalescence there is a collection of islands of varied size, and with time the larger ones grow or "ripen" at the expense of the smaller ones. The desire to minimize surface free energy of the island structure is the driving force. To understand the process, consider two isolated islands of surface tension γ and different size in close proximity as shown in Fig. 7-17a. For simplicity they are assumed to be spherical with radii r_1 and r_2, and therefore the free energy (G) of a given island is $4\pi r_i^2\gamma$ ($i = 1, 2$). The island contains a number of atoms n_i given by $4\pi r_i^3/3\Omega$, where Ω is the atomic volume. Defining the free energy per atom μ_i or chemical potential as dG/dn_i, after substitution,

$$\mu_i = \frac{d(4\pi r_i^2\gamma)}{d(4\pi r_i^3/3\Omega)} = \frac{8\pi r_i\gamma dr_i}{4\pi r_i^2 dr_i/\Omega} = \frac{2\Omega\gamma}{r_i}. \tag{7-42}$$

In chemical thermodynamics the chemical potential is often associated with the so-called "escaping tendency" of atoms. Where μ is large the effective atomic concentration is large, forcing them to escape to where μ is small. This is the case just outside very small particles. Thus, if $r_1 > r_2$ as shown, $\mu_2 > \mu_1$, and atoms will diffuse along the substrate from particle 2 (shrinking it in size) to particle 1 which grows at the former's expense. A mechanism has thus been established for coalescence without the necessity for islands to be in direct contact. In a multi-island array the kinetic details are

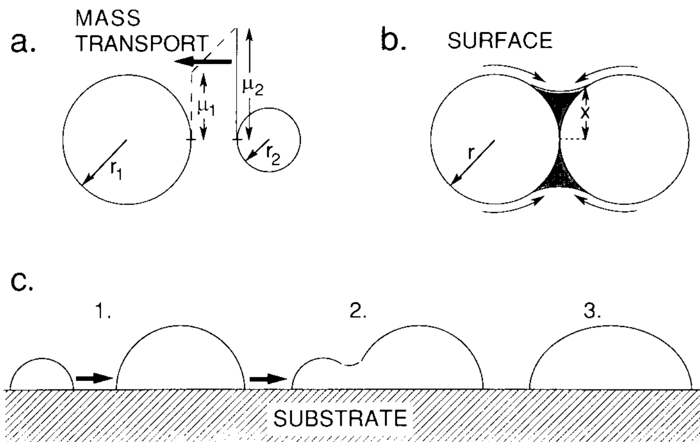


Figure 7-17 Coalescence of islands due to (a) Ostwald ripening, (b) sintering, (c) cluster migration.

complicated, but ripening serves to establish a quasi-steady-state island size distribution which changes with time. Ostwald ripening processes never reach equilibrium during film growth, and the theoretically predicted narrow distribution of crystallite sizes is generally not observed.

7.4.5.2 Sintering

Sintering is a coalescence mechanism involving islands in contact. It can be understood by referring to Fig. 7-18, depicting a time sequence of coalescence events between Au particles deposited on molybdenite (MoS_2) at 400°C and photographed within the TEM (Ref. 24). Within tenths of a second a neck forms between islands and then successively thickens as atoms are transported into the region. The driving force for neck growth is simply the natural tendency to reduce the total surface energy (or area) of the system. Since the algebraic magnitude of μ for atoms on the convex island surfaces ($r > 0$) exceeds that for atoms situated in the concave neck ($r < 0$), an effective concentration gradient between these regions develops.

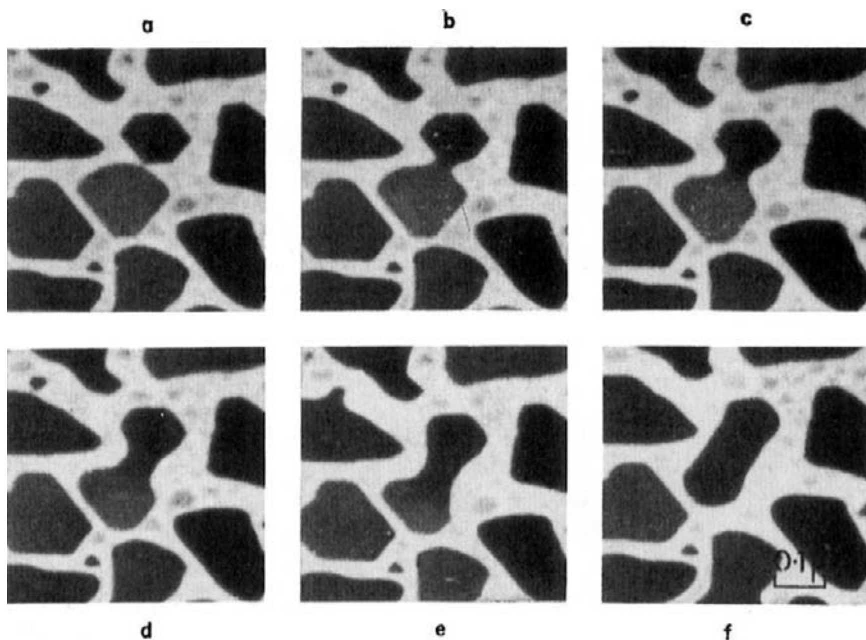


Figure 7-18 Successive electron micrographs of Au deposited on molybdenite at 400°C illustrating island coalescence by sintering. (a) Arbitrary zero time, (b) 0.06 s, (c) 0.18 s, (d) 0.50 s, (e) 1.06 s, (f) 6.18 s. (From Ref. 24.)

This results in the observed mass transport into the neck. Variations in island surface curvature also give rise to local concentration differences that are alleviated by mass flow.

In the case of sintering or coalescence of contacting spheres of radius r (Fig. 7-17b), theoretical calculations in the metallurgical literature (Ref. 25) have shown that the sintering kinetics is given by

$$X^n/r^m = A(T)t. \tag{7-43}$$

Here X is the neck radius, $A(T)$ is a temperature-dependent constant that varies with mass transport mechanism, n and m are constants, and t is the time. Of the several mechanisms available for mass transport in films, the two most likely ones involve diffusion either through the bulk or via the surface of the islands. For bulk diffusion $n = 5$, $m = 2$, whereas for surface diffusion $n = 7$, $m = 3$. Typical thumbnail calculations show that surface diffusion dominates sintering.

While surface energy and diffusion-controlled mass transport mechanisms undoubtedly influence liquid-like coalescence phenomena, sintering mechanisms are unable to explain the following:

1. Observed liquid-like coalescence of metals on substrates maintained at 77 K where atomic diffusion is expected to cease
2. Widely varying stabilities of irregularly shaped necks, channels, and islands possessing high curvatures at some points
3. The large range of times required to fill visually similar necks and channels
4. Enhanced coalescence in the presence of an applied electric field in the substrate plane

7.4.5.3 Cluster Migration

The last mechanism for coalescence considered deals with migration of clusters on the substrate surface (Fig. 7-17c). Coalescence occurs as a result of collisions between separate island-like crystallites (or droplets) as they execute random motion. Evidence provided by the field ion microscope, which has the capability of resolving individual atoms, has revealed the migration of dimer and trimer clusters. Electron microscopy has shown that crystallites with diameters of $\sim 50-100$ Å can migrate as distinct entities provided the substrate temperature is high enough. Interestingly, the mobility of metal particles can be significantly altered in different gas ambients. Not only do the clusters translate, but they have been observed to rotate as well as even jump on each other and sometimes reseparate thereafter! Cluster migration has been directly observed in many systems, e.g., Ag and

Au on MoS_2, Au and Pd on MgO, and Ag and Pt on graphite in so-called conservative systems, i.e., where the mass of the deposit remains constant because further deposition from the vapor has ceased. Coalescence in a conservative system is characterized by a decreased density of particles, increased mean volume of particles, a particle size distribution which increases in breadth, and a decreased coverage of the substrate.

The surface migration of a cap-shaped cluster with projected radius r occurs with an effective diffusion coefficient $D(r)$ having units of cm^2/s. Presently there exist several formulas for the dependence of D on r based on models assumed for cluster migration. The movement of peripheral cluster atoms, the fluctuations of areas and surface energies on different faces of equilibrium-shaped crystallites, and the glide of crystallite clusters aided by dislocation motion are three such models. In each case $D(r)$ is given by an expression of the form (Ref. 26)

$$D(r) = B(T)/r^s \exp - (E_C/k_B T) \tag{7-44}$$

where $B(T)$ is a temperature-dependent constant and s is a number ranging from 1 to 3. It comes as no surprise that cluster migration is thermally activated with an energy E_C related to that for surface self-diffusion, and that it is more rapid the smaller the cluster. However, there is a lack of relevant experimental data which can distinguish among the mechanisms. In fact, it is difficult to distinguish cluster mobility coalescence from Ostwald ripening based on observed particle-size distributions.

7.4.5.4 Coalescence and Grain Size

The vocabulary of polycrystalline film formation, particularly terms such as nucleation, growth, and coalescence, has a familiar ring in the metallurgical literature of solid-state phase transformations. For example, a cold-worked metal matrix transforms by nucleation of recrystallized grains (islands) followed by growth, thus consuming the surrounding matrix so that there is less area available for new nuclei. During recrystallization one region will start ahead of others because of some favorable local composition, temperature, defect, or bias in the system. Then nucleation occurs elsewhere, and later somewhere else, while growth continues independently in regions of prior nucleation. The situation is much like rain droplets falling on a pond. Where rain impinges on the pond, surface wave nuclei are created. The circular ripples grow outward and begin to impinge on one another. When all of the prior untransformed matrix disappears by grain growth and impingement or coalescence, recrystallization is complete. So, too, for grain formation during thin-film formation.

An important equation, based on an analysis by Avrami (Ref. 27) more than 60 years ago, is useful in describing the kinetics of such transformations. For the case of film nucleation and growth in two dimensions where circular disklike nuclei grow, the form of the equation is

$$f(t) = 1 - \exp - [\pi/3\dot{N}\dot{G}^2t^3], \tag{7-45}$$

where $f(t)$ is fractional amount of transformation and t is the time. Furthermore, $\dot{N}$ (nuclei/cm^2-s) and $\dot{G}$ (cm/s) are the nucleation and linear growth rates, respectively. As a function of log t, $f(t)$ is small during what may be thought of as an initial incubation period, then rises sharply when nucleation and growth rates overlap strongly, and finally saturates at $f = 1$ at long times (see Fig. 11-20a). Suppose when $f = 0.95$, cluster coalescence is complete and the film is virtually continuous. Substitution shows that the time required, $t_{0.95}$, is given by the relation $\frac{\pi}{3}\dot{N}\dot{G}^2t_{0.95}^3 = 3$. Provided the nucleation rate is constant, the number of grains nucleated per unit area in this time is $\dot{N}t_{0.95}$ or $\sim 1.42(\dot{N}/\dot{G})^{2/3}$. If the mean grain diameter is l_g, the number of grains per unit area is equivalently $(\pi l_g^2/4)^{-1}$, and therefore by equating terms,

$$l_g = 0.9(\dot{G}/\dot{N})^{1/3}. \tag{7-46}$$

That grain size is proportional directly to $\dot{G}$ and inversely to $\dot{N}$ is a fundamental idea common to bulk solids (e.g., castings, recrystallized metals) as well as thin films.

In a similar vein, Thompson (Ref. 28), using the Avrami equation, has simulated the processes of cluster nucleation and growth to the point of coalescence. Importantly, connections were drawn between the grain size at cluster impingement and the practical concerns of film stress and texture. The model assumed a substrate consisting of spherical cap clusters as in Fig. 7-11 such that when viewed from above the circular islands were each surrounded by a substrate annular ring of width X populated by adatoms. These diffuse by Eq. 7-30 to enlarge the island rather than form a new nucleus. The resultant film grain size was dependent on X and given by

$$l_g = 1.351X + 1.203(\dot{G}/\dot{N})^{1/3}. \tag{7-47}$$

Provided $(\dot{G}/\dot{N})^{1/3} > X$, Eqs. 7-46 and 7-47 are essentially the same.

It should be noted that the Avrami equation strictly applies to closed systems transforming under conditions where there is no matter exchange with the surroundings. Inclusion of the adatom ring apparently extends the equation's validity to open systems. An accessible and very readable derivation of the Avrami equation for the three-dimensional case can be found in Tu *et al.* (Ref. 17); later in Chapter 11 we shall again apply it to phase transformations in thin films.

7.5 EXPERIMENTAL STUDIES OF NUCLEATION AND GROWTH

Assorted microscopic and surface analytical techniques have been employed over the years to reveal the physical processes of nucleation and test the theories that model them. In this section we shall focus on the more important of these experimental methods and the results of illustrative studies.

7.5.1 ELECTRON MICROSCOPY

The traditional tool, particularly for island growth studies, has been the transmission electron microscope (TEM). In this technique metals such as Ag and Au are deposited onto substrates of interest such as cleaved alkali-halide single crystals, e.g., LiF, NaCl, and KCl, in an ultrahigh-vacuum system for a given time at fixed $\dot{R}$. The deposit is then covered with carbon and the substrate is dissolved away, leaving a rigid carbon replica which retains the metal clusters in their original crystal orientation. To enhance the very early stages of nucleation the technique of decoration is practiced. Existing clusters are decorated with a lower melting-point metal such as Zn or Cd which does not condense in the absence of prior noble metal deposition. This technique renders visible otherwise invisible clusters which may contain just a few atoms, thus making possible more precise comparisons with theory.

A disadvantage of the foregoing postmortem step-by-step observations is that much useful information on intermediate stages of film nucleation and growth is lost. For this reason, *in situ* techniques within both scanning (SEM) and transmission electron microscopes, modified to contain deposition sources and heated substrates, have been developed, but usually at the expense of decreased resolution. A recent *in situ* TEM study by Ross (Ref. 29) of the CVD deposition of Ge on Si (001) substrates illustrates the power of such methods. The Ge was grown at 650°C from digermane (Ge_2H_6) gas at pressures between 10^{-8} and 10^{-6} torr while dark-field images were continuously recorded. Epitaxial nucleation was deemed to occur after strain contrast was first detected at a dose of 50 L (or 19 monolayers; see Section 2.2.3) of Ge_2H_6. Two island shapes — square-bottomed pyramids and steep domes — then grew in a bimodal size distribution. Within minutes at a growth rate of 0.1 to 10 monolayers/min the smaller pyramids disappeared, leaving a sole distribution of dome-shaped islands $\sim$60 nm in size. An Ostwald ripening–like mechanism was suggested to be responsible

for the changes in nucleus density. Such observations have enabled a better understanding of the processing of "quantum dot" structures.

7.5.2 Auger Electron Spectroscopy (AES)

This technique is based on the measurement of the energy and intensity of the Auger electron signal emitted from atoms located within some 5–10 Å of a surface that is excited by a beam of incident electrons. The subject of AES will be treated in greater detail in Chapter 10. Here it is sufficient to note that the Auger electron energies are specific to, or characteristic of, the atoms emitting them and thus serve to fingerprint elemental composition. The magnitude of the AES signal is related directly to the abundance of the adatoms in question. Consider now the deposit/substrate combinations corresponding to the three growth modes (Ref. 30). If the AES signal from the film surface of each is continuously monitored during deposition at a constant rate, it will have the coverage or time dependence shown schematically in Fig. 7-19, assuming a sticking coefficient of unity.

In the case of island growth (Fig. 7-19a) the signal from the deposit atoms builds slowly while simultaneously that from the substrate atoms correspondingly falls. The interpretation of the AES signal in the case of layer growth (Fig. 7-19b) is more complicated. During growth of the initial monolayer, the Auger signal is proportional to the deposition rate and sticking coefficient of adatoms, as well as to the elemental detection sensitivity. For the second and succeeding monolayers the sticking coefficients change. This gives rise to slight deviations in slope of the AES signal each time a complete monolayer is deposited, and therefore the overall response is segmented as indicated. For S-K growth (Fig. 7-19c) the signal is ideally characterized by an initial linear increase up to one monolayer or sometimes a few monolayers. Then there is a sharp break and the Auger amplitude rises slowly as islands, covering a relatively small part of the substrate, emerge. It would be misleading to suggest that all AES data fit one of the categories in Fig. 7-19. Complications in interpretation arise from atomic contamination, diffusion and alloying between deposit and substrate, and transitions between two- and three-dimensional growth processes.

7.5.3 Some Results for Metal Films

The properties of metal particles and films on nonmetallic substrates have played an important role in optical coating and ceramic bonding technologies, and particularly in the performance and selectivity of catalysts (Ref. 31).

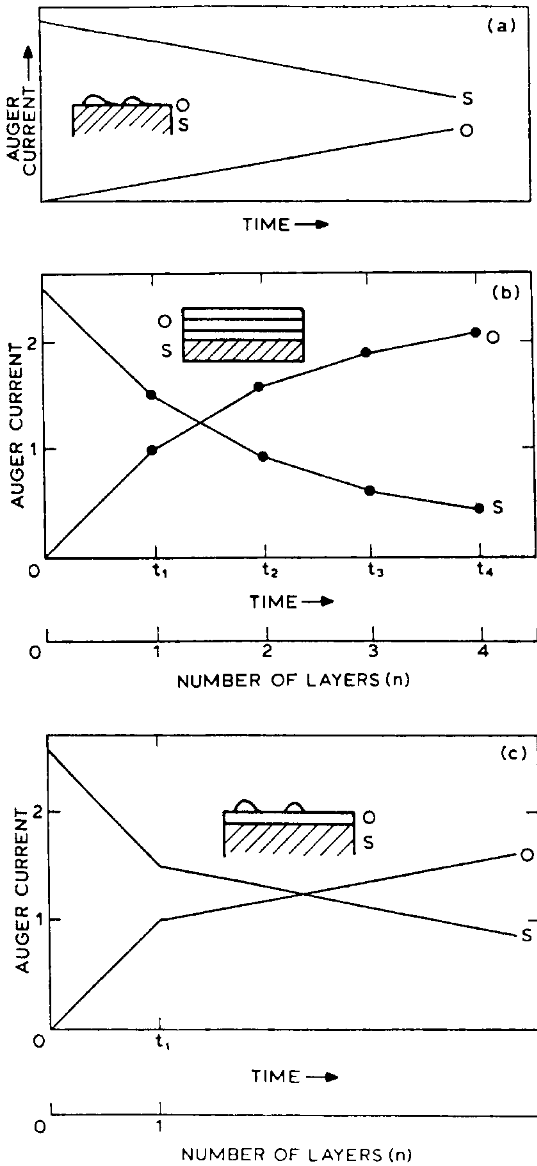


Figure 7-19 Schematic Auger signal currents as a function of time for the three growth modes: (a) island, (b) planar, (c) S-K. o = overlayer, s = substrate. (From Ref. 30.)

In addition, studies of the nucleation and growth of metals, especially the noble ones, on assorted alkali halide and oxide substrates have long provided the basis for understanding epitaxy and film formation processes. An appreciation of the scope of this past as well as present research activity on metal/substrate systems can be gained by noting that more than 300 references to the research literature refer just to epitaxial Au film/substrate combinations (Ref. 32). Other epitaxial metal-film/substrate systems have been comprehensively tabulated in the earlier literature (Ref. 33) together with deposition methods and variables. The sheer numbers and varieties of metals and substrates involved appears to point to the fact that epitaxy is a rather common phenomenon; island growth is involved in the overwhelming number of studies, but as a caveat the reader should see Sections 8.2.2 and 8.3.5. In the case of Au, epitaxial films have been deposited on metallic, covalent, and ionic substrate. Although the majority of substrate materials have cubic crystal structures, this is not an essential requirement for epitaxy, which occurs, for example, on hexagonal close-packed Zn as well as on monoclinic mica. The fact that epitaxy is possible between materials of different chemical bonding and crystal structure means that its origins are not simple. That a small difference in lattice constants between film and substrate is essential for epitaxy is apparently not true; small lattice mismatch is neither a necessary nor a sufficient condition for epitaxy. The lattice parameter of the metal can either be larger or smaller than that of the substrate. Having said this, it is also true that the defect density in these metal-film "island" epitaxial systems is very much larger than in the "planar" epitaxial semiconductor systems; in these very close lattice matching is maintained.

The following findings, now primarily of historical interest, briefly summarize numerous studies of nucleation and epitaxy of metal films deposited on ionic substrates (Ref. 34).

1. *Substrate*. The FCC metals generally grow with parallel orientations on (100), (110), and (111) surfaces of NaCl but with the (111) plane parallel to the (100) mica cleavage plane. Complex relative positioning due to translational and, more frequently, rotational movements appears to be the significant variable in epitaxy rather than lattice-parameter differences.

2. *Temperature*. High substrate temperatures facilitate epitaxy by: (a) lowering supersaturation levels, (b) stimulating desorption of impurities, (c) enhancing surface diffusion of adatoms into equilibrium sites, and (d) promoting island coalescence. The concept of an epitaxial temperature T_E has been advanced for alkali halide substrates. T_E depends on both the nature of the substrate and the deposition rate. For example, T_E for Ag on LiF, NaCl, KCl, and Kl was determined to be 340, 150, 130, and 80°C,

respectively. The progressive decrease in T_E correlates with increases in lattice parameter and enhanced ionic polarizabilities, which facilitate attractive forces between metal and substrate atoms.

3. *Deposition rate.* In general, low deposition rates, $\dot{R}$, foster epitaxy. It has been established that epitaxy occurs when $\dot{R} < \text{const} \exp - (E/k_B T_E)$. This inequality is satisfied physically when the rate at which adatoms settle into equilibrium sites exceeds the rate at which adatoms collide with each other. Such an interpretation requires that E be a surface diffusion activation energy rather than $E_{des} + E_2$ in Eq. 7-38b. The reader should compare this criterion for T_E with those proposed earlier.

7.5.4 Scanning Probe Microscopy Studies of Nucleation

Saved for last, we briefly note several more recent nucleation studies performed with the scanning tunneling (STM) and atomic force microscopes (AFM). These techniques, capable of unprecedented levels of resolution, have dramatically enhanced our understanding of surfaces and the atomic processes that occur on them. In particular, height variations at atomic steps are routinely measured with a resolution of 1 Å by STM. This makes it possible to determine whether islands have grown in only one atomic layer at a time, or atop one another in three dimensions. We start by considering the research by Lagally and co-workers (Refs. 8, 35, 36) on the earliest stages of Si atom homoepitaxy on a (001) Si substrate surface.

7.5.4.1 Silicon on (001) Si

Earlier in Section 7.2.4.1 it was pointed out that the (001) Si surface exhibits a (2×1) reconstruction and contains dimer rows. These rotate 90° as successive terraces are accessed giving rise to two types of monatomic steps. When silicon atoms are deposited on such surfaces the new layer of adatom islands form dimer rows. Occasionally, individual adatoms or pairs of atoms are visible. In addition to adatoms there are ad-dimers, which form when two adatoms pair up. Theory and experiment reveal that ad-dimers can translate as well as rotate. Although they exist predominantly on the dimer tops initially, with time they tend to populate the more stable trough configurations between dimers. Deposition on three successive terraces of the surface is shown in Fig. 7-20, each with their adatom island chains (light). Measurement of the island densities from such images recorded at a series of substrate temperatures yields the Arrhenius plot of Fig. 7-21. Because wandering adatoms either find an existing terrace edge or create

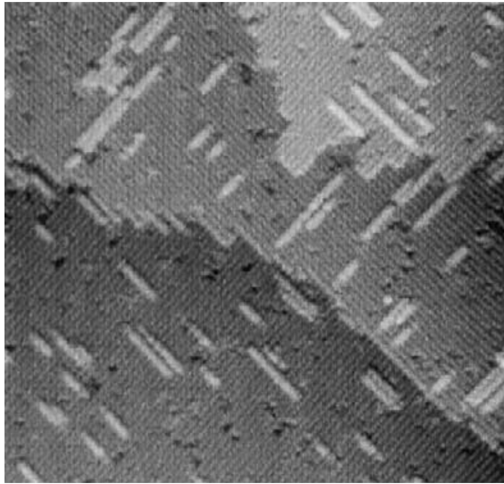


Figure 7-20 Scanning tunneling microscope image of Si deposition on three successive terraces. Field of view is 300 Å × 300 Å. (From Ref. 35.)

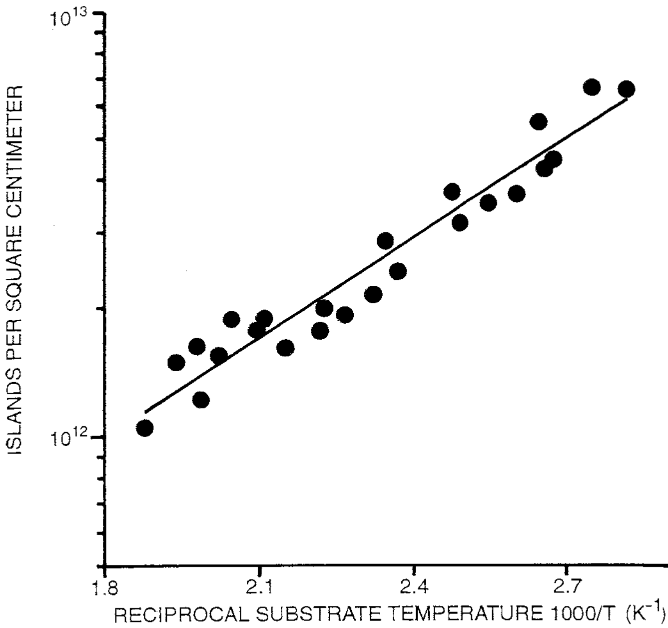


Figure 7-21 Arrhenius plot of island density vs the reciprocal of the substrate temperature. (From Ref. 35.)

new islands on the terrace by combining with other adatoms, the density of islands under specified deposition conditions is a measure of the diffusion coefficient. Thus, the energy barrier for surface diffusion determined from Fig. 7-21 is 0.65 eV, the value entered in Table 7-1. Direct STM observation has also revealed strongly anisotropic diffusion of Si adatoms. At 500 K adatom transport is 1000 times faster in the direction parallel to dimer rows than perpendicular to them.

7.5.4.2 Metals on Silicon

In a similar vein, Neddermeyer (Ref. 10) has reviewed studies of the initial stages of metal (Ag, Cu, Pd) film growth on (111) Si substrates. Atoms on this close-packed Si plane undergoes a very complex structural reconstruction that was mentioned in Section 7.2.4.2. In the case of silver, when only one-third of a monolayer deposits on this surface at 90°C, the STM image depicted in Fig. 7-22a is obtained. Individual adatoms of Ag in the form of a ringlike triangular entity denoted by A and B are seen to sit above the otherwise undisturbed substrate atoms. When additional Ag atoms attach to the initial nuclei, the larger faceted clusters C and D form.

7.5.4.3 Metals on Metals

In a study by Chambliss *et al.* (Ref. 37) on early stages of metal-film epitaxy, the roles of deposit–substrate strain and place-exchange intermixing were explored by STM. Growth traits of Au, Ag, Ni, and Fe on (111) Au are complicated by a herringbone-like reconstruction of the substrate surface. The latter arises because the smaller number of nearest neighbors on the surface (9) relative to the bulk (12) slightly shortens interatomic spacings so that a 4% greater atomic surface density arises. This means that the reconstructed Au surface is effectively a heteroepitaxial layer lying atop bulk Au. The resultant lattice-mismatch strain creates a complex array of dislocation and stacking fault defects that is manifested in herringbones. A 0.16 monolayer deposit of Ag on (111) Au yields the image shown in Fig. 7-22b. For the most part the Ag deposits are flat and have erased or disrupted the reconstructed Au surface. When the roles of Au and Ag were exchanged, however, it was observed that many Au atoms were buried a few angstroms beneath the (110) Ag substrate surface. This atomic intermixing occurs even at low temperatures. After the surface is covered with such a mixed Au–Ag layer, further Au deposition leads to 3-D Stranski–Krastanov growth of islands.

Interestingly, a similar striped pattern was observed during the electrodeposition of copper on a gold substrate (38). Apparently 10 monolayers of

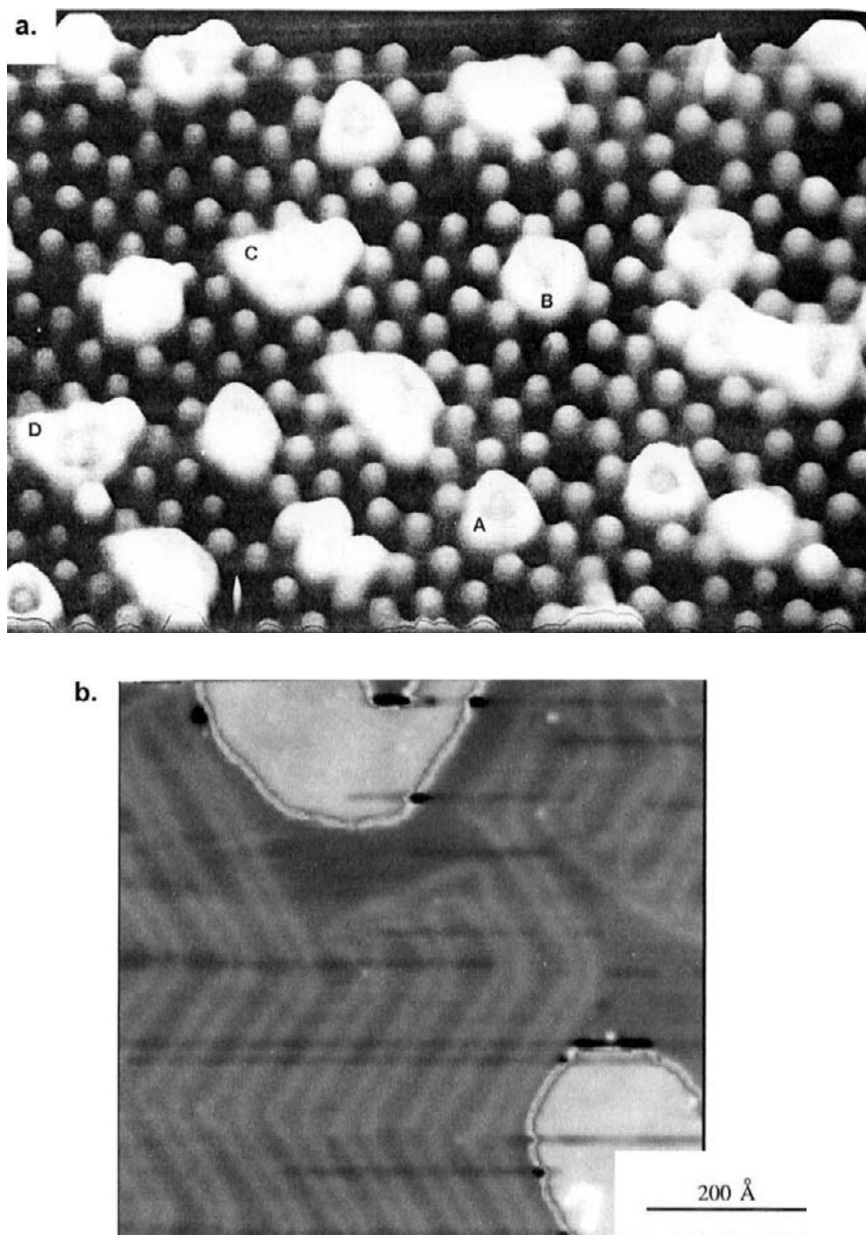


Figure 7-22 Scanning tunneling microscope images of deposited metal clusters. (a) Ag nuclei on (111) Si (Ref. 10). (b) A 0.16 monolayer deposit of Ag on (111) Au. (From Ref. 37.) (c) The 11th monolayer of Cu electrodeposited on Au. (From Ref. 38.) (Reprinted with the permission of the authors.)

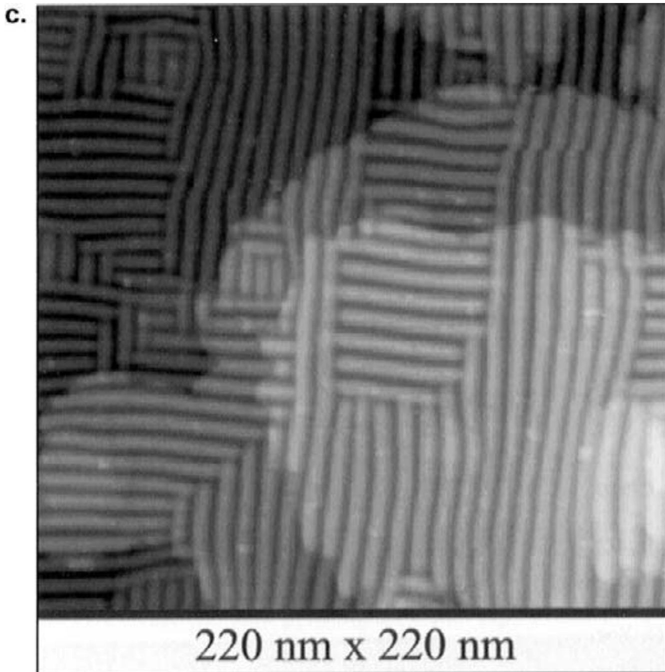


Figure 7-22 Continued.

metastable BCC Cu grow pseudomorphically on (100) FCC Au. But a transformation occurs in the 11th layer such that the Cu surface buckles and a new surface structure emerges containing small domains rotated 90° to one another (Fig. 7-22c). This marks the beginning of the deposition of the stable FCC Cu structure.

7.5.4.4 W_xN on SiO_2

In a last example, let us consider the use of the atomic force microscope (AFM) as a tool in studying film formation of W_xN on amorphous SiO_2. The function of this instrument is described in Section 10.3.5.4. Here we simply note that the AFM also traces surface topography, perhaps not with the spatial resolution of the STM, by detecting the van der Waals force interaction between tip and surface. In a study by Li *et al.* (Ref. 39), nucleation and growth of W_xN films by plasma-enhanced CVD at a deposition temperature of 340°C and rate of 1 nm/s was recorded during the first few seconds. The film topography is shown in Figs. 7-23a,b,c accom-

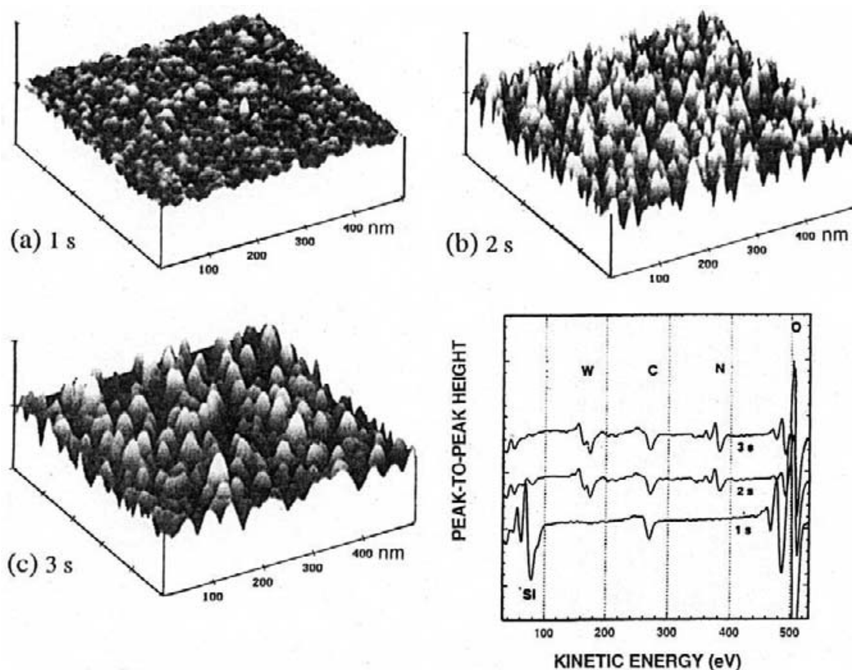


Figure 7-23 Atomic force microscope images of W_xN on SiO_2 after (a) 1 s, (b) 2 s, and (c) 3 s. The respective surface rms roughnesses are 0.13 nm (the same as for the original SiO_2), 0.50 nm, and 0.48 nm. Each division represents a length of 100 nm along the horizontal axis and a height of 3 nm on the vertical axis. Corresponding Auger spectra appear in the inset. (From Ref. 39. Reprinted with the permission of the authors.)

panied by information on film chemistry as revealed by Auger spectroscopy (Section 10.4.4). After 1 s there was no detectable W or N Auger signal, so this period may be taken as the deposition incubation time. A W_xN layer formed after 2 s consisting of 320 grains in an area measuring 500 nm $\times$ 500 nm, where each grain was $\sim$25 nm in diameter. Cluster coalescence occurred within 3 s, leaving 180 grains within the same area, where the size of each grain was now $\sim$45 nm. Clearly, the lateral grain dimensions are much larger than the nominal film thickness.

7.6 CONCLUSION

The earliest stages of film formation have long held a fascination for the scientific community. As a result, much direct observation via electron

microscopy was carried out over the years to reveal the behavior of small atomic clusters on substrate surfaces. In parallel there was a broad effort to theoretically model film structures and the processes that led to them, even though individual atoms could not be directly resolved. However, these traditional studies have been dramatically augmented in recent years by scanning probe techniques, particularly scanning tunneling microscopy, possessing the power to directly image adatom motion during deposition. In this way the crystallographic and kinetic details of nucleation events in a number of film/substrate systems, particularly semiconductor/semiconductor, but also metal/semiconductor as well as metal/ceramic, have been exposed.

The bulk of this chapter has been devoted to modeling nucleation ($\dot{N}$) and growth ($\dot{G}$) rates in terms of classical thermodynamic and kinetic descriptions, adopting both macroscopic and microscopic points of view. Starting with concepts of surface tension and capillarity theory, equilibrium sizes of nuclei and the energies to form them were thermodynamically estimated. Upon incorporating atomic surface-diffusion effects while allowing for loss of adatoms due to desorption, estimates for the macroscopic nucleation rate were obtained in terms of deposition variables and the materials properties involved. Atomistic theories have similarly yielded analogous expressions for nucleation rates. Theoretical modeling of nucleation and growth phenomena has aided our understanding of issues related to thin-film grain size and epitaxy. For example, in the former case large grains are fostered at high substrate temperatures, where ($\dot{G}$) is large, and at low deposition rates, where ($\dot{N}$) is small. Both trends are consistently incorporated into a quantitative relationship which predicts that film grain size ought to vary as $(\dot{G}/\dot{N})^{1/3}$.

Exercising practical control over thin-film structures means integrating an understanding of the detailed atomistic processes on real surfaces with the macroscopic implications of nucleation and growth theories.

EXERCISES

1. Among the surface electronic properties affected by differences in work function are thermionic emission and photoemission. The defining equations for these phenomena are $j = AT^2 \exp -(q\phi/k_B T)$ and $KE = h\nu - q\phi$, respectively. Here j is the thermionic current density, A is a constant, T is the temperature, q is the electronic charge, KE is the photoelectron kinetic energy, h is Planck's constant, and ν is the frequency of the incident light. By alloying the surface of tungsten with cesium it is found that ϕ is reduced from 4.5 to 1.4 eV.

(a) By what factor is current changed in both thermionic and photoemission processes as a result? *Note:* Current is proportional to electron velocity.
(b) Relative to a pure W emitter at 2000 K, what temperature is required to produce an equivalent thermionic current using a W–Cs emitter?

2. An adatom overlayer structure on a close-packed substrate surface, e.g., the (111) plane where atoms have hexagonal symmetry, is defined as $(3^{1/2} \times 3^{1/2})$R 30°. Sketch probable substrate and adatom positions for this case.

3. Consider a (110) substrate mesh containing adatoms lying along [001] and $[1\bar{1}0]$ directions.
   (a) If adatoms sit directly above every other substrate atom, identify the overlayer in the notation used in this chapter.
   (b) Suppose the adatoms are strung out along every other $[1\bar{1}0]$ direction. Identify the overlayer.

   Sketch the substrate and adatom positions for both cases.

4. Just as it is thermodynamically impossible to produce a perfectly pure material, explain why it is thermodynamically impossible to produce zero coverage ($\theta = 0$) or unity coverage ($\theta = 1$) adatom surface layers. *Hint:* Evaluate $d(G/n_s)/d\theta$ at $\theta = 0$ and 1.

5. A gas molecule with initial energy E_i approaches a surface.
   (a) Referring to Fig. 7-10, write expressions for the following quantities:
      1. Binding energy of the molecule in the physisorbed state.
      2. Binding energy of the molecule in the chemisorbed state.
      3. Energy required to desorb a chemisorbed molecule.
      4. Rate at which physisorbed molecules become chemisorbed. (Natural jump frequency is v_0.)
      5. Ratio of chemisorbed to physisorbed molecules in equilibrium.
   (b) Sketch Fig. 7-10 in a way that distinguishes the adsorption behavior between argon and nitrogen at a titanium surface.
   (c) Similarly, distinguish the adsorption behavior of oxygen at gold and iron surfaces.

6. (a) Cube-shaped nuclei are observed to form homogeneously in the gas phase, and heterogeneously on a flat substrate surface as well as at a step on this surface. For each of the three nucleation events described calculate the critical nucleus size and energy barrier for nucleation.
   (b) Under the same gas-phase supersaturation, how do these values compare with those for spherical and cap-shaped nuclei?

7. A cylindrical pill-like cluster of radius r nucleates on a dislocation that emerges from the substrate. The free-energy change per unit thickness is given by

$$\Delta G = \pi r^2 \Delta G_V + 2\pi\gamma r + A - B \ln r,$$

where $A - B \ln r$ represents the dislocation energy within the cluster.
   (a) Sketch ΔG vs r (note at $r = 0$, $\Delta G = \infty$).
   (b) Determine the value of r^*.
   (c) Show that when $\Delta G_V B/\pi\gamma^2 > \frac{1}{2}$, ΔG monotonically decreases with r, but when $\Delta G_V B/\pi\gamma^2 < \frac{1}{2}$, there is a turnaround in the curve. (The latter case corresponds to a metastable state and associated energy barrier.)

8. During e-beam evaporation, suppose that small metal particles (several nanometers in diameter) deposit on a substrate consisting of a metal covered with a thin dielectric oxide film. The particles are found to be charged so that a capacitor-like structure is formed.
   (a) Qualitatively suggest how nucleation theory might be altered in the presence of charged nuclei. Devise a simple procedure to change the film formation process.
   (b) Explain possible differences if oxidized n- or p-type silicon wafers were the substrates. In this case MOS structures are formed.

9. Young's equation plays a role in soldering technology.
   (a) Under what conditions will a spherical cap of solder wet substrates?
   (b) What conditions promote balling up of solder?
   (c) Fluxes are often used to promote good soldering. What terms in Young's equation are likely to be affected by fluxes?

10. Cap-shaped nuclei on substrates grow both by direct impingement of atoms from the vapor phase and by attachment of adatoms diffusing across the substrate surface.
   (a) In qualitative terms, how will the ratio of the two mass fluxes depend on nucleus size, area density of nuclei, and deposition rate?
   (b) Write a quantitative expression for the flux ratio, making any reasonable assumptions you wish.

11. (a) Two spherical nuclei with surface energy γ having radii r_1 and r_2 coalesce to form one spherical nucleus. If mass is conserved, calculate the energy reduction in the process.
   (b) Suppose two spherical caps of different radii coalesce on a planar substrate to form one cap-shaped nucleus. Calculate the energy reduction.

12. Two spherical nuclei of radii r_1 and r_2 are separated by a distance L. If $r_1 \gg r_2$, derive an expression for the time it will take for the smaller

nucleus to disappear by sequential atomic dissolution and diffusion to the larger nucleus by Ostwald ripening. Assume that the diffusivity of atoms on the surface is D_S. Make simplifying assumptions as you see fit.

13. Assume that the two nuclei in Fig. 7-18 coalesce by a sintering mechanism.
    (a) By carefully measuring the neck width and plotting it as a function of time, determine the value of n in the general sintering kinetics formula.
    (b) From these data, estimate an approximate value for the diffusivity. Assume $\gamma = 1\ \text{J/m}^2$, $T = 400°\text{C}$, and $\Omega = 17 \times 10^{-24}\ \text{cm}^3/\text{atom}$.

14. A film is deposited on a substrate by means of evaporation. In the expression for the rate of heterogeneous nucleation (Eq. 7-33), identify which terms are primarily affected by:
    (a) Raising the temperature of the evaporant source.
    (b) Changing the substrate material.
    (c) Doubling the source-substrate distance.
    (d) Raising the substrate temperature.
    (e) Improving the system vacuum.
    In each case qualitatively describe the nature of the change.

15. An atomic force microscope study of the nucleation and growth of polycrystalline silicon on SiO_2 substrates by rapid-thermal CVD methods revealed the following:

Substrate temperature	620°C	660°C	750°C
Incubation time (s)	90	40	20
Nucleus density (cm^{-2})	1.25×10^{10}	1.1×10^{10}	4×10^{9}
Nucleus spacing (nm)	94	108	161
Vertical growth rate (nm/min)	5.2	13.2	84.9
Horizontal growth rate (nm/min)	4	5.6	2.3
Nucleus shape (height/radius)	0.54	0.43	0.28
Surface roughness (nm)	8.8	9.8	15.6

Data from C. Basa, M. Tinani, and E. A. Irene, *J. Vac. Sci. Technol. A* **16**(4), 2466 (1998).

    (a) In each case rationalize the effect of substrate temperature on the indicated measured quantity.
    (b) Estimate the activation energies for Si nucleation and growth.
    (c) Sketch the apparent shape and size of the cluster deposits at the three temperatures.

16. A number of years ago a commercial product was introduced that enabled a metal to deposit only in a selected masked area of a substrate and nowhere else. In operation, a porous metal pellet containing a

proprietary ingredient was heated to a high temperature where a colorless substance presumably coated all surfaces within the chamber except that which was masked. Then the mask was removed from the substrate and the desired metal was evaporated. A thin metal film only deposited on the substrate surface previously masked. Remarkably, the area deposited on did *not* have to be in the line of sight of the evaporated metal source. Propose a scenario that would scientifically explain this apparently magical behavior.

REFERENCES

1. B. Lewis and J. C. Anderson, *Nucleation and Growth of Thin Films.* Academic Press, London, 1978.
2. R. W. Vook, *Int. Metals Rev.* **27**, 209 (1982).
3. C. B. Duke, ed., *Surface Science: The First Thirty Years*, *Surf. Sci.* **299/300** (1994).
4. W. A. Harrison, *Solid State Theory.* McGraw-Hill, New York, 1970.
5. M. Prutton, *Introduction to Surface Physics.* Clarendon Press, Oxford, 1994.
6. C. A. Wert and R. M. Thomson, *Physics of Solids*, 2nd ed. McGraw-Hill, New York, 1970.
7. D. P. Woodruff and T. A. Delchar, *Modern Techniques of Surface Science*, 2nd ed. Cambridge University Press, Cambridge, 1994.
8. Z. Zhang, F. Wu, and M. G. Lagally, *Ann. Rev. Mater. Sci.* **27**, 525 (1997).
9. H. E. Farnsworth, R. E. Shlier, and J. A. Dillon, *J. Chem. Phys. Solids* **8**, 116 (1959).
10. H. Neddermeyer, *Crit. Rev. Solid State Mater. Sci.* **16**(5), 309 (1990).
11. J. A. Kubby and J. J. Boland, *Surf. Sci. Rep.* **26**, 21 (1996).
12. G. Ertl and H. J. Freund, *Physics Today* **52**(1), 32 (1999).
13. M. Konuma, *Film Deposition by Plasma Techniques.* Springer Verlag, Berlin, 1992.
14. A. H. Cottrell, *Theoretical Structural Metallurgy.* Edward Arnold, London, 1948.
15. K. Jackson, D. Uhlmann, and J. Hunt, *Crystal Growth* **1**, 1 (1967).
16. J. Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy.* Academic Press, Boston, 1993.
17. K.-N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science for Electrical Engineers and Materials Scientists.* Macmillan, New York, 1992.
18. C. A. Neugebauer, in *Handbook of Thin-Film Technology*, eds. L. I. Maissel and R. Glang. McGraw-Hill, New York, 1970.
19. K. Reichelt, *Vacuum* **38**, 1083 (1988).
20. B. W. Wessels, *J. Vac. Sci. Technol.* **B(15)**, 1056 (1997).
21. D. Walton, T. N. Rhodin, and R. W. Rollins, *J. Chem. Phys.* **38**, 2698 (1963).
22. H. M. Yang and C. P. Flynn, *Phys. Rev. Lett.* **62**, 2476 (1989).
23. J. A. Venables, *Surf. Sci.* **299/300**, 798 (1994).
24. D. W. Pashley and M. J. Stowell, *J. Vac. Sci. Technol.* **3**, 156 (1966).
25. R. M. German, *Powder Metallurgy Science.* Metal Powder Industries Federation, Princeton, NJ, 1984.
26. D. Kashchiev, *Surf. Sci.* **86**, 14 (1979).
27. M. Avrami, *J. Chem. Phys.* **7**, 1103 (1937); **8**, 212 (1940); **9**, 177 (1941).
28. C. V. Thomspon, *J. Mater. Res.* **14**, 3164 (1999).
29. F. M. Ross, *IBM J. Res. Develop.* **44**, 489 (2000).
30. G. E. Rhead, *J. Vac. Sci. Technol.* **13**, 603 (1976).

31. L.-C. Dufour and M. Perdereau, in *Surfaces and Interfaces of Ceramic Materials*, eds. L.-C. Dufour, C. Monty, and G. Petot-Ervas. Kluwer, Dordrecht, The Netherlands, 1989.
32. R. W. Vook and B. Oral, *Gold Bull.* **20**(1/2), 13 (1987).
33. E. Grunbaum, in *Epitaxial Growth B*, ed. J. W. Matthews. Academic Press, New York, 1976.
34. K. L. Chopra, *Thin-Film Phenomena.* McGraw-Hill, New York, 1969.
35. M. G. Lagally, *Physics Today* **46**(11), 24 (1993).
36. Y.-W. Mo, J. Kleiner, M. B. Webb, and M. G. Lagally, *Surf. Sci.* **268**, 275 (1992).
37. D. D. Chambliss, R. J. Wilson, and S. Chiang, *IBM J. Res. Develop.* **39**, 639 (1995).
38. R. J. Randler, M. Dietterle, and D. M. Kolb, *Z. Phys. Chem.* **208**, 43 (1999).
39. H. Li, S. Jin, H. Bender, F. Lanckmans, I. Heyvaert, K. Maex, and L. Froyen, *J. Vac. Sci. Technol.* **B18**(1), 242 (2000).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 8

Epitaxy

8.1 INTRODUCTION

Two ancient Greek words, $\varepsilon\pi\iota$ (epi, placed or resting upon) and $\tau\alpha\xi\iota\zeta$ (taxis, arrangement), are the root of the modern word *epitaxy*, which refers to extended single-crystal film formation on top of a crystalline substrate. Epitaxy is arguably the most important phenomenon in semiconductor thin-film device technology. Interestingly, however, epitaxy is usually of little concern in most other major thin-film applications, e.g., optical, hard, and protective coatings and films used for recording, storage, and display of information. Although epitaxy was probably first observed to occur in alkali halide crystals more than a century ago, the actual word epitaxy was apparently introduced into the literature by the French mineralogist L. Royer in 1928 (Ref. 1). For many years the phenomenon of epitaxy continued to be of scientific interest to numerous investigators employing vacuum evaporation, sputtering, and electrodeposition methods. A sense of some of this early work on island growth systems, e.g., metal films on alkali halide substrates, was given in Chapter 7.

Actually, a hierarchy of epitaxial effects is exhibited by films and substrates having widely different crystal and electronic structures. The chapter starts by discussing them so that issues related to epitaxial crystallography, interfacial atomic interactions, and film stress and defects can be better appreciated.

Over the past several decades epitaxy has left the laboratory and assumed crucial importance in the processing of electronic and optoelectronic devices. For these purposes, defect-free epitaxial films exhibiting layer growth

are essential, and that is where our interest will be centered in this chapter. At the outset two types of epitaxy can be distinguished and each has important scientific and technological implications. *Homoepitaxy* refers to the case where the film and substrate are the same material. Epitaxial (epi) Si deposited on Si wafers is the most significant example of homoepitaxy. In fact, one of the first steps in the fabrication of integrated circuit transistors is CVD vapor-phase epitaxy of Si on Si (see Section 6.3.2). The reader may well ask why the underlying Si wafer is insufficient; why must single-crystal Si be extended by means of the epi-film layer? The reason is that the epilayer is generally freer of defects, is purer than the wafer substrate, and can be doped independently of it. A dramatic improvement in the yield of early bipolar transistors was the result of incorporating the epi-Si deposition step. The second type of epitaxy is known as *heteroepitaxy* and refers to the case where films and substrates are composed of different materials, e.g., AlAs deposited on GaAs. Heteroepitaxy is the more common phenomenon and the one the chapter almost exclusively treats. Optoelectronic devices such as light-emitting diodes (LEDs) and lasers utilizing compound semiconductors, are based on heteroepitaxial film structures.

Differences between the two basic types of epitaxy are schematically illustrated in Fig. 8-1. When the epilayer and substrate crystal are identical, the lattice parameters are perfectly matched and there is no interfacial-bond straining. In heteroepitaxy the lattice parameters are necessarily unmatched, and depending on the extent of the mismatch, we can envision three distinct epitaxial regimes. If the lattice mismatch is very small, then the heterojunction interfacial structure is essentially like that for homoepitaxy. However,

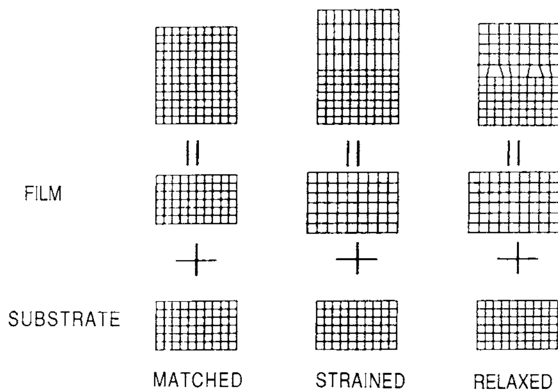


Figure 8-1 Schematic illustration of lattice-matched, strained, and relaxed heteroepitaxial structures. Lattice-matched heteroepitaxy is structurally similar to homoepitaxy.

differences in film and substrate chemistry and coefficient of thermal expansion can strongly influence the electronic properties and perfection of the interface. Small lattice mismatch is universally desired and actually achieved in a number of important applications through careful composition control of the materials involved.

When the film and substrate lattice parameters differ more substantially, we may imagine the other cases in Fig. 8-1. Either the two lattices strain to accommodate their crystallographic differences or, if this is not possible, dislocation defects form at the interface (relaxed epitaxy). *Strained-layer epitaxy* generally prevails during the early film formation stages irrespective of crystal structure or mismatch in lattice parameter. It usually occurs between film–substrate pairs composed of dissimilar materials which have the same crystal structure. Lattice parameter differences are generally an order of magnitude larger than in the case of lattice-matched heteroepitaxy. Near the interface the first few film layers strain to match the substrate crystallography; we say that such film growth is *pseudomorphic* and it gives rise to coherent epilayers. The outstanding examples of strained layer heteroepitaxy are the structures and devices consisting of Ge_xSi_{1-x} films grown on Si.

In recent years there has been a rising swell of research devoted to both the basic science of epitaxy and its engineering applications. A sampling of this activity includes:

1. Creating new wide-bandgap energy optoelectronic devices, i.e., LEDs and lasers, based on GaInN materials (Ref. 2), and improving the reliability of various photon sources and detectors used in optical communications systems.
2. Fabrication and use of assorted high-speed microelectronic digital, high-frequency, and wireless communications devices composed of Ge_xSi_{1-x} (Ref. 3) and compound-semiconductor materials (Ref. 4).
3. The dream of creating three-dimensional integrated circuits possessing intrinsically high device-packing densities. Rather than a single level of processed devices, a vertical multifloor structure can be imagined with each level of devices separated from neighboring ones by insulating films (Fig. 8-2). What is crucial is the ability to grow epitaxial semiconductor films on top of or bury them beneath insulating substrates, e.g., Si on SiO_2. This will require selective epitaxial growth at existing crystalline Si, and nowhere else, followed by lateral growth across surfaces that are ill suited to epitaxy (Ref. 5).
4. The fabrication of multilayer heterojunction composites. These epitaxial film structures include quantum wells and superlattices. By engineering bandgap structures some remarkable nanoscale quantum electronic and optical devices have emerged (Refs. 6, 7).

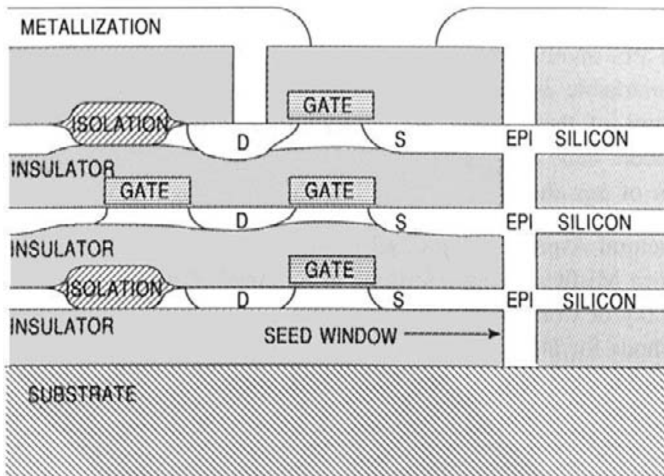


Figure 8-2 Cross-sectional model of a three-dimensional integrated circuit.

5. Deposition of oriented electroceramic films and multilayers for varied purposes, e.g., high-temperature superconductors, magnetic garnets for magnetooptical devices, and ferroelectric films for memory applications (Ref. 8).

In the previous chapter the earliest stages of homoepitaxy and *island epitaxy* were discussed. Building on this prior introduction and emphasizing thicker films, the present chapter has an ambitious agenda. Its purpose is to first present the physical issues related to film crystallography and defect structures that stem from epitaxial deposit–substrate interactions. The major focus of the chapter, however, is *layered-film epitaxy* of compound semiconductors. Assorted processes to deposit epitaxial compound-semiconductor films and the techniques employed to characterize them will be discussed. Properties of important epitaxial structures and recent applications in electronics and optoelectronics will be stressed throughout.

8.2 MANIFESTATIONS OF EPITAXY

8.2.1 INTRODUCTION

The concept of epitaxy spans a wide latitude of meaning with a hierarchy of structural and physical distinctions. On the one hand what has been called epitaxy can mean a loose structural accommodation between totally

different film and substrate materials. For example, *graphoepitaxy* exemplified by crystalline metals deposited on amorphous substrates lies at one end of the spectrum in this regard; the resulting films are quite defective. Epitaxy can also mean the intimate interaction between materials that are very similar electronically and chemically, possessing the same crystal structure and virtually identical lattice parameters, e.g., certain binary and ternary III–V compound semiconductors; in this case the films are relatively free of defects. This section addresses a variety of fundamental structural, crystallographic and property issues that arise in both common as well as extreme forms of epitaxial behavior.

8.2.2 Graphoepitaxy

This unusual form of epitaxy employs artificial microstructures, fabricated on a substrate surface, to manipulate the nucleation, growth, and orientation of film overlayers. For graphoepitaxy to occur, films must deposit with a given texture, e.g., (100). This means that (100) planes of grains lie parallel to the substrate surface, but their orientation is random and can rotate about the [001] axis. If the grains encounter a substrate pattern raised in relief, e.g., a grating with a period smaller than the grain size, the film will be constrained to perpetuate the original orientation. The effect can be understood by referring to the schematic depiction in Fig. 8-3 of a deposited polycrystalline film with (100) texture. When the substrate surface is either textured with a square (Fig. 8-3b) or triangular (Fig. 8-3c) wave relief, the growing film engulfs the otherwise randomly oriented grains that would form on a flat surface (Fig. 8-3a). Since no crystallographic or even bonding demands are made of the substrate, the latter can even be amorphous; for that matter so can the initial film deposit. Thus, starting with a deposit of amorphous Si (0.5 μm thick) on a fused (amorphous) SiO_2 substrate containing a square relief grating (3.8 μm spatial period, 0.1 μm deep), scanned laser-beam annealing recrystallized and enlarged the resulting Si grains to a size of 100 μm. But there were considerable angular spreads in the [100] direction lying parallel to the grating surface ($\pm 15^\circ$) as well as perpendicular to the film surface ($\pm 2.5^\circ$).

The phenomenon of graphoepitaxy was described in some detail because it reveals that films can be aligned or oriented even in the presence of a large misfit strain and lack of any apparent bonding between the film and substrate. Machlin (Ref. 10) has persuasively argued that many reports of epitaxy are, in fact, manifestations of graphoepitaxy, where assorted substrate terraces, cleavage steps, and defects are the artificial structures that foster oriented film growth. This is probably true of metal-island epilayer

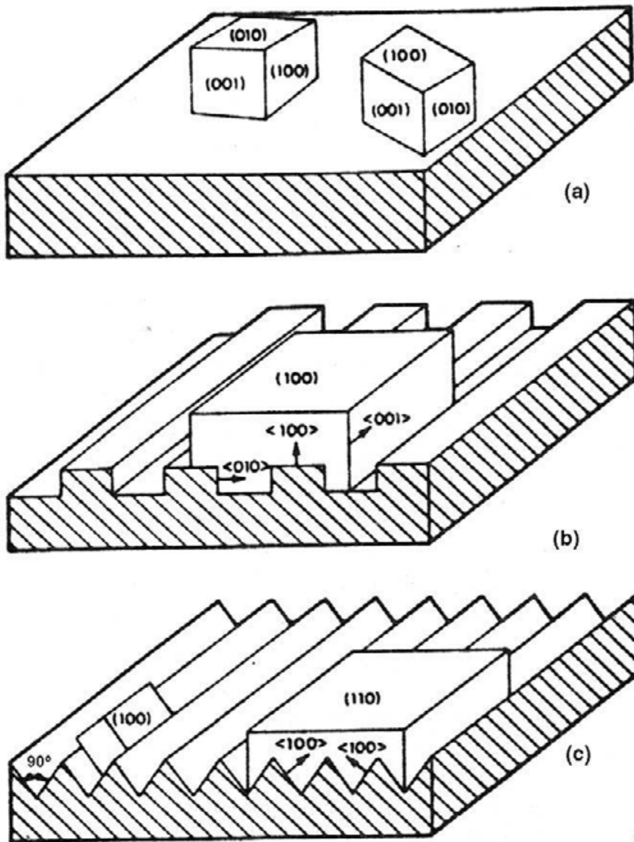


Figure 8-3 Examples of graphoepitaxy. (a) Polycrystalline film exhibiting a (100) texture deposited on an amorphous substrate. (b) Uniform (100) oriented film deposited on a substrate containing a square-wave relief. (c) Similar (100) oriented film deposited on a substrate containing a triangular-wave structure. (From Ref. 9. Reprinted with the permission of Dr. D. C. Flanders.)

films on alkali halide substrates (Section 7.5.3). In these systems there is generally weak bonding, large lattice misfits, and different film–substrate crystal structures and symmetries.

An important example of graphoepitaxy appears to occur in aluminum interconnect metallizations. Normally, Al grains exhibit (111) texture when deposited on blanket SiO_2 surfaces. However, films deposited within confining submicron channels, etched into the surrounding SiO_2, are oriented in a bamboo-like grain structure. Such a linear grain array is resistant to electromigration degradation (see Section 11.6.3) because the damage-prone

grain boundaries are oriented normal to the current flow. In effect, the Al stripe behaves like a long, thin, sequentially misoriented "single crystal."

8.2.3 HETEROEPITAXY

For the purposes of this chapter we can distinguish between two types of heteroepitaxial systems. The first deals with film–substrate materials that differ structurally, chemically, and electronically from one another, e.g., metal–semiconductor. In contrast there are the (compound) semiconductor–(compound) semiconductor heteroepitaxial structures which share a common crystallography and electronic structure, and often similar chemistry. It is the former case that is treated in this section.

8.2.3.1 Crystallographic Notation

In cases other than graphoepitaxy involving dissimilar materials, the relative orientations between the film and substrate are generally coupled. To understand the epitaxial relationship and the nature of the interfacial structure it is necessary to identify the crystallographic orientations between the film and substrate. Unlike the notation used to describe the two-dimensional surface crystallography in Section 7.2.3.2, the traditional (3-D) Miller indices are employed assuming there is no reconstructed substrate surface. For this purpose the indices of the overgrowth plane are written as (HKL) while those of the parallel substrate plane at the common interface are taken as (hkl). The corresponding parallel directions in the overgrowth and substrate planes, denoted by $[UVW]$ and $[uvw]$, respectively, must also be specified. This tetrad of indices, written by convention as $(HKL)//(hkl)$; $[UVW]//[uvw]$, serves to define the epitaxial geometry. As an example, for parallel epitaxy of Ni on Cu, the notation would read (001)Ni//(001)Cu; [100]Ni//[100]Cu. In this case both planes and directions coincide. For (111)PbTe//(111)$MgAl_2O_4$; [211]PbTe//[101]$MgAl_2O_4$ the interfacial plane is common but the directions are not. Sometimes, the epitaxial relationships can be predicted on the basis of lattice fitting arguments. Those planes and directions which give the best lattice fit often, but certainly not always, determine the film–substrate orientation.

An important quantity which characterizes epitaxy is the lattice *misfit* f, which was defined in Section 7.3.4.1 as

$$f = [a_0(s) - a_0(f)]/a_0(f), \tag{8-1}$$

where $a_0(f)$ and $a_0(s)$ refer to the unstrained lattice parameters of film and substrate, respectively. Epitaxy is often a matter of accommodation where

elimination of defects requires some bond stretching at the interface by both film and substrate atoms. A compromise is struck, creating a common interfacial lattice parameter. Thus, a positive f implies that the initial layers of the epitaxial film will be stretched in tension while the substrate is compressed. Similarly, a negative f means film compression and an extended substrate.

8.2.3.2 Geometrical Features of Metal/Semiconductor Heteroepitaxy

Metal/semiconductor structures are used for contact applications in all solid-state devices, and although epitaxial films may be desirable they are generally not essential. However, as device and contact dimensions continue to shrink the issue of metal/semiconductor heteroepitaxy assumes a greater urgency; not only does epitaxy affect electrical transport across the junction, but through heteroepitaxy the potential to fabricate three-dimensional silicon circuits is facilitated. Examples of metal- (and metal-silicide–) semiconductor heteroepitaxy are briefly treated in turn.

8.2.3.2.1 Metal/Silicon

We have already noted the sharp, defect-free interface between cobalt silicide ($CoSi_2$) and Si in Fig. 1-4. Similar epitaxial relationships hold between $NiSi_2$ and Si. What makes for excellent epitaxy in the cases of $CoSi_2$ and $NiSi_2$ is the fact that both silicides have cubic CaF_2 structures with respective lattice parameters of 5.365 Å and 5.406 Å. These are close to the a_0 value for Si, i.e., 5.431 Å, and therefore, the resulting lattice misfit for $CoSi_2$ is simply $f = (5.431\text{ Å} - 5.365\text{ Å})/5.365\text{ Å} = 0.0123$. Despite the larger misfit and poorer epitaxy in $TiSi_2$ this metal is currently the contact material of choice in silicon technology because of higher electrical conductivity.

8.2.3.2.2 β-$FeSi_2$/Silicon

Other metal silicide films do not exhibit the same epitaxial quality described above. The much-studied β-$FeSi_2$ is nevertheless interesting because it is a semiconductor and not a metal like other silicides (11). Imposing its low-symmetry orthorhombic structure ($a = 9.86$ Å, $b = 7.79$ Å, $c = 7.88$ Å) on the threefold symmetric (111) Si substrate causes oriented domains of epitaxial β-$FeSi_2$ films to form. As shown in Fig. 8-4a the domains are rotated relative to one another. For the indicated overgrowth the epitaxial notation for β-$FeSi_2$ grown on (111) Si is (101)β-$FeSi_2$//(111)Si:[010]β-$FeSi_2$//[$\bar{1}$10]Si. Similarly, when β-$FeSi_2$ films form on (100) Si substrates, the epitaxial ori-

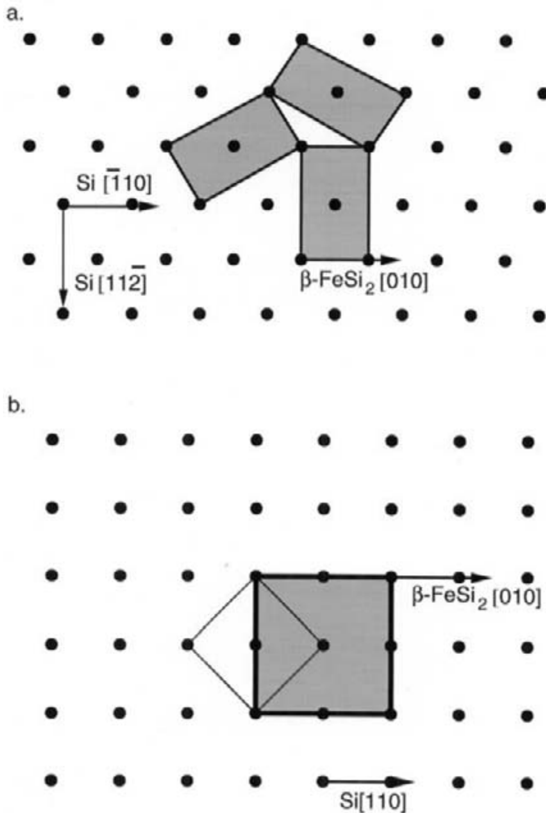


Figure 8-4 Epitaxial relationships between β-$FeSi_2$ and Si. (a) β-$FeSi_2$ (shaded) grown on (111) Si. (b) β-$FeSi_2$ grown on (100) Si. (From Ref. 11. Reprinted with the permission of the authors.)

entation shown in Fig. 8-4b is (100)β-$FeSi_2$//(100)Si:[010]β-$FeSi_2$//[110]Si. If there is no surface reconstruction in these diagrams, calculation of the misfit is relatively straightforward. For example, in the case of the latter, assuming the tetragonal structure approximates the orthorhombic lattice,

$$f = \frac{\sqrt{2} \times 5.431\,\text{Å} - 7.79\,\text{Å}}{7.79\,\text{Å}} = -0.0140\,.$$

8.2.3.2.3 Metal–GaAs Heteroepitaxy

Similar epitaxial relationships between assorted metals and compounds deposited on GaAs and other compound-semiconductor surfaces have also

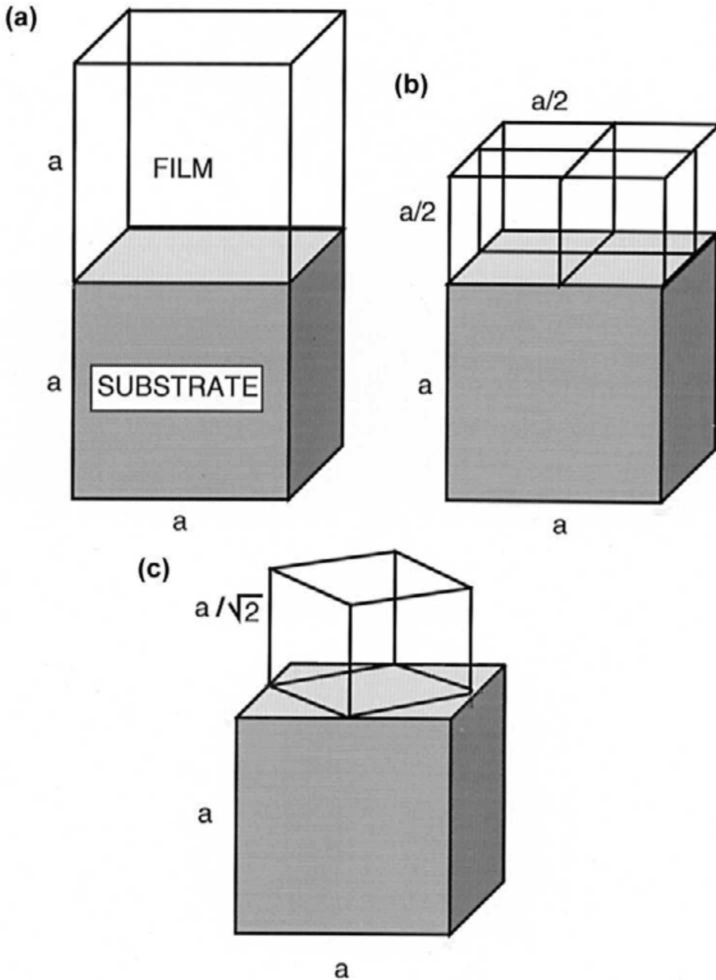


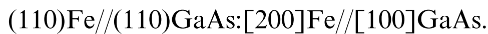
Figure 8-5 Simple epitaxial alignments for cubic films (f) on a cubic substrate (s). (a) $a_0(f) = a_0(s)$; (b) $2a_0(f) = a_0(s)$; (c) $2^{1/2}a_0(f) = a_0(s)$. (From Ref. 12. Reprinted with the permission of the author.)

been investigated (Ref. 12). Based on simple geometric matching we can distinguish three different epitaxial arrangements in Fig. 8-5, where the cubic (zinc blende) GaAs substrate ($a_0(s) = 5.653$ Å) is covered by films of materials that also have a cubic crystal structure (e.g., FCC, BCC, CsCl, NaCl).

The simplest epitaxial alignment is the cube-on-cube (Fig. 8-5a) where the lattice parameters are equal ($a_0(s) = a_0(f)$). For (100) GaAs only the

impractical elements Ca (FCC) and Rb (BCC) provide the closest lattice matches. However, rare-earth arsenide (RE-V) films, e.g., ErAs, YbAs, and $Er_xSc_{1-x}As$ films, exhibit reasonably good epitaxy. These compounds have the NaCl structure so that the As sublattice is continuous across the (100) interface, thus promoting chemical stability.

There are cases where $a_0(s)$ is almost twice $a_0(f)$, and such metals are expected to grow epitaxially (Fig. 8-5b). For example, consider the growth of an (110) Fe (BCC) film on (110) GaAs. Because the lattice parameter for Fe is 2.866 Å it appears that two Fe unit cells could be accommodated by one of GaAs. The resulting epitaxial geometry is denoted by



$$(110)\text{Fe}//(110)\text{GaAs}:[200]\text{Fe}//[100]\text{GaAs}.$$

In this system the misfit in the [001] direction is

$$f = [5.654 - 2(2.866)]/2(2.866) = -0.0138.$$

High-quality epitaxial Fe films have been deposited on GaAs, enabling fundamental studies of ferromagnetism (Ref. 13).

If $a_0(s)$ is close to $\sqrt{2}\,a_0(f)$ in magnitude, epitaxial deposits are expected to grow in 45° rotated alignment based on lattice-matching arguments (Fig. 8-5c). Candidate metals for such epitaxy include Al and Ag. Epitaxy does occur, but since other film orientations are also observed the situation is more complex. A few monolayers of intervening metal, e.g., Ga for Al, Fe for Ag, sometimes fosters the predicted epitaxy, perhaps by altering surface energies or by providing a growth template.

It is clear that a more complete understanding of dissimilar materials epitaxy must consider issues of interfacial bonding, stress, and thermodynamic stability in addition to simple geometric matching.

8.2.4 Tilted-Layer Epitaxy

In the various manifestations of epitaxy considered it has been tacitly assumed that the epitaxial films grow with their planes parallel to those of the substrate. But, as schematically illustrated in Fig. 8-6 (top), we have a situation where planes of the epitaxial film are coherently tilted with respect to the substrate planes (Ref. 14). This manifestation of epitaxy occurs on the vicinal surfaces of a miscut substrate. The exposed *vicinal* planes are slightly misoriented from some major low-index or *singular* plane. As a result the surface breaks up into an array of low-index terraces of generally variable length terminated by steps whose heights may also vary. When films are deposited on such slightly misoriented (~ 2–$3°$) surfaces at high temperatures and/or low fluxes, adatoms are mobile enough relative to their

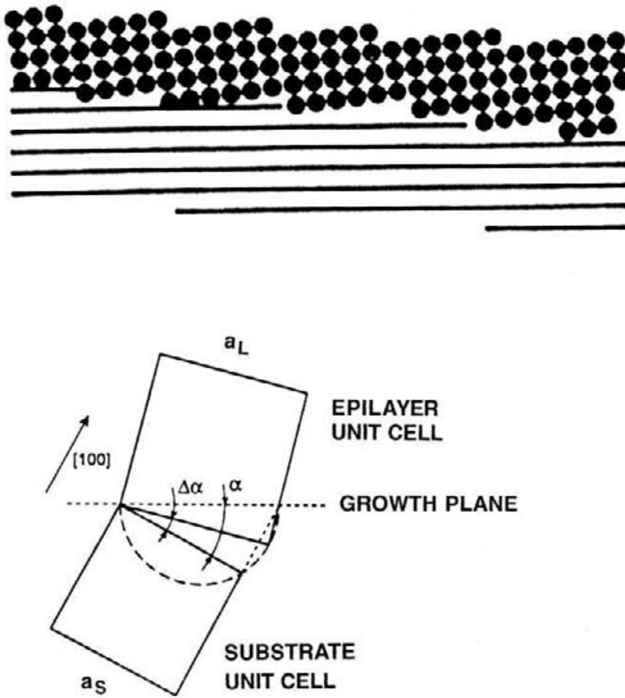


Figure 8-6 (Top) Illustration of tilted-layer epitaxy between film atoms (above) and vicinal planes of substrates (below). (From Ref. 14.) (Bottom) Geometry of strain relief by tilting lattice-mismatched vicinal heterosystems. (After Ref. 15.) (Reprinted with the permission of the authors.)

encounter probability and incorporate directly at step edges. In this regime, film growth occurs by step advancement along terraces, i.e., a "step-flow" process. Homoepitaxy of silicon on slightly tilted silicon surfaces, considered in Section 7.5.4.1, essentially proceeds by this mechanism.

Let us now quantitatively relate the angle of tilt necessary to preserve heteroepitaxial growth of films on a similarly stepped substrate with terraces of constant length. We consider cubic materials under the geometric conditions shown in Fig. 8-6 (bottom) where (100) substrate planes are misoriented from the exposed growth plane by angle α. With a tilt of angle $\Delta\alpha$ both film and substrate planes will essentially periodically coincide where they intersect at the interface. For a lattice misfit of f, elementary geometry reveals this will happen when

$$(1 + \mathrm{f})^{-1} \cos \alpha = \cos(\alpha - \Delta\alpha) \tag{8-2}$$

provided the epilayer is fully relaxed (Ref. 15).

8.3 LATTICE MISFIT AND DEFECTS IN EPITAXIAL FILMS

8.3.1 Equilibrium Theory of Lattice Misfit

In this section we explore some implications of lattice misfit with respect to the perfection of epitaxial films. The basic theory which accounts for the elastic/plastic changes in the bilayer was originally proposed by Frank and van der Merwe (Ref. 16). It attempts to account for the accommodation of misfit between two lattices rather than being a theory of epitaxy per se. The theory predicts that any epitaxial layer having a lattice-parameter mismatch with the substrate of less than ~9% would grow pseudomorphically. Initially, very thin films strain elastically to have the same interatomic spacing as the substrate, making the interface coherent with atoms on either side lining up. With increasing film thickness the rising total elastic strain energy will eventually exceed the energy associated with a relaxed structure consisting of an array of misfit dislocations (discussed in Section 8.3.4) separated by wide regions of relatively good fit. At this point the initially strained film ideally decomposes to this relaxed structure where a portion of the misfit is relieved by dislocations. As the film continues to grow, more misfit is relieved until at infinite thickness the elastic strain is totally eliminated. In the case of epitaxial growth without interdiffusion, pseudomorphism exists only up to some critical film thickness d_c beyond which dislocations are introduced. Matthews and co-workers (Ref. 17) were the first to derive an expression for d_c using an approach that was presented in the first edition of this book.

A similar, but perhaps more instructive treatment due to Nix (Ref. 18) is reproduced here. Throughout it is assumed that film and substrate have the same Young's (Y) modulus and same shear (μ) modulus. The analysis is divided into two film thickness (d) regimes that are depicted in Fig. 8-7.

1. $d < d_c$. In the early stages of film growth, *elastic* strain energy E_e (per unit area) increases with d as

$$E_e = Y\,d\varepsilon^2/(1 - \nu) \tag{8-3}$$

(see Eq. 1-45 for comparison) where ε is the biaxial elastic strain and ν is Poisson's ratio. No dislocations are present in the film.

2. $d > d_c$. Now consider the formation of (misfit) dislocations at the film–substrate interface as a means of relieving the elastic strain that develops during further film growth. If the dislocations are assumed to be arrayed in a square grid of side S, the elastic strain in the film is reduced from its initial misfit value (f) to $\varepsilon = \text{f} - b/S$. The quantity b/S is propor-

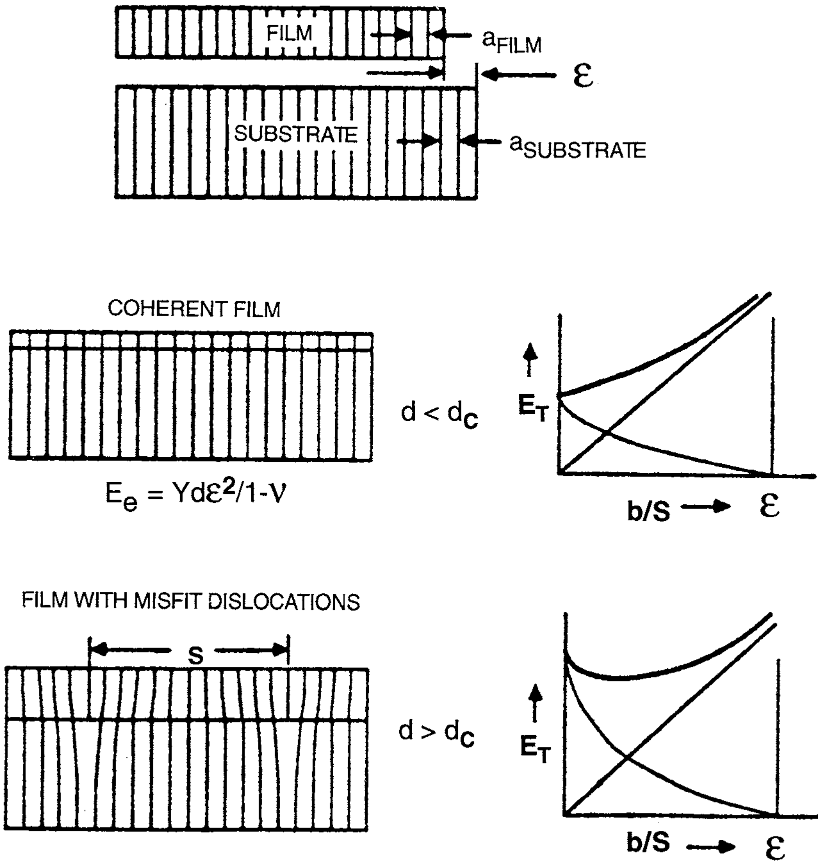


Figure 8-7 Illustration of the Matthews–Blakeslee equilibrium theory of misfit-dislocation formation. (Top) Coherent film is produced when $d < d_c$. (Bottom) Film with misfit dislocations result when $d > d_c$. (From Ref. 18. Reprinted with permission of W. D. Nix.)

tional to the number of misfit dislocations at the substrate interface and when $b/S = \text{f}$, the film strain vanishes. In essence each dislocation threads the entire film thickness and extends the lateral film length by the Burgers vector magnitude, b. The total strain energy E_T (per unit area) is now a sum of the elastic and dislocation energy E_d (per unit area) or

$$E_T = \underset{\text{(elastic)}}{\frac{Yd(\text{f} - b/S)^2}{(1-\nu)}} + \underset{\text{(dislocation)}}{\frac{\mu b^2\, 2\ln(\beta d/b)}{4\pi(1-\nu)S}}. \tag{8-4}$$

The second term deserves further comment. Theory shows that the energy per unit length of a dislocation is $[\mu b^2/4\pi(1-\nu)]\ln(R_0/b)+E_c$, where R_0 is a radius about the dislocation where the strain field terminates, and E_c is the dislocation core energy. After incorporating E_c and noting that R_0 scales as d, we may substitute $R_0 = \beta d$, where estimates of constant β range from 0.701 to 2.72. Also, the dislocation length per unit square area of grid is $\sim 2S/S^2$ or $2/S$, and by combining terms, Eq. 8-4 results. Physically the equation indicates that strain energy is a volume energy that increases linearly with film thickness. In contrast, dislocation energy is nearly constant with only a weak logarithmic dependence on d arising from R_0.

A plot of each energy contribution vs b/S is shown in Fig. 8-7. The fact that there is a minimum in E_T at a nonzero b/S value reveals that the structure is in mechanical equilibrium only if dislocations are present. By minimizing the total energy with respect to dislocation number, i.e., $dE_T/d(b/S) = 0$, and evaluating the resulting expression at $b/S = 0$, the critical film thickness (d_c) is determined to be

$$d_c = \frac{\mu b}{4\pi Y f}\ln(\beta d_c/b) \qquad \text{or} \qquad d_c = \frac{b}{8\pi(1+\nu)f}\ln(\beta d_c/b), \tag{8-5}$$

because $\mu = Y/2(1+\nu)$. For films thicker than d_c, misfit dislocations appear. In the region where d_c is approximately a few thousand angstroms, d_c is roughly $b/2f$. This means that the film will be pseudomorphic until the accumulated misfit $d_c f$ exceeds about half the unit cell dimension or $b/2$.

The validity of these ideas has been critically tested on several occasions in Si-based materials. By doping Si wafers with varying amounts of boron, whose atomic size is smaller than that of Si, the lattice parameter of the substrate can be controllably reduced. This affords the opportunity to study defect generation in subsequently deposited epitaxial films under conditions of very small lattice mismatch. As expected, an increase in f resulted in an increase in misfit dislocation density (Ref. 20).

8.3.2 DEFECTS in Ge_xSi_{1-x}/Si FILMS

Experimentation on Ge_xSi_{1-x}/Si epitaxial films has enabled the theory to be extended into the regime of large lattice misfits (Ref. 21). The results are shown in Fig. 8-8 where regions of lattice-strained but defect-free (commensurate) epitaxy are distinguished from those of dislocation-relaxed (discommensurate) epitaxy. Nature is kinder to us than the Matthews theory would suggest, and considerably thicker films than d_c predicted by Eq. 8-5 can apparently be deposited in practice. The reason is that Ge_xSi_{1-x}

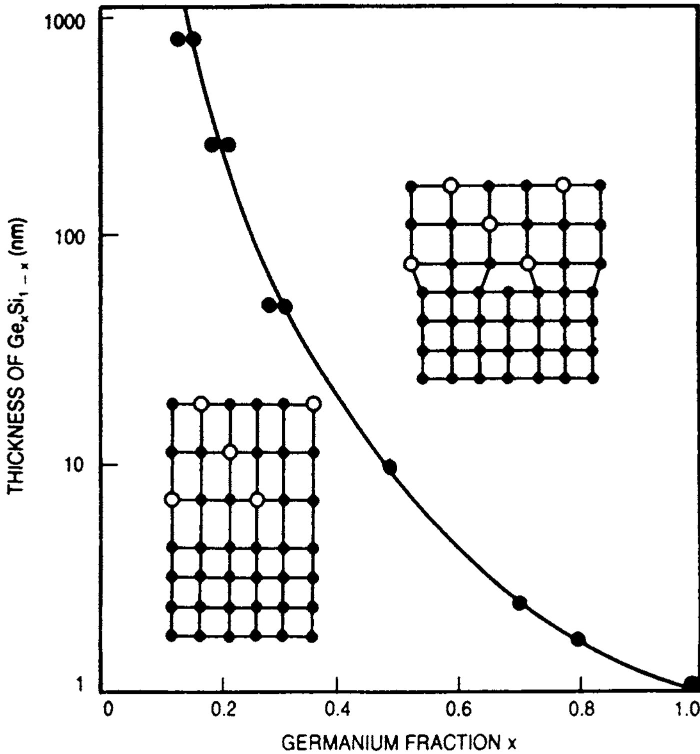


Figure 8-8 Experimentally determined limits for defect-free strained-layer epitaxy of Ge_xSi_{1-x} on Si. Note that f is proportional to Ge fraction. (From Ref. 22.)

strained-layer films are not in equilibrium. Extended dislocation arrays do not form instantaneously with well-defined spacings; rather, dislocations nucleate individually over an area determined by a width w and unit depth, over which atoms above and below the slip plane are displaced by at least $b/2$. For isolated screw dislocations, the energy per unit length is approximately equal to $\sim \mu b^2/4\pi \ln d/b$; when formed in a film of thickness d, its area energy density is then given by $(\mu b^2/4\pi w) \ln d/b$. By equating this to the elastic energy which drives dislocation nucleation, Bean has shown that

$$d_c = \frac{(1-\nu)b^2}{(1+\nu)8\pi w f^2} \ln \frac{d_c}{b}. \tag{8-6}$$

The solid line of Fig. 8-8 (Ref. 22) represents the excellent fit of this equation to the experimental data for the case where w is arbitrarily chosen to be five

[110] lattice spacings or ~19.6 Å. When the lattice misfit becomes large, the spacing between misfit dislocations decreases to the order of only a few lattice spacings. In such a case lattice misfit theory breaks down. There are no longer large areas of good fit separated by narrow regions of poor fit. Rather, poor fit occurs everywhere.

Epitaxial Ge_xSi_{1-x} films exhibit interesting strain-induced modulations in surface morphology (Ref. 23) that are shown in the TEM cross-sectional image of Fig. 8-9 (top). The surface ripples arise because the film is under compressive stress, a consequence of the fact that lattice parameters of Ge–Si solid-solution alloys necessarily exceed those for Si. As a result the lattice-plane spacing of the film shrinks near the cusplike troughs and expands at the rounded peaks as schematically depicted. A flat surface might be expected to represent the minimum energy configuration. However, when relieval of film strain-energy outweighs the tendency of surface energy (γ) to smooth steps arising from the relaxation distortion, roughening of the surface occurs. Analysis (Ref. 24) reveals that under these conditions the inequality $Y\varepsilon^2 > 4\gamma\pi^2 t_{rms}/\lambda^2$ holds, where λ and t_{rms} are the sinusoidal undulation wavelength and mean height, respectively, and ε is the mismatch strain.

Another manifestation of strain relaxation in compressed epitaxial films is the transition from a planar film to a three-dimensional morphology consisting of pyramidal islands (Ref. 25). This effect is shown in an annealed heteroepitaxial film of $Si_{0.7}Ge_{0.3}$ at the bottom of Fig. 8-9. The resultant faceted dome-shaped islands produce a surface roughening that is troublesome when fabricating coherently-strained device structures.

8.3.3 Types and Sources of Defects in Epitaxial Films

The issues just raised with respect to misfit dislocations and nonplanar surface morphologies are but a part of a larger concern for defects in epitaxial films. Semiconductor wafers are now largely "dislocation free"; in silicon there are fewer than 10 dislocations/cm^2 while for GaAs the dislocation density is typically less than 1000/cm^2 (Ref. 26). It is well known that dislocations, twins, and stacking faults degrade many device properties by lowering carrier concentrations and mobilities. They create states in the energy gap and serve to reduce the minority carrier lifetime and quantum efficiency of photonic devices. By acting as charge-recombination centers, defects drastically reduce the radiative processes of photon emission or absorption. For these reasons majority-carrier devices can tolerate higher defect levels (e.g., ~10^5/cm^2) than discrete minority-carrier devices (e.g.,

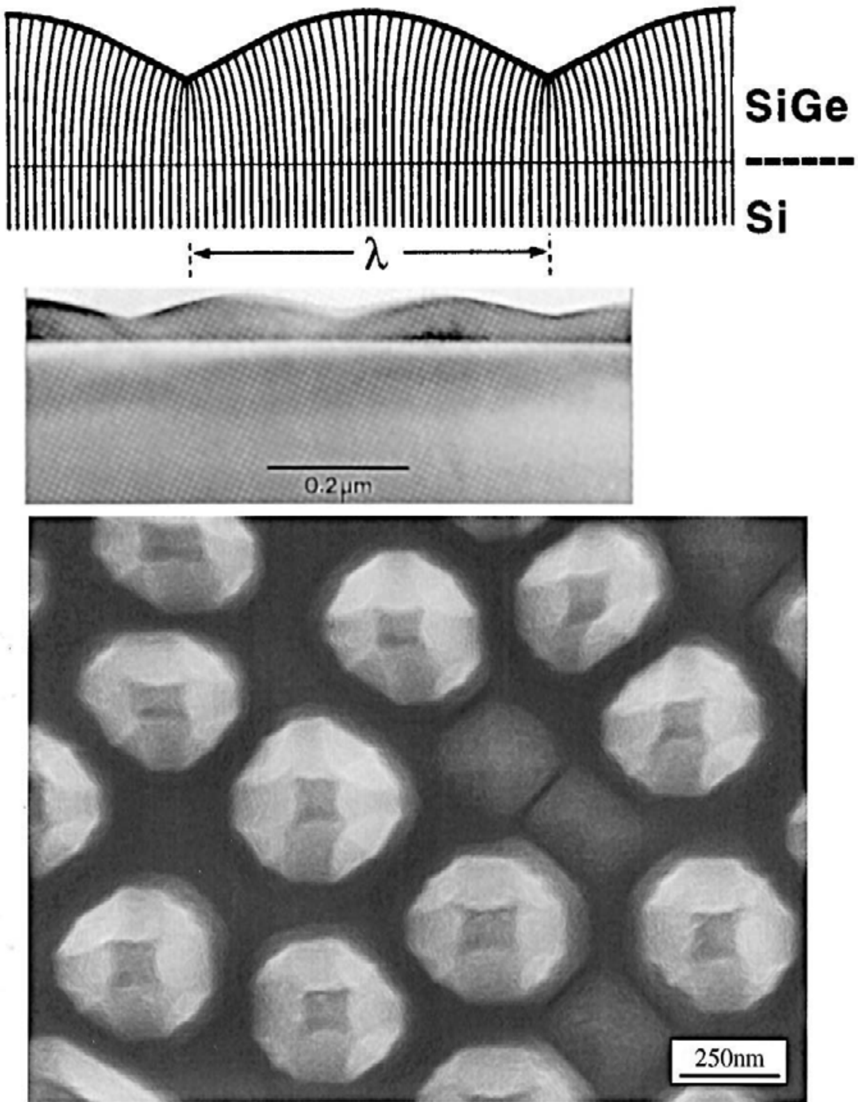


Figure 8-9 (Top) Cross-sectional [100] TEM image of strain-induced surface ripples on an uncapped $Ge_{0.81}Si_{0.19}$ film and accompanying schematic depiction of the distortion of vertical lattice planes. (After Ref. 23. Reprinted with the permission of the author.) (Bottom) Scanning electron micrograph top-view of faceted pyramid and dome-shaped islands formed during growth of a 40 nm thick $Si_{0.7}Ge_{0.3}$ alloy film on (001) Si. Film growth was by UHVCVD (Section 8.6.3.2) at 690°C. The largest islands are 70 nm high. (Courtesy of F. M. Ross, IBM, T. J. Watson Research Center.)

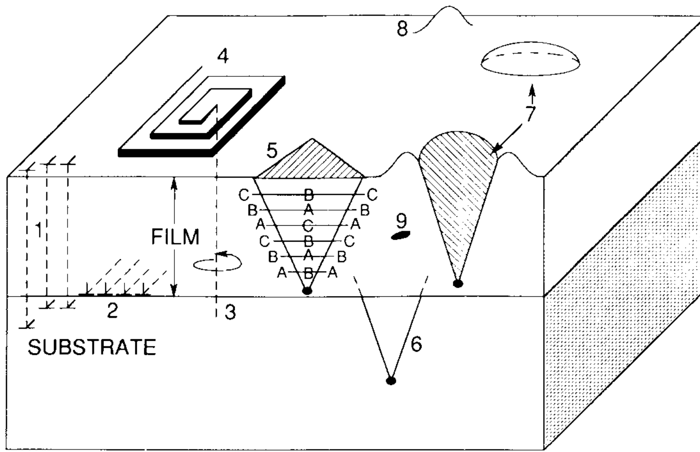


Figure 8-10 Schematic composite of crystal defects in epitaxial films. 1, Threading edge dislocations; 2, interfacial misfit dislocations; 3, threading screw dislocation; 4, growth spiral; 5, stacking fault in film; 6, stacking fault in substrate; 7, oval defect; 8, hillock; 9, precipitate or void.

$\sim 10^3/cm^2$). In a decade the Semiconductor Industry Association projects that defect counts in silicon wafers will (must) drop by a factor of between 15 to 20, e.g., 1 oxygen-induced stacking fault/cm^2 by 2010 (Ref. 27). Achieving such reductions in increasingly densely packed integrated circuits will require enormous efforts, first to uncover defect origins and then to devise methods for their elimination.

Several categories of defects in epitaxial films are shown schematically in Fig. 8-10 and it is instructive to briefly consider each in turn.

8.3.3.1 Defect Propagation from Substrate

A classic example of this is the propagation of an emergent screw dislocation spiral from the substrate surface into the growing film. Depositing atoms preferentially seek the accommodating ledge sites of the dislocation spiral staircase. Layers of defect-free film then radiate laterally to cover the substrate. Except for the screw dislocation, an epitaxial layer of otherwise good fit is possible. Occasional substrate dislocations which are present are apparently sources of such dislocations observed in homoepitaxial layers. Gross defects such as grain boundaries and twins are, however, rarely present.

8.3.3.2 Stacking Faults

Stacking faults are crystallographic defects in which the proper order of stacking planes is interrupted. For example, consider the first three atomic planes or layers of a (111) silicon film. Each of these planes may be imagined to be a close-packed array of atomic spheres (Fig. 1-2a), and each successive layer fits into the interstices of the previous layer. Atoms in the second layer (B) have no choice but to nest in one set of interstices of the first (A) layer. It now makes a difference which set of B layer interstices atoms of the third layer choose to lie in. If they lie above neither the A nor the layer atoms, the stacking sequence is ABC, and in a perfect epitaxial film the ABCABCABC, etc, order is preserved. If, however, a plane of atoms is missing from the normal sequence, e.g., ACABCABC, or a plane of atoms has been inserted, e.g., ABCBABCABC, then stacking fault defects are produced. It is established that they propagate from dislocations and oxide precipitates at the substrate interface. Misoriented clusters or nuclei containing stacking faults coalesce with normal nuclei and grow into the film in the manner of an inverted pyramid. Continuing growth causes the characteristic closed triangle shown in Fig. 8-10 to become progressively larger. For (100) growth the stacking faults form squares which can be revealed by appropriate etches. Although stacking faults are a vanishing breed in homoepitaxial Si films their density increases with decreasing growth temperature. In mismatched heteroepitaxial films, e.g., GaAs on Si, stacking faults are common.

Related to stacking faults are the ubiquitous "oval defects" observed during MBE growth (Section 8.6.2) of compound semiconductors. These defects, shown schematically in Fig. 8-10, are faceted growth hillocks which nucleate at the film–substrate interface and nest within the epitaxial layer. They usually contain a polycrystalline core bounded by four $\{111\}$ stacking-fault planes. With densities of $\sim 10-100\ \mathrm{cm}^{-2}$ and sizes ranging from 1 $\mu\mathrm{m}^2$ to $\sim 30\ \mu\mathrm{m}^2$, these defects are a source of considerable concern. There are a number of possible sources for oval defects in GaAs including carbon contamination from CVD precursors, incomplete desorption of oxides prior to growth, and spitting of Ga and Ga_2O from melts.

8.3.3.3 Dislocation Loops from Precipitates, Impurities, or Dopants

This category is self-explanatory. The precipitates and dislocations usually form during cooling and are the result of solid-state reactions subsequent to growth. Films containing high intentional or accidental dopants or impurity levels are susceptible to such defects.

8.3.3.4 Low-Angle Grain Boundaries and Twins

These defects arise from misoriented islands which meet and coalesce. When this happens small-angle grain boundaries or crystallographic twins result. The lattice stacking is effectively mirrored across a twin plane $\underline{C}$, i.e., . . . BCAB$\underline{C}$BACB. . . . Both types of defects may anneal out by dislocation motion if the temperature is high enough. During heteroepitaxial growth Matthews (Ref. 17) has suggested that there is some ambiguity in the exact orientation of small nuclei. He has formulated a rule of thumb that the relaxation of elastic misfit strain causes a variation in the orientation of crystal planes (in radians) roughly equal to the magnitude of the lattice misfit $\underline{f}$. Such an effect would naturally give rise to a network of small-angle grain boundaries.

8.3.4 Formation of Misfit Dislocations

More attention has probably been lavished on misfit dislocations than any other class of defect in high-performance epitaxial films. Although misfit dislocations lie in planes parallel to the substrate–film interface, they generally originate from threading dislocations (Ref. 28). These pierce through the film, the substrate, or both and lie in crystallographic planes that intersect the interface plane as shown in Fig. 8-11. Consider a threading

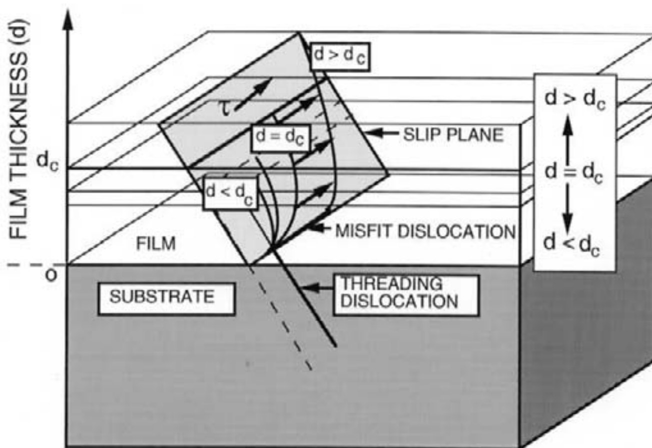


Figure 8-11 A depiction of how stress on thin film plane causes threading dislocation from substrate to form a misfit dislocation when $d > d_c$. (After Refs. 18 and 28.)

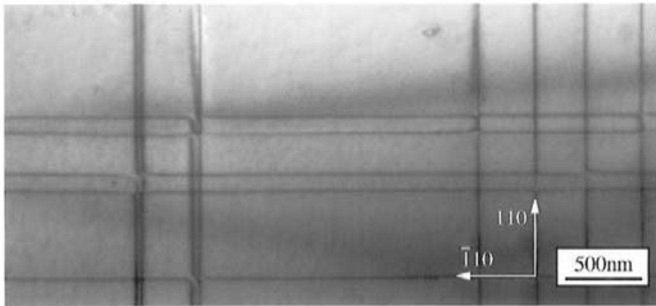


Figure 8-12 Misfit dislocation array generated during relaxation of a 200 nm thick $Si_{0.85}Ge_{0.15}$ alloy film on (001) Si. Bright field 220 transmission electron micrograph taken after MBE growth and 700°C annealing. (Courtesy of F. M. Ross, IBM, T. J. Watson Research Center.)

dislocation arising from the substrate. As it extends into the stressed film the threading component glides or bends in the slip plane. This dislocation segment bends more and more as the film thickens and becomes increasingly stressed. Correspondingly, the threading dislocation portion in the oppositely stressed substrate moves slightly in the opposite direction. Finally, at the critical thickness d_c, the film dislocation is able to glide infinitely, leaving behind a stable misfit dislocation at the interface. Threading dislocations can also nucleate from the top surface in the form of a small loop. Film stress expands the loop as the screw dislocation segments (parallel to Burgers vector **b**) are propelled outward while the edge component (perpendicular to **b**) glides to the interface and becomes the misfit dislocation. Since large misfits of $\sim$1–2% or more are necessary to nucleate dislocations, it is likely that stress-concentrating particulates, precipitates, or even atomic-scale defects are required to facilitate their formation.

Misfit dislocations have been often observed by transmission electron microscopy. For example, a misfit dislocation array, generated during relaxation of a SiGe alloy film on (100) Si, is shown in Fig. 8-12.

8.3.5 Epitaxial Defects in Perspective

In closing, it is appropriate to comment on the various levels of perfection that exist in epitaxial layers. Early epitaxial semiconductor films were judged to be single crystals based on standard X-ray and electron diffraction techniques that are relatively insensitive to slight crystal misorientations. However, subsequent electrical characterization of these films yielded significantly poorer electrical properties than anything imaginable in bulk melt-

grown single crystals. In general, thinned slices of bulk crystals were more structurally perfect than epitaxial films of equivalent thickness. This is certainly true of epitaxial films exhibiting island and SK growth and most of the films reported in the older literature. However, many of today's lattice-matched MBE films, grown under exacting conditions, are indeed structurally perfect when judged by the unambiguous standard of high-resolution TEM lattice imaging. In fact, the crystalline quality of epitaxial films generally exceeds that attainable in bulk grown crystals.

8.4 EPITAXY OF COMPOUND SEMICONDUCTORS

8.4.1 INTRODUCTION

Compound semiconductor films have been grown epitaxially on single-crystal insulators (e.g., Al_2O_3, CaF_2) as well as semiconductor substrates. However, it is the latter case that has attracted the overwhelming bulk of the attention and it will be our primary focus. In the exploitation of compound semiconductors for electronic and optoelectronic devices the creation of epitaxial heterojunctions is crucial (Refs. 29, 30). The term heterojunction refers to the interface between two single-crystal semiconductors of different composition and bandgap energy brought into contact, not differing doping levels of the same semiconductor. To better understand both the structure and electrical properties of these junctions and the devices that make use of them, an introduction to compound-semiconductor materials is an instructive way to start.

8.4.2 COMPOUND-SEMICONDUCTOR MATERIALS

8.4.2.1 Properties

Materials employed for epitaxial electronic and optoelectronic devices are largely drawn from a collection of direct-bandgap III–V semiconductor compounds. Although II–VI and IV–VI semiconductors are used to a much lesser extent, the discussion applies to them as well. Table 8-1 contains a list of important semiconductors together with some of their physical properties. Those that are pertinent to the possibility and quality of epitaxy are semiconductor nature, i.e., direct or indirect bandgap; bandgap energy; lattice constant; and thermal expansion coefficient. The implications of some of these properties on epitaxy/devices are detailed below.

Table 8-1
Semiconductor Properties

Material	Lattice parameter (Å)	Melting point (K)	P_{dis} (atm)	CTE ($10^{-6}\,°C^{-1}$)	Energy gap (eV at 25°C)	Electron mobility (cm^2/V-s)	Hole mobility (cm^2/V-s)
Diamond	3.560	– 4300		1.0	5.4	1800	
Si	5.431	1685		2.33	1.12I	1350	480
Ge	5.657	1231		5.75	0.68I	3600	1800
ZnS	5.409	3200		7.3	3.68D	120	
ZnSe	5.669	1790		7.0	2.58D	530	
ZnTe	6.101	1568		8.2	2.26D	530	130
CdTe	6.477	1365		5.0	1.56D	700	
HgTe	6.460	943		1.9	~0.15		
CdS				4.0	2.42D	340	
AlAs	5.661	1870	1.4	5.2	2.16I	280	
AlSb	6.136	1330	$<10^{-3}$	3.7	1.58I	900	400
GaP	5.451	1750	35	5.3	2.26I	300	150
GaAs	5.653	1510	1	5.8	1.43D	6500	300
GaSb	6.095	980	$<10^{-3}$	6.9	0.72D	5000	1000
InP	5.869	1338	25	4.5	1.35D	4500	100
InAs	6.068	1215	0.3	4.5	0.36D	30000	450
InSb	6.479	796	$<10^{-3}$	4.9	0.165D	80000	450
GaN	(*a*) 3.189 (*c*) 5.185	>1970			3.4D	1000	
AlN	(*a*) 3.112 (*c*) 4.982	>3000			6.2D		14
InN	(*a*) 3.54 (*c*) 5.705	~1400			1.89D		

D = direct; I = indirect. (*a*) and (*c*) refer to the lattice constants of the hexagonal α phases. From Refs. 11, 12, and M. S. Shur and M. A. Khan, *MRS Bull.* **22**(2), 44 (1997).

8.4.2.1.1 Direct and Indirect Energy Bandgaps

When light is emitted from or absorbed in a semiconductor, energy as well as momentum must be conserved. In a direct-bandgap semiconductor the carrier transitions between the valence and conduction bands occur without change in momentum of the two states involved. In the energy-

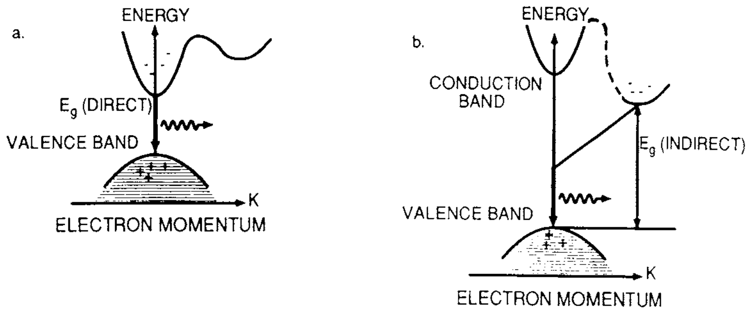


Figure 8-13 Depiction of electron transitions between conduction and valence bands in (a) direct- and (b) indirect-gap semiconductor materials.

momentum or equivalent energy-wave vector, parabola-like (E vs k) representation of semiconductor bands (Fig. 8-13), emission of light occurs by a vertical electron descent from the minimum conduction band energy level to the maximum vacant level in the valence band. This is what occurs in the direct energy gap materials GaAs and InP. However, in indirect-bandgap semiconductors such as Ge and Si the transition occurs with a change in momentum that is essentially accommodated by excitation of lattice vibrations and heating of the lattice. This makes direct hole–electron recombination with photon emission unlikely. But in direct-bandgap semiconductors such processes are more probable, making them far more efficient (by orders of magnitude) light emitters.

8.4.2.1.2 Bandgap Energy

This important semiconductor property can be understood by considering the variation of the absorption coefficient (α) as a function of photon energy as shown in Fig. 8-14. If light of intensity I_0 is incident on a semiconductor surface, the photon intensity at a depth x below the surface is attenuated to $I(x)$, such that

$$I(x) = I_0 \exp - \alpha x. \qquad (8\text{-}7)$$

In all semiconductors α becomes negligible once the wavelength exceeds the cutoff value. This critical wavelength, λ_c, is related to the bandgap energy E_g by a variant of the well-known relation $E = hv$ or $E_g = hc/\lambda_c$, where h, c, and v are Planck's constant and speed and frequency of light, respectively. Alternatively,

$$\lambda_c(\mu m) = 1.24/E_g(eV). \qquad (8\text{-}8)$$

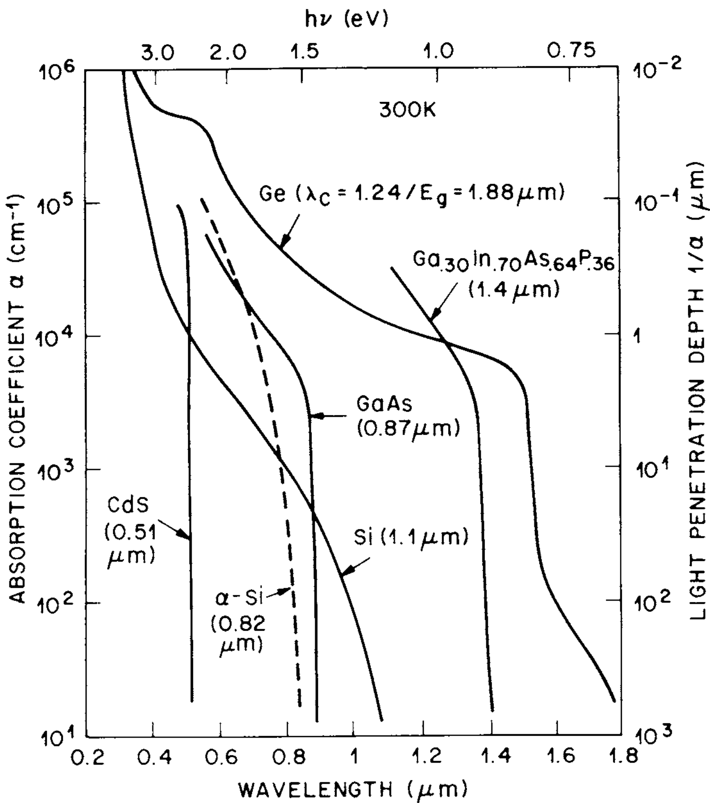


Figure 8-14 Optical absorption coefficients for various semiconductor materials. (Reprinted with permission from J. Wiley & Sons, S. M. Sze, *Semiconductor Devices—Physics and Technology*. McGraw-Hill, New York, 1985.)

For direct-bandgap semiconductors the value of α becomes large on the short-wavelength side of λ_c, signifying that light is absorbed very close to the surface. For this reason even thin-film layers of GaAs are adequate, for example, in solar cell applications. In Si, on the other hand, α varies more gradually with wavelength less than λ_c because of the necessity for phonon participation in light absorption/generation processes. Therefore, efficient solar-cell action necessitates thicker layers if indirect-energy-gap semiconductors are employed. In addition to a direct-bandgap semiconductor, photonic device operation generally requires a specific wavelength or value of E_g for light emission or absorption processes.

8.4.2.1.3 Lattice Parameter

To ensure defect-free interfaces in semiconductor film/substrate heterostructures, it is essential that the lattice parameters (a_0) of both be closely matched. For optical devices, lattice mismatches of less than 0.1% are sought. As an example consider GaAs and AlAs with respective a_0 values of 5.6532 and 5.6611 Å. The lattice mismatch between these compounds is $\Delta a_0/a_0$ or 0.16%. No hint of structural defects at the interface between epitaxial films of these compounds is observed in the lattice image of Fig. 8-15.

8.4.2.1.4 Thermal Expansion Coefficient

To prevent excessive thermal stress from developing at the film–substrate interface during deposition and processing, a match of thermal expansion coefficients is desirable. This condition is largely met in III–V devices based on GaAs, InP, and combinations of these materials.

8.4.2.2 Designing Epitaxial Film–Substrate Combinations

The first three of the semiconductor properties just mentioned converge in the extremely handy graphical representations of E_g vs a_0 shown in

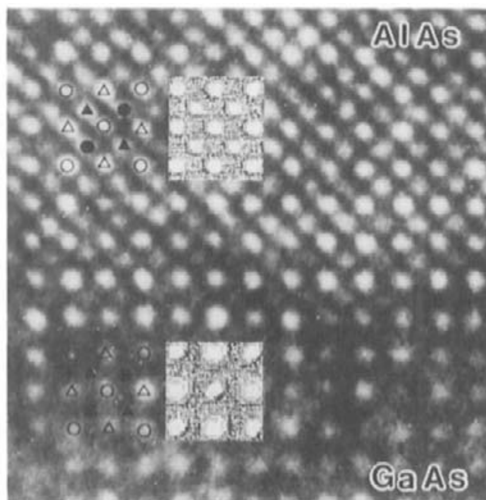


Figure 8-15 Electron microscope lattice image of GaAs/AlAs heterojunction taken with [100] illumination. (From H. Ichinose, Y. Ishida, and H. Sakaki, *JOEL News* **26E**(1), 8, 1988.)

Figs. 8-16a and 8-16b. These apply, respectively, to the common III–V compound semiconductors (Fig. 8-16a) where $E_g < 2\,\text{eV}$, and to the III–V and II–VI materials where $E_g > 2\,\text{eV}$ (Fig. 8-16b), together with their corresponding alloys. Through the use of these figures, the design and selection of complex semiconductor alloys with the desired properties may be visualized. Elements and binary compounds are represented simply as points. Ternary alloys are denoted by lines between constituent binary compounds. When one of the elements is common to both compounds, a continuous range of solid-solution ternary alloys generally forms upon alloying binaries. Thus the line between InP and InAs in Fig. 8-16a represents the collection of InP_xAs_{1-x} ternary solution alloys, with x dependent on the proportions mixed. Within the areas outlined by four binary compounds are quaternary alloys. Therefore, the $Ga_{1-x}In_xAs_{1-y}P_y$ system may be thought of as arising from suitable combinations of GaAs, GaP, InAs, and InP. However, there is no need to start with these four binary compounds when synthesizing a quaternary; it is usually only necessary to control the vapor pressures of the Ga-, In-, As-, and P-bearing species. Furthermore, solid lines represent direct-bandgap ternary compounds while the dashed lines refer to materials with an indirect bandgap.

A few examples will illustrate the use of these important diagrams. Suppose it is desired to create semiconductor materials for optoelectronic devices that have an energy gap of 1.0 eV. What are the compositions of some candidate materials and what are their lattice parameters? Extending a horizontal line across Fig. 8-16a at $E_g = 1.0\,\text{eV}$ (or $\lambda = 1.24\,\mu\text{m}$) we see that the alloys of the following pairs are crossed: GaAs–InAs; GaAs–GaSb; InP–InAs; AlAs–InAs; AlSb–GaSb; and AlSb–InSb. Although these all appear to be potentially useful, AlSb–GaSb has an indirect bandgap and is disqualified. Even though AlAs and AlSb are also indirect semiconductors, the indicated ternary compositions at 1 eV appear to be direct. Upon alloying these binary compounds we may assume, in the simplest approximation, that the resultant lattice constants and energy gaps of the ternaries are weighted averages of the binary values. As an example, let us determine the composition of the desired InGaAs alloy for which a_0 must equal 5.76 Å, the value at 1.0 eV. For a linear law of lattice-constant mixtures, i.e., Vegard's law, $a_0(In_xGa_{1-x}As) = (x)a_0(InAs) + (1-x)a_0(GaAs)$. Since $5.76 = 6.07x + 5.65(1-x)$, $x = 0.26$, and the predicted alloy has the composition $In_{0.26}Ga_{0.74}As$. Alternatively, linear averaging of energy gaps gives $E_g(In_xGa_{1-x}As) = (x)E_g(InAs) + (1-x)E_g(GaAs)$. Thus, $1.0 = 0.36x + 1.43(1-x)$ or $x = 0.40$. To resolve this discrepancy in x, let us consult Table 8-2 for experimental values of direct energy gaps in selected III–V ternary solid solutions. On this basis x is calculated to be 0.40, so the preferred ternary composition is $In_{0.40}Ga_{0.60}As$.

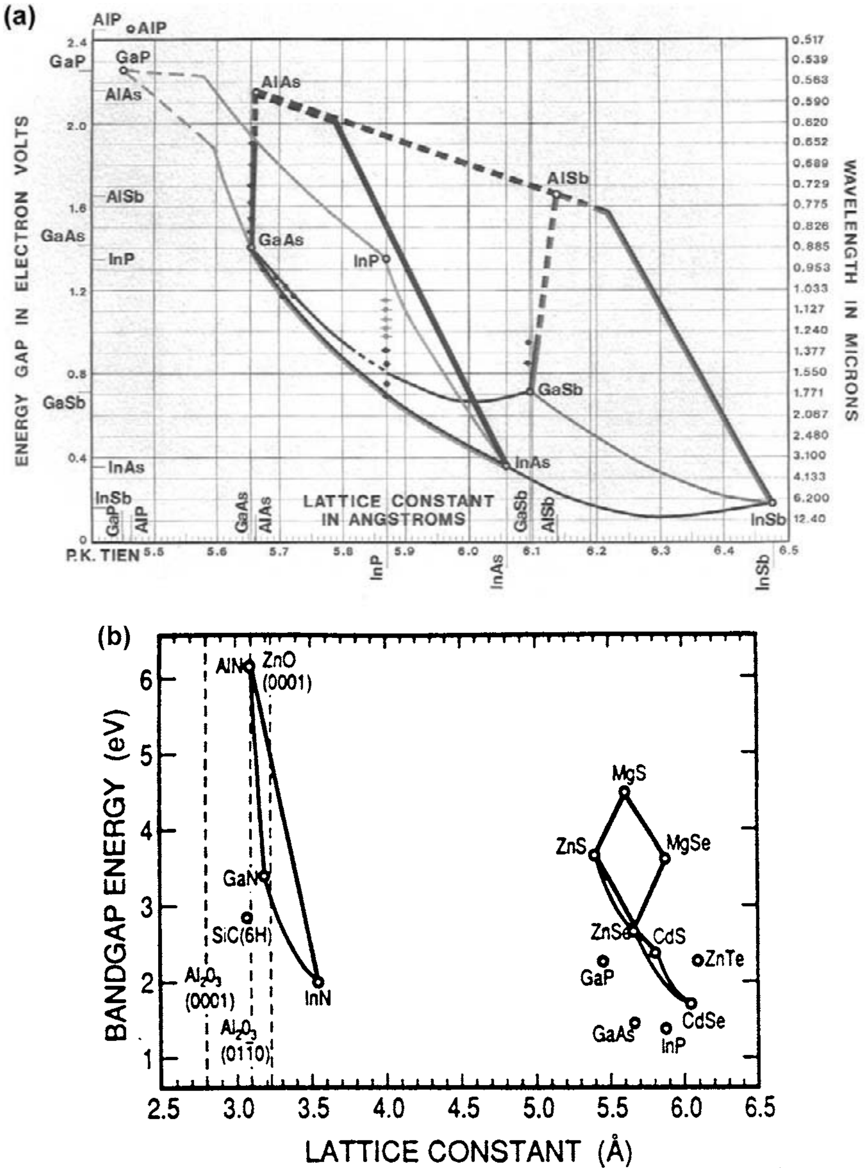


Figure 8-16 Energy gaps and corresponding lattice constants for (a) various III–V compound semiconductors with $E_g \lesssim 2\,\text{eV}$ (courtesy of P. K. Tien, AT&T Bell Laboratories); (b) various II–VI and III–V compound semiconductors with $E_g \gtrsim 2\,\text{eV}$. Only data for the hexagonal α AlN, GaN, and InN phases are plotted. (From T. Matsuoka, A. Ohki, T. Ohno, and Y. Kawaguchi, *J. Cryst. Growth* **138**, 727, 1994. Reprinted with the permission of Dr. T. Matsuoka.)

Table 8-2

Composition Dependence of Direct Energy Gaps in Selected III–V Alloys[a]

Alloy	Direct energy gap (eV)	
$Al_xGa_{1-x}As$	$1.424 + 1.455x$	$(0 < 0.37 < x)$
$Ga_xIn_{1-x}As$	$0.36 + 1.064x$	
$Al_xIn_{1-x}As$	$0.360 + 2.012x + 0.698x^2$	
$GaAs_xSb_{1-x}$	$0.726 - 0.502x + 1.2x^2$	
$InAs_xSb_{1-x}$	$0.18 - 0.41x + 0.58x^2$	
InP_xAs_{1-x}	$0.360 + 0.891x + 0.101x^2$	
$Al_xIn_{1-x}Sb$	$0.172 + 1.621x + 0.43x^2$	

[a] From J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*. Macmillan, New York, 1990.

Next, consider the practical problem of fabricating lasers emitting coherent light at 1.0 eV. In these devices, semiconductor films must be deposited on a readily available substrate, e.g., GaAs or InP, and, importantly, be lattice matched to it. Our ternary alloy is lattice-matched to neither GaAs nor InP, and none of the prospective 1 eV materials is lattice-matched to GaAs. A new semiconductor must therefore be found. Calculation using Vegard's law shows that $Ga_{0.47}In_{0.53}As$, which is commercially produced for optoelectronic devices, has the same lattice constant as InP; unfortunately, E_g for this ternary alloy is not 1 eV but rather 0.70 eV (from Fig. 8-16a). All is not lost, however, because by alloying $Ga_{0.47}In_{0.53}As$ with InP E_g for the new quaternary GaInAsP alloy can be raised to a value of 1.0 eV, all the while simultaneously maintaining $a_0 = 5.87$ Å. To estimate the required stoichiometry we assume $E_g(\text{GaInAsP}) = (1-x)E_g(Ga_{0.47}In_{0.53}As) + xE_g(\text{InP})$, where x is the fraction of P added. Substituting, $1.0 = (1 - x)0.70 + x1.27$ and $x = 0.53$; therefore, the desired semiconductor film composition is $Ga_{0.22}In_{0.78}As_{0.47}P_{0.53}$. Growing such a lattice-matched film may not be easy but is certainly not insurmountable.

Other ternary semiconductor properties are similarly altered upon blending binary compounds. For example, the index of refraction n, required for light-guiding properties in lasers, varies as (Ref. 31)

$$n(x) = 3.590 - 0.710x + 0.091x^2 \quad (8\text{-}9)$$

in the AlAs–GaAs system. Thus it is possible to design ternary alloys with a larger E_g and smaller n relative to GaAs, while maintaining an acceptable

lattice match for high-quality heterojunctions. These unique combinations of properties have led to the development of a family of injection lasers, light-emitting diodes, and photodiodes in the GaAs–AlAs system and other III–V semiconductors as well.

8.4.3 Scope of Devices and Applications

There have been two broad sets of applications based on compound semiconductor epitaxial thin-film structures. The first are optoelectronic devices where generally one, or more commonly, a limited number of heterojunctions is involved. Examples include light-emitting diodes (Fig. 8-17) and lasers (Fig. 8-18). Edge- as well as surface-emitting versions of these light sources exist for assorted purposes, e.g., fiber-optical communications systems, multiaccess optical disks, and optical computing. In solid-state optoelectronic devices, which also include photovoltaics and photodetectors, there are essentially no substitutes for compound semiconductors, i.e., they must stem from III–V, II–VI, or IV–VI materials and their combinations. The operation of the different optoelectronic devices can be simply understood on the basis of Fig. 8-19, which distinguishes among three different phenomena that occur at p–n junctions. In the first (Fig. 8-19a), *absorption* of a photon with energy of at least E_g results in electron promotion from the valence to conduction band. This is essentially what happens in diode detectors and solar cells, because once there are charge carriers in the conduction band they power and activate electronic circuits. *Spontaneous emission* of a photon having energy E_g is depicted in Fig. 8-19b.

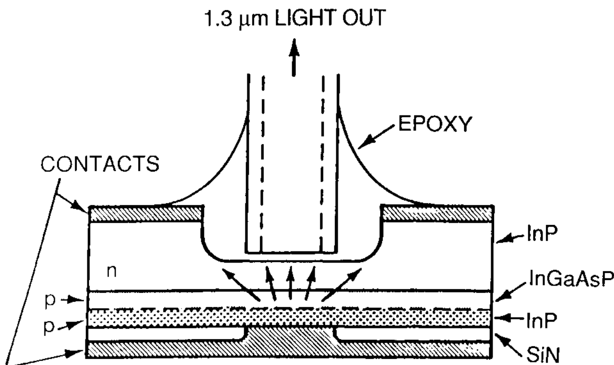


Figure 8-17 Schematic of a surface-emitting InGaAsP ($\lambda = 1.3\,\mu m$) light-emitting diode. (Courtesy N. K. Dutta, AT&T Bell Laboratories.)

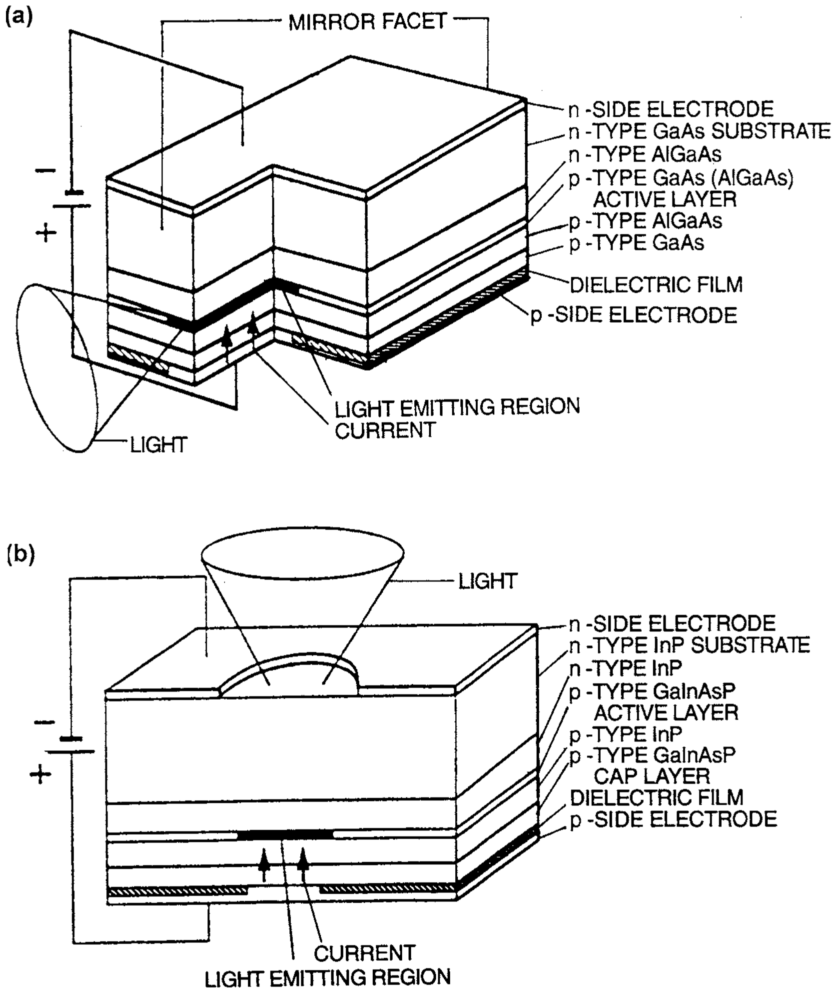


Figure 8-18 Schematic views of basic laser structures. (a) Edge-emitting planar type based on GaAs (or InP). (b) Surface-emitting GaAlAs/GaAs laser prepared by MOCVD. (From M. Fukuda, *Reliability and Degradation of Lasers and LEDs*. Reprinted with the permission of Artech House and the author.)

This process arises when current passes through a forward-biased junction, e.g., in a light-emitting diode. Lastly, *stimulated emission* occurs in lasers whose junctions are also forward-biased (Fig. 8-19c). The existence of an optical cavity and mirrors to confine and enable directed light amplification distinguishes lasers from LEDs. For a more extensive treatment of the

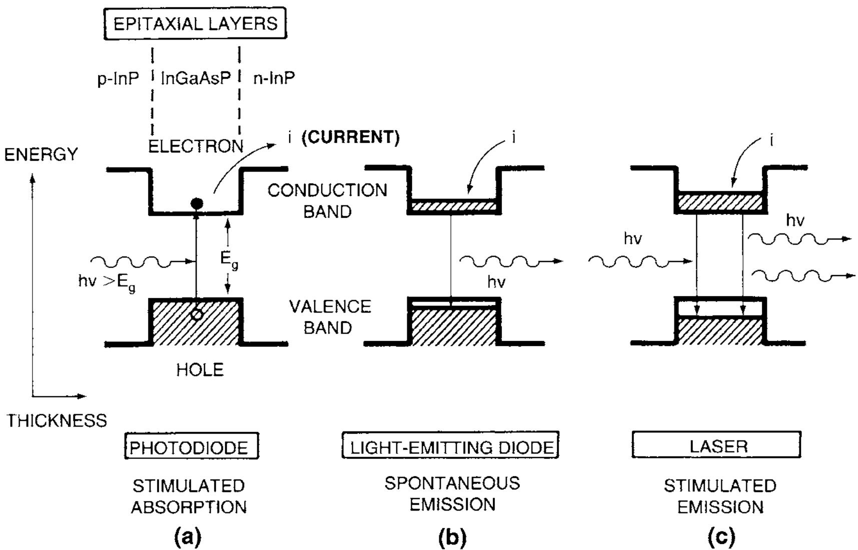


Figure 8-19 Three electron-transition processes between semiconductor valence and conduction bands shown within the context of photonic junction devices. (a) Absorption; (b) spontaneous emission; (c) stimulated emission.

physics, characteristics, and operation of photonic devices the reader is referred to the many excellent textbooks (Refs. 30, 32) on the subject.

A second important class of applications for these materials is in analog and digital electronic devices (Ref. 4). Into this category fall assorted discrete diodes and heterojunction transistors as well as integrated-circuit combinations for use in computer circuits and high-frequency communications systems. Although present computer technology is based almost exclusively on silicon devices and integrated circuits, there are other interesting high-speed electronics applications where compound semiconductor devices are indispensable. As examples we cite wireless communications of all kinds at RF and particularly microwave ($\sim$1 GHz to 1000 GHz) frequencies. They employ analog as well as digital devices and circuits made of these materials for such diverse purposes as mobile communications, satellite-based voice and data systems, and computer networking applications. The high charge-carrier mobilities evident in Table 8-1 are central to these applications. There are also advantages in using heterojunctions. For example, in transistor heterostructures with an emitter having a wider bandgap relative to both base and collector, there is a barrier to reverse injection of minority carriers, and this raises the gain/efficiency ratio of the device.

In addition to these examples of "traditional" heterojunctions, an exciting thrust of compound semiconductor epitaxy deals with superlattices or layered stacks of films. These structures not only have interesting scientific implications, but have been incorporated in assorted electronic and photonic devices where their unique properties have improved or extended traditional performance levels.

For reasons of interest and book balance (there is already much coverage of electronic devices), the subsequent discussion will focus on optical devices.

8.4.4 Lighting the IR to Visible Spectrum with Compound Semiconductors

The development of materials and devices for operation in the spectral domain extending from the visible to the infrared has been one of the great triumphs of solid-state science and engineering in the past three decades (Ref. 33). High-brightness light-emitting diode sources have been a particular quest and the record of achievement based on quantum efficiencies, i.e., the number of photons generated per electron–hole pair, has been impressive. For example, the following quantum efficiencies have been recorded for LEDs in the visible region: 17% for red (0.629 μm); 11.5% for orange (0.604 μm); 6.3% for green; and 9% for blue. Some sample materials issues and applications are provided next to provide a sense of this technological odyssey, which has now made blue lasers and white LEDs a reality.

8.4.4.1 Optical Communications

Optical communication systems transmit information optically through fibers. This is done by converting the initial electronic signals into light pulses employing laser or light-emitting diode light sources. The light launched at one end of an optical fiber is confined to the fiber core and propagates along it over long distances. At the other end of the system the light pulses are detected by photodiodes and converted back into electronic signals, which, in telephone applications, finally generate sound. In such a system it is crucial to transmit the light with minimum attenuation or low optical loss. Great efforts have been made to use the lowest-loss fiber possible and minimize loss at both the source and detector ends. If optical losses are high, it means that the optical signals must be reamplified with additional, costly repeater stations. The magnitude of the problem can be appreciated when transoceanic communications systems are involved. In silica-based fibers it has been found that minimum transmission losses occur at a wavelength in the range 1.3–1.55 μm. The necessity to operate within

this IR-wavelength window bears directly on the choice of suitable semiconductors and epitaxial deposition technology required to fabricate the required sources and detectors.

Reference to Table 8-1 shows that InP is transparent to 1.3–1.55 μm light. This means that InP wafers can serve both as a substrate on which epitaxial films can be grown and as a mechanical support facilitating the coupling of fibers to devices (Fig. 8-17). As we have seen earlier a very close lattice match to InP ($a_0 = 5.869$ Å) can be effected by alloying $Ga_{0.47}In_{0.53}As$ and InP. In this way high-performance lasers based on the lattice-matched GaInAsP/InP system have emerged for optical communications use.

8.4.4.2 Light-Emitting Semiconductor Devices

Bright visible-light-emitting devices with $\lambda > 0.5\ \mu$m have been available for well over 30 years. The ubiquitous red lasers and light-emitting diodes are based on epitaxial AlGaAs on GaAs, while orange, yellow, and green LEDs are fabricated from indirect bandgap materials, e.g., GaAsP doped with nitrogen. However, practical sources for the violet-blue-green end of the spectrum, i.e., 0.4–0.5 μm, have been notably slow in appearing. They require semiconductors with wide energy bandgaps ranging from somewhat below 2.5 to 3 eV and greater. Table 8-1 reveals that all compound semiconductor classes have potential candidate materials in this E_g range, e.g., (II–VI) ZnS, ZnSe, and ZnTe; (III–V or III–N) GaN, AlN, and InN; and (IV–VI) SiC, an indirect semiconductor. Despite much research and development over the years, particularly in II–VI materials (Ref. 34), devices derived from them were not commercialized even though sizable wafers of ZnS, ZnSe, ZnTe, and SiC have been available for some time (Ref. 35). Two major obstacles to the fabrication of II–VI and III–N diode devices have been the difficulty in doping these materials p-type, and the unacceptably large dislocation and stacking fault densities at heteroepitaxial interfaces. The former has limited junction formation while the latter has led to short-lived devices. With advances in p-doping ($\sim 10^{18}/\text{cm}^3$ in ZnSe) coupled with a reduction in defect density (to $\sim 10^5/\text{cm}^2$) due to epitaxial growth first on GaAs (where $f = 0.25\%$) and then on ZnSe wafers, device performance improved (laser lifetimes of ~ 1000 h); however, it still fell short of what was demanded. Blue-light device prospects for the wurtzite or hexagonal III–N materials appeared to be even more dim. For example, in the case of GaN the substrate of choice is sapphire (Al_2O_3) but the lattice mismatch of 14% has led to dislocation densities of $\sim 10^{10}/\text{cm}^2$. In comparison, defect densities of less than $\sim 10^3/\text{cm}^2$ are typical in lattice-matched III–V devices grown on GaAs and InP.

8.4.4.3 GaN Light-Emitting Semiconductor Devices

The situation just described prevailed until 1993 when Nichia Chemical Industries, a small family-owned business in Japan, astonished the world by demonstrating a very bright blue GaN LED (Ref. 36). This revolutionary advance as well as subsequent ones based on it were spearheaded by S. Nakamura and enabled Nichia to leap ahead of Western efforts in this area (Refs. 37–40). Two key discoveries contributed to this progress. First, satisfactory p-type doping was achieved using Mg. Secondly, a compliant, recrystallized amorphous buffer film of GaN on top of the sapphire substrate reduced the harmful effects of dislocations and lattice and thermal expansion coefficient mismatches. This enabled device-quality epitaxial GaN layers to be grown by MOCVD methods, despite initially high dislocation densities (Ref. 2).

Today, highly efficient UV, blue, and green InGaN LEDs have been fabricated with (external) quantum efficiencies of 7.5% at 0.371 μm (UV), 11.2% at 0.468 μm (blue), and 11.6% at 0.520 μm (green). Some of these blue-light sources have brightnesses of more than 1000 mcd and are comparable to superbright red LEDs; they are more than a hundred times brighter than SiC blue LEDs. There are many actual as well as potential applications for this new generation of light sources. The largest LED-display sky screen in the world, located in Tianjin, China, measures 40 meters in length and 14 meters in height and contains some 680,000 blue and 1,700,000 green LEDs. Traffic lights composed of blue LED sources have already appeared.

An important recent advance has been the development of long-lived InGaN blue-light lasers (Ref. 40) such as that shown in Fig. 8-20. Laser

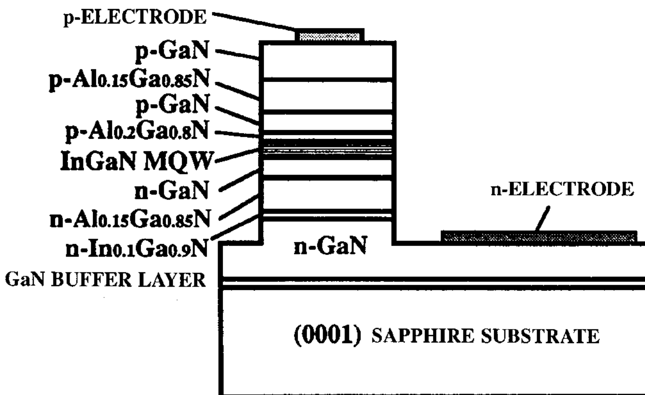


Figure 8-20 Structure of the first InGaN multiquantum-well laser diode. (From Ref. 40. Reprinted with the permission of the author.)

lifetimes critically depend on the dislocation density in the active diode region, and this has been dramatically reduced by fabricating devices on epitaxial layers that are seeded and grown over a buffer layer (see Section 8.5.2.2). Upon commercialization of blue lasers, an important application promises to be optical storage for memories. Since information storage densities vary as the inverse square of the wavelength of light used, increases by a factor of about 4 may be anticipated relative to the present red AlInGaP lasers. A big prize would be an LED or laser-based replacement for the incandescent lamp. For more detailed information the reader should consult the now extensive literature on this subject.

8.5 HIGH-TEMPERATURE METHODS FOR DEPOSITING EPITAXIAL SEMICONDUCTOR FILMS

8.5.1 Scope

In both this and the next section methods for depositing epitaxial semiconductor films will be described. For convenience of organization, Section 8.5 focuses primarily on elevated-temperature processes where chemistry and chemical equilibria play important roles. Thus, liquid-phase epitaxy involving semiconductor melts, and CVD methods are primarily stressed here. On the other hand, generally lower-temperature epitaxial deposition processes, characterized perhaps more in physical than chemical terms, are emphasized in Section 8.6. The prime example of this case is molecular beam epitaxy (MBE). Nevertheless, a stark division into two categories is somewhat arbitrary because deposition processes overlap and exist in hybrid combinations. For example, MOCVD, a relatively low-temperature process, is treated in this section, but MOMBE, which is essentially MBE but employs MOCVD precursors, is deferred to the next. Throughout, distinctions between these processes will be raised, and their applications and relative merits stressed.

8.5.2 Epitaxial Films from Melts

8.5.2.1 Liquid-Phase Epitaxy

In liquid-phase epitaxy (LPE) melts rather than vapors are in contact with the growing films and substrates. Introduced in the early 1960s LPE is still widely used to produce compound-semiconductor devices. For example, of commercially produced devices presently over 40% are light-emitting diodes, and of these almost 50% are LPE-grown GaP devices; another 20%

are LPE-grown AlGaAs/GaAs LEDs (Ref. 41). However, for greater layer uniformity and atomic abruptness, LPE has been supplanted by various vapor-phase epitaxy methods, notably MOCVD. LPE involves the precipitation of a crystalline film from a supersaturated melt onto a substrate that serves as both the template for epitaxy and the physical support for the heterostructure. The process can be understood by referring to the GaAs binary phase diagram (Fig. 1-15). Consider a Ga-rich melt containing 10 at.% As. When heated above $\sim 920^{\circ}$C all of the As dissolves. If the melt is cooled below the liquidus temperature into the two-phase field, it becomes supersaturated with respect to As. Only a melt of lower than the original As content can now be in equilibrium with GaAs. The excess As is, therefore, rejected from solution in the form of GaAs which grows epitaxially on a suitably placed substrate. Readers will appreciate that the crystals they grew as children from supersaturated aqueous solutions essentially formed by this mechanism.

Through control of cooling rates, different kinetics of layer growth apply. For example, the melt temperature can either be lowered continuously together with the substrate (equilibrium cooling) or separately reduced some 5–20°C and then brought into contact with the substrate at the lower temperature (step cooling). Theory backed by experiment has demonstrated that the epitaxial layer thickness increases with time as $t^{3/2}$ for equilibrium cooling and as $t^{1/2}$ for step cooling (Ref. 29). Correspondingly, the growth rates or time derivatives vary as $t^{1/2}$ and $t^{-1/2}$, respectively. These diffusion-controlled kinetics respectively indicate either an increasing or decreasing film growth rate with time depending on mechanism. Typical growth rates range from ~ 0.1 to $1\,\mu$m per minute. A detailed analysis of LPE is extremely complicated in ternary systems because it requires knowledge of thermodynamic equilibria between solids and solutions, nucleation and interface attachment kinetics, solute partitioning, diffusion, and heat transfer. LPE offers several advantages over other epitaxial deposition methods, including low-cost apparatus capable of yielding films of controlled composition and thickness, with lower dislocation densities than the parent substrates.

To grow multiple GaAs/AlGaAs heterostructures, the seed substrate is sequentially translated past a series of crucibles holding melts containing various amounts of Ga and As together with such dopants as Zn, Ge, Sn, and Se as shown in Fig. 8-21. Each film grown requires a separate melt. Growth is typically carried out at temperatures of $\sim 800^{\circ}$C with maximum cooling rates of a few degrees Celsius per minute. Limitations of LPE growth include poor thickness uniformity and rough surface morphology, particularly in thin layers. The CVD and MBE techniques are distinctly superior to LPE in these regards.

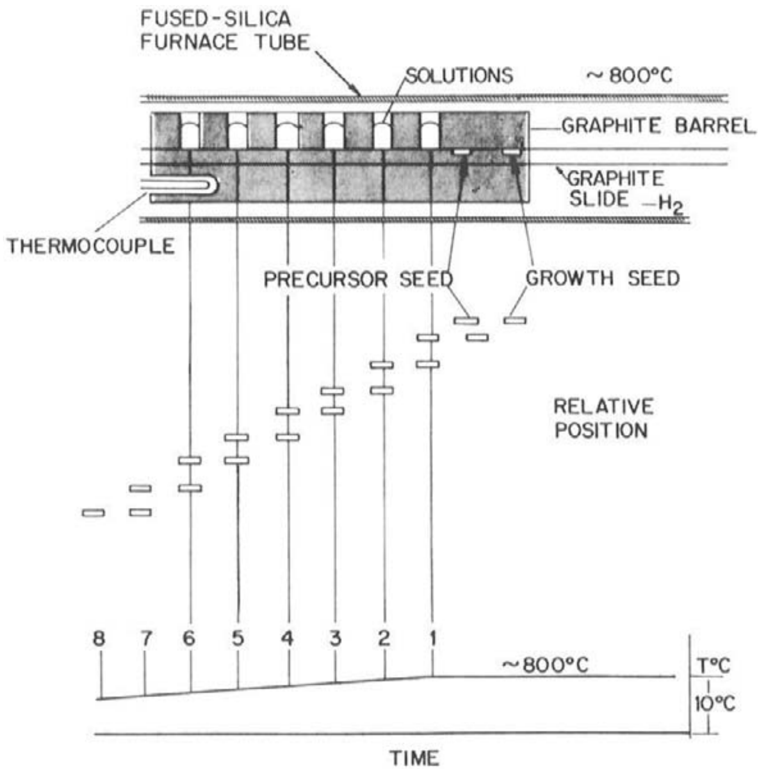


Figure 8-21 Schematic of LPE reactor. (Courtesy of M. B. Panish, AT&T Bell Laboratories.)

Although semiconductor LPE applications have been stressed here, the technique is also widely practiced to prepare epitaxial films of niobates and garnets.

8.5.2.2 Seeded Lateral Epitaxial Film Growth over Insulators

The methods we describe here briefly have been successfully implemented in Si but not in GaAs or other compound semiconductors. Inclusion of this subject here is suggested by the use of melts. Technological needs for three-dimensional integrated circuits and isolation of high-voltage devices have spurred the development of techniques to grow epitaxial Si layers over such insulators as SiO_2 or sapphire. The LEGO (Lateral Epitaxial Growth over Oxide) process (Ref. 42) is one such technique whose intent is to form isolated surface tubs of high-quality Si surrounded on all sides by a moat of

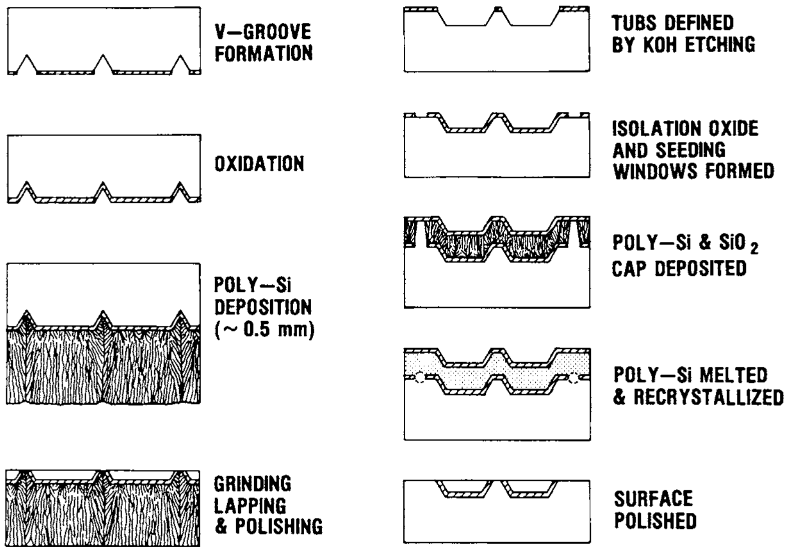


Figure 8-22 Process sequence employed to isolate single-crystal Si tubs. (Left) Conventional dielectric isolation process. (Right) LEGO process. (Courtesy of G. K. Celler, AT&T Bell Laboratories.)

SiO_2. Because of this excellent insulation, devices fabricated within the tubs are also radiation hardened or immune from radiation-induced charge effects originating in the underlying bulk substrate. The process, shown schematically in Fig. 8-22, starts with patterning and masking a Si wafer to define the tub regions followed by etching of deep slanted wall troughs. A thick SiO_2 film is grown, and seed windows are opened down to the substrate by etching away the SiO_2. Then a thick polycrystalline Si layer (100 μm thick) is deposited by CVD methods. This surface layer is then melted by the unidirectional radiant heat flux from incoherent light emitted by tungsten halogen-arc lamps (lamp furnace). The underlying wafer protected by the thermally insulating SiO_2 film does not melt except in the seed windows. Crystalline Si nucleates at each seed and grows vertically and then laterally across the SiO_2, leaving a single-crystal layer in its wake upon solidification. Lastly, mechanical grinding and lapping of the solidified layer prepares the structure for further microdevice processing. Conventional dielectric isolation processing also employs a thick CVD Si layer. But the latter merely serves as the mechanical handle that enables the bulk of the Si wafer to be ground away.

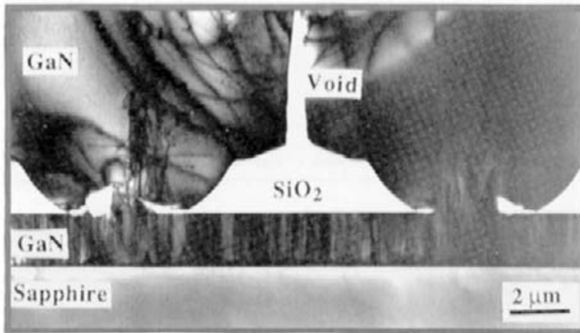


Figure 8-23 Cross-sectional TEM micrograph of laterally overgrown GaN film on a SiO_2 mask and window areas. (From Ref. 2. Reprinted with the permission of the author.)

An alternate process for broad-area lateral epitaxial growth over SiO_2 (i.e., Si on SiO_2 or SOI) employs a strip heat source in the form of a hot graphite or tungsten wire, scanned laser, or electron beam. After patterning the polycrystalline or amorphous Si on the oxide, the strip sweeps laterally across the wafer surface. Local zones of the surface then successively melt and recrystallize yielding, under ideal conditions, one large epitaxial Si film layer.

Lateral-overgrowth epitaxy is not only limited to silicon technology. An interesting example occurs in the processing of blue GaN lasers. As shown in Fig. 8-23, the buffer layer of GaN deposited on (0001) Al_2O_3 (sapphire), although smooth, contains too high a dislocation density for the fabrication of reliable devices. However, by utilizing a SiO_2 mask containing an array of windows, subsequently deposited GaN films seed or nucleate within them as single crystals. These grow into relatively thick crystalline overgrowths that laterally coalesce to yield a continuous layer. Importantly, a relatively defect-free epitaxial layer is created since the propagation of threading dislocations is effectively curtailed at the constricted seed windows. Diode lasers (Fig. 8-20) fabricated on this top layer have displayed 10,000 h lifetimes during continuous operation at room temperature (Ref. 2).

8.5.3 CVD-Based Epitaxy

In the CVD-based epitaxial film deposition processes considered now, two gaseous precursors, one for each atomic group, are generally employed, and compound-semiconductor films are produced at pressures ranging anywhere from $\sim$0.5 to 760 torr. Deposition is carried out at temperatures

many hundreds of degrees lower than the melting point of the semiconductor. This fact underscores an important distinction that should be noted between prior processes to produce epitaxial films and the ones discussed in subsequent sections. For example, the epitaxy of Si (Section 6.3.2), LPE growth from the melt (Section 8.5.2.1), and even the reversible transport vapor-phase epitaxy (VPE) of GaAs by chloride or hydride CVD processes (Section 6.2.6), are all carried out at elevated temperatures, and thus close to thermodynamic equilibrium. These deposition processes therefore can usually be reversed from film growth to film etching or dissolution through change in temperature. Since the net film-deposition rate is the difference between the rates of adatom–substrate bonding and atomic desorption or reevaporation, near thermodynamic equilibrium the film growth rate is a small difference between large numbers.

In contrast, the processes now considered, particularly MOCVD and later MBE (Section 8.6.2), are conducted at relatively low temperatures and further from thermodynamic equilibrium. Atomic desorption and reevaporation rates are reduced while vapor impingement and adatom–substrate bonding rates remain unchanged, leading to high film-deposition rates. The difference between reversible and a more nearly irreversible film deposition has important ramifications with respect to film-thickness uniformity. In the former case, difficult-to-control high rates of film growth and etching tend to make for rougher films. On the other hand, smoother film surfaces and sharper interfaces between successive deposited layers are a consequence of low-temperature deposition. This is one of the advantages of MOCVD relative to VPE and LPE in the growth of high-performance epitaxial films.

8.5.3.1 Metalorganic CVD Processes for Semiconductor Epitaxy

We have already discussed various MOCVD processes in connection with the deposition of metal and oxide films (Section 6.6.4); whereas the former invariably deposit in polycrystalline form, oxide films assume structures ranging from amorphous to polycrystalline. However, MOCVD was invented to deposit III–V and II–VI compound semiconductor films where it is always the intent to grow high-quality single-crystal or epitaxial films (Refs. 43–45). For this reason the acronyms MOVPE (metalorganic) or OMVPE (organometal) vapor-phase epitaxy are equivalently used when speaking of processes to deposit epitaxial films. In overcoming the shortcomings of the earlier chloride and hydride CVD methods, MOCVD has been particularly effective in preparing films for a variety of visible and long-wavelength lasers as well as quantum-well devices.

8.5.3.1.1 Precursors

Availability of suitable precursors is critical to the success of MOCVD processes (Ref. 46). Over the years a workable number have been synthesized and some of these are entered in Table 8-3. Among the attributes precursors must possess are adequate volatility, a sufficiently large temperature window between evaporation and decomposition, clean decomposition at relatively low temperatures, good shelf life, manufacturability (yield, cost), and, importantly, low toxicity. Chemically, precursors largely consist of metal alkyl compounds with methyl (M) and ethyl (E) groups present in twofold (di, D) or threefold (tri, T) coordination. Vapor pressures for some precursors are also entered together with typical semiconductor deposition-temperatures.

Table 8-3

Metalorganic Precursors and Semiconductors Grown by MOCVD[a]

Compound	Reactants	Vapor pressure[b] of OM precursor		Growth temperature (°C)
		a	b	
AlAs	TMAl + AsH_3	8.224	2135	700
AlN	TMAl + NH_3			1250
GaAs	TMGa + AsH_3	8.50	1824	650–750
GaN	TMGa + NH_3			800
GaP	TMGa + PH_3			750
GaSb	TEGa + TMSb	9.17 7.73	2532 1709	500–550
InAs	TEIn + AsH_3			650–700
InP	TEIn + PH_3			725
ZnS	DEZn + H_2S	8.28	2190	
ZnSe	DEZn + H_2Se			
CdS	DMCd + H_2S	7.76	1850	
HgCdTe	Hg + DMCd + DMTe	7.97	1865	
CdTe	DMCd + DMTe			

[a] Adapted from Ref. 46.

[b] $\log P\,(\text{torr}) = a - \dfrac{b}{T\,(\text{K})}$

8.5.3.1.2 Gas Reactions

These metalorganic precursors undergo two main types of CVD reactions. In the first they are reacted with very pure and inexpensive Group V gaseous trihydrides (AsH_3, PH_3, SbH_3), e.g., TMGa and AsH_3 to deposit GaAs (Eq. 6-13). However, as hydrides are extremely toxic, safer but more costly Group V alkyl substitutes, e.g., TMAs, TEAs (triethylarsenic), TMP, and TESb, have been developed. This makes possible a second type of reaction, an all-alkyl gas pyrolysis such as

$$\mathrm{TMGa}_{(g)} + \mathrm{TMSb}_{(g)} \rightarrow \mathrm{GaSb}_{(s)} + 3\mathrm{C_2H}_{6(g)} \text{ (ethane).}$$

In addition, metalorganic compounds such as DEZn and bicyclodipentadienylmagnesium (Cp_2-Mg) are available for doping purposes. A constant concern with organic gases, however, is incorporation of carbon dopant. Variations in film growth rates are relatively small for alternative MOCVD precursors delivering the same element.

8.5.3.1.3 Reactors

The design of MOCVD reactors means confronting issues frequently absent in other CVD systems. These include the efficient delivery of expensive precursors, the generally hazardous nature of the gases, the often severe temperature gradients in the system, and precise pressure control and valving necessary to deposit multilayer films. The latter capabilities are critical in order to maintain atomically abrupt interfaces in high-performance optoelectronic devices, e.g., quantum-well lasers. To ensure film growth uniformity, substrates usually face upward and lie horizontally on a rotating platform where they are exposed to showerheads or gas nozzles that supply the Group III and V precursors from above as shown in Fig. 8-24. Under such conditions (Refs. 47, 48) the boundary-layer thickness (δ) varies as $\delta(\mathrm{cm}) \sim [P\omega]^{-1/2}$, where P is the reactor pressure and ω is the substrate angular velocity. Certain critical gas-flow parameters are also dependent on ω. For example, the Reynolds number, $\mathrm{Re} \simeq r_d \rho\omega/\eta$, where r_d is the substrate disk radius, ρ is the density, and η is the viscosity. In addition, a mixed convection parameter (MCP) has been defined (Ref. 48) as $\mathrm{MCP} = g\{(T_d/T_\infty - 1)/(\omega^{3/2}\eta^{1/2})\}$, where g is the gravitational constant, and T_d and T_∞ are the temperatures of disks with radii r_d and infinity, respectively. Increasing ω causes Re to rise and MCP to decline. Avoidance of unstable flows, therefore, requires a trade-off between these two dimensionless numbers, e.g., $\mathrm{Re} < 2000$, $\mathrm{MCP} < 2$. It is found that rotational speeds of several hundred to a thousand rpm are optimum in order to reduce δ in a controlled way while avoiding turbulent gas flow. Utilizing

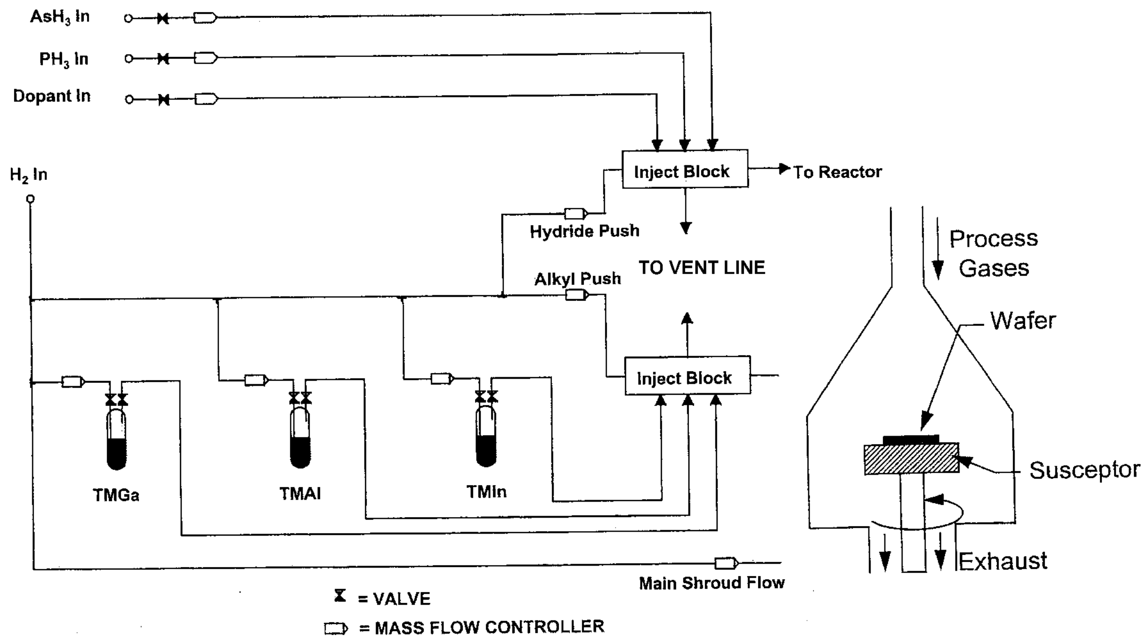


Figure 8-24 Simplified schematic of a III–V MOCVD system showing the gas delivery system and vertical rotating disk reactor. (From A. G. Thompson, *Materials Letters* **30**, 255 (1997). Reprinted with the permission of the author.)

computer-controlled gas exchange and supply systems, high-quality, epitaxial multilayer semiconductor structures have been grown in reactors containing a 42-cm diameter rotating disk that accommodates thirty-eight 2.5 cm or nine 10 cm diameter wafers. In addition to GaAs, other III–V as well as II–VI and IV–VI semiconductor compound films have been synthesized this way.

The deposition of GaN-based films for blue-green-violet photonic devices (Section 8.4.4.3) represents an important triumph for MOCVD methods. Because of the very high equilibrium nitrogen vapor pressures for GaN and InN, growth of InGaN films poses a great challenge. Unusually high nitrogen/Group III (e.g., TMGa) partial pressure ratios of 1000 or more are required, and the nitrogen source gases (N_2, NH_3) must be efficiently delivered to substrates typically maintained at $\sim 550°C$ for InN and $\sim 900°C$ for GaN (Ref. 49). Special reactors with high gas flow capabilities have been built for this purpose.

8.5.3.1.4 Film Growth Rate

Film growth rates ($\dot{G}$) and compositions for CVD processes are directly dependent on gas partial pressures (P) and volume flow rates ($\dot{V}$). Let us consider the case of III–V compound semiconductor MOCVD in the mass-transport limited regime. It is assumed there is an abundance of the group V element and that pyrolysis of group III alkyl molecules is complete, yielding monomers. The perfect gas law suggests that the growth rate for a binary compound film is given by (Ref. 50)

$$\dot{G}_{\mathrm{B}} = \frac{MP\dot{V}}{k_{\mathrm{B}}T\rho A} \qquad (\mathrm{cm/s}) \tag{8-10}$$

where M and ρ are its molecular weight and density, respectively, A is the area of deposition, and $k_{\mathrm{B}}T$ has the usual meaning. In the case of ternary column III sublattice alloys, the growth rate ($\dot{G}_{\mathrm{T}}$) is given by the sum of two such terms, one for each element, i.e.,

$$\dot{G}_{\mathrm{T}} = \frac{M(1)P(1)\dot{V}(1)}{k_{\mathrm{B}}T\rho(1)A} + \frac{M(2)P(2)\dot{V}(2)}{k_{\mathrm{B}}T\rho(2)A}. \tag{8-11}$$

If, for example, we were considering $Al_xGa_{1-x}As$ films, 1 would refer to Al, 2 to Ga, $M(1)$ to AlAs, and $M(2)$ to GaAs. The stoichiometric constant x is simply given by

$$x = \frac{M(1)P(1)\dot{V}(1)/\rho(1)}{M(1)P(1)\dot{V}(1)/\rho(1) + M(2)P(2)\dot{V}(2)/\rho(2)}. \tag{8-12}$$

There may be other integer prefactors for the *PV* terms if any of the alkyl sources is other than a monomer; thus for TMAl, a dimer, all terms involving Al would be multiplied by 2. Also, in the reaction-controlled regime for film growth, equations 8-10 and 8-11 would be multiplied by a temperature-dependent rate constant.

8.5.3.2 Atomic Layer Epitaxy (ALE)

This CVD-based technique enables the growth of epitaxial compound-semiconductor films an atom layer at a time (Refs. 51, 52). For this reason ALE has also been called digital epitaxy. To see how it works let us consider Fig. 8-25, where the binary compound AB is being deposited on an AB substrate surface that is terminated by a layer of A– and B–atoms (– denotes a bond). We assume that the precursors for atoms A and B, molecular AC and BD, respectively, have high vapor pressures. Furthermore, the chemical bonding between A and B is assumed to be stronger than that between A–A and B–B pairs. Upon introducing the AC precursor (1) the reaction with B atoms on the substrate effectively create a surface layer of B–A– together with weakly held C. If the growth temperature is sufficiently high, excess AC merely reevaporates and film growth is self-limiting (2). To add another monolayer the reactor is purged of AC, and BD precursor is introduced (3).

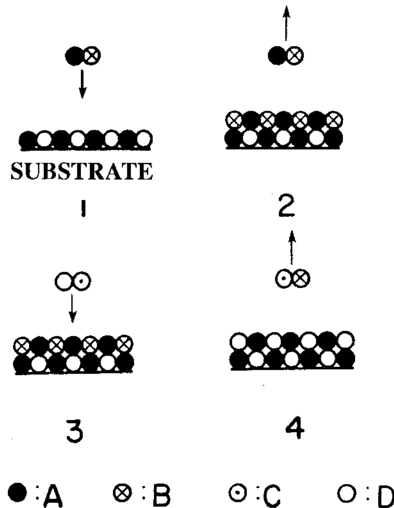


Figure 8-25 Schematic of the atomic layer epitaxy process. The process steps are described in the text. (From T. F. Kuech, *Proc. IEEE* **80**, 1609, 1992. Reprinted with the permission of the author.)

The B atoms will now combine with the A– atoms to form the B–A–B– structure, and as before, excess BD evaporates. But in addition, C and D react and the CD molecular by-product desorbs leaving an AB film layer behind (4). Through successive cycles of stepwise deposition, films of the desired thickness can be deposited without precise control of process variables.

In the II–VI compounds ZnS, ZnTe, and ZnSe epitaxial film growth by ALE is even simpler than described above; the precursors can simply be the individual atomic species, e.g., Zn and S_2. But for the III–V semiconductors, ALE growth is more difficult because the group III elements, e.g., Ga and In, have much lower vapor pressures than corresponding elements from group V. Raising the temperature to promote group III atom volatility results in group V atom loss and film degradation. Therefore, instead of elemental sources, atomic layer epitaxy in III–V semiconductors is accomplished using molecular precursors as in Fig. 8-25, particularly metalorganics for the group III metal (e.g., $Ga(CH_3)_3$) and a hydride, e.g., PH_3, AsH_3, for the group V element. In the case of GaAs growth the first step involves thermal cracking of hydrides which generates AsH_x (with $x = 1, 2$) and As_4 products; the As (i.e., A) then directly reacts with the substrate-surface terminated Ga- (i.e., B). Reaction of remaining hydrogen with the metalorganic precursor gas liberates Ga, which bonds to the monolayer of As-, as well as the volatile by-product CH_4 (i.e., DC). It is clear that narrow processing temperature windows for condensation, reaction, precursor decomposition, and reevaporation must be optimized. The adsorption–desorption kinetics presented earlier (Section 7.2.5), extended to include surface chemical reactions, can help to model such ALE processes.

In addition to compound semiconductors, other materials have been grown by ALE methods, including elements (e.g., Si, Ge); oxides (e.g., Al_2O_3, SnO_2, In_2O_3); nitrides (e.g., TiN, TaN); and superconductors (e.g., δ-NbN, $YBa_2Cu_3O_7$). Employing ALE-grown materials, devices such as monolayer (delta) doped field-effect and heterojunction transistors, as well as quantum-well lasers and superlattice structures, have been demonstrated. Selective epitaxy, side-wall coverage, and monolayer etching are some of the unique features of ALE. On the other hand, low overall film-deposition rates and the complexity of purging and sequencing precursors are decided disadvantages.

8.5.3.3 Additional CVD Epitaxy Processes

In closing, a couple of novel concepts to produce epitaxial films will be briefly addressed. The first, known as vapor-levitation epitaxy, or VLE, is schematically shown in Fig. 8-26. In this process a heated substrate is

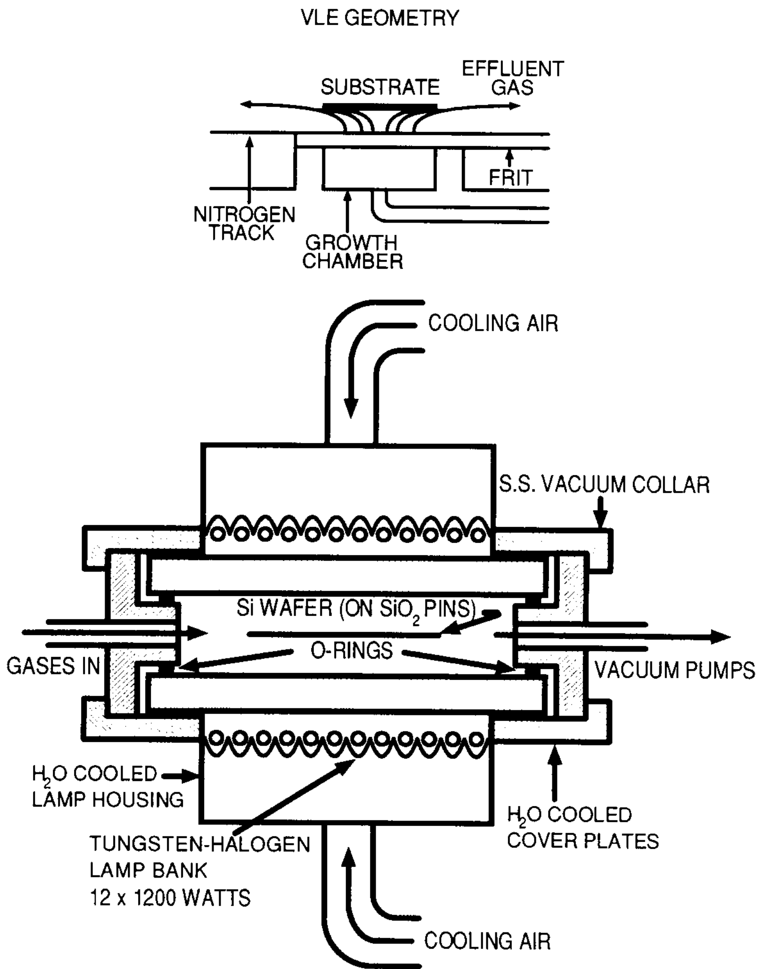


Figure 8-26 (Top) Illustration of VLE process. (Bottom) Schematic of RTCVD process. (Courtesy of M. L. Green, AT&T Bell Laboratories.)

levitated above a nitrogen track close to a porous frit through which the hot gaseous reactants pass. Upon impingement on the substrate, chemical reactions and film deposition occur while product gases escape into the effluent stream. The gas velocity increases while the gas-concentration profile exhibits depletion as a function of radial distance from the center of the circular substrate. These effects cancel one another and uniform films are deposited. The VLE process was designed for the growth of epitaxial III–V

semiconductor films and has certain advantages worth noting:

1. There is no physical contact between substrate and reactor.
2. Thin layer growth is possible.
3. Sharp transitions can be produced between films comprising multilayer stacks.
4. Commercial scale-up appears to be feasible.

The second method, known as rapid-thermal CVD processing (RTCVD), enables epitaxial deposition through rapid, controlled variations of substrate temperature. Source gases (e.g., halides, hydrides, metalorganics) react on low thermal-mass substrates heated by radiation from external high-intensity lamps (Fig. 8-26). The latter undergo rapid temperature excursions, and heating rates of hundreds of degrees Celsius per second are possible. In the case of III–V semiconductors, high-quality epitaxial films have been deposited by first desorbing substrate impurities at elevated temperatures followed by immediate lower temperature growth (Ref. 53).

Lattice-matched, atomically sharp heterojunction interfaces can be grown by RTCVD methods, particularly with metalorganic precursors. Only molecular beam epitaxy, considered next, can match or exceed these capabilities.

8.6 LOW-TEMPERATURE METHODS FOR DEPOSITING EPITAXIAL SEMICONDUCTOR FILMS

8.6.1 SCOPE

Descriptors such as low temperature, physical, and high vacuum serve to distinguish some of the epitaxial-film deposition processes treated here from those discussed in the previous section. In addition to the well-known and widely used ultrahigh-vacuum MBE methods, a number of hybrid MBE–CVD processes will be discussed and compared. The intent of these is to usually grow lattice-matched III–V compound-semiconductor films. In contrast, a novel ultrahigh-vacuum, low-temperature CVD process devoted to the epitaxial deposition of Ge_xSi_{1-x} films will also be described. These strained epitaxial films are now the basis of high-speed electronic devices and have assumed considerable importance. This section ends with a brief discussion of some lesser-known processes for depositing epitaxial films.

8.6.2 Molecular Beam Epitaxy

This conceptually simple single-crystal film-growth technique represents the state of the art attainable in deposition processes from the vapor phase (Ref. 54). Molecular beam epitaxy essentially involves highly controlled evaporation in an ultrahigh vacuum ($\sim 10^{-10}$ torr) system. The reaction of one or more evaporated beams of atoms or molecules with the single-crystal substrate yields the desired epitaxial film. High-quality film-growth only ensues if the surface-diffusion-incorporation time (τ_{di}) is less than for the deposition of a monolayer. If these two times are reversed, unincorporated atoms will be physically buried by the incoming monolayer and give rise to defective layers. Since τ_{di} is thermally activated, a low growth temperature limit is implied for good epitaxy. For example, at typical MBE growth rates in the range of 1 μm/h, excellent quality AlGaAs can be grown at temperatures as low as $\sim 680°$C, GaAs as low as $\sim 475°$C, and InAs as low as $\sim 350°$C. In these results it appears that the Group III species governs the growth temperature. Thus, higher temperatures are required for $Al_xGa_{1-x}As$ because AlAs is thermally more stable than GaAs. For InP growth from In and P_2 beams on (100) InP, substrate temperatures of 350–380°C have been used. Similarly, $In_xGa_{1-x}As$ films, lattice matched to InP, have been grown between 400 and 430°C. Recommended substrate temperatures for MBE growth of GaAs range from 500 to 630°C and are higher than the minimum ones quoted. The MBE deposition window (Section 3.2.3.2), film structure, and morphology have been comprehensively and authoritatively treated by Tsao (Ref. 55) in terms of fundamental thermodynamic, kinetic, and defect issues.

A modern MBE system is displayed in the photograph of Fig. 8-27. It features independent control of beam sources and film deposition, cleanliness, and real-time structural and chemical characterization capability. In such systems precise fabrication of semiconductor heterostructures from a fraction of a micron thick down to a single monolayer is possible. The heart of an MBE deposition system is shown schematically in Fig. 8-28a. Arrayed around the substrate are semiconductor and dopant atom heating sources consisting of either effusion cells or electron-beam guns. The latter are employed for the high-melting Si and Ge materials. On the other hand, effusion cells are used to evaporate compound-semiconductor elements and their dopants. An effusion cell is essentially an isothermal cavity containing a hole through which the evaporant exits; it behaves like a small-area source and exhibits cos ϕ vapor emission. Vapor pressures of important compound semiconductor species are displayed in Fig. 3-2.

Consider now a substrate positioned a distance L from a source aperture of area A, with $\phi = 0$. An expression for the number of evaporant species

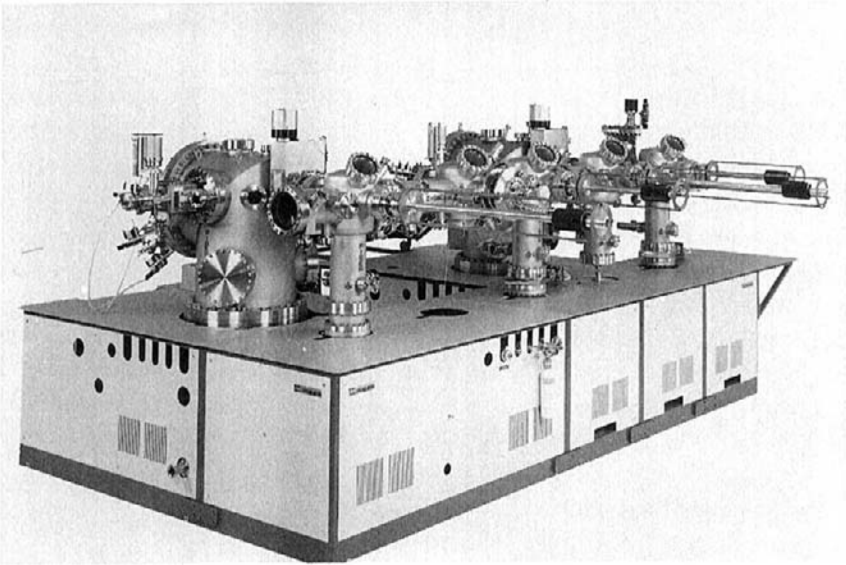


Figure 8-27 Photograph of multichamber MBE system. (Courtesy of Riber Division, Inc. Instruments SA.)

striking the substrate is

$$\dot{R} = \frac{3.51 \times 10^{22} PA}{\pi L^2 (MT)^{1/2}} \frac{\text{molecules}}{\text{cm}^2\text{-s}}. \tag{8-13}$$

As an example consider a Ga source in a system where $A = 5\,\text{cm}^2$ and $L = 12\,\text{cm}$. At $T = 900°\text{C}$ the vapor pressure $P_{\text{Ga}} \simeq 4.2 \times 10^{-4}$ torr and substituting $M_{\text{Ga}} = 70$, the arrival rate of Ga at the substrate is calculated to be 5.69×10^{14} atoms/cm^2-s. The As arrival rate is usually much higher and, therefore, film deposition is controlled by the Ga flux. An average monolayer of GaAs is 2.83 Å thick and contains 6.3×10^{14} Ga atoms/cm^2. Hence, the growth rate is calculated to be $(5.69 \times 10^{14} \times 2.83 \times 60)/6.3 \times 10^{14} =$ 153 Å/min, a value typically an order of magnitude lower than for CVD growth rates.

In many applications GaAs/$Al_xGa_{1-x}As$ multilayers are required. For this purpose the Ga and As beams are on continuously while the Al source is operated intermittently. Actual growth rates are determined by the measured layer thickness divided by the deposition time. The fraction x can

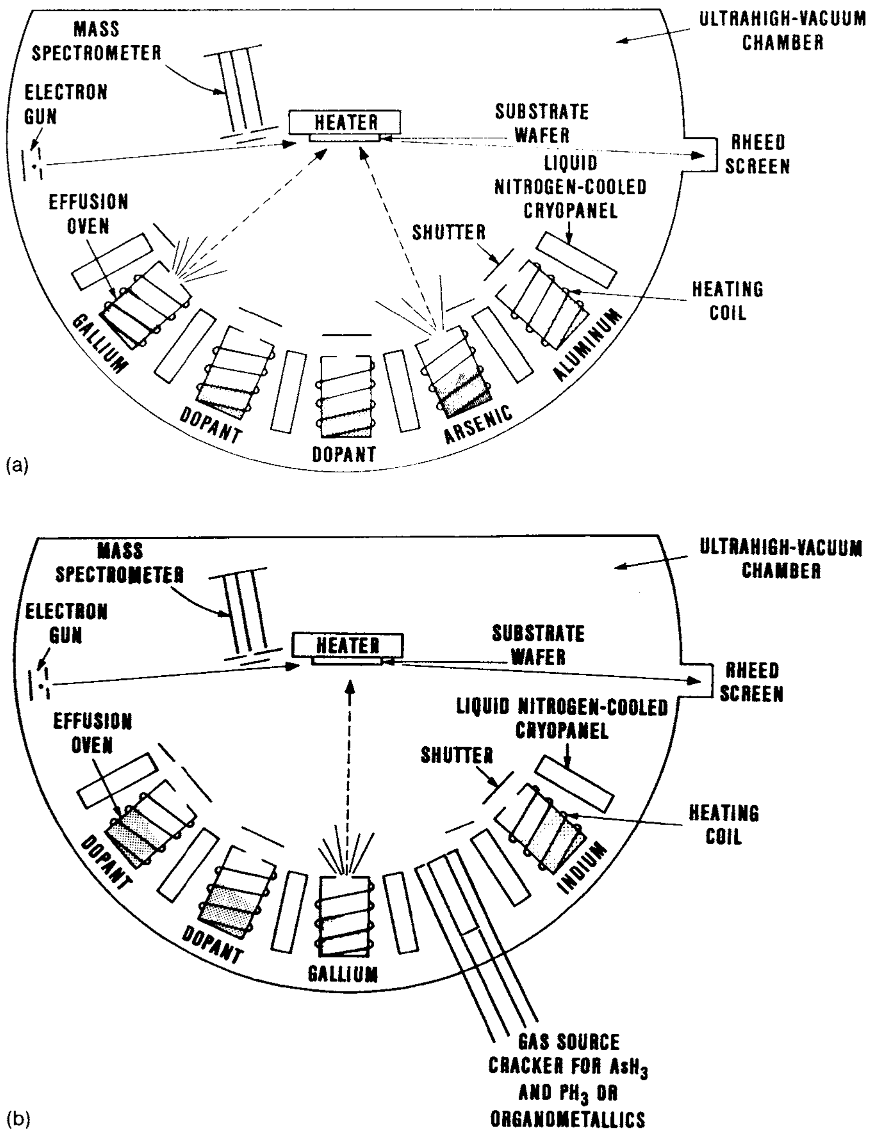


Figure 8-28 Arrangement of sources and substrate in (a) conventional MBE system; (b) MOMBE system. (Courtesy of M. B. Panish, AT&T Bell Laboratories.)

be determined from the relation

$$x = \frac{\dot{G}(\text{Al}_x\text{Ga}_{1-x}\text{As}) - \dot{G}(\text{GaAs})}{\dot{G}(\text{Al}_x\text{Ga}_{1-x}\text{As})}, \tag{8-14}$$

where the respective growth rates ($\dot{G}$) must be known.

8.6.2.1 Gas Source MBE

Although MBE, which uses solid sources for both group III and V elements, has been extremely successful for the growth of arsenides and antimonides, it has been less useful for phosphides (Ref. 54). The very high vapor pressure of phosphorus has meant excessive desorption relative to incorporation in growing films. To offset this desorption, the V/III flux ratio is an order of magnitude higher than for arsenides. The large quantities of P employed means physically big sources which stretch the capacity of pumping systems. It was primarily as a response to difficulties with phosphorus that gas-source MBE was introduced into compound semiconductor technology (Ref. 56). Because the process shown in Fig. 8-28b incorporates a gas source to supply group V elements, terms such as gas-source molecular beam epitaxy (GSMBE) or, less commonly, chemical beam epitaxy (CBE) have been coined. Organometallics, often used for this purpose, are thermally cracked, releasing the group V element as a molecular beam into the system. Hydride gas sources, e.g., AsH_3, PH_3, have also been similarly employed for this purpose. Excellent epitaxial-film quality has been obtained in these hybrid MBE–MOVPE processes. A final process known by the acronym MOMBE employs metalorganics for the group V elements and solid group III elements.

MBE systems have been employed in manufacturing environments since 1984 to fabricate GaAs laser diodes (Ref. 57) and since 1986 to make monolithic microwave integrated circuits and high-power heterojunction bipolar transistors (Ref. 58). Surface defect densities less than $10\,\text{cm}^{-2}$ have been attained by optimizing the function of the group III Knudsen cell. More recently, GSMBE systems using an AsH_3 cracker source have been built to continuously deposit 1.7 μm thick metal–semiconductor field-effect transistor (MESFET) structures with growth occurring at 540°C and 10^{-5} torr. Multiwafer throughput of four 10 cm diameter or seven 3 cm diameter wafers per 2.5 h, using a vacuum load-locked capability of recharging 500 cm^3 Knudsen cells of Ga, Al, and In, gives an idea of the operation scale. Electrical (sheet) resistance variations of less than 2% (over a 27 cm

diameter area), GaAs electron mobilities as high as 124,000 cm^2/V-s (77 K), and an oval defect density of $\sim 40\,cm^{-2}$ are indicative of the excellent quality of the epitaxy.

8.6.2.2 MBE vs OMVPE: Advantages and Disadvantages

These two epitaxy techniques have emerged as the ones of choice in high-performance applications (Ref. 54). Relative to LPE and VPE they both readily enable near-atomic film/substrate interface abruptness, high levels of doping uniformity, and growth of multilayered structures having different compositions and doping levels. When compared to OMVPE, MBE's former advantage in interface abruptness no longer exists. However, MBE growth is further removed from thermodynamic equilibrium so that films are produced at lower substrate temperatures. Perhaps the greatest advantage of MBE is the vacuum environment that makes possible the *in situ* diagnostic capabilities of reflection high-energy electron diffraction (RHEED) and mass spectrometry. Discussed later in Section 8.7.4.2, the former technique yields precise monolayer growth rates and provides a measure of film smoothness; the latter is used to monitor desorption of species from chamber wall or depositing film surfaces. Relative safety is another advantage of MBE as the metalorganic sources are often toxic and or pyrophoric. Advantages of OMVPE include two to three times greater film growth rates, less maintenance downtime, and little difficulty in depositing phosphorus. Although the situation has occasionally flip-flopped, transistor and laser diode device characteristics from MBE-grown III–V materials have generally exceeded those of devices prepared by OMVPE.

8.6.3 Silicon Heteroepitaxy

Since the early 1960s Si has been the semiconductor of choice. Its dominance cannot, however, be attributed solely to its electronic properties, for it has mediocre carrier mobilities and only average breakdown voltage and carrier saturation velocities. The absence of a direct bandgap rules out light emission and severely limits its efficiency as a photodetector. Silicon does, however, possess excellent mechanical and chemical properties. A high modulus of elasticity and high hardness enable Si wafers to withstand the rigors of handling. Its great natural abundance, ability to be readily purified, possession of a highly inert and passivating oxide, and ease of device processing have all helped to secure a dominant role for Si in solid-state technology. Nevertheless, Si is being increasingly supplanted in high-speed applications by compound semiconductors.

The idea of combining semiconductors which can be epitaxially grown on low-cost Si wafers is very attractive. Monolithic integration of III–V devices with Si integrated circuits offers the advantages of combined photonic–electronic functionality and higher speed signal processing distributed over larger substrate areas. Furthermore, Si wafers are more robust and dissipate heat more readily than GaAs wafers. Unfortunately, there are severe crystallographic as well as chemical compatibility problems that limit Si-based heteroepitaxy. Reference to Table 8-1 reveals that Si is only closely lattice matched to GaP and ZnS. Furthermore, its small lattice constant precludes conventional epitaxial matching to III–V semiconductor alloys.

8.6.3.1 Wafer Bonding

Direct wafer bonding refers to the phenomenon of making mirror-polished, flat and clean wafers of almost any materials adhere to each other by bringing their surfaces into contact; van der Waals forces are the source of the bonding. Special treatments are required to make the surfaces hydrophilic (with one or two monolayers of oxygen or water) or hydrophobic (hydrogen-covered due to a HF dip); such surfaces facilitate subsequent bonding, e.g., by hydrogen bridge bonds. Assorted combinations such as Si–Si, Si–SiO_2–Si for silicon-on-insulator (SOI) technology (Section 8.5.2.2), and Si-based heteroepitaxial structures with large substrate mismatches have been fabricated by these methods (Ref. 59).

In wafer bonding the epi-film layer (X) is first deposited on a lattice-matched substrate (Y). Then a clean, atomically flat Si wafer is bonded, usually with applied heat and pressure, to the similarly flat Y–X– combination so as to produce a Y–X–Si sandwich. After substrate Y is chemically etched off, the desired X–Si pair remains. By these means wafer-bonded, III–V semiconductor/Si heteroepitaxial structures of large lattice mismatch have been fabricated containing misfit dislocations but apparently no threading dislocations.

Despite the exciting possibilities of wafer bonding, Ge_xSi_{1-x}/Si strained-layer epitaxy appears to be the primary basis for future silicon heterostructures.

8.6.3.2 Ultrahigh-Vacuum Chemical-Vapor Deposition (UHV/CVD)

This relatively new process pioneered by Meyerson (Refs. 60, 61) challenges our conventional notions regarding CVD processes, particularly the high-temperature epitaxial deposition of silicon. As noted in Section 6.3.2,

Si epitaxy is normally practiced well above 1000°C in atmospheric or sometimes in low-pressure (10–100 torr) reactors. Desorption of water vapor and oxygen contaminants as well as native oxide from the silicon surface, so necessary for high-quality epitaxy, is a result of the high temperatures employed. But, such temperatures also cause troublesome redistribution of dopants between the substrate and epitaxial layer being grown (autodoping). The evolution of silicon technology to ever-smaller dimensions coupled with heavily doped patterned substrates has made it imperative to limit diffusion by reducing the so-called thermal budget, e.g., 1000°C for 30 min, for epitaxy. Rather than high-temperature surface cleaning a novel approach was to passivate the Si surface *ex situ* by etching in HF. This simple step produces a monolayer of hydrogen that passivates the Si surface and reduces its tendency to react with H_2O and O_2 by some 13 orders of magnitude (Ref. 60)! As a result, native oxide does not form; instead the passivated surface can now be exposed to SiH_4, which displaces the adsorbed hydrogen and directly reacts with the substrate to produce epitaxial Si. By this strategy an atomically clean Si surface can be obtained at deposition temperatures of ~500°C. To maintain surface cleanliness, the UHV/CVD reactor is typically first pumped to the 10^{-9} torr range while film growth is carried out at a pressure of ~10^{-3} torr or lower. In addition to silane, simultaneous gas flows of GeH_4, B_2H_6, and PH_3 are used, the first to produce Ge_xSi_{1-x} films and the latter two to dope them. With respect to the latter, excellent control of boron content in Si is achieved through variation of the diborane flow. In addition to Ge_xSi_{1-x}/Si superlattices shown in Fig. 8-29, similar defect-free, highly doped, metastable 10 at.% B–Si/Si multilayer structures have been grown.

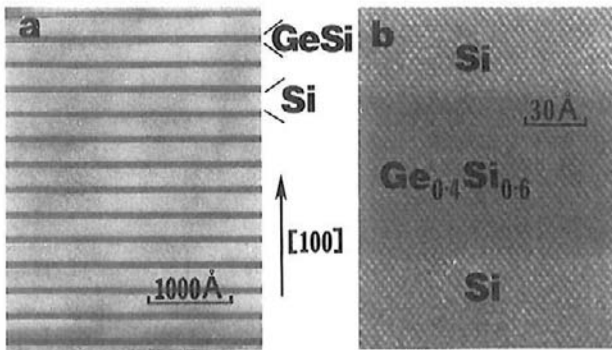


Figure 8-29 (a) Cross-sectional TEM image of $Ge_{0.4}Si_{0.6}$–Si strained-layer superlattice. (b) High-resolution lattice image. (Courtesy of J. C. Bean and R. Hull, AT&T Bell Laboratories.)

The major application for SiGe technology is the base region of heterojunction bipolar transistors (HBTs). In this device the highly doped Ge_xSi_{1-x} base is pseudomorphically grown over an *n*-doped Si collector and then covered with a polysilicon emitter. A typical configuration merges these HBTs with their high-frequency capability and Si CMOS for digital applications. Other potential applications for SiGe ranging from wireless communications to optoelectronics including traditional (CMOS) and quantum (resonant tunneling diode) devices have been discussed in the literature (Refs. 61, 62). It now appears that high-speed devices based on SiGe are competitive with those derived from III–V compound semiconductors. The often quoted observation that "GaAs is the material of the future . . . and it always will be" may have added significance now that SiGe has boosted the performance of Si devices.

8.6.4 Less Common Epitaxy Processes

Just as there are different manifestations of epitaxy, a variety of approaches have been undertaken to produce epitaxial structures and architectures. In this section, several of these unusual techniques are described. In general they are limited to selected materials but include some interesting and potentially important applications.

8.6.4.1 Molecular Beam Allotaxy (MBA)

In this method a buried heteroepitaxial layer is formed upon annealing a subsurface precipitate layer. The MBA process (Ref. 63) can be understood by considering the formation of an epitaxial $Si/CoSi_2/(100)Si$ structure. Because of the small lattice mismatch between $CoSi_2$ and Si noted earlier (Section 8.2.3.2), this is a model MBA system. To produce the trilayer epitaxial structure a buffer homoepitaxial Si film is first deposited on a Si wafer within an MBE system at 500–700°C. Then as the Si flux is maintained constant, Co metal is coevaporated first at a steadily increasing rate, then at a constant rate for some time, and finally at a rate that ramps down to zero. A Si top layer completes the structure. In effect a trapezoidal depth profile of Co in the form of small-to-large-to-small $CoSi_2$ precipitates buried within the Si matrix results. When these precipitates are annealed to temperatures of 1100–1200°C they coalesce into $CoSi_2$ layers with flat and atomically abrupt interfaces. The growth mechanisms, i.e., Ostwald ripening and sintering, already discussed (Section 7.4.5) are responsible for the enlargement of the bigger precipitates at the expense of the smaller ones. A

similar continuous buried β-$FeSi_2$ layer has been prepared by MBA. However, the Si cap layer contained many twins and is considerably more defective than is the case for $CoSi_2$.

8.6.4.2 Mesotaxy

The same objective, buried disilicide layers embedded in silicon, can also be realized by an ion-implantation process known as mesotaxy. In the case of $CoSi_2$ 200 keV Co^+ ions are implanted into Si at fluences of $\sim 2 \times 10^{17}$ ions/cm^2 resulting in a Gaussian distribution of metal atoms. This is followed by annealing treatments at 1000°C during which time the Co concentration profile narrows by similar coalescence phenomena while the surrounding lattice damage is significantly reduced. Abrupt epitaxial silicide layers buried 1000 Å beneath the Si surface have been produced this way (Ref. 64).

Through similar ion implantation of oxygen ions into silicon at high energies and fluences, followed by annealing, highly insulating, subsurface layers of amorphous SiO_2 films have been produced. Ion-beam synthesis of buried SiO_2 is known as SIMOX (separation by implanted oxygen). Together with MBA and mesotaxy they represent important potential processes for the eventual realization of three-dimensional integrated circuits.

8.6.4.3 van der Waals Epitaxy

We have already noted the severe restriction on heteroepitaxy imposed by lattice mismatch between materials containing strong ionic or covalent bonds. However, one might imagine conditions for epitaxy to be relaxed on substrate surfaces that do not contain the dangling electrons which promote chemical bonding across interfaces. In the absence of the latter we speak of an epitaxy based on weak van der Waals interactions (Ref. 65). Suitable surfaces for such purposes stem from easily cleaved materials, or semiconductors whose bonds are terminated by specific atoms, e.g., S or Se on (111) GaAs, H on Si(111), and F on CaF_2. Among the inorganic film/substrate van der Waals epitaxy systems that have been studied are Se/Te; TMX_2/SnS_2; TMX_2/mica; TMX_2/S–GaAs(111); GaSe/Se–GaAs(111); and GaSe/H–Si(111), where TM is a transition metal, e.g., Mo, Nb, and X is S, Te, or Se. Epitaxial organic films, e.g., phthalocyanines/TMX_2 and C_{60}/TMX_2, have also been prepared. In virtually all cases large lattice mismatches are involved, but good epitaxial growth has been claimed.

8.6.4.4 Colloidal Epitaxy

The concept of epitaxy has extended beyond merely arranging atoms on a crystalline substrate, to depositing colloidal particles on a geometrically patterned substrate template. In this way, colloidal silica particles with a 0.525 μm radius have been made to controllably settle onto a substrate lithographically patterned with an FCC (100) planar array of accommodating holes (Ref. 66). The result is a periodic three-dimensional, microporous dielectric structure known as a *photonic crystal*. These crystals have unique optical properties and are claimed to "manipulate light the way superconductors manipulate electrons" (Ref. 67).

8.7 MECHANISMS AND CHARACTERIZATION OF EPITAXIAL FILM GROWTH

In this section we explore the interactions of depositing atoms with surfaces leading to the incorporation of atoms into growing epitaxial films. Mechanisms involving molecular beams have been most studied because characterization instrumentation is usually more readily incorporated into MBE chambers than CVD systems. For this reason there will be some discussion of MBE-based diffraction methods that have revealed so much about the structure of surfaces as well as films that grow on them. Finally the important subject of selective epitaxy will also be addressed to close the chapter.

8.7.1 HOMOEPITAXY OF SILICON

One may expect that the kinetics of MOMBE film growth (Ref. 68) of Si would parallel that of thermal CVD Si homoepitaxy treated in Section 6.3.2. For example, the deposition of epitaxial silicon on Si(001) employing a Si_2H_6 precursor yielded an Arrhenius curve of growth rate vs temperature quite similar to that of Fig. 6-13. However, in MOMBE a transition from reaction control to flux (gas transport) control occurred at about 600°C compared to 1000°C in CVD. Below this temperature the rate of silicon formation was limited by desorption of reaction product hydrogen in a process occurring with an activation energy of 2.0 eV. Above this temperature desorption is very rapid and film growth is limited only by the precursor flux.

In the case of UHV/CVD deposition of Si, the critical hydrogen adlayer begins to desorb above 400°C with an activation energy of 2.58 eV. Film

growth must commence before appreciable surface hydrogen loss occurs. Interestingly, however, SiH_4 itself forms an ordered self-terminating adlayer on Si, or an effective epitaxial monolayer at room temperature. The presence of Ge on the Si surface accelerates epitaxial film growth by lowering the activation energy for desorption of the hydrogen by-product.

8.7.2 Film Growth of GaAs and Related Films

8.7.2.1 Kinetics of Adsorption and Desorption

Since the first step in MBE film growth involves surface adsorption, let us first consider a beam of Ga atoms incident on a GaAs surface. Below about 480°C Ga atoms readily physisorb on the surface but above this temperature Ga both adsorbs and desorbs. Time-resolved mass-spectroscopy measurements of the magnitude of the atomic flux desorbing from the substrate have revealed details of GaAs film-growth mechanisms (Ref. 69). The instantaneous Ga surface concentration, n_{Ga}, is increased by the incident Ga beam flux, $\dot{R}(\mathrm{Ga})$, and simultaneously reduced by a first-order kinetics desorption process. Therefore,

$$\frac{dn_{Ga}}{dt} = \dot{R}(\mathrm{Ga}) - \frac{n_{Ga}}{\tau(\mathrm{Ga})} \tag{8-15}$$

where $\tau(\mathrm{Ga})$ is the Ga adatom lifetime and $n_{Ga}/\tau(\mathrm{Ga})$ represents the Ga desorption flux, $\dot{R}_{des}(\mathrm{Ga})$. Integrating Eq. 8-15 yields

$$\dot{R}_{des}(\mathrm{Ga}) = \dot{R}(\mathrm{Ga})\{1 - \exp[t/\tau(\mathrm{Ga})]\} \tag{8-16}$$

For a rectangular pulse of incident Ga atoms the detected desorption flux closely follows the dependence of Eq. 8-16. Similarly, when the Ga beam is abruptly shut off, the desorption rate decays as $\exp - (t/\tau(\mathrm{Ga}))$. The exponential rise and decay of the signal is shown schematically in Fig. 8-30a.

In the case of As_2 molecules incident on a GaAs surface, the lifetime is extremely short so that the desorption pulse profile essentially mirrors that for deposition (Fig. 8-30b), i.e., $\dot{R}_{des}(As_2) = \dot{R}(As_2)$. However, on a Ga-covered GaAs surface, $\tau(As_2)$ becomes appreciable with desorption increasing only as the available Ga is consumed (Fig. 8-30c). These observations indicate that in order to adsorb As_2 on GaAs, Ga adatoms are essential.

8.7.2.2 Mechanisms

Detailed models for epitaxial growth on (100) GaAs with both As_2 and As_4 precursors have been reviewed more recently by Joyce (Ref. 70). In the

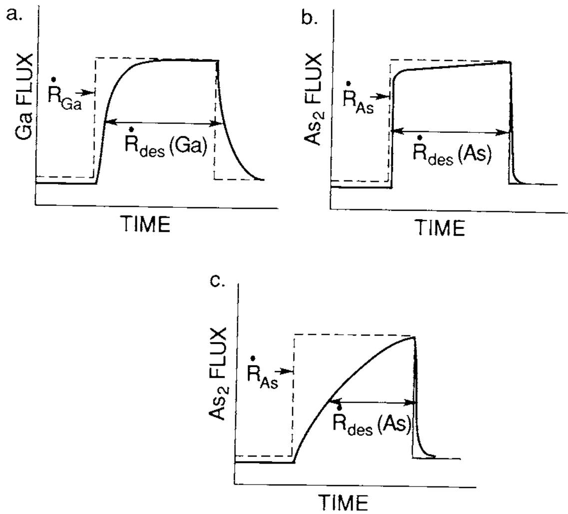


Figure 8-30 Deposition and desorption pulse shapes on (111) GaAs for (a) Ga, (b) As_2, (c) As_2 on a Ga-covered surface. (After Ref. 69.)

case of incident Ga plus As_2 fluxes on a Ga-stabilized GaAs surface (defined later), the As_2 molecules are first adsorbed into a mobile, weakly bound precursor state that has a residence time of $\sim 10^{-5}$ s. First-order dissociative chemisorption on Ga atoms enables As incorporation into the growing film. If the Ga flux is less than half the As_2 flux, one As atom sticks to each Ga atom. At low substrate temperatures, association of As_2 produces As_4 molecules which then slowly desorb. Similarly, when films are grown employing an As_4 flux, these molecules are also first adsorbed as a mobile, weakly bound precursor state. This entity migrates along the surface with a diffusive activation energy of 0.25 eV, while its residence time is temperature dependent with a desorption activation energy of 0.4 eV. The sticking coefficient of As never exceeds 0.5, even with a completely covered Ga surface or when the Ga flux exceeds the As_4 flux. This fact is explained by assuming a pairwise dissociation of the As_4 molecules chemisorbed on adjacent Ga atoms in a second-order reaction. During GaAs growth it is suggested that when two As_4 molecules react, four As atoms are incorporated while the other four desorb as As_4. For either As precursor under conventional deposition conditions the film growth-rate determining step is the migration rate of Ga adatoms to appropriate lattice sites where reaction occurs. In summary these adsorption/desorption effects strongly underscore the kinetic rather than thermodynamic nature and control of MBE growth.

8.7.2.3 Stabilized and Reconstructed Surfaces

Normally, III–V compound-semiconductor films are grown with a two- to tenfold excess of the group V element. This maintains the elemental V/III impingement flux ratio >1. For GaAs and $Al_xGa_{1-x}As$, this condition results in stable stoichiometric film growth for long deposition times. In contrast to the preceding so-called As-stabilized growth, there is Ga-stabilized growth which occurs when the flux ratio is approximately 1. An excess of Ga atoms is generally avoided, however, because they tend to cluster into molten droplets.

The (100) and (111) surfaces of GaAs and related compounds exhibit a variety of reconstructed surface geometries dependent on growth conditions and subsequent treatments (Ref. 55). For As- and Ga-stabilized growth, (2×4) and (4×2) reconstructions, respectively, have been observed on (100) GaAs. Other structures, i.e., C(2×8) As and C(8×2) Ga, have also been reported for the indicated stabilized structures. To complicate matters further, intermediate structures, e.g., (3×1), (4×6), and (3×6), as well as mixtures also exist within narrow ranges of growth conditions. The complex issues surrounding the existence and behavior of these surface reconstructions are continuing subjects of research.

8.7.2.4 Mechanism of Growth in GaInAsP

During epitaxial film deposition of multicomponent semiconductors the mechanisms of substrate chemical reactions and atomic incorporation can be quite complex. For example, a proposed model of sequential deposition of the first two monolayers of GaInAsP on an InP substrate is depicted in Fig. 8-31 for the hydride process (Ref. 71). The first step is suggested to involve adsorption of P and As atoms. Then GaCl and InCl gas molecules also adsorb in such a way that the Cl atoms dangle outward from the surface. Next, gaseous atomic hydrogen adsorbs and reacts with the Cl atoms forming HCl molecules, which then desorb. Now the process repeats with P and As desorption and when the cycle is completed, another bilayer of quaternary alloy film deposits. This picture accounts for single-crystal film growth and the development of variable As/P and Ga/In stoichiometries.

8.7.3 Selective Epitaxy

Compared to actual and potential applications for selective deposition of metal films (Section 6.8.6), selective epitaxial growth (SEG) (also known as

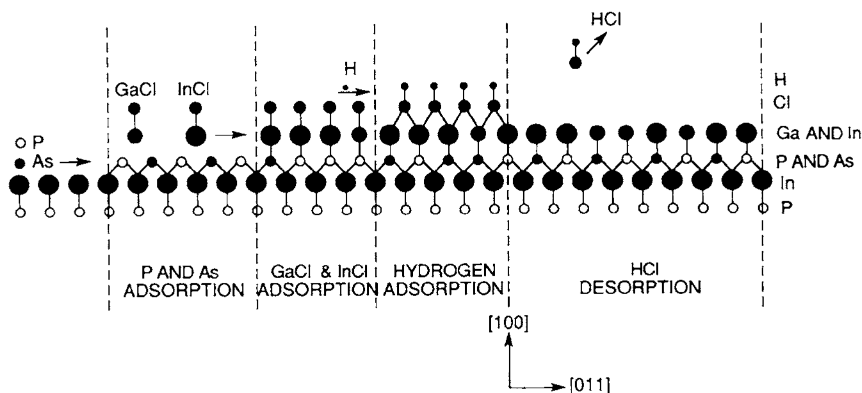


Figure 8-31 Atomic mechanisms involved in the sequential deposition of GaInAsP on InP. (Reprinted with permission from J. Wiley & Sons. From Ref. 71.)

selective area epitaxy (SAE), or selected area growth (SAG) of semiconductors promises to have a far greater impact (Ref. 50). By selective epitaxy we mean the growth of single-crystal films over a desired patterned area and nowhere else. The technique has been practiced in both elemental and compound semiconductor materials. Strategies for SEG differ depending on the type of semiconductor and nature of film deposition, i.e., physical or chemical. Most often, selective epitaxy is associated with differential film deposition in windows, opened in masks, relative to the mask itself. In addition, it is required that selective epitaxial films have the same electrical and optical properties as similarly deposited blanket or broad-area films. Surfaces for SEG are not always planar. If, for example, the substrate is textured with a triangular topography, localized epitaxial growth may be concentrated at the V-shaped bottoms as adsorbed species migrate from the peaks, down the slopes, and into the valleys. In such cases crystallography is important in defining the growth morphologies.

8.7.3.1 Selective Epitaxy in Silicon Materials

There is a history of SEG in Si dating back to the early 1960s (Ref. 72). This effort has almost entirely been devoted to fostering nucleation and growth of epitaxial Si on Si surfaces and suppressing these processes on adjoining SiO_2 or silicon nitride mask surfaces. It is usually accomplished by CVD processes involving H_2 reduction of a chlorosilane, e.g., $SiCl_4$, $Si_2Cl_2H_2$, to deposit Si, and with simultaneous HCl additions to etch away spurious Si nuclei before they coalesce and grow. In an LPCVD-based SEG

study using a $Si_2Cl_2H_2/HCl/H_2$ mixture, as many as 10^5 hemispherical, 0.5 μm sized cap-shaped Si nuclei per cm^2 deposited on dielectric rather than Si surfaces (Ref. 73). Elimination of such an intolerable defect density is critical before any of the applications suggested for selective epitaxy, e.g., filling deep trenches and contact holes, formation of emitter windows in bipolar transistors, processing of DRAM cells and three-dimensional integrated circuits, can be realized. More recently (Ref. 74) selective hetero-epitaxial deposition of SiGe films has been studied and the effects of SiO_2/Si area ratio, HCl flow rates, temperature, and doping on film growth explored. In contrast to Si the growth of SiGe at low temperatures is strongly dependent on oxide coverage.

8.7.3.2 Selective Epitaxy in Compound Semiconductors

In general, compound semiconductor devices require fewer processing steps than their silicon-based counterparts. A great advantage of selective deposition is the further reduction in the number of steps. In fact, the fabrication of a semiconductor laser wholly by selective epitaxy processing has already been demonstrated (Ref. 75).

In the remainder of this section only III–V semiconductors used for photonic devices will be considered and selective MBE, MOCVD, and MOMBE techniques for their deposition briefly reviewed.

8.7.3.2.1 MBE

Through the imposition of a shadow mask between the molecular beam source and substrate, patterned epitaxial features have been readily produced in GaAs, AlGaAs, and InAs materials. However, the growth of complicated structures necessitates manipulating numerous shadow masks. Furthermore, mask contact with the substrate generates defects, and for these reasons SEG by this method is not feasible. By using deposited SiO_2 masks, the latter shortcoming is eliminated. Although high-quality epitaxial films grow in the mask window, a polycrystalline deposit often coats the mask (pseudoselective epitaxy); in addition, the single and polycrystalline regions at the mask edge are generally defective. A key to preventing GaAs nucleation on SiO_2 masks is to enhance desorption of Ga adatoms. Elevated growth temperatures and a reduced As flux tend to enhance selectivity. In general, less deposition on the mask and hence greater selectively occurs with GaAs relative to AlGaAs because SiO_2 reacts more with Al than with Ga.

8.7.3.2.2 MOCVD

Although selective epitaxy by MBE means is limited by physical variables, MOCVD takes advantage of the greater number of possibilities inherent in varying system chemistry. For example, by using a metalorganic chloride source of Ga, i.e, $Ga(C_2H_5)_2Cl$ instead of, say, $(CH_3)_3Ga$ (TMG), selective growth of GaAs over SiO_2, Si_3N_4, and $SiON_x$ is more readily achieved (Ref. 76). Apparently GaCl does not adsorb on the mask material, nor is this chloride dissociated to free Ga for deposition. This suggests that Al and In halides may also promote selective growth of AlGaAs and InGaAs. In a similar vein the reaction of $AsCl_3$ together with $(CH_3)_3Ga$ inhibited deposition of GaAs on W and Si_3N_4 masks (Ref. 77). Furthermore, HCl adsorption on the mask may aid etching of any deposit there.

8.7.3.2.3 MOMBE

Although the chemical precursors are the same as for MOCVD, high-vacuum operation eliminates boundary-layer transport and suppresses gas-phase reactions, both of which are conducive to high selectivity. Fewer gas-phase collisions of precursor molecules mean less decomposition and reduced impingement on and interaction with the mask. As a result, greater selectivity is found in MOMBE than in atmospheric MOCVD. A study of selective GaAs deposition employing mass spectroscopy is particularly revealing (Ref. 78). At about 350°C TMG molecules were observed to adsorb, immediately decompose, and release Ga atoms on the GaAs surface. But, on the SiO_2 mask surface molecular decomposition did not occur below 535°C. This suggests a processing window for selective epitaxy in this system based on temperature.

8.7.4 In Situ Film Characterization of MBE Films

This section deals with two techniques that are capable of monitoring the structure and composition of epitaxial films during *in situ* growth. Low-energy electron diffraction (LEED) and reflection high-energy electron diffraction (RHEED), which are distinguished in Fig. 8-32, have provided much of our knowledge of epitaxial film growth mechanisms. An ultrahigh-vacuum environment is a necessity for both methods because of the sensitivity of diffraction to adsorbed impurities and the need to eliminate electron-beam scattering by gas molecules. This is why they are limited to use in MBE systems.

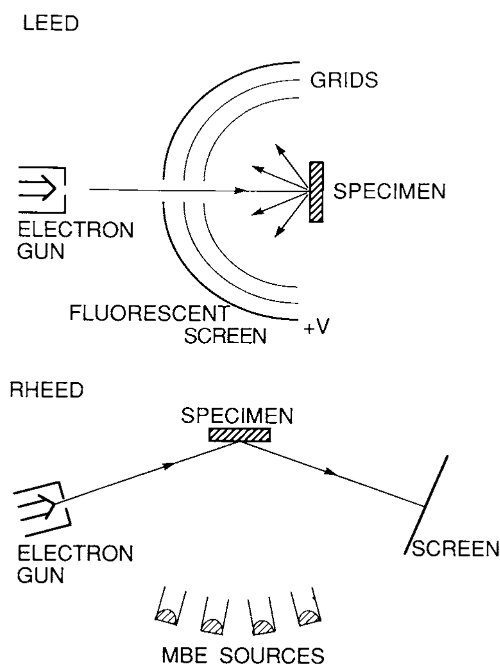


Figure 8-32 Experimental arrangements for LEED and RHEED techniques.

8.7.4.1 LEED

In LEED a low-energy electron beam (10–1000 eV) impinges normally on the film surface and only penetrates a few angstroms below the surface. Bragg's law for both lattice periodicities in the surface plane results in cones of diffracted electrons emanating along forward and backscattered directions. Simultaneous satisfaction of the diffraction conditions means that constructive interference occurs where the cones intersect along a set of lines or beams radiating from the surface. These backscattered beams are intercepted by a set of grids raised to different electric potentials. The first grids encountered retard the penetration of low-energy inelastic electrons. Diffracted (elastic) electrons of higher energy pass through and, accelerated by later grids, produce illuminated spots on the hemispherical fluorescent screen.

Both LEED and RHEED patterns of the (7×7) structure of the Si (111) surface are shown in Fig. 8-33. To obtain some feel for the nature of these diffraction patterns it is useful once again to think in terms of the reciprocal space. Arrays of reciprocal latice points form rods or columns of reciprocal-lattice planes shown as vertical lines pointing normal to the real surface.

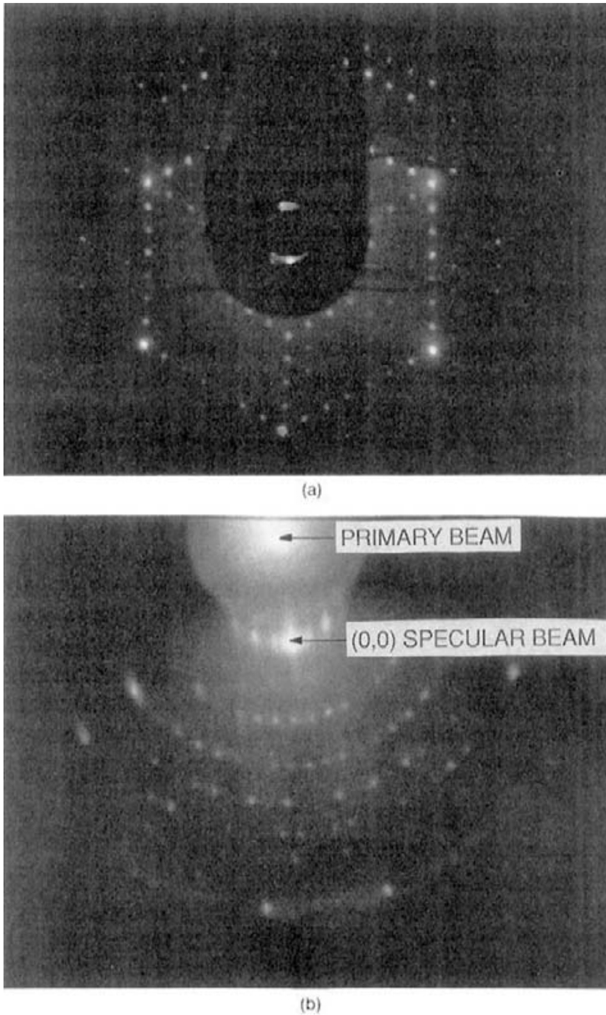


Figure 8-33 (a) LEED pattern of Si surface (38 eV electron energy, normal incidence). (b) RHEED pattern of Si surface (5 keV electron energy, along $\langle 112 \rangle$ azimuth). (Courtesy of H. Gossmann, AT&T Bell Laboratories.)

They are indexed as (10), (20), etc., in Fig. 8-34. Consider now an electron wave of magnitude $2\pi/\lambda$ propagating in the direction of the incident radiation and terminating at the origin of the reciprocal lattice. Following Ewald we draw a sphere of radius $2\pi/\lambda$ about the center. A property of this construction is that the only possible directions of the diffracted rays are

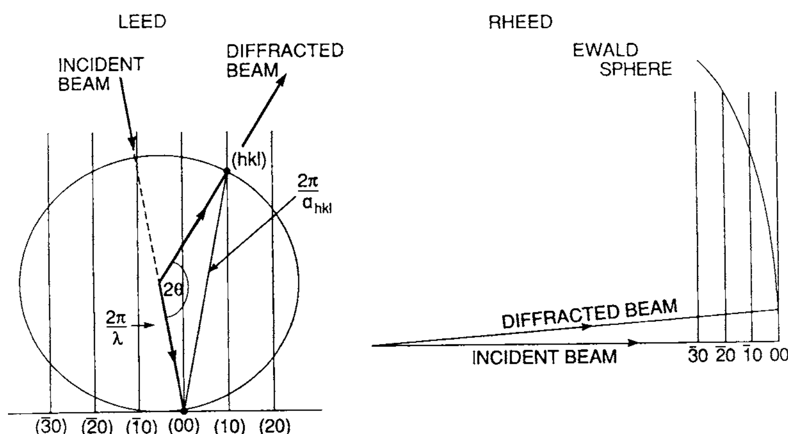


Figure 8-34 Ewald sphere construction for LEED and RHEED methods. The film plane is horizontal, and reciprocal planes are vertical lines.

those which intersect the reflecting sphere at reciprocal lattice points as shown. To prove this we note that the normal to the reflecting plane is the vector connecting the ends of the incident and diffracted rays. But this vector is also a reciprocal lattice vector. Its magnitude is $2\pi/a$, where a is the interplanar spacing for the diffracting plane in question. It is obvious from the geometry that

$$\frac{2\pi}{a} = \frac{2 \times 2\pi \sin\theta}{\lambda}, \tag{8-17}$$

which reduces to Bragg's law, the requisite condition for diffraction. When the electron energies are small as in LEED, the wavelength is relatively large, yielding a small Ewald sphere. A sharp spot diffraction pattern is the result. The intense hexagonal spot array of Fig. 8-33a reflects the sixfold symmetry of the (111) plane while the six fainter spots in between are the result of the (7×7) surface reconstruction.

8.7.4.2 RHEED

An immediate advantage of RHEED is that the measurement apparatus does not physically interfere with deposition sources in an MBE system the way LEED does. This is one reason why RHEED has become the overwhelmingly preferred real-time film characterization accessory in MBE systems. In RHEED the electron beam is incident on the film surface at a

grazing angle of a few degrees at most. Electron energies are much higher than for LEED and range from 5 to 100 keV. From the standpoint of diffraction, the high electron energies lead to a very large Ewald sphere (Fig. 8-34). The reciprocal lattice rods have finite width due to lattice imperfections and thermal vibrations; likewise, the Ewald sphere is of finite width because of the incident electron-energy spread. Therefore, the intersection of the Ewald sphere and rods occurs for some distance along their height, resulting in a streaked rather than spotty diffraction pattern. During MBE film growth both spotted and streaked patterns can be observed; spots occur as a result of three-dimensional volume diffraction at islands or surface asperities while streaks characterize smooth layered film growth. These features can be seen in the RHEED patterns obtained from MBE-grown GaAs films (Fig. 8-35).

8.7.4.3 RHEED Oscillations

An important attribute of the RHEED technique is that the diffracted beam intensity is relatively immune to thermal attenuation arising from lattice vibrations. This makes it possible to observe the so-called RHEED oscillations during MBE growth of epitaxial semiconductor films. In the case of GaAs the intensity of the specular RHEED beam undergoes variations which track the step density of Ga on the growing surface layer. If we consider Fig. 8-36 and associate the maximum beam intensity with the flat surface where the fractional coverage $\overline{\theta} = 0$ (or $\overline{\theta} = 1$), then the minimum intensity corresponds to $\overline{\theta} = 0.5$. During deposition of a complete monolayer, the beam intensity, initially at the crest, falls to a trough and then crests again. Film growth is, therefore, characterized by an attenuated, sinelike wave with a period equal to the monolayer formation time. Under optimal conditions the oscillations persist for many layers and serve to conveniently monitor film growth with atomic resolution. The temperature above which RHEED oscillations are expected can be easily estimated. To allow a few atomic jumps to occur in order to smooth terraces before they are covered by a monolayer (per second) requires a diffusivity of roughly 10^{-15} cm^2/s. By the example given in Section 7.4.3, RHEED oscillations are predicted to occur above $0.2T_M$, $0.12T_M$, and $0.03T_M$ on group IV element, metal, and alkali halide substrates, respectively, in rough qualitative agreement with experiment.

What we have said strictly applies to MBE growth on unstepped or singular planes. But what about growth on vicinal surfaces roughened by steps and terraces of length h? At low temperatures, the surface-diffusion length of adatoms is less than h and two-dimensional nucleation of monolayer islands occurs as on a singular plane. These are the conditions that

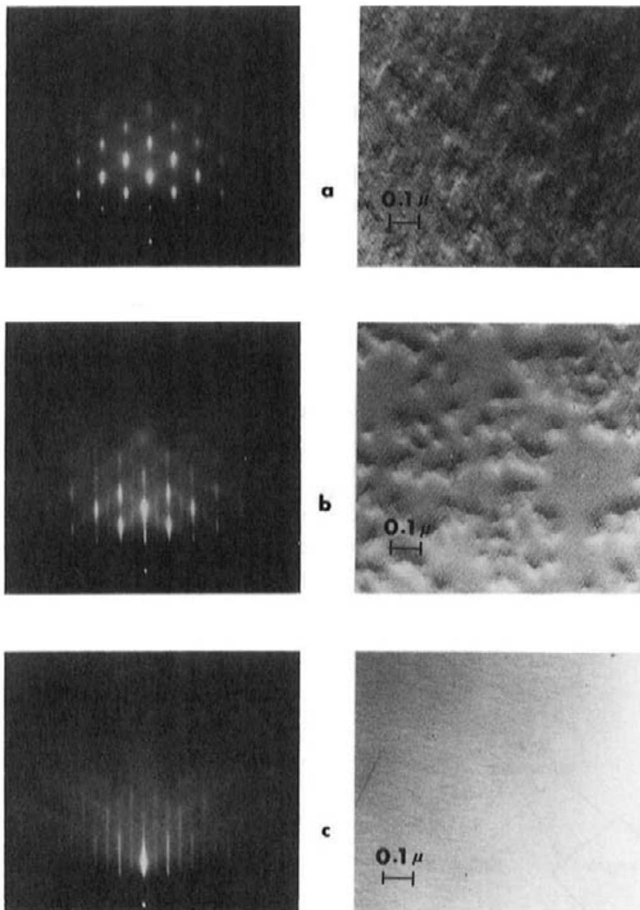


Figure 8-35 RHEED patterns [40 keV, $(\bar{1}\bar{1}0)$ azimuth] and corresponding electron micrograph replicas (38,400 ×) of same GaAs surface. (a) Polished and etched GaAs substrate heated in vacuum to 580°C for 5 min. (b) 150 Å film of GaAs deposited; (c) 1 μm of GaAs deposited. (From Ref. 69. Courtesy of A. Y. Cho, AT&T Bell Laboratories.)

lead to RHEED oscillations. However, at elevated temperatures the adatom migration length exceeds h and step edges now become a favored sink for adatoms (Ref. 70). Instead of island formation and coalescence with a progressive buildup of $\bar{\theta}$, film growth now proceeds by step propagation across a featureless terrain ($\bar{\theta} = 0$). Under these conditions RHEED oscillations vanish. By noting the disappearance of RHEED oscillations (Ref. 66)

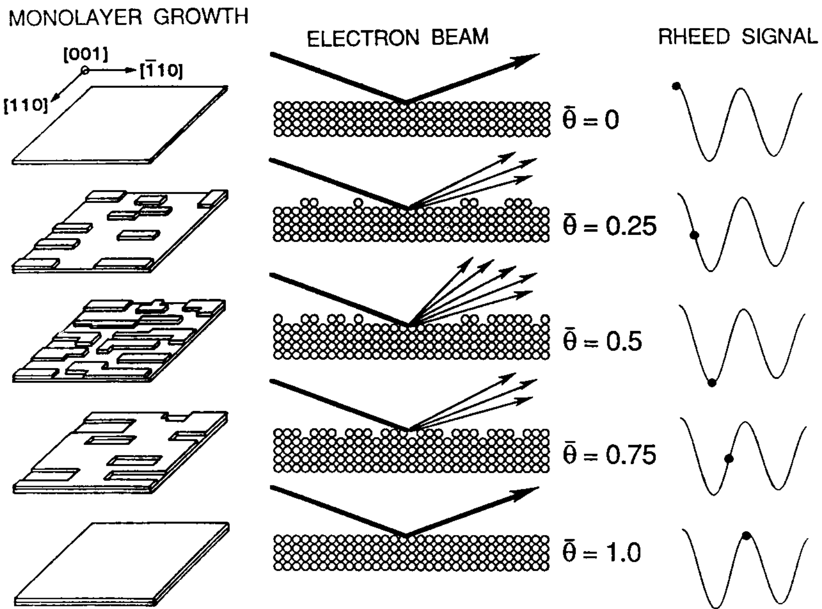


Figure 8-36 Real space representation of the formation of a single complete monolayer. $\bar{\theta}$ is the fractional layer coverage. Corresponding RHEED oscillation signal is shown.

in GaAs, the transition temperature (T_c) separating the two-dimensional nucleation and step propagation regimes was experimentally measured to be 590°C. Estimates of T_c based on random walk, i.e., $h^2 = 2D\tau$, where D and τ are the adatom diffusion coefficient and lifetime, respectively, are consistently ~150–200 K lower than observed. A probable reason for the underestimate is the effective reduction in D due to adatom reactions at step edges.

8.8 CONCLUSION

Synthesized thin films and bulk single crystals share much in common. Both were initially grown in response to scientific interest in the properties of solids that were crystallographically oriented, had controlled compositions, were devoid of grain boundaries, and were freer of complicating defects than their polycrystalline counterparts. After years of development, bulk single crystals emerged to critically alter existing technologies as well as create new ones, e.g., metal superalloy turbine blades for jet engines,

electroceramics such as quartz, niobates, and garnets for electronic, magnetic, and optical applications, and silicon for microelectronics. So, too, with epitaxial thin films. But, by nature, heteroepitaxial thin films have proven a greater challenge to grow and exploit than bulk crystals. Intrinsically small film grain sizes, and the necessity of a substrate often having a different composition and structure, adversely influence crystalline perfection through the presence of grain boundaries, generation of film stress, and perpetuation of interfacial defects. Furthermore, electronic and optoelectronic applications requiring epitaxial films are extremely intolerant of defects.

It has been a long odyssey from the early so-called epitaxial metal films deposited on alkali-halide crystals to the lattice-matched heteroepitaxial compound-semiconductor films. Along the way a novel collection of epitaxial vapor-phase deposition processes based on chemical (CVD, MOCVD) and physical (MBE) methods as well as hybrid combinations (MOMBE, high-vacuum CVD) have been developed. These have enabled extraordinary compositional, structural, and film-thickness control of layered heteroepitaxial structures, resulting in an impressive array of GaAs- and InP-based LEDs and lasers for display, recording, and optical communications applications. These same III–V semiconductors as well as Ge_xSi_{1-x} materials have also been exploited in high-speed transistors used for assorted computer and communications purposes. In overcoming a most complex set of materials and processing barriers, the remarkable development of GaInN-based blue lasers has been a supreme triumph of epitaxy and a testament to its possibilities.

EXERCISES

1. For the geometry shown in Fig. 1-4, what are the crystallographic indices of the epitaxial orientations between the $CoSi_2$ film and Si substrate?

2. Suppose grains of a polycrystalline deposit exhibit a (111) texture, i.e., (111) planes are parallel to the substrate but with a random orientation in the surface plane.
   (a) Sketch the morphology of isolated grains deposited on an amorphous substrate.
   (b) A triangular grating is etched into the substrate. The grating angle is 109.5° (instead of the 90° angle in Fig. 8-3c). Describe the Miller indices of relevant directions and planes in the deposited graphoepitaxial layer.

(c) How would the answers to part (b) change if the triangular grating angle were 70.5°?

3. Sketch an array of squares of length a_f on a sheet of paper. On a second sheet of paper sketch a second array of squares of length a_s ($a_s \neq a_f$). Tilt these sheets relative to each other to create an interface along which corners of both sets of squares are coincident. Measure the involved angles. Do they agree with Eq. 8-2?

4. X-ray measurements on an epitaxial $YBa_2Cu_3O_7$ superconductor film deposited on $SrTiO_3$, yields interplanar spacings $d_{001} = 11.647$ Å and $d_{110} = 2.733$ Å. For strain-free $YBa_2Cu_3O_7$ the interplanar spacings are $d_{001} = 11.680$ Å and $d_{110} = 2.724$ Å. Determine the strain magnitudes and signs in the [001] and [110] directions. Describe the state of stress.

5. An epitaxial film is deposited with a lattice mismatch of 2% relative to a substrate whose diameter is 10 cm. The epilayer grown at elevated temperature is thick enough that the residual strain is zero.
   (a) If the Burgers vector is 2 Å, what is the interfacial misfit-dislocation spacing?
   (b) A misfit dislocation grid covers the entire film/substrate interface. What is the total length of dislocation line?
   (c) Estimate the dislocation density.

6. (a) Physically explain the ordinate intercepts and curve shapes that lead to E vs b/S variations for $d < d_c$ and $d > d_c$ in Fig. 8-7.
   (b) Plot the expected equilibrium misfit dislocation spacing S vs film thickness.
   (c) In general the experimental misfit dislocation spacing is larger than the equilibrium spacing. Why?

7. Increasing the molecular weight (M) within a given family of materials usually increases the lattice constant (a_0), reduces the binding energy, and lowers the melting point (T_M). Check the validity of such trends in the elemental semiconductors Si, Ge, and Sn, as well as assorted III–V and II–VI semiconductors within a binary compound family (one common element) by plotting a_0 and T_M vs M. Also plot the bandgap energy vs M for different semiconductor families. Is there a trend?

8. On the basis of the discussion in Section 3.2.3.2, are the MBE deposition windows for InP, GaP, InAs, and GaN expected to be wider or narrower relative to GaAs? Assume that the phase diagrams in these binary systems resemble those for GaAs.

9. (a) Show that for $Ga_xIn_{1-x}As_yP_{1-y}$ lattice matched to InP, $x = 0.190y/(0.418–0.0125x)$.

(b) Derive an expression that yields the elastic mismatch strain for Ge_xSi_{1-x} on (100) Si in terms of the alloying level x. Repeat for $In_xGa_{1-x}As$ on GaAs.

10. The quaternary $Ga_xIn_{1-x}As_yP_{1-y}$ alloy semiconductors have an energy gap and lattice parameter given, respectively, by $E_g(\mathrm{eV}) = 1.35 - 0.72y + 0.12y^2$ and $a_0(x, y)(\text{Å}) = 0.1894y - 0.4184x + 0.0130xy + 5.869$. For light-wave devices operating at 1.32 and 1.55-μm, calculate the values for x and y, assuming lattice matching to InP.

11. It is desired to make diode lasers that emit coherent radiation with a wavelength of 1.0 μm. For this purpose, III–V compounds or ternary solid-solution alloys derived from them can be utilized. At least four possible compound combinations (alloys) will meet the indicated specifications. For each alloy specify the original pair of binary compounds, the composition, and the lattice constant. (Assume linear mixing laws.)

12. Speculate on the implications of SiGe heteroepitaxy on single-crystal Ge wafers. How do you expect it to differ from SiGe epitaxy on Si? Specifically address the questions of misfit, critical film thickness d_c, strain-induced modulations in surface morphology of SiGe films, deposition methods, and possible uses.

13. (a) What lattice mismatch can be expected between an $Al_{0.4}Ga_{0.6}As$ film deposited on GaAs?
    (b) A heteroepitaxial structure consisting of $Al_yIn_{1-y}As$ lattice-matched to $Ga_{0.47}In_{0.53}As$ is deposited on InP. What are the values for y and E_g for the ternary AlInAs alloy?

14. Sequential layers of GaAs and AlGaAs films were grown by MBE. The GaAs beams were on throughout the deposition, which lasted 1.5 h. However, the Al beam was alternately on for 0.5 min and off for 1 min during the entire run. Film thickness measurements showed that 1.80 μm of GaAs and 0.35 μm of AlAs were deposited.
    (a) What are the growth rates of GaAs and $Al_xGa_{1-x}As$?
    (b) What is x?
    (c) What are the thicknesses of the GaAs and $Al_xGa_{1-x}As$ layers?
    (d) How many layers of each film were deposited? (From A. Gossard, AT&T Bell Laboratories.)

15. (a) If the temperature regulation in effusion cells employed in MBE is $\pm 1°$C, what is the percent variation in the flux of atoms arriving at the substrate for the deposition of InP films? (In evaporated at 1100 K, P_2 evaporated at 420 K.)
    (b) Similar InP films are grown by MOCVD methods in a reactor maintained at 725°C. Roughly estimate the expected percent vari-

ation in the growth rate of films if the reactor temperature is regulated to $\pm 1°C$. Assume the activation for film growth is 0.5 eV.

16. RHEED oscillations from epitaxial films grown on a vicinal surface oriented $1°$ from the (100) crystallographic plane damp out at a temperature of 550°C. The activation energy for atomic diffusion of depositing species is known to be 1.9 eV. If the vicinal surface is oriented $3°$ from (100), at what temperature are the oscillations likely to damp out?

REFERENCES

1. L. Royer, *Bull. Soc. Fr. Mineral Cristallogr.* **51**, 7 (1928).
2. S. Nakamura, *J. Mater. Res.* **14**, 2716 (1999).
3. D. J. Paul, *Adv. Mater.* **11**(3), 191 (1999).
4. S. M. Sze, ed., *Modern Semiconductor Device Physics*. J. Wiley & Sons, New York, 1998.
5. J.-P. Colinge, *SOI Technology: Materials to VLSI*, 2nd ed. Kluwer, Boston, 1997.
6. Special issue, *Proc. IEEE, Quantum Devices and Applications* **87**(4). 535 (1999).
7. F. Capasso and G. Margaritondo, eds., *Band Discontinuities: Physics and Device Applications*, Elsevier, Amsterdam, 1990.
8. O. Auciello, C. M. Foster, and R. Ramesh, *Ann. Rev. Mater. Sci.* **28**, 501 (1998).
9. M. W. Geis, D. C. Flanders, H. I. Smith, and D. A. Antoniadis, *J. Vac. Sci. Technol.* **16**, 1640 (1979).
10. E. S. Machlin, *Materials Science in Microelectronics*, Vol. 1. Giro Press, Croton-on-Hudson, NY, 1995.
11. H. von Känel, N. Onda, and L. Miglio, in *Science and Technology of Thin Films*, ed. F. C. Matacotta and G. Ottaviani. World Scientific, Singapore, 1995.
12. C. J. Palmstrom, *Ann. Rev. Mater. Sci.* **25**, 389 (1995).
13. G. A. Prinz, *MRS Bull.* **12**(6), 28 (1988).
14. C. P. Flynn, *MRS Bull.* **15**(6), 30 (1991).
15. F. Riesz, *Vacuum* **46/8-10**, 1021 (1995).
16. F. C. Frank and J. H. van der Merwe, *Proc. Roy. Soc.* **A189**, 205 (1949).
17. J. W. Matthews, in *Epitaxial Growth*, ed. J. W. Matthews. Academic Press, New York, 1975.
18. W. D. Nix, *Metall. Trans.* **20A**, 2224 (1989).
19. J. W. Matthews and A. E. Blakeslee, *J. Cryst. Growth* **29**, 272 (1975).
20. G. B. Stringfellow, *Rep. Prog. Phys.* **45**, 469 (1982).
21. J. C. Bean, in *Silicon-Molecular Beam Epitaxy*, eds. E. Kasper and J. C. Bean. CRC Press, Boca Raton, FL, 1988.
22. J. C. Bean, *Physics Today* **39**(10), 2 (1986).
23. A. G. Cullis, *MRS Bull.* **21**(4), 21(1996).
24. A. J. Pidduck, D. J. Robbins, and A. G. Cullis, in *Microscopy of Semiconductor Materials*, eds. A. G. Cullis, J. L. Hutchison, and A. E. Staton-Bevan. IOP Publishing, Bristol, 1993.
25. D. E. Jesson, K. M. Chen, and S. J. Pennycook, *MRS Bull.* **21**(4), 31 (1996).
26. R. Hull and J. C. Bean, *Crit. Rev. Solid State Mater. Sci.* **17**(6), 507 (1992).
27. M. Ohring, *Failure and Reliability of Electronic Materials and Devices*. Academic Press, Boston, 1998.
28. L. B. Freund, *MRS Bull.* **17**(7), 53 (1992).

29. S. M. Sze, *Semiconductor Devices—Physics and Technology*. McGraw-Hill, New York, 1985.
30. J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*. Macmillan, New York, 1990.
31. H. C. Casey, D. D. Sell, and M. B. Panish, *Appl. Phys. Lett.* **24**, 63 (1974).
32. P. Bhattacharya, *Semiconductor Optoelectronic Devices*. Prentice Hall, Englewood Cliffs, NJ, 1994.
33. G. B. Stringfellow, in *Semiconductors and Semimetals*, eds. J. I. Pankove and T. B. Moustakas, Vol. 50, Academic Press, Boston, 1998.
34. L. A. Kolodziejski, R. L. Gunshor, and A. V. Nurmikko, *Ann. Rev. Mater. Sci.* **25**, 711 (1995).
35. Directory of Substrate Suppliers, *Compound Semiconductor* **4**(5), 36 (1998).
36. B. Johnstone, *Compound Semiconductor* **1**(1), 36 (1995).
37. S. Nakamura and G. Fasol, *The Blue Laser Diode*. Springer Verlag, Berlin, 1997.
38. J. I. Pankove and T. B. Moustakas, in *Semiconductors and Semimetals*, eds. J. I. Pankove and T. B. Moustakas, Vol. 50, Academic Press, Boston, 1998.
39. T. Matsuoka, *Adv. Mater.* **8**(6), 469 (1996).
40. S. Nakamura, *Ann. Rev. Mater. Sci.* **28**, 125 (1998).
41. M. Meyer, *Compound Semiconductor* **5**(9), 26 (1999).
42. G. K. Celler, McD. Robinson, and L. E. Trimble, *J. Electrochem. Soc.* **132**, 211 (1985).
43. M. Razeghi, *The MOCVD Challenge*. Adam Hilger, Bristol, 1989.
44. T. F. Kuech, *Proc. IEEE* **80**, 1609 (1992).
45. R. F. Hicks, *Proc. IEEE* **80**, 1625 (1992).
46. G. B. Stringfellow, *Organometallic Vapor-Phase Epitaxy*. Academic Press, San Diego, 1999.
47. A. G. Thompson, *Mater. Lett.* **30**, 255 (1997).
48. A. G. Thompson, R. A. Stall, P. Zawadzki, and G. H. Evans, *J. Electronic Matls.* **25**, 1487 (1996).
49. O. Ombacher, *J. Phys. D: Appl. Phys.* **31**, 2653 (1998).
50. J. J. Coleman, *Proc. IEEE* **85**, 1715 (1997).
51. A. Usui and H. Watanabe, *Ann. Rev. Mater. Sci.* **21**, 185 (1991).
52. L. Niinisto and M. Leskela, *Thin Solid Films* **225**, 130 (1993).
53. M. L. Green, D. Brasen, H. Luftman, and V. C. Kannan, *J. Appl. Phys.* **65**, 2558 (1989).
54. G. W. Wicks, *Crit. Rev. Solid State Mater. Sci.* **18**(3), 239 (1993).
55. J. Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy*. Academic Press, Boston, 1993.
56. M. B. Panish and H. Temkin, *Ann. Rev. Mater. Sci.* **19**, 209 (1989).
57. H. Mataki and H. Tanaka, *Microelectron. J.* **25**, 619 (1994).
58. S. Izumi, Y. Kouji, and N. Hayafuji, *J. Vac. Sci. Technol.* **B17**(3), 1011 (1999).
59. U. Gösele and Q.-Y. Tong, *Ann. Rev. Mater. Sci.* **28**, 215 (1998).
60. B. S. Meyerson, *Proc. IEEE* **80**, 1592 (1992).
61. D. J. Paul, *Adv. Mater.* **11**(3), 191 (1999).
62. R. A. Metzger, *Compound Semiconductor* **1**(3), 21 (1995).
63. S. Mantl, *J. Phys. D: Appl. Phys.* **31**, 1 (1998).
64. A. E. White and K. T. Short, *Science* **241**(8), 930 (1988).
65. A. Koma, *Thin Solid Films* **216**, 72 (1992).
66. A. van Blaaderen, R. Ruel, and P. Wiltzius, *Nature* **385**, 321 (1997).
67. J. D. Joannopoulos, R. D. Meade and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*. Princeton Univ. Press, Princeton, NJ, 1995.
68. B. A. Joyce, T. Shitara, M. R. Fahy, K. Sato, J. H. Neave, P. N. Fawcett, I. Kamiya, and X. M. Zhang, *Mater. Sci. Eng.* **B30**, 87 (1995).

69. A. Y. Cho and J. R. Arthur, *Prog. Solid State Chem.* **10**, 157 (1975).
70. B. A. Joyce, in *Critical Issues in Semiconductor Materials and Processing Technologies*, eds. S. Coffa, F. Priolo, E. Rimini, and J. M. Poate, Kluwer, The Netherlands, 1992.
71. G. H. Olsen, *GaInAsP Alloy Semiconductors*, ed. T. P. Pearsall. J. Wiley & Sons, New York, 1982.
72. B. J. Ginsberg, J. Burghartz, G. B. Bronner, and S. R. Mader, *IBM J. Res. Develop.* **34**, 816 (1990).
73. J. T. Fitch, *J. Electrochem. Soc.* **141**, 1046 (1994).
74. S. Bodnar, E. de Berranger, P. Bouillon, M. Mouis, T. Skotnicki, and J. L. Regolini, *J. Vac. Sci. Technol.* **B15**, 712 (1997).
75. Y. L. Wang, H. Temkin, R. A. Hamm, R. D. Yadvish, D. Ritter, L. H. Harriot, and M. B. Panish, *Electron. Lett.* **27**, 1324 (1991).
76. T. F. Keuch, M. A. Tischler, and R. Potemski, *Appl. Phys. Lett.* **54**, 910 (1989).
77. R. Azoulay and L. Dugrand, *Appl. Phys. Lett.* **58**, 128 (1991).
78. Y. Ohki, Y. Hirantani, and M. Yamada, *Jpn. J. Appl. Phys.* **28**, L1486 (1989).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 9

Film Structure

9.1 INTRODUCTION

There is scarcely a more important variable than structure in influencing the properties of both as-deposited and subsequently processed films. In particular, controlling grain size, morphology, and crystallinity are primary concerns of the processing used to generate practically all thin films and coatings. Irrespective of whether they arise from solidified liquids, from solid-state reactions, or from condensed vapors, grain structures are nature's way of organizing countless numbers of atoms that crystallize from numerous growth centers. And as is the case with virtually all phase transformations in materials, we have seen how *nucleation* and *growth* processes are involved in thin-film formation. The initially nucleated film grains provide the template that often programs the evolution of the fully developed grain structure.

Interestingly, similar structural morphologies cut across all material classes and the different processing methods used to produce them. In this regard it is instructive to compare the grain structures of films with those of bulk phases derived from liquid-solid transformations. For example, just as vapor supersaturation (specifically ΔG_V) drives thin-film nucleation, so, too, the free energy associated with supercooling below the melting temperature (ΔG_S) is the thermodynamic driving force for melt solidification (Ref. 1). And in both films and castings the temperature of the bounding substrate or mold has an important influence on the evolution of grain structure. At a chilled mold wall, for example, ΔG_S is large and this reduces ΔG^*, so that the grain nucleation rate is highest, and the grain size smallest. Further into the melt the extent of supercooling diminishes, the nucleation rate declines,

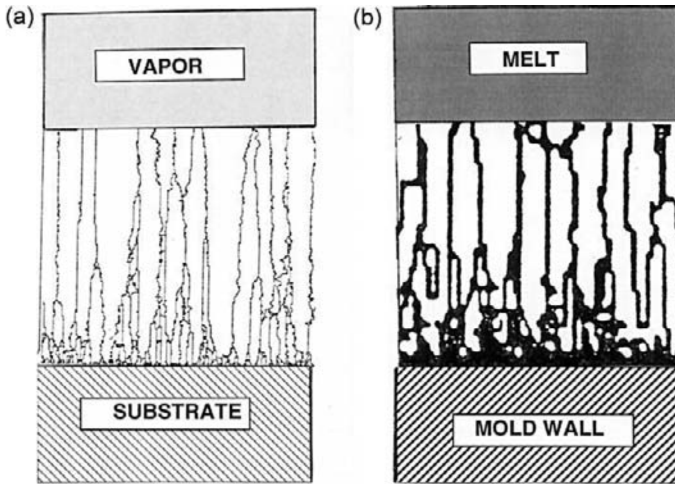


Figure 9-1 Comparison of simulated grain morphologies (a) at the thin film–substrate interface (from Ref. 2) and (b) at the interface between solidified cast grains and the mold wall. (After F. Zhu and W. Smith, *Acta Metall.* **40**, 3369, 1992.)

and the grain size increases, elongating in a direction normal to the mold-wall plane; this is also the direction in which the evolved latent heat of fusion flows. Perhaps not unexpectedly, the simulated grain morphology of evaporated thin films (Ref. 2) shown in Fig. 9-1a resembles that for castings (Fig. 9-1b). In both cases fine grains at the substrate interface are normally topped by columnar grains. Such analogies cannot be extended too far because temperatures and atom concentrations as well as movements differ at vapor–solid and liquid–solid interfaces. Thus, whereas three-dimensional grain growth often characterizes the solidification of melts, film growth from the vapor reflects the directionality of condensing atoms. Despite the fact that film grains may perhaps be $\sim$10–1000 times smaller, both structures exhibit the phenomenon of preferred texture or preferential grain growth in specific crystallographic directions.

Although it occurs seamlessly during deposition, distinguishing film nucleation from the development of fully realized polycrystalline film and coating grain structures represents a great leap. In the first topic treated we shall see that films display a continuum of microstructures ranging from columnar to equiaxed to amorphous, and mixtures in between. The columnar morphology, which, interestingly, develops in amorphous as well as crystalline films, often has internal defect and intercolumnar void or porosity distributions that adversely affect properties, e.g., moisture absorption and performance degradation of optical films. Computer simulations of

deposition, one atom at a time, will help us understand the structural evolution of such columnar structures. When deposited at high temperatures the grain structure tends to be more equiaxed and reminiscent of microstructures of annealed solids. At the other extreme, low-temperature deposition is conducive to the formation of fine-grained or even amorphous structures. Furthermore, very novel film morphologies and architectures can be synthesized when deposition geometries and conditions are controllably altered.

Although much of our attention will be directed to films on flat substrates, we shall also consider the structure and coverage of films deposited on stepped and deeply trenched surfaces. The most important examples of these arise in integrated circuit films where microstructures are often constrained by features having submicron dimensions. Because metallization reliability is at stake, issues of interest include grain morphology dependence on deposition variables, and how some measure of grain-size control and modification can be exercised. These topics are addressed not only here, but also elsewhere in this book.

Since they are technically important and raise interesting bonding and property issues as well, the chapter ends with a discussion of amorphous thin films. Reconciling the macroscopic and microscopic views of film structure derived from various experimental observations and computer simulations is a theme and challenge that pervades the discussion.

9.2 STRUCTURAL MORPHOLOGY OF DEPOSITED FILMS AND COATINGS

Since structural morphologies of metal, semiconductor, and ceramic films possess similar features to a first approximation, they can be largely treated in a generic manner. The influence of deposition variables on the structural features that develop in physically deposited films (Ref. 3) has been universally depicted in terms of structure-zone diagrams (SZDs). In this section a number of these representations will be presented with the objective of revealing the intimate interplay between deposition and film-growth processes.

9.2.1 Structure-Zone Models for Evaporated and Sputtered Coatings

9.2.1.1 Basic Film Condensation Processes

As we have seen, condensation from the vapor involves incident atoms becoming bonded adatoms which then diffuse over the film surface until

they desorb or, more commonly, are trapped at low-energy lattice sites. Finally, incorporated atoms reach their equilibrium position in the lattice by bulk diffusive motion. This atomic odyssey involves four basic processes: (1) shadowing, (2) surface diffusion, (3) bulk diffusion, and (4) desorption. The last three are quantified by the characteristic diffusion and sublimation activation energies whose magnitudes scale directly with the melting point T_M of the condensate. Shadowing is a phenomenon arising from the geometric constraint imposed by the roughness of the growing film and the line-of-sight impingement of arriving atoms. The dominance of one or more of these four processes as a function of substrate temperature, T_S, is manifested by different structural morphologies. This is the basis of structure-zone models (SZMs), which have been devised to characterize film and coating grain structures.

9.2.1.2 SZM for Evaporated Films

The earliest of the structure zone models was proposed by Movchan and Demchishin (Ref. 4) based on observations of very thick evaporated coatings (0.3 mm to 2 mm) of metals (Ti, Ni, W, Fe) and oxides (ZrO_2 and Al_2O_3) deposited at rates ranging from 12,000 to 18,000 Å/min. The structures were identified as belonging to one of three zones (1, 2, 3) based on the ratio T_S/T_M. In this scheme, zone 1 structures ($T_S/T_M < 0.3$) are columnar, consisting of inverted conelike units capped by domes and separated by voided boundaries that are several nanometers wide. They arise from shadowing effects and frozen or very limited adatom motion. Sometimes the structure takes on a "cauliflower-type" appearance. The zone 2 structure ($0.3 < T_S/T_M < 0.45$) is also columnar but with tighter metallurgical grain boundaries that are $\sim$0.5 nm wide. Surface or grain-boundary diffusion apparently plays a role in the evolution of this structure, because the columnar grain size increases with T_S/T_M in accord with the activation energies for these mass-transport mechanisms. Equiaxed grains that grow in size by bulk diffusion characterize zone 3 structures ($T_S/T_M > 0.5$).

Sanders (Ref. 5) suggested a five-zone description for evaporated films based on the dominant physical processes involved. Below the melting point there are three zones paralleling those of Movchan and Demchishin. Thus, in zone 1 ($T_S/T_M < 0.1$) atomic shadowing dominates film growth, resulting in isolated columns. Similar columnar grains but with impervious boundaries resulting from appreciable surface diffusion describe zone 2 ($0.1 < T_S/T_M < 0.3$). In zone 3 ($0.3 < T_S/T_M < 1$) it was observed that accelerated recrystallization, surface faceting, grain growth, and twinning resulted from more rapid surface diffusion coupled with mobile dislocations and grain boundaries.

9.2.1.3 SZM for Sputtered Films

A similar zone scheme was introduced by Thornton (Ref. 6) for sputtered metal deposits but with four zones (1, T, 2, 3). This structural model is based on morphologies developed in 20 to 250 μm thick magnetron-sputtered coatings of Ti, Cr, Fe, Cu, Mo, and Al deposited at rates ranging from 50 to 20,000 Å/min. The exploded view of Fig. 9-2a illustrates the effect of the individual physical processes on structure and how they depend on substrate temperature, the ever-present variable, and inert sputtering gas pressure (P), a new variable that has been introduced. Pressure affects film microstructure through several indirect mechanisms (Ref. 3). For example, if the pressure is increased to the point where the mean free path for elastic collisions between the sputtered atoms and sputtering gas becomes of the order of the source–substrate distance, the oblique component of the deposition flux is increased because of gas scattering; this results in a more open zone 1 structure. On the other hand, a reduction in gas pressure results in increased energetic-particle bombardment which in turn densifies the film. In addition, it has been suggested that increased gas pressure decreases adatom mobilities.

A comparison between zone structures and properties for evaporated and sputtered coatings is made in Table 9-1. In general, analogous structures evolve at somewhat lower temperatures in evaporated films than in sputtered films. Zone 1 structures, which appear in amorphous as well as crystalline deposits, are the result of shadowing effects which overcome limited adatom surface diffusion. In the zone 2 regime, structures are the result of surface diffusion-controlled growth. Lattice and grain-boundary diffusional processes dominate at the highest substrate temperatures, giving rise to the equixed recrystallized grains of zone 3. Of note is zone T, consisting of a dense array of poorly defined fibrous grains, which may be viewed as the transition region between zones 1 and 2. No clear trends distinguish evaporated from sputtered coatings insofar as the temperatures at which structural zone transitions occur.

In contrast to metals, ceramic materials tend to have low hardness at low values of T_S/T_M, indicating that their strength is adversely affected by lattice and grain-boundary imperfections. Ceramics also become harder, not softer, in zones 2 and 3 despite being heated to higher temperatures.

9.2.1.4 Revised and New Structure-Zone Diagrams

In subsequent years portions of existing structure-zone diagrams have been revised and new variables have been suggested in order to create new SZDs. A sampling of these is offered next.

(a)

ISOLATED ZONE 1
GROWTH DEFECTS
SUBSTRATE INDUCED
ZONE 1 STRUCTURE
SHADOWING
PROCESSES
GAS INDUCED
ZONE 1 STRUCTURE
ZONE 2
ZONE T
SURFACE
DIFFUSION
PROCESSES
ZONE 3
RECRYSTALLIZATION &
GRAIN GROWTH
BULK DIFFUSION
PROCESSES
COMPOSITE
STRUCTURE
DIAGRAM
4
3
2
1
0.15
INERT GAS
PRESSURE (Pa)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
SUBSTRATE
TEMPERATURE (T/Tm)

(b)

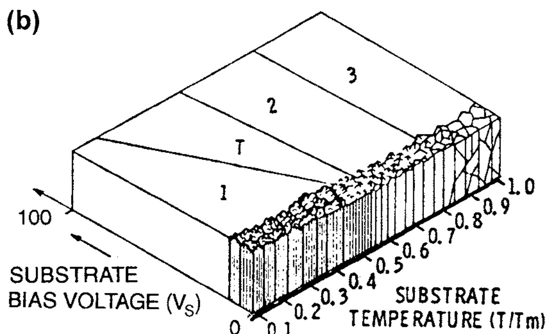


Figure 9-2 (a) Schematic representation of sputtered-film structures showing the superposition of shadowing, surface-diffusion, and bulk-diffusion processes that establish structural zones. (Reprinted with permission from Ref. 6, © 1977 Annual Reviews Inc.) (b) SZM for thick RF-sputtered films showing effects of both ion bombardment and thermal induced mobility. (From Ref. 8. Reprinted with the permission of Dr. Russell Messier.)

Table 9-1
Zone Structures in Thick Evaporated and Sputtered Coatings

Zone	T_S/T_M	Structural characteristics	Film properties
1(E)	<0.3	Tapered crystals, dome tops, voided boundaries.	High dislocation density, hard.
1(S)	<0.1 at 0.15 Pa to <0.5 at 4 Pa	Voided boundaries, fibrous grains. Zone 1 is promoted by substrate roughness and oblique deposition.	Hard.
T(S)	0.1–0.4 at 0.15 Pa, 0.4–0.5 at 4 Pa	Fibrous grains, dense grain boundary arrays.	High dislocation density, hard. High strength, low ductility.
2(E)	0.3–0.5	Columnar grains, dense grain boundaries.	Hard, low ductility.
2(S)	0.4–0.7		
3(E)	0.5–1.0	Large equiaxed grains, bright surface.	Low dislocation density, soft recrystallized grains.
3(S)	0.6–1.0		

Note: (E) refers to evaporated, (S) refers to sputtered.

9.2.1.4.1 Revised SZD for Evaporated Metal Films

A systematic study (Ref. 7) of 10 evaporated elemental films 1000 Å thick has revealed much about the systematics of film grain structure. The different metals were deposited on NaCl and Si_3N_4 substrates at a background vacuum pressure of 4×10^{-8} torr. Electron microscopy revealed the maximum and minimum grain size variation with T_S that is shown in Fig. 9-3a, as well as the grain morphology leading to the SZD pictured in Fig. 9-3b. It is seen that for $T_S/T_M < 0.2$ ($T_M/T_S > 5$) the grains are equiaxed with a diameter of less than 200 Å. Within the range $0.2 < T_S/T_M < 0.3$ some grains larger than 500 Å appear surrounded by smaller grains. Columnar grains make their appearance at $T_S/T_M > 0.37$, and still higher temperatures promote lateral growth with grain sizes larger than the film thickness. Although the same zone classification scheme has been used for both sputtered and evaporated films, it should be noted that the grain morphology differs. Zones I and T (a transition zone) possess structures produced by continued renucleation of grains during deposition and subsequent

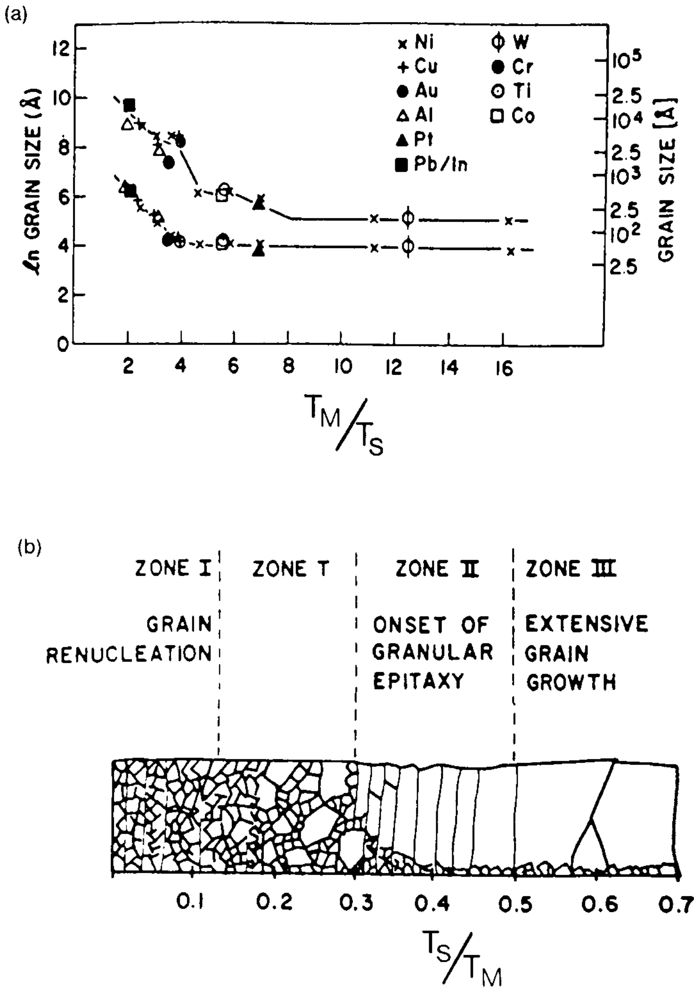


Figure 9-3 (a) Plot of maximum and minimum grain size variation with homologous substrate temperature for 10 different evaporated metals. (b) Zone model for evaporated metal films. (From Ref. 7.)

grain growth. As a result the bimodal grain structure of zone T develops. Zone II structures are the result of granular epitaxy and grain growth. The variation in grain structure in zones I, T, and II presumably arises because different grain boundaries become mobile at different temperatures. In zone I virtually all grain boundaries are immobile, whereas in zone II they

are all mobile. Consequently, at higher temperatures the probability of any boundary sweeping across a grain and reacting to form another mobile boundary is increased. Coupled with enhanced surface diffusion a decrease in porosity results in zone III. Bulk grain growth and surface recrystallization occur at the highest temperatures with the largest activation energies. This is evident in Fig. 9-3a, showing the steep dependence of grain size with T_S for $T_M/T_S < 4$.

9.2.1.4.2 SZD Incorporating the Effects of Ion Bombardment

Messier *et al.* (Ref. 8) proposed the structure zone diagram shown in Fig. 9-2b to describe RF sputtered films. This SZD differs from that presented for magnetron sputtering because the floating bias potential at the substrate (V_S) replaces pressure (P) as a variable. We note, however, that raising V_S enhances the extent of ion bombardment, a situation that can also be accomplished by reducing P. Thus V_S varies inversely with P and we may expect some distortion in the zone 1/zone T domains relative to Fig. 9-2a. Specifically, zone T is widened relative to zone 1 because ion bombardment promotes adatom motion and therefore has the same effect as raising substrate temperatures. While 1 to 3 nm is the columnar grain diameter at the substrate interface, near the film surface it typically increases to several hundred nanometers, a value that depends on overall film thickness. No systematic experimental data have been accumulated to test this revised SZD.

A study by Tang *et al.* (Ref. 9) has revealed that during argon-ion beam-assisted deposition, an ion radiation shadow at the angle of beam incidence produces an interface around initially nucleated crystalline nodules. In the case of evaporated TiO_2 films, either crystalline or amorphous inclined columnar structures grow above the nodules, depending on substrate temperature.

9.2.1.4.3 SZDs for Magnetron Sputtered Films

New structure-zone diagrams have by no means been exhausted. For example, one has been proposed for reactive pulsed-magnetron-sputtered Al_2O_3 (Ref. 10). The major variable in this SZD is oxygen content, which in turn is a complex function of the electrical and chemical conditions within the discharge. Only zone 2 and 3 structures appeared. The former were substoichiometric oxides with columnar grains that had metallic properties, while zone 3 oxides were stoichiometric and ceramic in nature. Between these zones the transition structure was amorphous and contained between 42 and 48 at.% oxygen.

These same authors (Ref. 11) have suggested yet another SZD, this time for closed-field unbalanced magnetron sputtering (Section 5.3.3.2). Coatings of Al, Zr, and W appeared to develop zone 2 microstructures at homologous temperatures (T_S/T_M) lying between 0.15 and 0.45. This compares with a T_S/T_M window between 0.5 and $\sim$0.77 in both the Thornton and Messier schemes. Similarly, zone 3 structures appeared for $T_S/T_M = 0.45$–0.68 while Thornton and Messier suggested a range between $\sim$0.77 and 1. Clearly, lower temperatures are required for equivalent structural development in unbalanced magnetron relative to traditional sputtering. Porous-column formation is evidently suppressed by film compaction effects due to the enhanced ion bombardment of the substrate.

9.2.2 COLUMNAR GRAIN STRUCTURE: THE TANGENT RULE

The columnar grain structure of thin films has been a subject of interest for several decades. This microstructure consisting of a network of low-density material surrounding an array of parallel rod-shaped columns of higher density has been much studied by transmission and scanning electron microscopy. As previouly noted, columnar structures are observed when the mobility of deposited atoms is limited and therefore their occurrence is ubiquitous. For example, columnar grains have been observed in high-melting-point materials (Cr, Be, Si, and Ge), in compounds of high binding energy (TiC, TiN, CaF_2, and PbS), and in nonnoble metals evaporated in the presence of oxygen (Fe and Fe–Ni). Amorphous films of Si, Ge, SiO, and rare earth–transition metal alloys (e.g., Gd–Co) whose very existence depends on limited adatom mobility are frequently columnar when deposited at sufficiently low temperatures. Inasmuch as grain boundaries are axiomatically absent in amorphous films, it is perhaps more correct to speak of columnar morphology in this case. This columnar morphology is frequently made visible by transverse fracture of the film because of crack propagation along the weak, low-density intercolumnar regions. Magnetic, optical, electrical, mechanical, and surface properties of films are affected, sometimes strongly, by columnar structures. In particular, the magnetic anisotropy of seemingly isotropic amorphous Gd-Co films is apparently due to its columnar structure and interspersed voids. A collection of assorted electron micrographs of film and coating columnar structures is shown in Fig. 9-4. Particularly noteworthy are the structural similarities among varied materials deposited in different ways suggesting common nucleation and growth mechanisms.

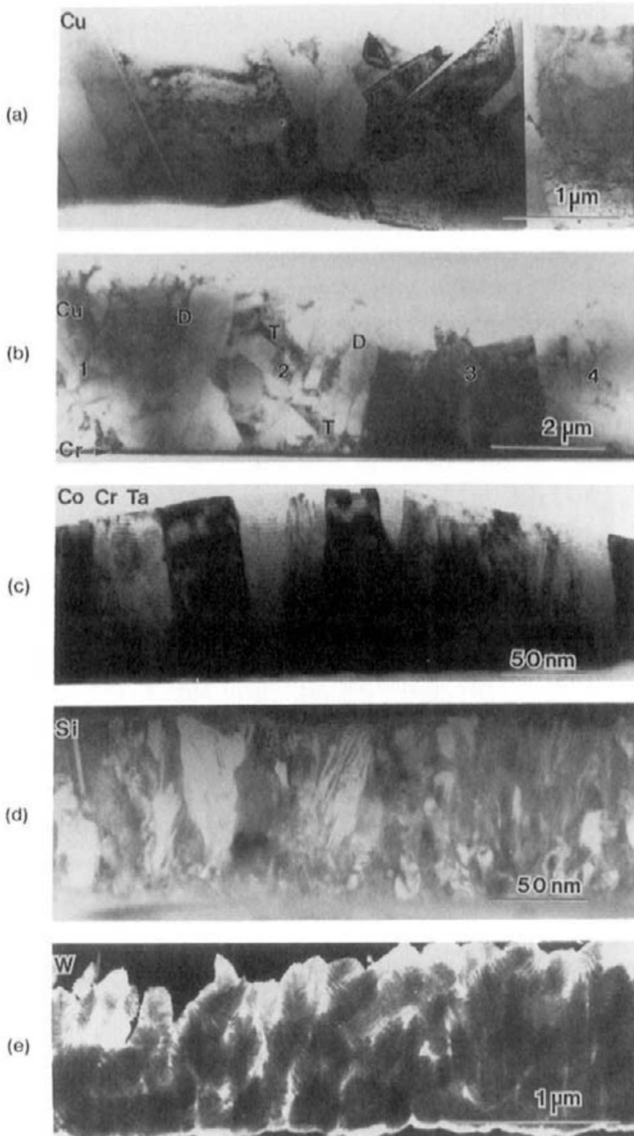


Figure 9-4 Representative set of cross-sectional transmission electron micrographs of thin films illustrating variants of columnar microstructures. (a) Acid-plated Cu, (b) sputtered Cu, (c) sputtered Co–Gd–Ta alloy, (d) CVD silicon (also Fig. 6-23), (e) sputtered W. D = dislocation, T = twin. (Courtesy of D. A. Smith, IBM, T. J. Watson Research Center. From M. F. Chisholm and D. A. Smith, in *Advanced Techniques for Microstructural Characterization*, eds. R. Krishnan, T. R. Anantharaman, C. S. Pande, and O. P. Arora, Trans-Tech Publications, Switzerland, 1988. Reprinted with permission from Trans-Tech Publications.)

An interesting observation (Ref. 12) on the geometry of columnar grains has been formulated as the so-called Tangent Rule expressed by Eq. 9-1. Careful measurements on obliquely evaporated Al films reveal that the columns are oriented toward the vapor source as shown in the microfractograph of Fig. 9-5. Angle β between the column axis and substrate normal is universally observed to be somewhat less than angle α, defined by the source direction and substrate normal. By varying the incident vapor angle over a broad range ($0 < \alpha < 90^\circ$), it was experimentally found that the relation

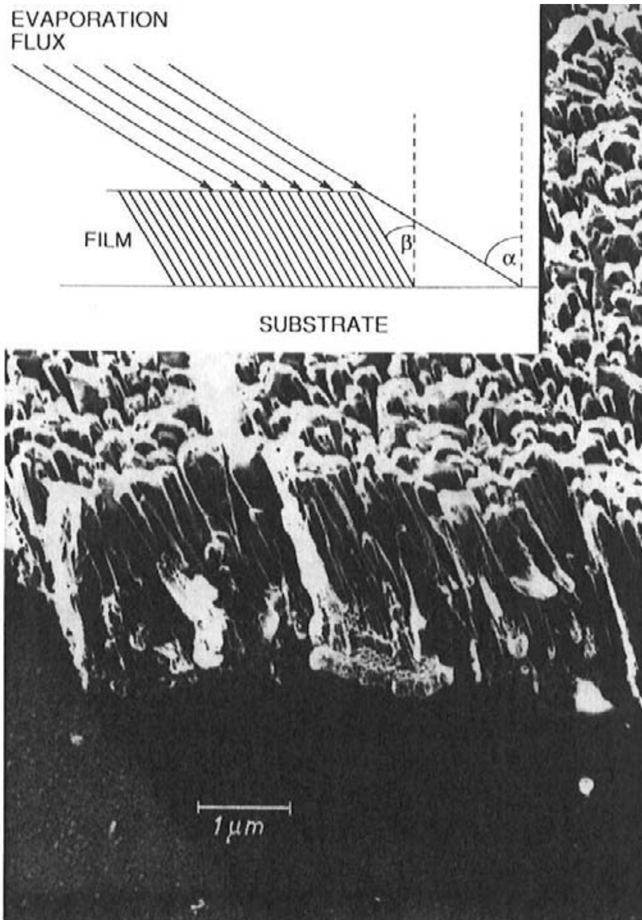


Figure 9-5 Electron micrograph of a replica of a 1 μm thick Al film (Ref. 12). Inset shows deposition geometry that leads to the Tangent Rule.

connecting α and β closely approximated

$$\tan \alpha = 2 \tan \beta. \tag{9-1}$$

The very general occurrence of the columnar morphology implies a simple nonspecific origin such as geometric shadowing which affords an understanding of the main structural features.

Physically, it appears that the reason that columns tilt toward the film normal rather than being exactly parallel to the evaporation flux plane is due to the dynamics of surface diffusion. Quite simply, adatoms tend to migrate directionally a bit further under oblique deposition in an attempt to conserve atomic momentum parallel to the film plane. This idea was advanced by Hara *et al.* (Ref. 13) and reviewed by Abelmann and Lodder (Ref. 14) in connection with oblique evaporation and surface diffusion effects. Interestingly, since surface diffusion is influenced by contaminants (e.g., H_2O, O_2), alteration of column inclination and even changes in the transition temperatures between structural zones may be expected. Indeed, careful analysis of Fe and Co films deposited under controlled conditions has revealed such effects.

In a study by Kuratani *et al.* (Ref. 15), an interesting connection between grain column tilt and the energy of ions bombarding the film was reported. Although it was observed in Cr films, it is sufficiently general to have broader applicability. When atoms deposit without benefit of ion bombardment, grain tilt occurs in accordance with the Tangent Rule geometry. But with ion impact normal to the surface, mobile adatoms receive additional randomized momentum transfer in the film plane. This has the effect of minimizing (see Fig. 4-17) or even eliminating column tilting if the ion energy is sufficiently high.

In Section 9.3.2 we shall see that the Tangent Rule is not only consistent with computer models of structural evolution, but in fact can be derived from a hard-sphere atomic model of the deposition geometry.

Let us now take a closer look at the detailed microstructure of columnar grains. In sputtered amorphous Ge and Si, as well as TiB_2, WO_3, BN, and SiC thin films, an evolutionary development of columnar structures ranging in size from ~ 20 Å to 4000 Å unfolds as outer layers are peeled away. When prepared under low adatom mobility conditions ($T_S/T_M < 0.5$), three general structural units are recognized; nano-, micro-, and macrocolumns, together with associated nano-, micro-, and macrovoid distributions. A schematic of these interrelated, nested columns is shown in Fig. 9-6. It is very likely that the columnar grains of zones 1 and T are composed of such nano- and microcolumns.

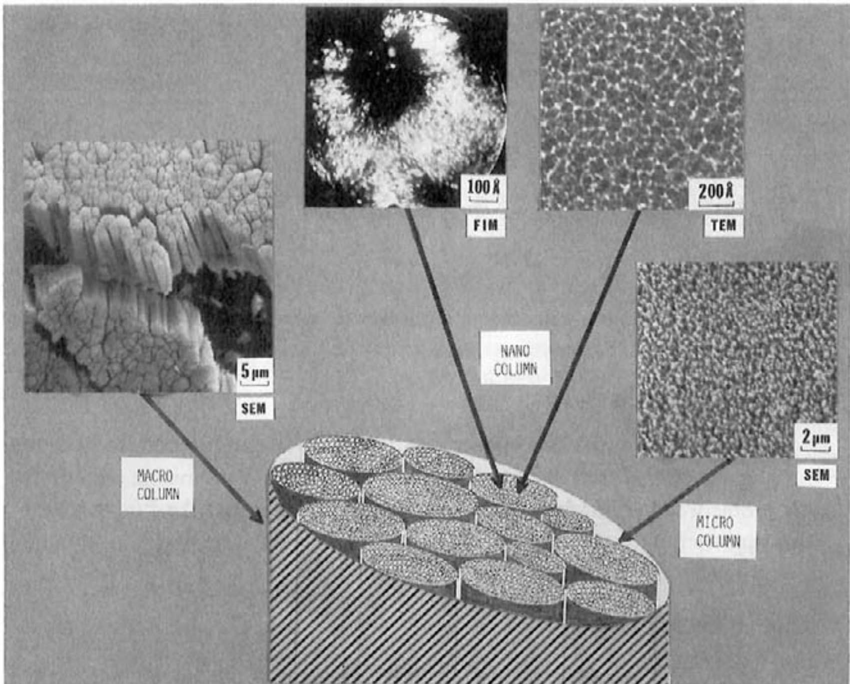


Figure 9-6 Schematic representation of macro-, micro-, and nanocolumns for sputtered amorphous Ge films. (From Ref. 8.)

9.2.3 FILM DENSITY

A reduced film density relative to the bulk density is not an unexpected outcome of the zone structure of films and its associated porosity. Because of the causal structure–density and structure–property relationships, density is expected to strongly influence film properties. Indeed, we have already alluded to the deleterious effect of lowered overall film densities on optical and mechanical properties. A similar degradation of film adhesion and chemical stability as well as electrical and magnetic properties can also be expected. Measurement of film density generally requires a simultaneous determination of both film mass per unit area and thickness. Among the experimental findings (Ref. 16) related to film density are:

1. The density of both metal and dielectric films increases with thickness and reaches a plateau value which asymptotically approaches that of the

bulk density. Depending on material deposition method and conditions, the plateau occurs at different film thickness. The gradient in film density is thought to be due to several causes, e.g., higher crystalline disorder, formation of oxides, greater trapping of vacancies and holes, pores produced by gas incorporation, and special growth modes which predominate in the early stages of film formation.

2. Metal films tend to be denser than dielectric films because of the larger void content in the latter. A quantitative measure of the effect of voids on density is the packing factor p, which is defined as

$$p = \frac{\text{Volume of solid}}{\text{Total volume of film (solid + voids)}}. \quad (9\text{-}2)$$

Typical values of p for metals are greater than 0.95, whereas for fluoride films (e.g., MgF_2, CaF_2) p values of approximately 0.7 are realized. However, by raising T_S for the latter, p can be increased to almost unity.

3. Thin-film condensation is apparently accompanied by the incorporation of large nonequilibrium concentrations of vacancies and micropores. Whereas bulk metals may perhaps contain a vacancy concentration of 10^{-3} at the melting point, freshly formed thin films can have excess concentrations of 10^{-2} at room temperature.

In addition, microporosity on a scale much finer than imagined in zones 1 and T has been detected by transmission electron microscopy phase (defocus) contrast techniques (Ref. 17). Voids measuring 10 Å in size, present in densities of about 10^{17} cm^{-3}, have been revealed in films prepared by evaporation as shown in Fig. 9-7. The small voids appear as white dots surrounded by black rings in the underfocused condition. Microporosity is evident both at grain boundaries and in the grain interior of metal films. In dielectrics a continuous network of microvoids appears to surround grain boundaries. This crack network has also been observed in Si and Ge films where closer examination has revealed that it is composed of interconnecting cylindrical voids. Limited surface diffusion, micro-self-shadowing effects, and stabilization by adsorbed impurities encourage the formation of microporosity. In addition to reducing film density, excess vacancies and microvoids may play a role fostering interdiffusion in thin-film couples where the Kirkendall effect (Section 11.5.4.2) has been observed. The natural tendency to decrease the vacancy concentration through annihilation is manifested by such film changes as stress relaxation, surface faceting, adhesion failure, grain growth and recrystallization, formation of dislocation loops and stacking faults, and decrease in hardness.

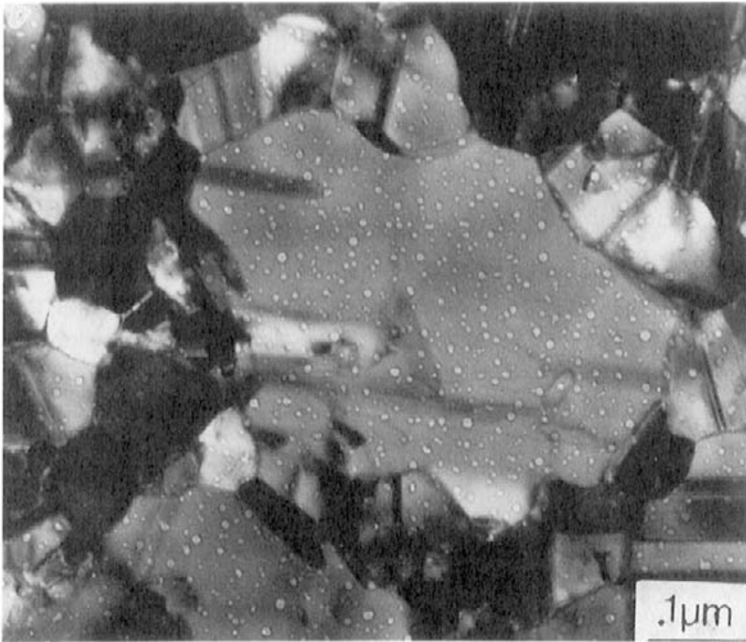


Figure 9-7 Transmission electron micrograph showing microvoid distribution in evaporated Au films. (Courtesy of S. Nakahara, AT&T Bell Laboratories.)

9.3 COMPUTATIONAL SIMULATIONS OF FILM STRUCTURE

9.3.1 SCOPE

In this section we shall describe attempts to simulate the features of film morphologies and grain structures by computational methods. Computer modeling of the atomic-scale processes of film growth from the vapor phase has been a research topic for over a quarter of a century. Structures that develop normal to the film plane have been emphasized with columnar growth a particular focus. As we shall see there is a continuum of computer approaches and techniques that differ depending on the film growth processes being modeled. Among the latter we shall consider (or reconsider) structures derived from vapor-phase deposition, epitaxial growth, ion-beam bombardment, and cluster-beam deposition. In simulating these processes the major variables that must be accounted for are the energy, size, and

composition of the depositing species, and the temperature and nature of the substrate.

The energy of the depositing species often establishes the division as to what will be modeled and how. For low energies the equilibrium structural morphology and arrangement of atoms is evaluated through a combination of geometric and probabilistic considerations. The *Monte Carlo* stochastic lattice–gas and solid-on-solid simulations are of this type (Ref. 18). On the other hand, a *molecular dynamics* approach is used for modeling the effects of impinging particles having higher energy; here ultrashort-time kinetic interactions between the particle and surface as well as atomic displacements are often of interest. Both kinds of simulations reflect the spirit of statistical mechanics because they attempt to model macroscopic behavior by summing over the collective dynamical behavior of large numbers of individual atoms.

Reconsidering the Tangent Rule is an excellent place to start because its analytic derivation complements the more complex ballistic-aggregation computer simulations of columnar film growth considered later.

9.3.2 The Tangent Rule Revisited

Machlin (Ref. 19) has derived an elegant statistical-geometric analysis for understanding the physical basis of the Tangent Rule. The model assumes hard sphere atoms of the vapor phase impinging on a substrate at an angle α measured from the surface normal. Deposition occurs on a close-packed surface lattice as shown in Fig. 9-8. As soon as the first atom lands in site a, the probability of an atom subsequently depositing at adjacent site w (on the "windward" side, i.e., facing the wind) exceeds the probability of landing in the pocket of site l (on the "leeward" side). Quite simply, the atomic asperity casts a shadow of the incident particle "illumination" that hinders occupation of shadowed leeward sites. The relative site occupation probabilities are proportional to the ratio of the indicated line lengths L_l and L_w, because it is assumed dashed adatoms that impinge on top of atoms A and B will respectively fall into l and w pockets. Only total length $L_w + L_l$ need be considered because it corresponds to a unit pair of one windward and one leeward site. The probabilities p_w and p_l, for condensation on w and l sites, are then given by $p_w = L_w/(L_w + L_l)$ and $p_l = L_l/(L_w + L_l)$, respectively.

To start the calculation we note that the shortest distance between rays 1 and 3 is $6r \cos \alpha$, where r is the atomic radius. By simple geometry, distance CD, passing through the atom centers, is equal to $3r/\cos 30^\circ$. For $\alpha > 30^\circ$, the shortest distance between incident rays 2 and 3 is $L_w = (3r/\cos 30^\circ) \cos(\alpha - 30^\circ)$. Therefore, $L_l = 6r \cos \alpha - (3r/\cos 30^\circ) \cos(\alpha - 30^\circ)$;

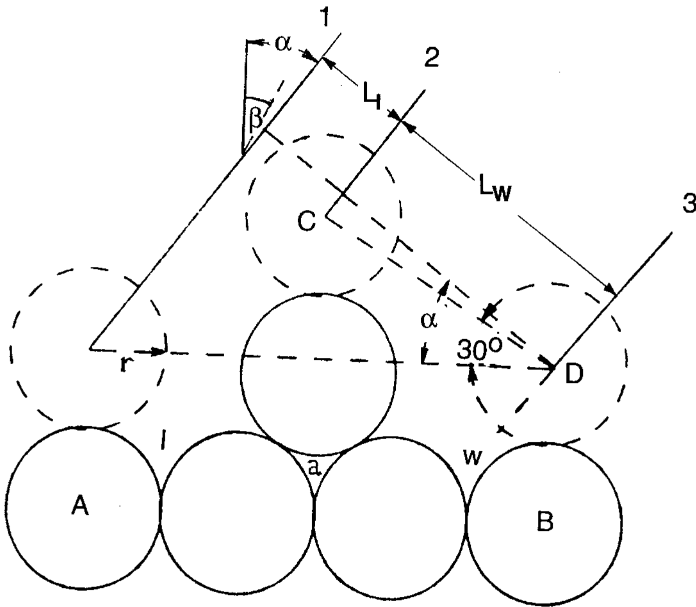


Figure 9-8 Hard-sphere geometric model for the development of columns during oblique deposition. (After Ref. 19. Reprinted by permission.)

however, if $\alpha < 30°$, $\cos(30° - \alpha)$ is the appropriate term. The probabilities p_w and p_l, for occupying w and l sites, are then related by

$$\frac{p_w}{p_l} = \frac{L_w}{L_l} = \frac{(3r/\cos 30°)\cos(\alpha - 30°)}{6r\cos\alpha - (3r/\cos 30°)\cos(\alpha - 30°)} = \frac{3\cos\alpha + (3)^{1/2}\sin\alpha}{3\cos\alpha - (3)^{1/2}\sin\alpha}. \quad (9\text{-}3)$$

This result holds for $\alpha < 60°$, whereas for $\alpha > 60°$,

$$\frac{p_w}{p_l} = \frac{3\cos\alpha + (3)^{1/2}\sin\alpha}{1 - \sin\alpha} \quad (9\text{-}4)$$

because the shortest distance between rays 1 and 2 is $2r - 2r\sin\alpha$, while L_w remains unchanged.

Finally, we note that during deposition the top adatom adlayer of the growing columns shifts by an average amount proportional to $p_w - p_l$. The angle β made by the column with respect to the surface normal is given by

$$\tan\beta = (p_w - p_l)/3^{1/2}, \quad (9\text{-}5)$$

because condensation of each atom causes a horizontal displacement of the

adatom layer equal to r and a vertical displacement of $3^{1/2}r$. Upon substitution of the values for p_w and p_l we obtain

$$\tan \alpha = 3 \tan \beta \qquad (9\text{-}6)$$

instead of the widely accepted $\tan \alpha = 2 \tan \beta$. Despite the discrepancy, this equation yields better agreement with computer models for estimating β when $\alpha = 45°$.

9.3.3 Monte Carlo Simulations

Most of the crystal growth studies to date have been carried out by Monte Carlo or random "ballistic aggregation" simulations which superimpose the statistics of atomic placement on a geometric array of surface sites (Ref. 18). This includes the simple straight-trajectory, hard disk, stochastic lattice-gas, and solid-on-solid models. These are distinguished by limited impact mobility in the first model, but greater thermal mobility, i.e., adatom surface diffusion, in the latter three. In an intuitively simple simulation of film deposition, individual hard spheres (atoms) are serially "evaporated" onto a growing film at a fixed angle. The spheres are allowed to relax following impingement into the nearest triangular cradle formed by three prior deposited atoms, thus maximizing close atomic packing.

In other simulations the region above the substrate is assigned a space-lattice structure so that depositing atoms can only assume fixed positions in space. The structure of the substrate essentially defines the space-lattice of the film so that such modeling ensures well-oriented epitaxial growth. For low film-growth rates, impinging atoms are assumed to be statistically independent. First, a lattice site is randomly selected. If the site is vacant a vapor atom can be adsorbed. But if it is occupied, the atom desorbs or hops by concerted diffusional exchange, until a surrounding vacant site is filled. Activation energies (E) govern these processes so that relevant Boltzmann factors, (i.e., $\exp - (E/k_B T)$, are normally incorporated into the model. Despite its wide use, lattice-space modeling has serious limitations. In essence the constraint to growth provided by the assumed lattice geometry limits growth to a small subset of possible atomic configurations.

Despite their limitations, Monte Carlo computer simulations (Ref. 20) have contributed greatly to our understanding of the origin of columnar grain formation and the role played by shadowing. By serial deposition at angle α the structural features shown in Fig. 9-9 were obtained. The simulation shows that limited atomic mobility during low-temperature deposition reproduces film morphologies observed experimentally. As examples, film density decreases with increasing α, high-density column-like

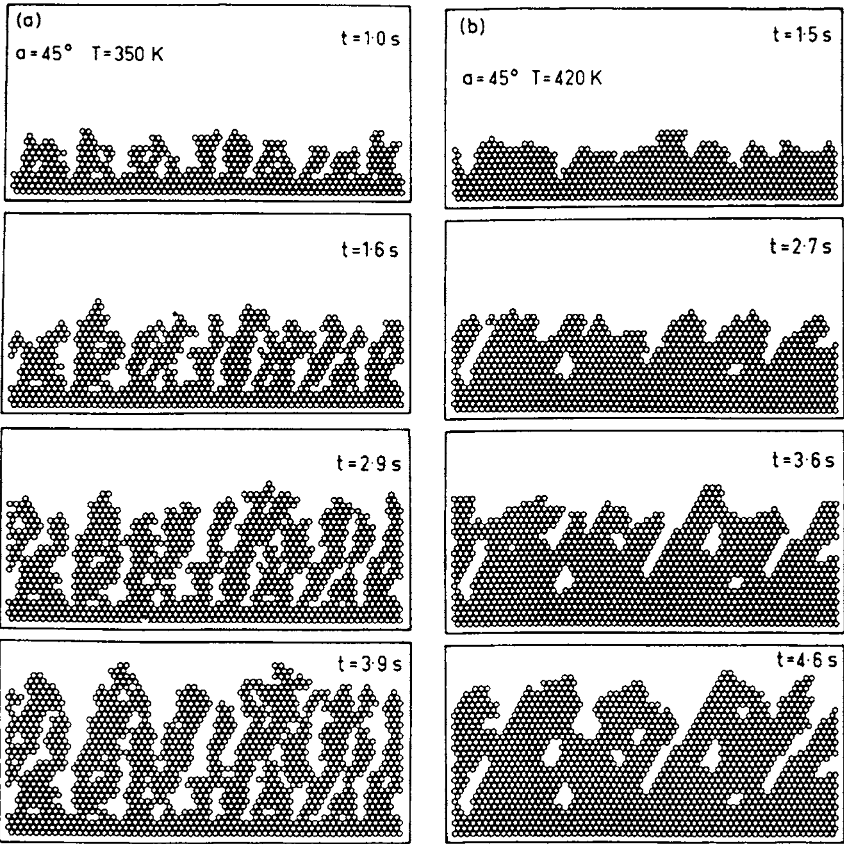


Figure 9-9 Computer-simulated microstructure of Ni film during deposition at different times for substrate temperatures of (a) 350 K and (b) 420 K. The angle of vapor deposition α is 45°. (After Ref. 20.)

regions appear at angles for which $\beta < \alpha$, and film densification is enhanced at elevated temperatures. Over a wide range of deposition rates, the simulated packing density (p) of the film remains roughly constant as a function of temperature. Then at the point where surface diffusion becomes appreciable, p rises rapidly over a short temperature interval and finally saturates at a value approaching unity at elevated temperature. Furthermore, the low-temperature value of p is higher at smaller α angles (Ref. 21). Lastly, the column orientations agree well with the Tangent Rule.

The evolution of voids occurs if those atoms exposed to the vapor beam shield or shadow unoccupied sites from direct impingement, and if post-

impingement atom migration does not suceed in filling the voids. This self-shadowing effect is thus more pronounced the lower the atomic mobility and extent of lattice relaxation. An important consequence of the columnar-void microstructure is the instability it engenders in optical coatings exposed to humid atmospheres. Under typical evaporation conditions ($\sim 10^{-6}$ torr, $T_S = 30–300°C$, and deposition rate of 300–3000 Å/s) dielectric films generally develop a zone 2 structure. Water from the ambient is then absorbed throughout the film by capillary action. The process is largely irreversible and alters optical properties such as index of refraction and absorption coefficient. Moisture-induced degradation has plagued optical films for many years. A remedy for this problem is to employ ion bombardment (Chapter 5), a process that serves to compact the film structure.

In addition to intercolumnar channels and void distributions that lower p, an important concern in optical coating structures is the presence of nodules. Believed to arise from particles "spit" from sputtering targets or electron-beam-heated materials that cannot tolerate thermal shock, nodules serve to increase the absorbence of otherwise transparent films. By incorporating substrate surface asperities or particles consisting of a collection of "atoms," the effect of nodules on film structure has been modeled (Ref. 22).

9.3.4 MOLECULAR DYNAMICS SIMULATIONS

In contrast to the Monte Carlo determinations of equilibrium structures, molecular dynamics (MD) simulations place a greater emphasis on the motion and physical behavior of the involved atoms. An attempt is made in MD to distinguish between specific atomic species, i.e., metals and semiconductors, by employing a potential energy function ($U(r)$) that describes their pairwise interaction. The two-body Lennard–Jones potential, i.e.,

$$U(r) = 4U_0[(a/r)^{12} - (a/r)^6], \tag{9-7}$$

is often chosen for the purpose since it displays an attractive component (second term) when atoms are far apart, but strong repulsion (first term) when they approach too closely. Terms U_0, a, and r represent the potential well depth, lattice dimension, and distance between atoms, respectively, and are selected to model the behavior of specific materials. Initial positions and momenta are assigned to the starting atoms. The many-body system is allowed to move for a small time and new forces, positions, and momentum values are recalculated at convenient intervals. In this way atom trajectories are tracked through the very short stages of vibration, bond breaking, and bond formation. These complex events are increasingly being modeled through incorporation of quantum-mechanical interactions so as to more

accurately determine the forces on atoms and electrons and their resulting velocities. In contrast to Monte Carlo modeling where film deposition occurs over a time scale of seconds, molecular dynamics simulations deals with picosecond events, and whereas $\sim 10^6$ atoms can be deposited in the former, the computational complexity of MD limits the simulation to $\sim 10^4$ atoms. Specific examples illustrating the use of MD follow.

9.3.4.1 Ion Bombardment of Films

Molecular dynamics modeling is particularly instructive in depicting the otherwise unobservable instant when an ion impacts a growing film. The structure shown in Fig. 9-10 (top), modeled by Müller (Ref. 23), illustrates an energetic 100 eV Ar ion impinging on a rough Ni surface. Atoms 1, 2, and 3 are of particular interest because the overhang they lie on would appear to enclose a void with further Ni atom deposition. Within ~ 1 ps of impact these atoms sputter directionally in a way that enhances adatom surface diffusion. After another ~ 10 ps a further collision-cascade sequence and surface relaxation ensue. In the process the protruding ledge disappears, lessening the probability of entrapping another void. Interestingly, the central void also decreased in size. All of these effects are consistent with the well-known densification of films during ion bombardment.

Extending the effects of ion bombardment, MD has been employed (Refs. 21, 24) to model sputtering effects during film growth. Simulation results include:

1. Greater structural densification occurs as the kinetic energy of atoms incident on a growing film increases.
2. The film density rises as the ratio of the ion flux to vapor flux increases. This effect is enhanced the higher the ion energy.
3. Intrinsic tensile stress develops in sputtered nickel films, particularly when the structure is rich in small closed voids that enable short-range attractive interatomic forces to act effectively. At higher ion energies the tensile stress decreases. In contrast, a more recent MD simulation (Ref. 24) of stress in sputtered inert-gas films revealed the development of compressive stress.

9.3.4.2 Cluster Beam Deposition

Shortly after the physical basis of ionized cluster beam (ICB) deposition (Section 5.5.5) was demonstrated, the structure of the resulting films was simulated by MD methods (Ref. 25). Instead of single atoms, clusters of

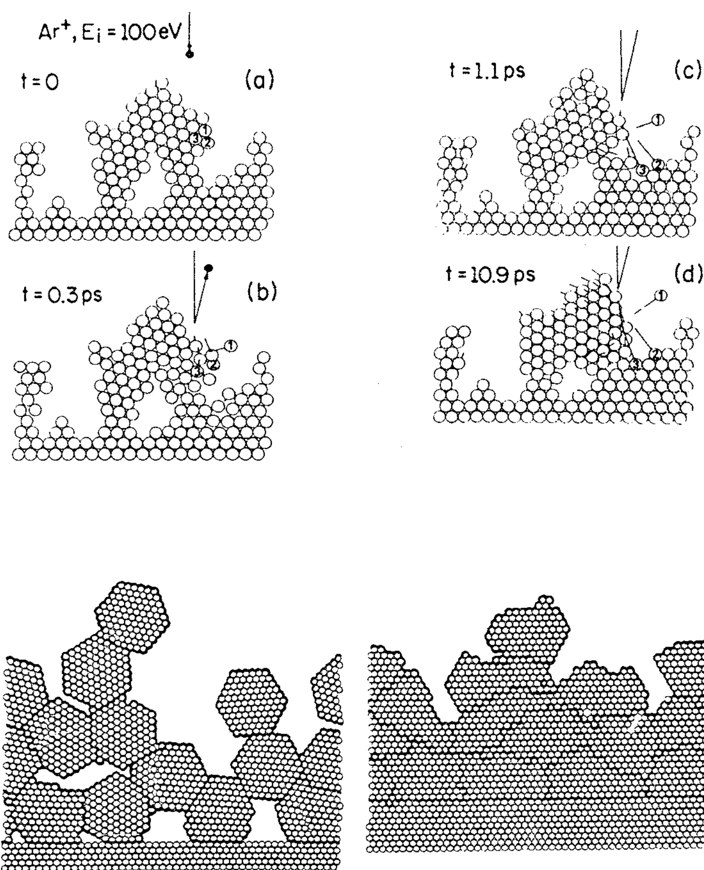


Figure 9-10 (Top) Molecular dynamics simulation of a Ni film deposited at 0 K at various times (t) after bombardment by a 100 eV Ar^+ ion. Atomic displacements are indicated by straight line segments. (From Ref. 23.) (Bottom) Film microstructure resulting from impingement of 2-D clusters consisting of 91 atoms of aluminum, a slice of a 700 atom 3-D cluster of Al. Left: Cluster kinetic energy is 1 keV. Right: Cluster kinetic energy is 1.5 keV. (From Ref. 25.) (Reprinted with the permission of the author.)

many atoms impacting a substrate was modeled. Simulations of film growth due to successive cluster impacts are shown in Fig. 9-10 (bottom) as a function of increasing incident kinetic energy. For impact energies less than 1 keV the cluster morphology does not dramatically change and porous films deposit. At 1.5 keV the cluster melts and flattens, producing a high-quality interface. Cluster impact at the still higher energies of 3 keV generates a considerable amount of atomic disorder and substrate damage

that does not anneal away. However, complete breakup of the cluster was never seen, nor was there any significant enhancement of surface diffusivity.

9.3.5 Analog Simulations of Film Growth

If Monte Carlo and molecular dynamics modeling of film structure may be viewed as the *digital* manipulation of individual atoms, then the use of continuum mathematics and geometry to describe the behavior of huge numbers of atoms may be termed *analog* simulation. The cone growth model due to Messier and co-workers (Refs. 26, 27) is one such analog simulation that provides an interesting way to view the cross-sectional morphology of thin films. In this simulation of film growth, structural development is viewed as a competition between evolving conelike clusters. The cone seeds nucleate with apexes at the substrate interface and grow in a geometrically programmed (fractal) way. As they widen toward the film growth surface, some cones die while others grow at their expense. This morphological evolution is schematically indicated in Fig. 9-11a and seen in

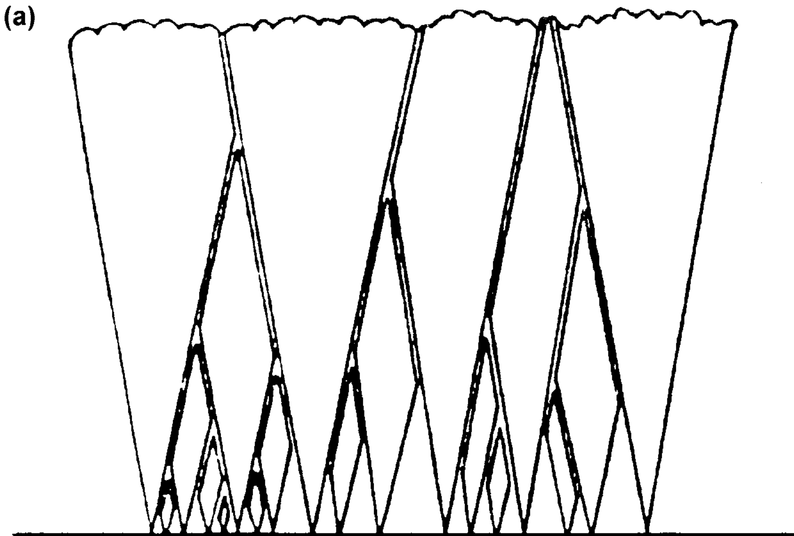


Figure 9-11 (a) Simulation of film morphology by cone growth. (From R. Messier, *J. Vac. Sci. Technol.* **A4**, 490 (1986). Reprinted with the permission of Dr. Russell Messier.) (b) Cross-sectional microstructure of a 100 μm thick polysilicon coating. (From D. Schwarcz and M. Ohring, unpublished research.) (c) Simulation of nodular growth. (After Refs. 26, 27. Reprinted with the permission of Dr. Russell Messier.)

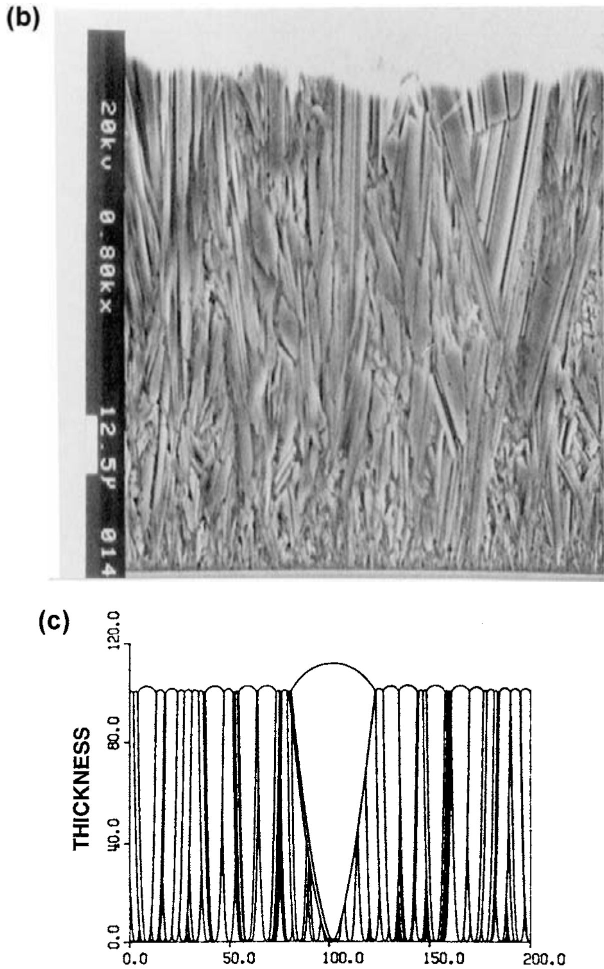


Figure 9-11 Continued.

the grain structure of a polysilicon coating deposited by chemical vapor deposition (Fig. 9-11b). Such a picture of growth is also consistent with observations that film roughness generally increases as some power of the deposition time.

To quantify models for the evolving cone geometry, three issues are addressed in the computation: (1) the location and properties of seeds, (2) the growth rule for cones, and (3) the selection of which cones survive and continue to expand after their intersection and combination. We shall

consider the case where a fixed number of pointlike seeds are randomly distributed across a substrate line. Each seed grows into a power-law cone via the equation $y = Ax^n$, where y and x are the vertical and horizontal dimensions of the cone, and A is a constant. Growth exponent n is also constant, but to better simulate the randomness of the growth process, n values for different cones are assumed to have a Gaussian distribution centered about a mean value n_0 with a standard deviation σ. Thus for the linear cones of Fig. 9-11a, $n = 1$. Parabolic cones develop for $n = 2$ and narrower cones evolve for $n > 2$. Cone caps are assumed to be circular with a radius proportional to cone width. Finally, when cones intersect it is assumed that the survival probability of a cone overtaking its competitor depends linearly on n, with wider cones (small n) being favored. In addition, a nodule-like defect emerges as the consequence of a bimodal distribution of n values (Fig. 9-11c). Other film structures reflecting renucleation of cones during film growth have also been modeled. Despite the success of the cone growth model in developing recognizable morphologies, it is not rooted in basic physics, cannot distinguish the behavior of specific atoms, and cannot predict void distributions or transitions between zone structures.

In other analog simulations film-growth dynamics in CVD (Ref. 28) as well as sputtered films (Ref. 29) has been modeled in a way that incorporates surface diffusion and capillarity effects. Under deposition, small sinusoidal perturbations on the substrate were predicted to evolve into columns that resemble corrugations seen on the surface of closely pleated fabric.

9.4 GRAIN GROWTH, TEXTURE, AND MICROSTRUCTURE CONTROL IN THIN FILMS

9.4.1 INTRODUCTION

Thus far the chapter has focused on film morphology issues (e.g., form, size, and shape) in the columnar structures that develop normal to the film plane. But very little has been said of grain morphology viewed in the film surface. In this section we shall explore how the grain structure of films seen in plan view evolves during both deposition and postdeposition processing. The issues raised here are applicable to all polycrystalline thin films irrespective of application. However, our major interest in grain microstructure stems largely from reliability concerns about metallizations and metal films important in microelectronics. This has provided the incentive for studies of contact and interconnect metal microstructures both in blanket

films and in films confined to line and fill trenches and vias. Two main topics, namely, grain size and film texture, will be addressed with a particular focus on factors that promote either stability or change in these structural attributes. Grain size is our first consideration.

9.4.2 Grain Growth in Films

We start with the widely known fact that when solids are heated, grains grow. Assuming grain boundary (GB) segments have a radius of curvature (R), the velocity of boundary motion (v) is related to the rate at which R changes with time t, or dR/dt. But the boundary moves when atoms hop across the interface between grains in an effort to lower the GB free energy per unit area (γ_{GB}). As shown previously in Section 7.4.5.1, the associated driving force or chemical potential is proportional to γ_{GB}/R so that $dR/dt \sim \gamma_{GB}/R$. Upon integration, the well-known dependence

$$R^2 - R_0^2 \simeq kt^{1/2} \tag{9-8}$$

emerges, where R_0 is the initial grain radius and k, which has absorbed γ_{GB}, is a thermally activated constant. Parabolic grain-growth kinetics is thus predicted and usually observed in bulk solids. Importantly, grain growth is *normal*, which means that the shape of the grain size distribution function does not change with time.

Unlike bulk where only GB free energy is involved, grain growth in thin films is generally more complex. This subject has been extensively modeled over the years by Srolovitz (Ref. 2) and Thompson (Refs. 30, 31). To simplify matters, we consider a film consisting of polygonal grains oriented as columns, i.e., grain boundaries normal to the substrate plane. Now free energies associated with the substrate interface (γ_{fs}) and upper film surface (γ_{fv}), as depicted in Fig. 9-12, play important roles. Following Thompson we may therefore assume that all of these energy densities, or forces per unit area, contribute additively to the overall boundary motion according to the equation

$$dR/dt = M\{(\langle\gamma_{fs}\rangle - \gamma_{fs})/d + (\langle\gamma_{fv}\rangle - \gamma_{fv})/d + \gamma_{GB}(1/\langle R\rangle - 1/R)\}. \tag{9-9}$$

The quantities in angle brackets are average values within the film of thickness d, and R is measured in the film plane. Physically, this equation links the time change in R to the *mean* energy density residing in atoms located at surfaces, interfaces, and GBs. Surface free-energy differences from average levels throughout the film serve as the driving force for grain growth. The quantity M is defined as the mobility and embodies the atomic kinetic factors involved in GB migration. In connection with M it is worth

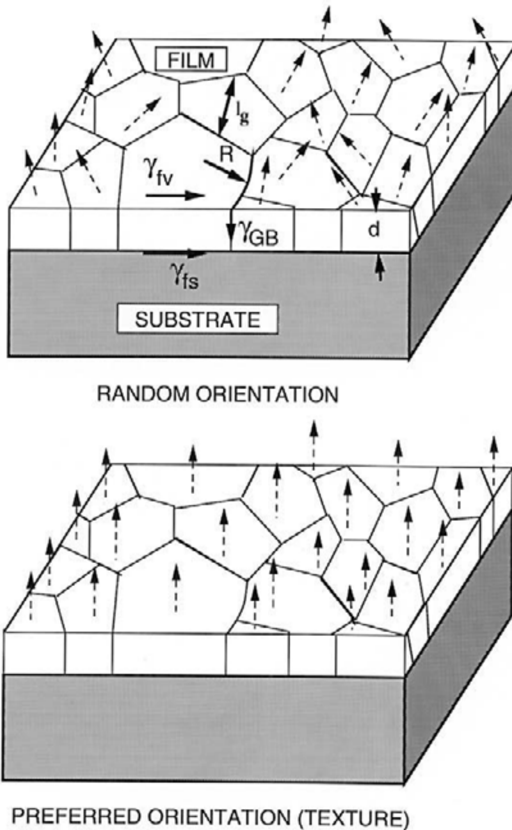


Figure 9-12 (Top) Model of a polycrystalline thin film consisting of randomly oriented polygonal grains. Surface energies associated with the substrate interface, grain boundary and upper film surface are shown. (After Ref. 30.) (Bottom) Same film displaying preferred orientation or texture. *Note:* Dashed arrows represent a measure of crystallographic orientation.

mentioning that the mobility of particles (atoms, ions, electrons) is normally defined as a velocity per unit force. Furthermore, we may view Eq. 9-9 as an application of the Nernst–Einstein equation (Eq. 1-35); on this basis $M = D/RT$, where D is the diffusion coefficient and RT has the usual meaning.

In applying Eq. 9-9 it is assumed (Ref. 31) that γ_{fs} and γ_{fv} are dependent on crystallographic orientation. Strain energy associated with film adhesion to the substrate as well as thermal stresses due to temperature differences is also considered in the analysis. The origins of film texture can then be

imagined because minimization of total film energy will cause selective grains to either shrink or expand. Computer modeling of the resulting grain growth and distribution of grain sizes in thin films has led to the following predictions and physical observations:

1. Normal grain growth is not expected in thin films. Instead, *abnormal* grain growth is predicted. This is often manifested by a bimodal grain size distribution caused by secondary grain growth in which some grains may grow excessively large (e.g., by a factor of $\sim 10^2$) relative to surrounding ones. An interesting example of abnormal grain growth in 1 μm thick electrodeposited copper films has been reported by Harper *et al.* (Ref. 31). Original grains 0.05 to 0.1 μm in size grew to several microns within 9 h at 21°C, while lowering the electrical resistivity and compressive stress in the process. A reduction in the grain boundary interfacial energy coupled with elimination of boundary pinning sites was suggested to be the cause of the "self-annealing."
2. Abnormal grain growth leads to an evolution in the average crystallographic orientations of grains. The effects are more pronounced the thinner the film.
3. At elevated deposition temperatures where low-strain growth occurs, surface and interfacial-energy minimization dominates grain growth. Under these conditions close-packed surface textures evolve, e.g., (111) in FCC metals and diamond-cubic elemental semiconductors.
4. Grain growth in films often stagnates when the grain size is 2 to 3 times the film thickness. Surface grooves where grain boundaries meet the film surface are the apparent cause of stagnation. Physically, boundary velocities go to zero when the driving forces for motion fall below critical levels. Solute segregated at grain boundaries often acts as a drag that inhibits growth.

9.4.2.1 Microstructural Evolution at Grain Surfaces

Recent experimental studies of the structure and topography of Al films sputtered on SiO_2 at 25°C have yielded additional information on the size and shape of grains and subgrain features. By employing transmission electron microscopy (TEM) on films 0.1 to 1 μm thick, imaged in cross section, it was found (Ref. 32) that the columnar morphology is maintained during film growth (Fig. 9-13, left). These results are roughly consistent with Fig. 9-3b, which shows the onset of columnar grains at about the same T_S/T_M value. Furthermore, in TEM plan view (Fig. 9-13, right) the grain size (l_g) varied with film thickness (d) as $l_g \sim d^{0.9}$; both results qualitatively agree with predictions from simulations, e.g., cone growth (Section 9.3.5), and

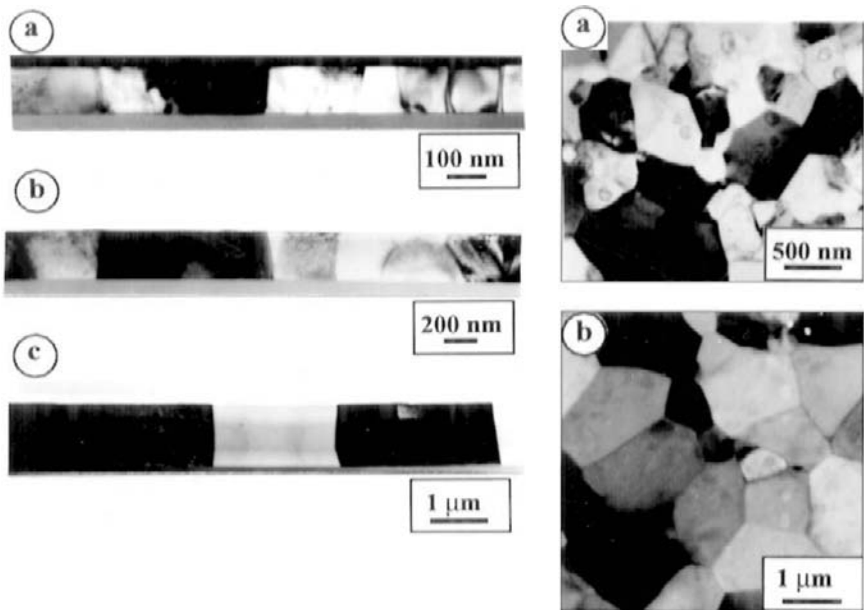


Figure 9-13 (Left) Cross-sectional TEM view of (a) 0.1 μm, (b) 0.3 μm, and (c) 1 μm thick Al films showing that a columnar grain morphology is maintained during film growth. (Right) Plan view TEM micrographs of (a) 0.3-μm and (b) 1-μm thick Al films. The grains are columnar and their size increases with thickness. (Courtesy of J. E. Sanchez. From Ref. 32.)

suggest a dynamic scaling of structural features. This was additionally confirmed by atomic force microscopy of the Al film surface roughness. Measurements were similar to those described for film nucleation (Section 7.5.4), except now the films are continuous. When the roughness topography was Fourier analyzed according to spatial frequency, three structural regions were identified. Variations at the lowest frequencies or longest wavelengths are measures of the mean grain size. Successively higher frequencies revealed smaller features associated with hillocks, grain boundary grooves, and ridges within grains. For example, 0.6 μm thick films have a grain size of roughly 1 μm and ridge spacings are less than 1000 Å in size.

9.4.3 Energetics of Microstructural Change

Potential mechanisms of energy release that can alter microstructure in thin films are shown in Fig. 9-14 together with typical energies (in

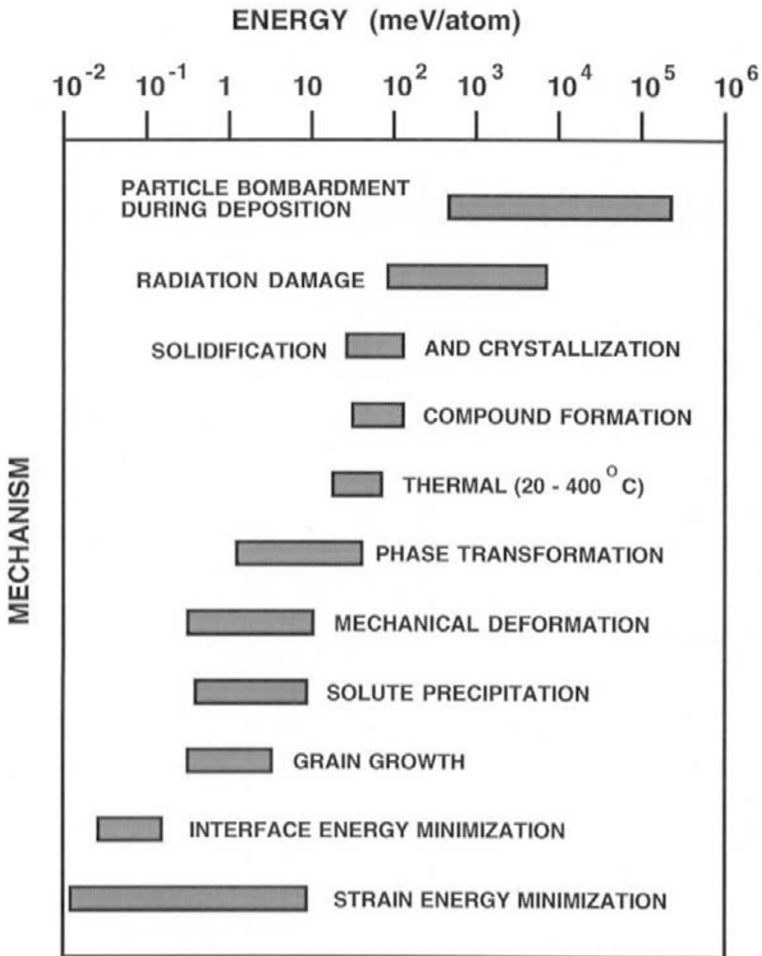


Figure 9-14 Sources of energy available to control thin-film microstructure. (After Ref. 33. Courtesy of J. C. M. Harper, IBM, T. J. Watson Research Center.)

meV/atom) for each. Following Harper and Rodbell (Ref. 33), the thumbnail calculations that follow estimate the magnitude of the energies for a few of the listed mechanisms. This is an instructive exercise and an essential first step in predicting the possibility of microstructural evolution in thin films.

1. *Thermal energy.* The available thermal energy is simply $k_B T$ and in the range 20–400°C it varies between 25 and 58 meV/atom. Uniform thermal

energy cannot induce anisotropic microstructures, but does serve to activate many of the mechanisms.

2. *Grain growth.* The energy released is the product of γ_{GB} and the area of eliminated boundaries. Completely surrounded cubic grains, initially of edge length l_g, are assumed to be much smaller than the volume of the final grains. The energy density difference is then essentially given by $3\gamma_{GB}/l_g$, where the factor of 3 arises from six square surfaces each shared by two grains. Starting with $l_g = 0.1\ \mu$m, which grows to 1 μm, and $\gamma_{GB} = 0.5\ \text{J/m}^2$ for a high-angle boundary in Cu, the difference in energy density is about 15 MJ/m^3 or 1.1 meV/atom.

3. *Interface energy minimization.* We have already seen in the previous section that the (111) texture apparently stems from energy minimization at the film top surface and substrate interface. In the case of Cu, where $\gamma_{fv} \sim 1.5\ \text{J/m}^2$ and assuming $\gamma_{GB} \sim \gamma_{fs}$, then for a 1 μm thick film the energy density is 0.15 meV/atom. But if the film is only 0.1 μm thick, 10 times more energy is available to control texture.

4. *Grain growth coupled to solute precipitation.* When grain boundaries sweep through grains that are supersaturated with respect to solute, precipitation is facilitated at the heterogeneous GB sites. This phenomenon is known as discontinuous precipitation. The magnitude of the energy change in this process is primarily dominated by the entropy difference between configurations of dissolved vs precipitated solute. This corresponds to an energy difference of $k_B(T_s - T_p)C \ln C$ where k_B is the Boltzmann constant, T_s and T_p refer to the solution and precipitation temperatures, respectively, and C is the solute concentration (see Eq. 7-5 for comparison). Typical energies of 7 meV can be released over small volumes so that discontinuous precipitation can serve as a source for grain growth and texture change. This is the case in thin Cu–Co alloy films containing 0.4–8.6 at.% Co which exhibit a strong (111) texture when coevaporated (Ref. 34). After annealing at 450°C for 2 h a strong (200) texture develops and many grains are 10 times the film thickness. Electron microscopy has revealed significant precipitation and grain growth even at 200°C. Furthermore, a desirable electrical resistivity decrease accompanies such microstructural change.

As a general rule, energy release for mechanisms higher in the indicated order of Fig. 9-14 is capable of driving change for mechanisms lower in the order. Thus thermal energy release and solute precipitation can both trigger grain growth.

9.4.3.1 Stable Grain Structures

Having just treated the energetics of microstructural evolution it is instructive to consider the limits of change in a film composed of columnar

grains (Ref. 35). To establish the equilibrium conditions that apply we must first enumerate the forces that act on grain boundaries. When the forces are in equilibrium a stable grain structure results. Some of the GB forces per unit area (or energy densities) were given earlier. For example, the force due to *grain-boundary energy* is equal to $2\gamma_{GB}/l_g$, where l_g is the grain diameter. Likewise, the *surface* and *interface energy* variations are $\Delta\gamma_{fv}/d$ and $\Delta\gamma_{fs}/d$.

Among the forces that restrain GB motion are those due to grain grooving and interaction with second-phase particles. In GB grooving the restraint stems from the V-shaped notch that effectively pins the boundary. The *grooving* force, a product of γ_{fv} and the surface curvature, has a magnitude given by $\gamma_{GB}^2/d\gamma_{fv}$. To obtain this form, we make use of the force balance at the groove notch, i.e., $2\gamma_{fv}\cos\theta = \gamma_{GB}$ ($\theta = \frac{1}{2}$ notch angle). Second-phase particles or precipitates, known for a long time to impede grain growth, deserve special attention. When the planar boundary encounters a spherical precipitate of radius r, the latter restrains the motion of the former. A cusp forms in the boundary plane such that the GB becomes attached to the precipitate along a circular contact perimeter. The perimeter length can vary from practically zero to a maximum of $2\pi r$ so that we may take πr to be a rough average. Thus, the force exerted by a single precipitate on the GB is expected to be $\pi r\gamma_{GB}$. But we wish to calculate the force when there are a number (n_p) of precipitates per cm^2 present corresponding to a volume fraction f. If we consider a disk of GB volume 1 cm^2 in area and $2r$ thick, then $f = n_p\frac{4}{3}\pi r^3/2r$. Thus, $n_p = 3f/2\pi r^2$ and the *precipitate force* restraining GB motion is the product of $\pi r\gamma_{GB}$ and $3f/2\pi r^2$ or $3\gamma_{GB}f/2r$. Solute atoms of diameter a and concentration C_s similarly exert a drag force when they segregate at mobile grain boundaries. Correspondingly, a *solute force* of magnitude $3\gamma_{GB}C_s/a$ may then be imagined.

When the driving forces for growth due to grain curvature and surface energy are counterbalanced by restraining forces, a condition of grain stability ensues. Grain size *stagnation* is the consequence of the stable equilibrium expressed by

$$\frac{2\gamma_{GB}}{l_g} + \frac{\Delta\gamma_{fv}}{d} + \frac{\Delta\gamma_{fs}}{d} = \frac{\gamma_{GB}^2}{d\gamma_{fv}} + \frac{3\gamma_{GB}f}{2r} + \frac{3\gamma_{GB}C_s}{a}. \tag{9-10}$$

Various conditions of stagnation can arise depending on the magnitude of the terms in this equation. For example, if curvature-driven grain growth is primarily resisted by grooving, $2\gamma_{GB}/l_g = \gamma_{GB}^2/d\gamma_{fv}$. Typically, $\gamma_{GB} \sim \frac{1}{3}\gamma_{fv}$, and therefore the limiting grain size is $l_g \sim 6d$, or six times the film thickness. Smith (Ref. 35) has found that the largest grains in pure gold and aluminum films annealed to stagnation were $l_g(\mathrm{Au}) \sim 10d$ and $l_g(\mathrm{Al}) \sim 3d$, in rough agreement with theory.

9.4.4 FILM TEXTURE

9.4.4.1 Introduction: Definitions and Crystallography

Until now the chapter has dealt solely with film morphology issues (e.g., form, size, and shape), first in the columnar structures viewed perpendicular to the film plane, and then in grains observed lying parallel to the film surface. Notably absent has been any mention of crystallography and its influence on polycrystalline film properties. We start by noting that each grain in a polycrystalline film has a specific crystallographic orientation relative to a fixed reference direction, and the orientation varies from grain to grain (see Fig. 9-12). It is often the case that the distribution of grain orientations is not random. When this happens we speak of preferred orientation or texture, a condition usually depicted graphically in terms of crystallographic pole figures (Ref. 36), two of which will be shown and discussed subsequently. Sheet metal, for example, might assume a (100)[100] deformation texture. This means that grain (100) planes preferentially lie in the sheet plane, and [100] directions in the (100) plane coincide with the rolling direction. In addition to *deformation* texture arising from slip and grain rotation, texture may develop or change as a result of annealing or *recrystallization*. A variety of physical and mechanical properties depend on texture including elastic modulus, yield strength, magnetic permeability, etch rate, and diffusion.

It will come as no surprise that thin polycrystalline films also often display a texture in as-deposited and as-treated conditions. As we shall see, film texture is influenced by deposition method and variables, nature of the substrate, energetic ion bombardment, and geometrical confinement by surface features. Diffraction, whether by X-rays or electrons, has been the traditional way to determine film crystallography and extent of texture. Thus, if a disproportionate diffracted intensity from a given (*hkl*) plane occurs relative to its intensity in a powder sample of random grain orientations, a preferred (*hkl*) texture exists.

Traditional definitions of texture have been expanded to include misorientation between adjacent grains; i.e., grain-boundary texture. We may thus imagine the involved grain boundaries separating grains that are tilted relative to one another. A twin boundary is a common example where atom positions on one side of the boundary are exactly mirrored on the other. There is an interesting crystallographic, but more complex correlation that occurs when the two lattices are misoriented. To observe it, sketch the same square lattice on two semitransparent sheets of paper. If both lattices are pinned at a common origin and the top latttice is rotated relative to the bottom one, certain lattice points will be shared or coincident at specific

rotation angles. The two interpenetrating lattices form a coincidence-site lattice or CSL with a unit-cell volume some multiple of the original lattice unit-cell volume. For example, a 36.9° rotation about the [001] direction generates a so-called $\Sigma5$ CSL because there are 5 lattice points between coincident sites; furthermore, the CSL cell volume is 5 times that of the original unit cell (Ref. 37). Also increasing the strength of (111) fiber texturing leads to greater frequencies of occurrence of type $\Sigma3$, $\Sigma7$, $\Sigma13$, etc., CSL boundaries. And the greater the extent of lattice coincidence, the more stable the grain boundary.

9.4.4.2 Texture in Thin Films

There is a great body of literature reporting texture determinations as part of the obligatory complement of film-characterization information, as well as fewer, but more focused studies of the phenomenon itself. The fact that [111] texture enhances resistance to electromigration damage (Section 11.6.3) in Al interconnects has spurred many of these investigations. General aspects of film texture have been discussed by Machlin (Ref. 19), while Knorr (Ref. 38) has focused on its implications with regard to metallization reliability. Conclusions regarding texture are summarized as follows.

9.4.4.2.1 General Observations

Manifestations of film texture are quite varied and not easily predicted. Obviously no texture develops in amorphous films. In general, evaporated films are more weakly textured than sputtered films. Whereas Ti (0002), Al (111), and Au (111) exhibit strong textures, sputtered Cu shows three-component textures of (111), (200), and random; these as well as (220) and (511) are also observed in electrodeposited and CVD Cu films (Ref. 34). Texture is barely discerned in metals deposited at very low temperatures ($T/T_M < 0.15$) where there is effectively no adatom mobility or grain-boundary (GB) migration. At higher homologous temperatures increased atom movements foster the development of texture. Sometimes, texture evolves from random in initial deposits, to strong orientation of low-energy planes parallel to the film surface, and finally, to changes in preferred texture as the film thickens further. Thus changes from (111) to (100) film orientation were observed in Ag films after they reached a thickness of 2 μm (Ref. 39). Similarly, 0.6 μm Ag films deposited at 80 K changed texture from (111) to (100) upon heating to 300 K (Ref. 40). Crystallographic twinning during film growth together with the changing competition between surface and strain energy all contribute to the observed texture and changes in it. For example, exposing planes having minimum surface energy drives the initial

epitaxy that is invariably present in the film. But the elastic strain continually builds, and at some critical thickness stress relaxation may reduce strain energy more than the increase in surface energy due to texture change; at this point texture is predicted to change. Anisotropy of the elastic moduli will then favor the growth of low strain-energy oriented grains at the expense of those grains possessing higher strain energy (Ref. 19).

9.4.4.2.2 Substrate Effects

Both the substrate composition and roughness affect film texture. To improve the reliability of Al–Cu metallizations on SiO_2, thin titanium films are often interposed between them. In this SiO_2/Ti/Al–Cu structure a (0002) texture is present in Ti as shown in the X-ray pole figures of Figs. 9-15a and 9-15b. To understand these pole figures imagine that the hexagonal prism cells comprising the Ti lattice in one of the film grains are standing upright

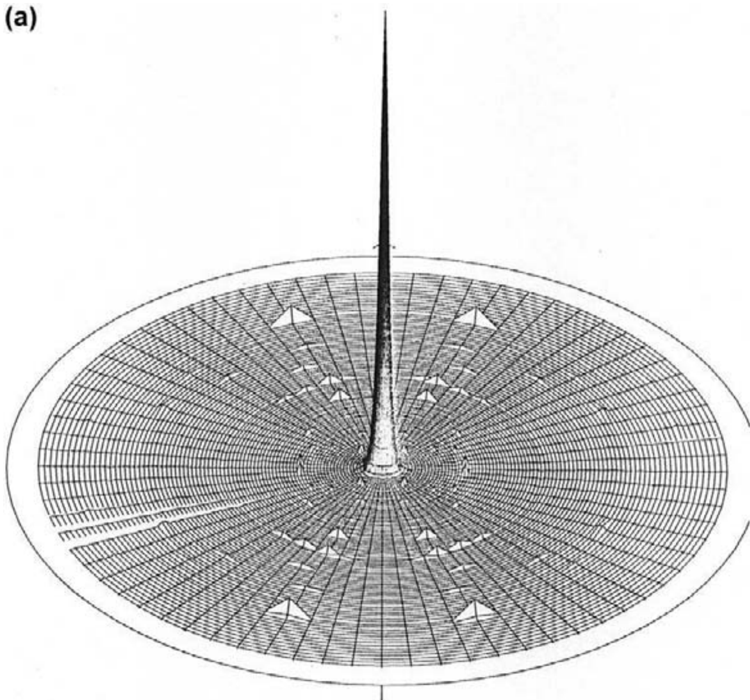


Figure 9-15 Comparison of texture developed in 200 nm thick titanium films deposited on (a) smooth SiO_2 and (b) rough SiO_2 substrates. The (0002) X-ray pole figures are shown. (Ref. 33. From R. Murphy and K. P. Rodbell, IBM Corp., unpublished. Courtesy of J. C. M. Harper, IBM, T. J. Watson Research Center.)

on an SiO_2 substrate lying in the plane of the page. Further imagine that this grain is in the center of a sphere that we view from above at some distance so that what appears is a projected stereographic circle. When a vector parallel to the unit cell prism axis is directed upward it pierces this circle at its center. If all of the film grains were concentrated at the sphere center and their orientation vectors radiated similarly, the pattern shown in Fig. 9-15a would emerge where peak heights reflect the relative strength of the film texture. While the overwhelming majority of Ti grains possess a strong (0002) texture, there are occasional grains oriented in other directions as indicated by the smaller triangular peaks. Furthermore, a comparison between the two figures reveals a much narrower peak width and stronger texture in Ti films deposited on the smoother SiO_2 substrate.

With regard to substrate morphology, a determining factor seems to be the ratio of the adatom diffusion distance ($\sim 2(Dt)^{1/2}$) to the characteristic length of the surface roughness. The greater surface mobility associated with longer diffusion lengths enables adatoms to more readily establish low

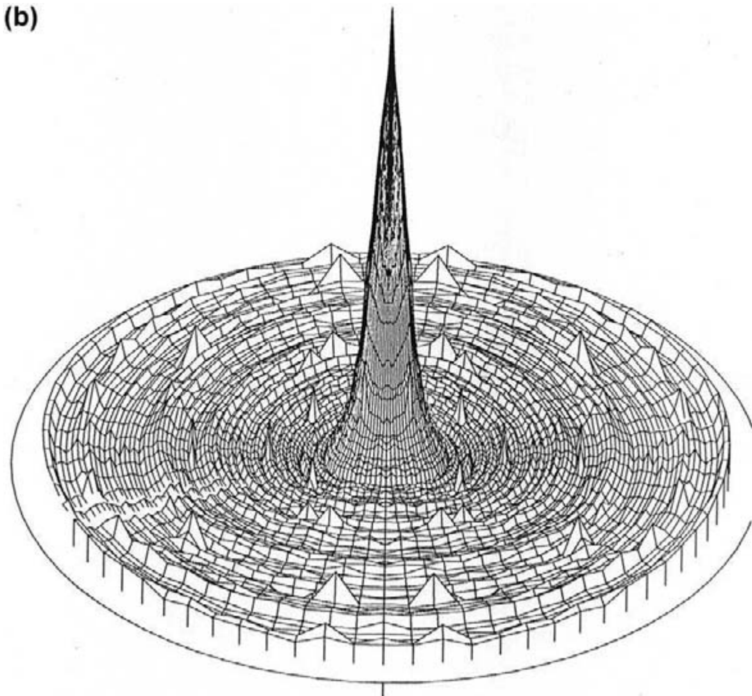


Figure 9-15 Continued.

interface-energy orientations. In addition, more directed deposition achieved during collimated sputtering of Ti was found to be more effective in strengthening the texture relative to conventional magnetron sputtering (Ref. 41).

9.4.4.2.3 Effect of Energetic Particle Bombardment

The modification of film structure and texture due to particle bombardment is universal and can be quite pronounced. During sputtering, ion and neutral atoms bombard growing films under normal incidence and typically possess energies from fractions of an eV to upwards of 100 eV. Increasing target power and changes in the gas-discharge composition can strengthen the film texture or even change it. For example, the preferred orientation of TiN on SiO_2 has been controllably changed from purely (111) to purely (002) by increasing the ion/neutral ratio from 1 to 5 during 20 eV nitrogen bombardment (Ref. 42).

Off-normal incidence particle bombardment is also influential in aligning grains relative to the impingement direction. The texture arising from normal incidence remains but added preferred directions evolve within the film plane. For example, cubic yttria-stabilized ZrO_2 (YSZ) deposits on amorphous or crystalline substrates at elevated temperatures with a strong (001) texture. Strong in-plane (220) and (111) orientations also arise as a complex response to the ion bombardment angle and ratio of ion flux to atom flux (Ref. 43). Other bombardment effects include the tilting of the YSZ columnar microstructure toward the ion beam direction. Enhanced adatom mobility due to the higher effective deposition temperature and resputtering of unfavorably oriented crystallites are among the reasons for the effects of particle bombardment on texture. Ion channeling effects also influence film texture by mechanisms that were discussed in Section 4.6.4.2.

9.4.4.2.4 Patterning Blanket Films

Even though we have been exclusively concerned with blanket films, the discussion also applies to patterned and geometrically confined films, but with modification. For example, originally sputtered Al–Cu films exhibit a strong (111) texture. In submicron-wide and -thick Al–Cu interconnections that are lithographically patterned from such films, a bamboo microstructure emerges. Based on geometric chance, the larger the initial size of equiaxed grains the greater the probability that all GBs will span the patterned stripe axis. Significantly, single-crystal-like conductors punctuated by normally oriented GBs are particularly resistant to electromigration damage (Section 11.6.3.3). If the grain-size distribution is mixed, however, less desirable Y-shaped GB-triple points will also populate the bamboo structure of the conductor.

9.4.5 THIN-FILM MICROTEXTURE

Thus far we have really only treated *macrotexture* because X-ray diffraction methods normally determine the collective orientation of large numbers of grains simultaneously. Imagine now electron backscatter diffraction (EBSD), which determines *microtexture* because crystallographic orientation is measured one grain at a time. This important new technique is based on the capabilities of a scanning electron microscope (Refs. 37, 44). The finely focused electron beam of the SEM acts as a probe capable of measuring texture over a region measuring about $0.5 \times 0.5\ \mu m^2$. When the incident electron beam penetrates a grain, a complex planar backscattered electron-diffraction pattern is generated (Section 10.3.2.2). Computer analysis of this diffraction image enables the relative crystallographic orientation of this grain to be extracted. The beam then moves to the next grain where the process is repeated. In this way as the beam scans across a polycrystalline film the pole figure sequentially builds. One variant of the technique, known as orientation imaging microscopy (OIM), is demonstrated in Fig. 9-16. Shown in three levels of gray are the orientations of each grain in a blanket aluminum film. Although the overall texture is strongly (111), most grains are oriented within $\pm 15^\circ$ of this plane normal.

Texture in Al and Al–Cu films has raised a number of puzzling issues (Ref. 45). Why, for example, is there a 5° misorientation from perfect (111) texture when films are deposited on SiO_2, but this anomaly is absent when Ti is the substrate? At the earliest stages of deposition on both substrates the films are randomly oriented. But apparently, Al wets Ti leading to greater heterogeneous nucleation and film continuity at thinner film thicknesses. Then normal grain growth ensues and surface energy minimization promotes exact (111) texture. Alternatively, the granular epitaxy that develops seems to demonstrate a "texture inheritance" from the underlying Ti grains. But for Al on SiO_2 there is delayed film continuity. Competition develops between (faceted) islands containing a stepped and ridged surface oriented 5° from (111). In addition to texture misorientation, abnormal grain growth occurs.

9.5 CONSTRAINED FILM STRUCTURES

9.5.1 INTRODUCTION

In this section we consider two largely unrelated examples where film structure, architecture, or morphology is constrained either by the unusual substrate topology or by the deposition geometry. The first and most important example concerns the confinement imposed by the substrate

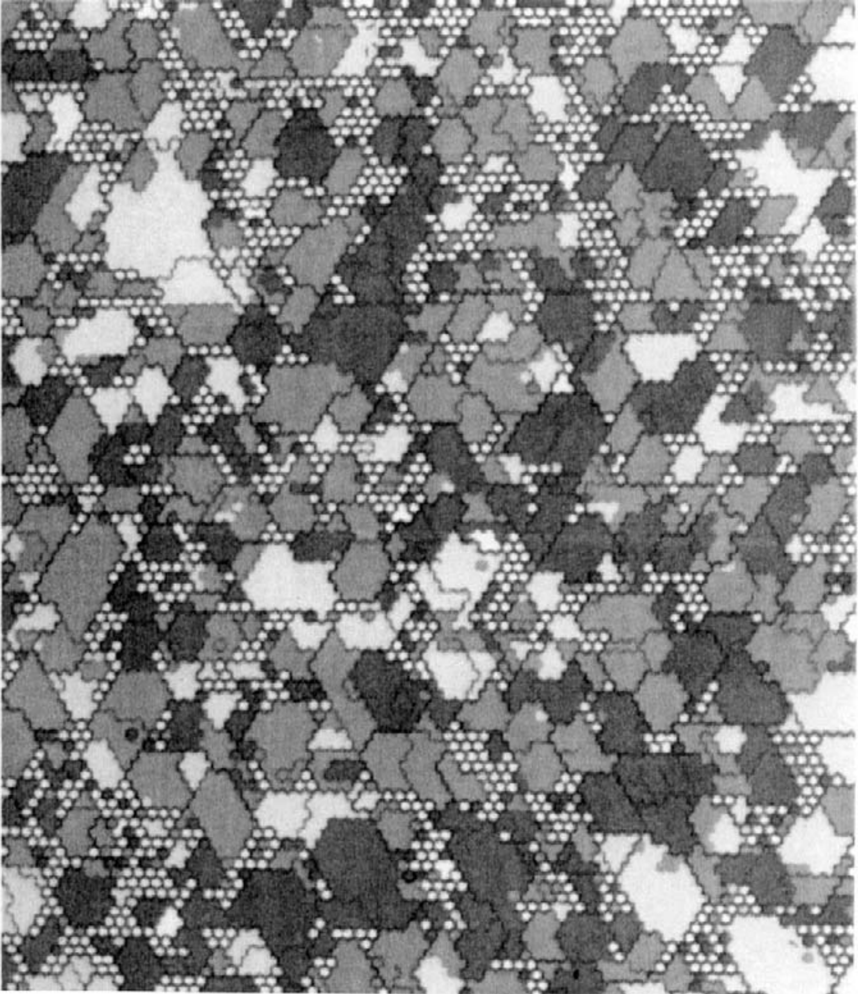


Figure 9-16 OIM image of a polycrystalline thin Al film. Grain orientations within 5° of [111] texture axis are dark, 10° gray, and 15° light. Data were taken in a hexagonal array with a 0.4 μm spacing between pixels. (After Ref. 37.) (Also see Fig. 10-17b.)

surface topography. Step coverage, a major issue of concern in this regard, was addressed previously (Sections 3.3.3 and 5.3.4.3). Here, this subject is treated again but with a focus on the grain microstructures assumed by films deposited on surfaces containing vias and trenches. In particular, the filling of these features with metal is an important ULSI technology issue, and

strategies for accomplishing it will be discussed. A second set of examples deals with the intriguing morphology and structure of films that either are sculptured or contain geometrically modified columnar grains.

9.5.2 LINING AND FILLING TRENCHES BY PVD METHODS

9.5.2.1 Conventional Approaches

Creation of multilevel-metallization via and trench structures means first conformally lining the surrounding dielectric (e.g., SiO_2) with a thin diffusion barrier (e.g., TiN, WN). Then aluminum, and now recently copper, is deposited over the barrier layer to completely fill the space in high-aspect-ratio cavities. Both steps pose challenges (Ref. 46). The microstructure of a TiN via or trench lining deposited by collimated sputtering is similar to that schematically shown in Fig. 9-17a. Dense zone T columnar grains prevail

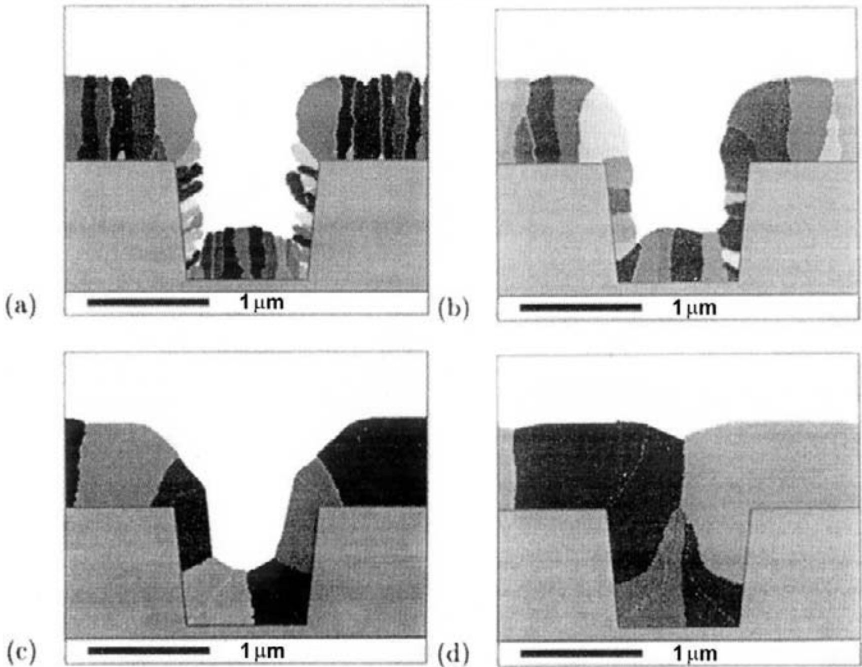


Figure 9-17 Calculated aluminum film microstructures deposited at 250°C using the SIMBAD simulation program. Diffusion lengths are (a) 0.06 μm, (b) 0.18 μm, (c) 0.6 μm, and (d) 1.2 μm. (From Ref. 46.)

over all horizontal surfaces, but the side-wall grains, reflecting the directionality of the sputtering process, are rough and tilted upward toward the via opening. This less dense side-wall structure containing many grain boundaries is not an effective barrier to potential diffusion by metal-fill atoms. Postdeposition annealing in nitrogen or oxygen resulting in additional nitride or oxide formation is a way to densify the structure.

Simple sputtering of Al and Cu does not generally fill via features which are now less than 0.25 μm wide and more than $\sim 0.5\,\mu$m deep. Instead, the film often assumes a bread-loaf morphology that overhangs the hole and sometimes pinches to trap a void (Fig. 3-10b). Causes of the problem include lack of directional atomic impingement in the first place, and insufficient metal mobility and wetting of the lining surface. Raising the substrate temperature in what amounts to hot PVD is an obvious prescription to promote better filling. The intent is to induce metal transport from the large planar reservoir of deposited film (or field) surrounding the via to flow into the hole, fill it completely and eventually yield a planarized surface. In the simulation (Ref. 47) of Al grain-structure evolution shown in Figs. 9-17b, 9-17c, and 9-17d, the effect of different diffusion lengths (L) on these events is shown. From diffusion theory, $L \cong (D_S \tau_S)^{1/2}$, where expressions for the surface diffusivity D_S and lifetime of a mobile surface adatom τ_S were defined earlier (see Eqs. 7-25, 7-30). Filling of the void space in the middle of the trench is very much like the diffusional mass transport that fills the narrow channel between neighboring metal particles during sintering (Section 7.4.5.2). The driving force to minimize surface energy establishes atom movements from regions of high (convex) to low (concave) surface curvature; alternatively, vacancy motion is oppositely directed. In either case, increasing L values promote better trench filling.

9.5.2.2 Two-Step and Pressure Filling

Reducing multilevel-metallization deposition temperatures and times is an important trend. The purpose is to minimize damage to interconnections generated by harmful thermal stresses that develop during processing. Thus, a two-step trench-filling process has been proposed involving initial film deposition at low temperature followed by reflow at elevated temperatures under vacuum. For Al metallization, a typical process consists of depositing an 8000 Å film at 50°C with reflow occurring at 535°C and 1.5 mtorr (Ref. 46). The structural change that occurs during reflow of copper films deposited on a Ta barrier has been computer modeled (Refs. 48, 49) with results shown in Fig. 9-18. In the GROFILMS grain-oriented film-microstructure simulator program, a line segment/nodal method tracks changes in grain shape and size during deposition and subsequent annealing.

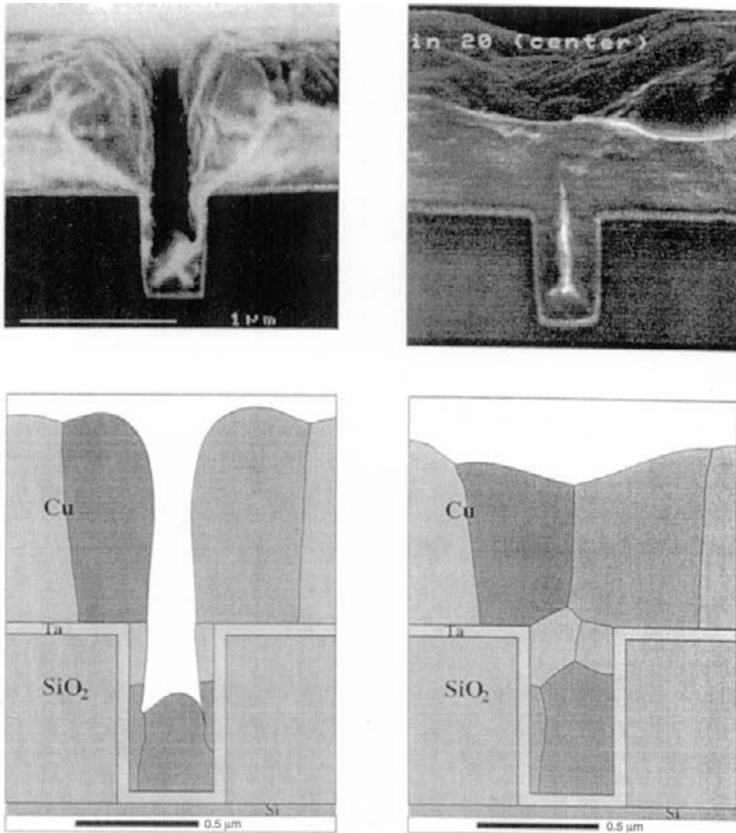


Figure 9-18 (Top) SEM images of sputtered copper microstructures in a 0.35 μm wide, 2:1 aspect ratio trench as deposited (left) and after annealing at 450°C for 25 min. (Bottom) Simulation of Cu microstructures using GROFILMS program. Property values assumed include Cu surface diffusion activation energy = 0.83 eV, Cu surface energy = 1.78 J/m^2, Cu grain-boundary energy = 0.625 J/m^2, Cu/Ta interfacial energy = 0.477 J/m^2, Ta surface energy = 2.15 J/m^2. (From Ref. 48. Reprinted with the permission of the authors.)

Through exposure to elevated temperature, surface, grain-boundary (GB) and lattice diffusional processes are unleashed enabling interfacial energy-induced changes in surface curvature, as well as GB migration and grooving. Excellent agreement between the simulated and reflowed grain structure is evident. Interestingly, simulations have also revealed that reflowed Cu dewetted and agglomerated over a tungsten barrier film yielding incomplete side-wall coverage. This is apparently a case where tungsten's gain as a diffusion-barrier material is offset by its loss as a wettable substrate.

There is yet an additional postdeposition treatment that has proven successful in promoting the necessary Al metal flow to fill sub-0.5 μm wide vias. This novel Forcefill or Hi-Fill process (Ref. 46) makes use of high isostatic pressures ($\sim$60 MPa) applied at relatively low film temperatures ($\sim$400°C). Under these conditions the metal that covers the hole flows plastically and fills the cavity below. The process is reminiscent of the traditional mechanical forming operation of extrusion where metal is forced under pressure to plastically flow through a die (via) opening. In addition to reducing the yield stress of Al, the elevated temperature promotes dynamic recrystallization which leads to very large ($\sim$10 μm) strain-free grains compared to smaller 2 μm grains in reflowed films. Sometimes these grains appear to be cracked or separated from one another raising the possibility of a reliability problem. Clearly a relatively thick field film is necessary for good filling. Pressure filling has been applied in the fabrication of 64-Mb DRAMs and multilevel (0.35 μm) logic gates.

Several important trends will define the future of thin-film metallizations. First, copper is expected to broadly replace aluminum alloys in interconnects. Secondly, the stripe width will continue to shrink. Thirdly, the aspect ratio (or film thickness to width) of interconnects, presently equal to $\sim$1.5:1, is projected to increase to 4:1 by 2010. And finally, in a severe form of geometric confinement, lines will be fabricated within patterned channels of the surrounding SiO_2. Excess metal from such lines is removed by a damascene process employing chemical–mechanical polishing (CMP), which creates a planar surface of interconnects embedded within insulation. As a result of these trends, atom diffusion lengths will easily exceed film dimensions so that grain boundary pinning, precipitation, and phase nucleation at interfaces rather than in bulk grains will dominate microstructural change. Furthermore, there is evidence (Ref. 50) that geometric confinement weakens the strong (111) texture. Thus, in the competition at different thin-film growth centers, side walls may be more influential than bottoms in promoting oriented growth.

9.5.3 SCULPTURED FILMS

As a challenge, the reader is invited to examine the film structures shown in Figs. 9-19a,b and, without reading further, guess how they were produced (Ref. 51). The helicoidal pillar structures are particularly fascinating and resemble those of chiral media, e.g., liquid crystals. When fabricated from optical-film materials such as SiO_2 and MgF_2 they display similar optical activity, i.e., rotation of the plane of polarized light, to those of cholesteric liquid crystals. In addition, these films can be made containing a

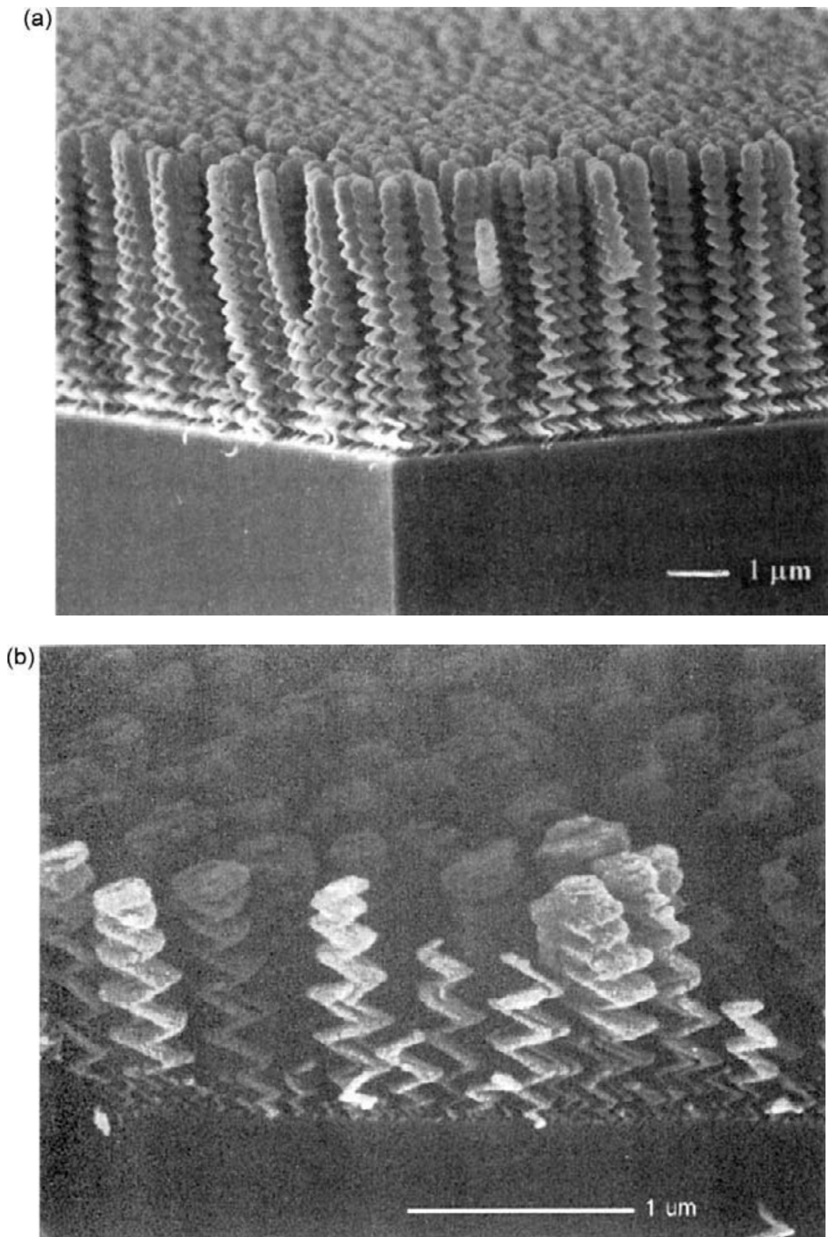


Figure 9-19 Sculptured film structures produced by an 85° oblique flux of evaporant. (a) Helical columns produced by rotary substrate motion. (b) Zigzag columns produced by alternating tilting. (From Ref. 51. Reprinted with the permission of the authors.)

controllable amount of porosity with pillar diameters and densities as small as 30 nm and 3 per μm^2, respectively (Ref. 52). Both film structures evolved under shadowing-dominated aggregation due to oblique deposition, and reduced adatom mobility due to low substrate temperatures ($T_S/T_M < 0.3$). These dual conditions favor zone 1 columnar growth where a matchstick-like morphology develops.

Evaporation onto substrates oriented at glancing angles ($\theta > 75^\circ$) as seen in Fig. 9-20a is the technique used to prepare these films. The C- or S-shaped film of Fig. 9-20b was simply the result of continuously tilting the substrate relative to the evaporation source, first in one direction from 5° to 175° and then back again to 5° (Ref. 53). Even when deposition becomes almost totally glancing, contrary to the Tangent Rule (Eq. 9-1), the column angle relative to the substrate is roughly 30°. This means that the cusps of the C and S shapes cannot be too pronounced. It is also probably the reason that directionally deposited thin films lining trench walls have a tilted grain structure (see Fig. 9-17a). With additional substrate rotation there is greater flexibility in deposition geometry. It was the combination of substrate tilt and rotation that enabled the sculptured films of Fig. 9-19 to be created. Metals (Cu, Al, Mn, Cr) as well as nonmetals ($MgCl_2$, SiO) have been evaporated in this way.

Among the potential applications of these interesting film structures, biocompatable surfaces, optical media, catalysts, and chemical sensors have been mentioned. For example, relative to currently plasma-sprayed porous coatings for surgical prostheses, the sculptured films offer the advantages of low-temperature processing and a controlled void volume for bone in-growth. Optical properties of these films appear to have received the greatest attention. Since the refractive index of materials is proportional to their density, films with variable and even graded index are possible. Film densities ranging from only 10% to as much as ~90% of the bulk solid value have been produced in this way. Porosity in optical films, however, invites moisture permeation and is particularly deleterious to their reliability. A strategy to prevent this is to seal the surface with dense capping layers produced by deposition normal to the substrate.

9.6 AMORPHOUS THIN FILMS

9.6.1 INTRODUCTION

For a long time it appeared that amorphous solids were limited to such materials as bulk silica glasses, slags, and polymers whose viscous-melt structures could be retained upon solidification. It was thought that amorphous

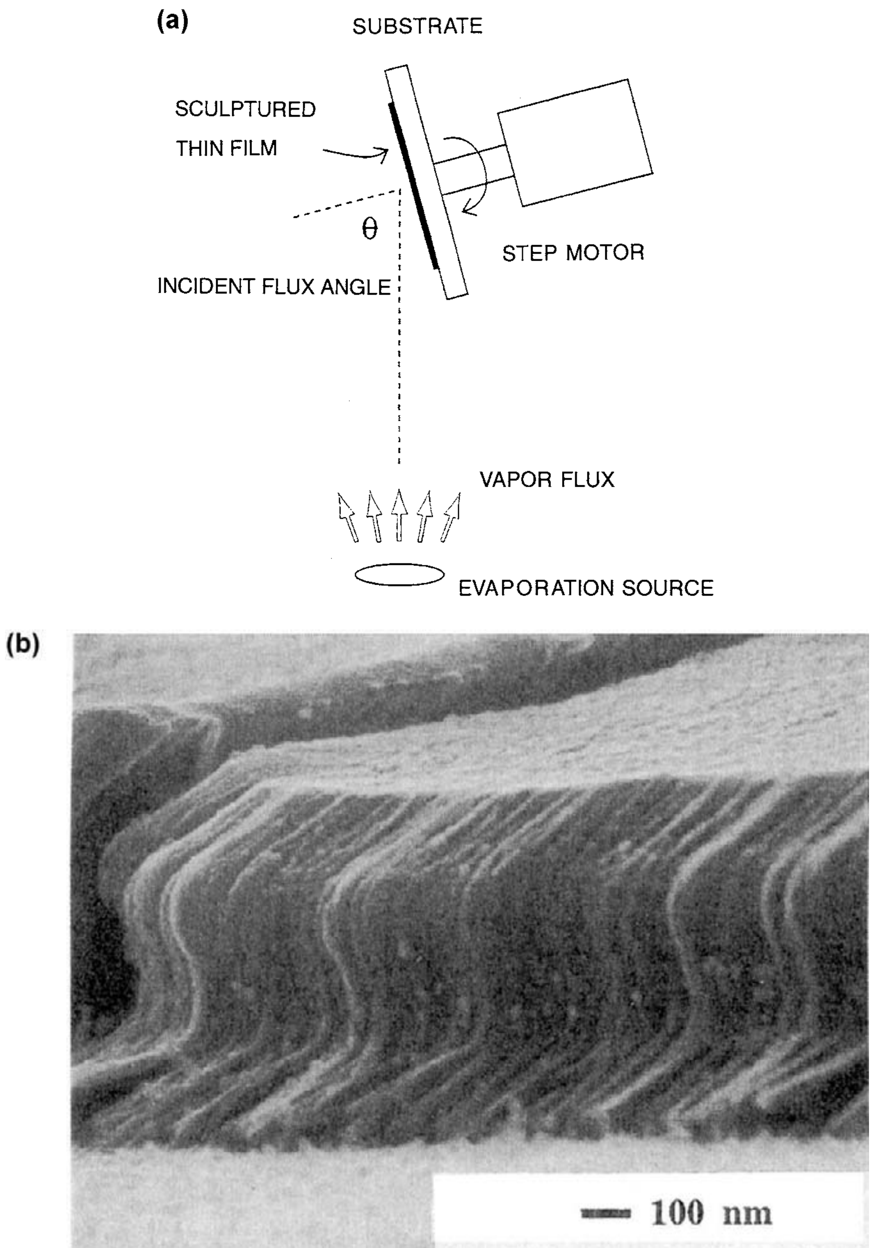


Figure 9-20 (a) Schematic of the glancing angle deposition process. (From Ref. 51. Reprinted with the permission of the authors.) (b) S-shaped film morphology. (From Ref. 53. Reprinted with the permission of Dr. Russell Messier.)

metals could not be prepared; the mobility of metal atoms in melts was so great that crystallization could not be suppressed. The challenges of proving widely accepted common wisdoms wrong often provide powerful incentives for experimenters. Thus, in the mid-1950s Buckel (Ref. 54) first produced pure amorphous metals by evaporating Ga and Bi onto glass substrates cryogenically cooled to liquid-helium temperatures. The extraordinarily high cooling rates during vapor condensation, estimated to be in excess of 10^{10}°C/s, immobilize or freeze adatoms on the substrate where they impinge. This prevents atoms from diffusing and seeking out equilibrium lattice sites, yielding a disordered, liquid-like, but solid matrix. Even so, amorphization of pure metals is apparently limited to those having some covalent bonding character.

Subsequent interest is amorphous metals focused on alloys. Alloy metal films proved easier to deposit in amorphous form because each component effectively interferes with the mobility of the other. This meant that higher substrate temperatures ($\sim$77 K) could be tolerated, and that vapor quench rates did not have to be as high as those required to produce pure amorphous metal films. Important fruits of the early *thin-film* work were realized in the later research and development activities that led to the synthesis of *bulk* amorphous metals by quenching melts. Today, continuously cast ribbons and strips of metallic glasses (Metglas) are commercially produced for such applications as soft magnetic transformer cores and brazing materials. Cooling rates of $\sim 10^6$ °C/s are required to prevent nucleation and growth of crystals. Heat transfer limitations restrict the thickness of these metal glasses to less than 0.1 mm.

Amorphous thin films of some of these materials as well as other metal alloys and virtually all elemental and compound semiconductors, semimetals, oxides, and chalcogenide (i.e., group VI–containing elements) glasses have been prepared by a variety of PVD and CVD techniques. From laboratory curiosities amorphous Se, GdCo, and GeSe thin films have been exploited for such applications as xerography, magnetic bubble memories, and high-resolution optical lithography, respectively. Most recently, interest has centered on semiconducting film materials, primarily amorphous silicon and carbon. The former is used in solar cells and the latter for hard, wear-resistant coatings. Unlike metal alloys where atomic packing is of primary concern, the electronic nature of amorphous semiconductors is of interest. Exploring the structure of amorphous solids is a good place to start.

9.6.2 Amorphous Materials at the Atomic Scale

While the equilibrium locations of all atoms in crystalline solids can be inferred from just a few atomic positions within a unit cell, this is not true

for amorphous solids which do not possess long-range order (Ref. 55); instead, the latter exhibit two types of short-range order. *Topological* short-range order is characterized by an average number of nearest neighbors or z, the coordination number, as well as by the mean separation of atoms. The radial distribution function (RDF), which can be determined by X-ray diffraction, provides values for z as well as nearest-neighbor distances. For example, in amorphous metal alloys and elemental semiconductors, z is $\simeq 12$ and $\simeq 4$, respectively. In addition, some amorphous alloys also display *compositional* short-range order where nearest-neighbor compositions differ from average bulk values.

Since they contain a huge number of different atomic configurations, amorphous solids are usually described in terms of statistical distribution models that fall into two categories, namely, dense random (close) packing (DRP) and continuous-random-network (CRN). The DRP model is better suited to amorphous alloys, whereas amorphous elemental semiconductors are more closely described in CRN terms. General features of the DRP model include: (1) atoms tend to fall into regular as well as distorted tetragonal groupings that are further packed into larger units; (2) clusters with fivefold symmetry exist; and (3) the DRP density is about 0.64. Similarly, according to the CRN model: (1) z for each atom is the same as in the crystalline solid, (2) the bond length is nearly constant, and (3) bond angles display a significant spread about a mean value.

9.6.3 Structure of Amorphous Semiconductor Films

Amorphous semiconductor thin films are perhaps the most intriguing amorphous solids of all. Defects such as dangling bonds stemming from incompletely coordinated group IV atoms, plus distorted bond lengths and angles, all conspire to promote electronic as well as structural disorder in these covalently bonded materials. Because of atoms in three- and fivefold coordination, dangling bond defects with distorted bond lengths are readily sustained and promote electronic as well as structural disorder in these materials. Full coordination of the four bonds between atoms in the amorphous network is impossible, so we say that the network is "overconstrained." Dangling bonds reaching a density greater than 10^{20} cm^{-3} can apparently occur in a-Si (Ref. 56). This value should be compared with the atomic density of 5×10^{22} cm^{-3} in crystalline Si. The structural disorder creates localized, defect electronic states that tail into the energy gap between the top of the valence band and bottom of the conduction band. Because they are so numerous these states effectively blur the energy gap. One might expect the electrical conductivity to be enhanced as a result.

However, dangling bonds trap charge and reduce charge mobility, rendering the material more insulating and device behavior less reliable. For this reason, such bonds are passivated or saturated (terminated) by hydrogenating them. When this happens the dangling bond density may decrease several orders of magnitude, e.g., to $\sim 10^{16}\ \mathrm{cm}^{-3}$, while the bandgap energy rises (e.g., to 1.7 eV with $\sim$10% H). Solar cells and thin-film transistors used in active-matrix liquid crystal displays are among the commercial applications of amorphous silicon.

9.6.4 Amorphous Silicon Nitride and SiO_2

Silicon nitride is normally prepared by reacting silane with ammonia in an argon plasma, but an N_2/SiH_4 discharge can also be used. Like a-Si, amorphous silicon nitride is overconstrained because there are four bonding electrons on Si and three on N. During plasma deposition however, as much as 30 at.% hydrogen can be incorporated, and by terminating Si and N bonds the overconstraint is relieved. It is in this sense that silicon nitride is often described as a ternary solid-solution alloy, SiNH. In contrast, plasma-deposited SiO_2 films contain very little hydrogen. Because divalent oxygen atoms can readily locate silicon atoms within the amorphous network, there is little overconstraint. Uncompensated oxygen bonds are terminated by hydrogen, leaving a relatively small concentration of OH groups behind.

9.6.5 Amorphous Metal Alloy Systems

Even though thin-film deposition techniques can independently facilitate high deposition rates and low substrate temperatures, not all metal alloys can be amorphized. Necessary, but not sufficient, for this purpose are atoms of different size, which alloys obviously provide. In addition to atomic size and achieving the required quench rates, alloy compositions are critical. Most of the presently known glass-forming binary alloys fall into one of four categories (Ref. 57):

1. Transition metals and 10–30 at.% semimetals
2. Noble metals (Au, Pd, Cu) and semimetals
3. Early transition metals (Zr, Nb, Ta, Ti) and late transition metals (Fe, Ni, Co, Pd)
4. Alloys consisting of IIA metals (Mg, Ca, Be)

In common, many of the actual glass compositions correspond to where deep or low-temperature eutectics are found on the phase diagram. Two

different amorphous alloy–film system types will be considered. They shed different lights on aspects of these interesting materials and their properties.

9.6.5.1 Au–Co Amorphous Films

We first consider amorphous Co–30 Au films since they were among the first to be well characterized structurally and through electrical resistivity (ρ) measurements (Ref. 58). The Au–Co equilibrium phase diagram reveals simple eutectic behavior with critical eutectic composition and temperature values of 23.5 at.% Co and 996°C, respectively; the alloy composition selected thus bears little relation to these eutectic features. Amorphous films were evaporated from independently heated Co and Au sources onto substrates maintained at 80 K. Dark-field electron microscope images and corresponding diffraction patterns are shown side by side in Fig. 9-21. The as-deposited film is rather featureless with a smooth topography, while the broad-halo diffraction patterns cannot be easily and uniquely assigned to known lattice spacings of crystalline alloy phases in this system. Both pieces of evidence point to the existence of an amorphous phase. The question of whether so-called amorphous films are in reality microcrystalline is not always easy to resolve. In this case, however, the subsequent annealing behavior of these films was quite different from what is expected of fine-grained crystalline films. Heating to 470 K resulted in the face-centered cubic diffraction pattern of a single metastable phase, whereas at 650 K, lines corresponding to the equilibrium Co and Au phases appeared.

Resistivity changes accompanying the heating of Co–38 at.% Au (an alloy similar to Co–30 at.% Au) revealed a two-step transformation as shown in Fig. 9-22. Beyond 420 K there is an irreversible change from the amorphous structure to a metastable FCC crystalline phase which subsequently decomposes into equilibrium phases above 550 K. The final two-phase structure is clearly seen in Fig. 9-21. Prior to this, the high resistivity of the amorphous films is due to enhanced electron scattering by the disordered solid solution. Crystallization to the FCC structure reduces the resistivity, and phase separation, further still.

Both the amorphous and metastable phases are stable over a limited temperature range in which the resistivity of each can be cycled reversibly. Once the two-phase structure appears, it of course can never revert back to less thermodynamically stable forms. This amorphous–crystalline transformation apparently proceeds in a manner first suggested by Ostwald in 1897. According to the so-called Ostwald rule, a system undergoing a reaction proceeds from a less stable to a final equilibrium state through a succession of intermediate metastable states of increasing stability. In this sense the amorphous phase is akin to a quenched liquid phase. Quenched films

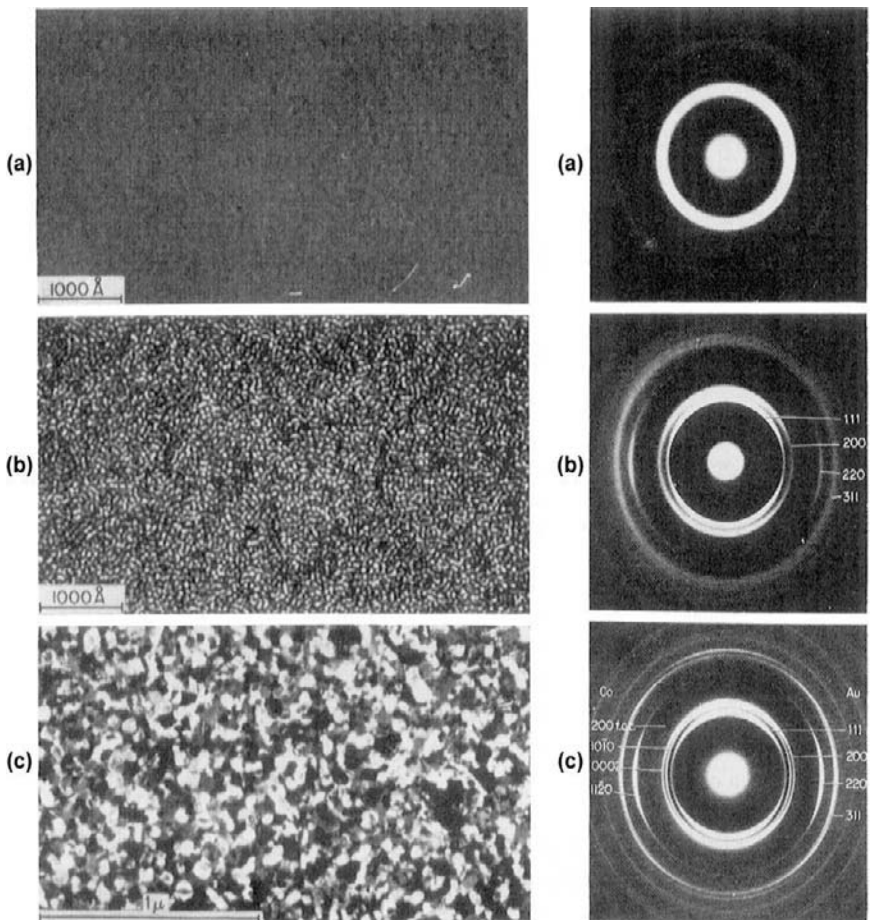


Figure 9-21 Electron micrographs and diffraction patterns of Co–30 at.% Au. (Top) As deposited at 80 K, warmed to 300 K (amorphous); (middle) film warmed to 470 K (single-phase FCC structure); (bottom) film heated to 650 K (two-phase equilibrium). (From Ref. 58.)

exhibit other manifestations of thermodynamic instability. One is increased atomic solubility in amorphous or single-phase metastable matrices. For example, the equilibrium phase diagram for Ag–Cu is that of a simple eutectic with relatively pure terminal phases of Ag and Cu which dissolve less than 0.4 at.% Cu and 0.1 at.% Ag, respectively, at room temperature. These limits can be extended to 35 at.% on both sides by vapor quenching the alloy vapor. Similar solubility increases have been observed in the Cu–Mg, Au–Co, Cu–Fe, Co–Cu, and Au–Si alloy systems.

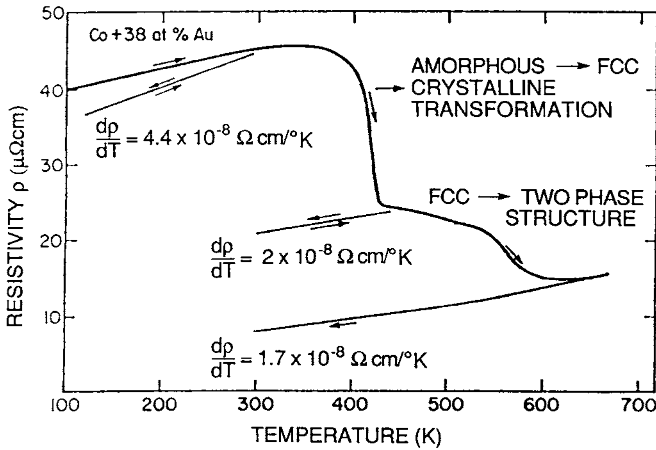


Figure 9-22 Resistivity of a Co–38 at.% Au film as a function of annealing temperature. Reversible values of $d\rho/dT$ in various structural states of the film are shown together with changes in ρ during phase transformations. (From Ref. 58.)

9.6.5.2 Ni–Zr and Co–Zr Amorphous Films

Confounding the notion that rapid quenching of liquids or vapors is required to produce amorphous alloy films is the startling finding that they can also be formed by solid-state reaction (Ref. 59). Consider Fig. 9-23, which shows the result of annealing a bilayer couple consisting of pure polycrystalline Ni and Zr films to 300°C for 4 h. The phase diagram predicts negligible mutual solid solubility and extensive intermetallic compound formation; surprisingly, an amorphous NiZr alloy film is observed to form. Clearly, equilibrium compound phases have been bypassed in favor of amorphous-phase nucleation and growth as kinetic considerations dominate the transformation. The effect, also observed in Rh–Si, Si–Ti, Au–La, and Co–Zr systems, is not well understood. Apparently the initial bilayer film passes to the metastable amorphous state via a lower energy barrier than that required to nucleate stable crystalline compounds. However, the driving force for either transformation is similar. Unlike other amorphous films extensive interdiffusion can be tolerated in Ni–Zr without triggering crystallization.

The very earliest stage of the solid-state amorphization reaction in the Co–Zr system was detected with atomic resolution employing field ion microscopy (Ref. 60). It appears that the large negative heat of mixing plus the decrease of interface energy at the incoherent Co/Zr interface provide

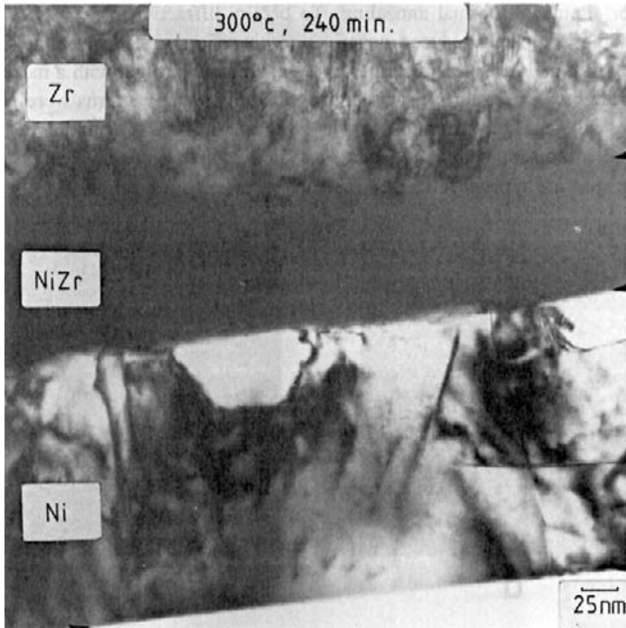


Figure 9-23 Cross-sectional electron micrograph of an amorphous Ni–Zr alloy film formed by annealing a crystalline bilayer film of Ni and Zr at 300°C for 4 h. (Courtesy of K. N. Tu, IBM, T. J. Watson Research Center. From Ref. 59.)

the driving forces for the reaction. Initially a 2.5 nm amorphous region forms at the interface even at room temperature. The process starts with the supersaturation of large-angle Zr grain boundaries by Co. Beyond a critical Co concentration the Zr lattice is destabilized and the amorphous phase grows spontaneously. Evidently the increase in supersaturation is due to the presence of both immobile Zr atoms and defects. However, point defects and dislocations in Zr help to nucleate the amorphous phase because in their absence no amorphization occurs.

In Section 11.4.2.1 the kinetics of the amorphous-to-crystalline transformation in CoSi films will be described.

9.6.6 A Model to Simulate Structural Effects in Thin Films

One of the outcomes of their research on quenched amorphous alloy films was the engaging mechanical model Nowick and Mader (Ref. 61)

developed to better explain the experimental results. Many phenomena observed in pure and alloy thin-film structures are qualitatively simulated by this model. For this reason it is valuable as a pedagogic tool and worth presenting here. The "atoms" comprising the thin films were acrylic plastic balls of different sizes. They were rolled down a pinball-like runway tilted at 1.5° to the horizontal to simulate the random collisions of evaporant atoms. A monolayer of these atoms was then "deposited" on either an "amorphous" or "crystalline" substrate. The former was a flat sheet of plastic while the latter contained a perfect two-dimensional periodic array of depressions into which atoms could nest. Provision was made to alter the alloy composition by varying the ball feed. A magnetic vibrator simulated thermal annealing. To obtain diffraction patterns from the arrays, reduced negatives (with an array size of about 4 mm square) were prepared. The balls appeared transparent on a dark background with a mean ball separation of ~ 0.13 mm. Fraunhofer optical-diffraction patterns were generated by shining light from a He–Ne laser ($\lambda = 6328$ Å) on the negative mounted in contact with the 135 mm lens of a 35 mm camera. Several resulting photographs are reproduced here.

The perfect array of spheres of one size is shown together with the corresponding diffraction pattern in Fig. 9-24a. A hexagonal pattern of sharp spots, reminiscent of electron diffraction patterns of single-crystal films, reflects the symmetry of the close-packed array. When a stacking fault is created in the structure, the diffraction pattern shows streaks (Fig. 9-24b). These run perpendicular to the direction of the fault in the structure.

The effect of deposition rate is shown in Figs. 9-24c and 9-24d. When the film is deposited "slowly" there are a number of grains, vacancies, and stacking faults present in the array. Relative to Fig. 9-24a the diffraction spots are broadened, a precursor to ring formation. In Fig. 9-24d the film is deposited at a "high" rate and the grain structure is considerably finer and more disordered with numerous point defects, voids, and grain boundaries present. Now, semicontinuous diffraction rings appear and these are very much like the common X-ray Debye–Scherer rings characteristic of polycrystals. Interestingly, the intensity variation around the ring is indicative of preferred orientation. When the rapidly deposited films are annealed through vibration, the array densifies, vacancies are annihilated, faults are eliminated, and grains reorient, coalesce, and grow. The larger grains mean a return to the spotted diffraction pattern.

We now turn our attention to alloy films. In the case of "concentrated" alloys containing equal numbers of large and small spheres with a size difference of 27%, the as-deposited structure is amorphous, as indicated in Fig. 9-25. The diffraction pattern contains broad halos. Upon vibration annealing, the film densifies slightly but the atomic logjam cannot be broken

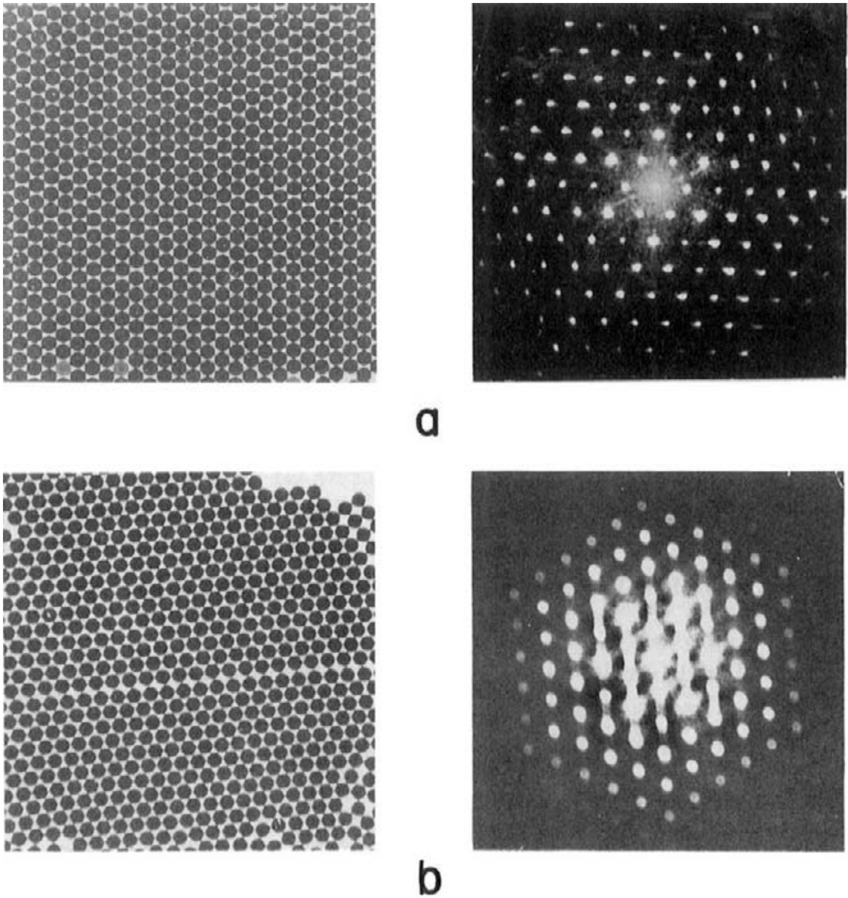


Figure 9-24 Atomic sphere film structures and corresponding Fraunhofer diffraction patterns for: (a) perfect array; (b) stacking fault; (c) pure film, slow deposition rate; (d) pure film, high deposition rate. (Reprinted with permission from the IBM Corp. From Ref. 61.)

up. There is no appreciable change in its structure or diffraction pattern — it is still amorphous. For less concentrated alloys ($\sim$17%), however, the as-deposited structure is very fine grained but apparently crystalline.

All of the foregoing results were for films deposited on the smooth substrate. The "crystalline substrate" affords the opportunity to model epitaxy effects. Pure films deposit in almost perfect alignment with the substrate when deposited slowly. Imperfect regions are readily eliminated upon annealing and nearly perfect single crystals are obtained. Rapidly

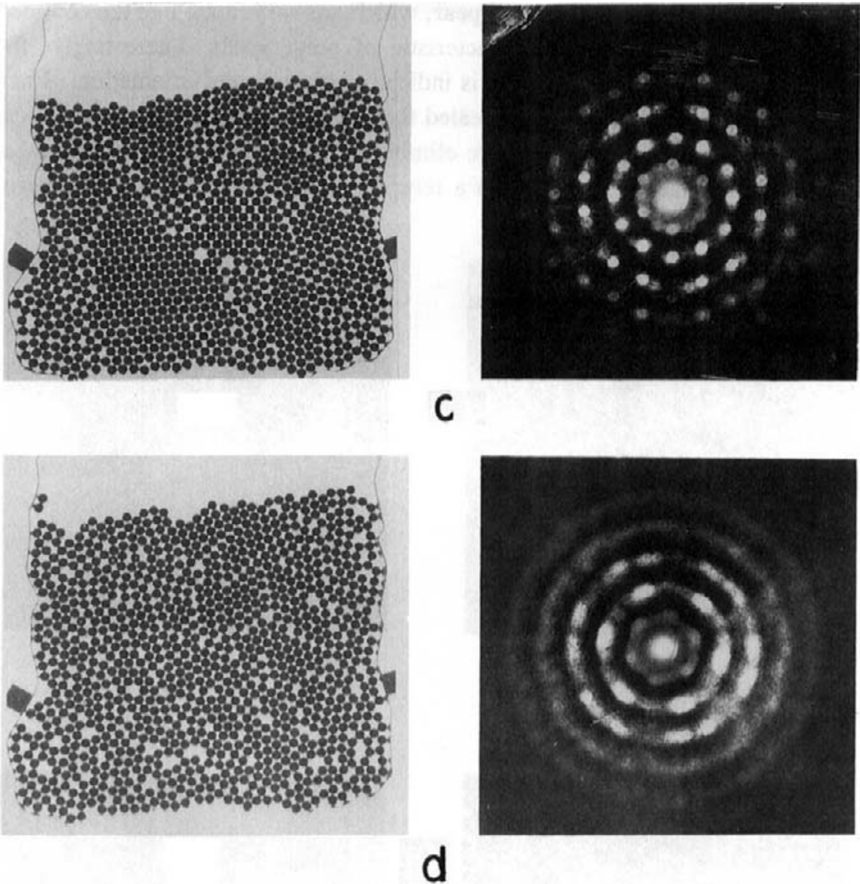


Figure 9-24 Continued.

deposited films are less influenced by the underlying substrate and remain polycrystalline after annealing. Clearly epitaxial growth is favored by low deposition rates. The presence of alloying elements impeded epitaxy from occurring in accord with experience.

Perhaps this mechanical model is antiquated in an age of computer simulation, but much insight into the dependence of film structure as a function of deposition variables can be gained from it. Readers interested in this as well as other such simulations of planar atomic arrays, such as the celebrated Bragg bubble-raft model (Ref. 62), should consult the literature on the subject.

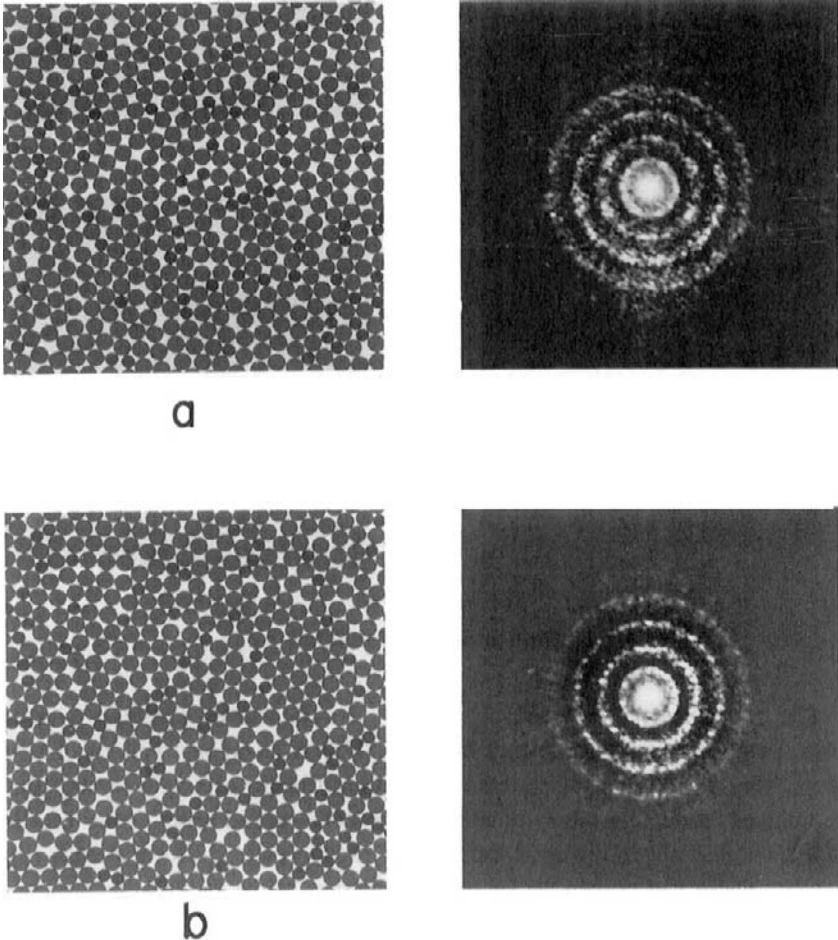


Figure 9-25 Atomic sphere film structure and corresponding diffraction patterns for concentrated alloy (50A–50B, 27% size difference). (a) As-deposited (amorphous); (b) vibration annealed. (Reprinted with permission from the IBM Corp. From Ref. 61.)

9.7 CONCLUSION

Unlike polycrystalline bulk materials where grain structures evolve from numerous competing growth centers distributed in three dimensions, film microstructures reflect the unidirectional transfer of atoms from the gas phase to a substrate. Although greater directionality of atom transport

occurs in sputtering and evaporation than in CVD processing, films from all three methods share common zone structures and morphological features. Perhaps the most striking is columnar grain-growth normal to the substrate plane. Shadowing effects, substrate temperature, and energy of depositing atoms are the influential factors that affect grain morphology. At low temperatures, diffusional transport of depositing atoms is limited so that they basically aggregate according to line-of-sight constraints. Columns develop that are often tilted and separated by voided boundaries; computer simulations have convincingly predicted such features. The structural template provided by the substrate does not strongly influence film growth at low temperatures. Thus columns are observed on both crystalline and amorphous substrates, and in crystalline as well as amorphous films. In extreme cases involving near grazing-angle deposition and nonplanar substrates, sculptured film structures, wavy columns, nonconformal coverage, and nonuniform film thicknesses can be expected.

At elevated temperatures both surface and bulk diffusion allow atoms to access equilibrium lattice sites, fill voids, and enlarge grains. These effects now dominate geometric shadowing. Evolution of grain and film surface morphology is determined by a balance of assorted driving forces derived from surface, stress, and chemical energy released during deposition and postdeposition annealing. In plan view, film grains tend to be small and equiaxed and often display a bimodal grain-size distribution.

Amorphous thin films stand in stark contrast to the epitaxial and polycrystalline films of the previous and present chapters. Both elemental and alloy materials have been amorphized and the atomic arrangement in such structures has been a continuing subject of interest. When heated, the structure transforms to conventional crystalline forms. Paradoxically, however, some crystalline alloy film bilayers have been heated to yield amorphous layers.

EXERCISES

1. (a) A thin-film layer of thickness d is composed of contacting upright cylindrical grains of radius r. If these columnar grains pack in a square array (in plan view) what is the packing factor for this film?
   (b) Suppose smaller cylindrical grains exactly fill in all the interstices between the larger original grains to create a denser bimodal grain-size structure. What is the packing factor for such a film?
   (c) Consider a film of thickness d composed of contacting cone-shaped grains of equal volume. Each cone is inverted with the apex at the substrate interface and the circular base of radius r at the film

surface. For a tightly packed array of circular bases, write an expression for the packing factor in terms of d and r.

2. The transition temperature (T_T) between zones 1 and 2 can be estimated by equating the rate at which film voids are produced during deposition to the rate at which they are filled by adatom diffusion. In the former, the characteristic time τ is inversely proportional to the deposition rate, whereas in the latter τ is the diffusion time for random walk.
   (a) Estimate the value of T_T for Al, Au, and Cu. Assume that the metals are deposited at a rate of 100 nm/s and have a column diameter of 1 nm, and that adatoms migrate with diffusivity equal to $10^{-7} \exp - (6.5T_M/T)\ m^2/s$, where T_M is the melting point.
   (b) If the deposition rate is 25 nm/s, calculate the values of T_T/T_M for these metals.

3. (a) A planar substrate containing a trench in the shape of an isosceles triangle must be coated with metal. For this purpose a point evaporant source is located directly below the triangle apex (whose angle is 30°) in order to symmetrically deposit metal on the sloped cavity walls as well as substrate. Sketch the resultant expected tilt variation of the columnar grain deposit by drawing vapor trajectories from the source to representative points on the substrate and cavity, and directly applying the Tangent Rule.
   (b) Repeat this for a rectangular trench cavity.

4. (a) A film with an as-deposited zone 1 structure was sputtered at temperature T_1 while a film of identical composition and thickness with zone 3 structure was sputtered at temperature T_3. The first film was annealed at temperature T_3 while the second film was cooled to temperature T_1. Describe the expected final grain structures.
   (b) Describe a postdeposition treatment that would convert an amorphous film structure to a fine grained polycrystalline one.
   (c) Suggest a postdeposition treatment that would amorphize a fine-grained film.
   (d) Devise a method that might possibly convert a film stripe 5 μm wide containing a polycrystalline array of 0.5 μm grains to a bamboo structure having large grains that span the stripe width.

5. A suggested three-dimensional structure zone diagram (Ref. 11) for unbalanced magnetron sputtering has as its axes bias voltage, homologous temperature T/T_M, and ratio of ionic to atomic fluxes. If the zone boundaries are spherical (octant) surfaces, schematically indicate where zone 1, 2, and 3 structures are expected to dominate on such a diagram.

6. During examination of the grain structure of a film evaporated from a point source onto a large planar substrate, the following observations were made as a function of position:

   1. There is a film thickness variation.
   2. There is a grain size variation.
   3. There is a variation in the angular tilt of columnar grains.
   4. There is a film density variation.

   Give physical reasons for these observations.

7. (a) Explain how the Lennard–Jones potential might be different for atoms located on the surface relative to atoms within the bulk of a solid.
   (b) Using Eq. 9-7, calculate the equilibrium values of r in terms of a, and U in terms of U_0.
   (c) Does a film modeled with the Lennard–Jones potential exhibit thermal expansion? Explain.

8. (a) How do the grain structures of castings and thin films resemble and differ from each other?
   (b) What are the similarities and differences between grain structures of recrystallized deformed metals and thin films?

9. Atoms on either side of a curved grain boundary (GB) reside on surfaces of different curvature, establishing a local chemical potential gradient that will drive GB migration. Use the Nernst–Einstein equation to derive the kinetics of grain growth.

10. Verify the approximate energy density (or energy per atom) values given in Fig. 9-14 for the processes of (1) solidification, (2) compound formation, (3) solute precipitation, (4) phase transformation, (5) ion bombardment during deposition, and (6) mechanical deformation. Make any reasonable assumptions about material properties.

11. Develop an equation for the grain size of evaporated metal films as a function of substrate temperature. Calculate the maximum and minimum grain sizes expected for a 1000-Å thick Bi deposited at 0°C? Similarly, what maximum and minimum grain sizes can be expected for a 1000-Å thick Ag deposited at 300°C?

12. Extrusion theory suggests that the pressure (P) required for frictionless flow of a metal billet of area A_0 through a die of cross-sectional area A_f is given by $P = \sigma_0 \ln(A_0/A_f)$, where σ_0 is the yield stress of the extrudate. Assume A_f is associated with a via cross-sectional area and A_0 is the field film area providing metal. If $A_0 = 4\ \mu m^2$, and σ_0 for Al is 10,000 psi, estimate the pressure required to fill a via whose diameter is 0.4 μm.

13. If trench filling is like sintering, roughly estimate the time required for copper to fill a trench that is 0.25 μm wide and 2 μm long at a temperature of 500°C. Make any assumptions you wish.

14. Consider the reaction between Ni and Zr films in the diffusion couple of Section 9.6.5.2. Assume two potential products are the metastable amorphous phase (A) and a thermodynamically stable compound (C), e.g., NiZr.
    (a) Using the free energy-reaction coordinate representation for chemical reactions shown in Fig. 1-19, sketch the paths taken to yield either A or C products at 300°C. Clearly identify the key free energies that appear. Note that for Ni–Zr the heats of formation for A and C are necessarily slightly different.
    (b) Write an expression for the ratio of the amounts of amorphous to compound phase [A]/[C] that nucleate in terms of the relevant free energies from part (a). Assume the involved surface energies for both A and C formation are the same.

15. An amorphous (a) film of $CoSi_2$ is deposited over a polycrystalline (c) $CoSi_2$ film and the combination is then heated to an elevated temperature where the amorphous layer is consumed by the advancing polycrystalline layer. It is observed that the velocity at which the interface between films translates is higher, by 2 to 3 orders of magnitude, than the velocity at which c-$CoSi_2$ grains grow laterally. Explain the disparity in these two velocities if the a → c transformation energy is 290 kJ/mol, the activation energy for diffusing species is 97 kJ/mol, the c-$CoSi_2$ grain size is 1000 Å, and the grain-boundary interfacial energy is 1 J/m^2. Make any reasonable assumptions you wish.

REFERENCES

1. M. Ohring, *Engineering Materials Science*. Academic Press, San Diego, 1995.
2. D. J. Srolovitz, *J. Vac. Sci. Technol.* **A4**, 2925 (1986).
3. J. E. Greene, in *Multicomponent and Multilayered Films for Advanced Microtechnologies: Techniques, Fundamentals and Devices*, eds. O. Auciello and J. Engemann. Kluwer, The Netherlands, 1993.
4. B. A. Movchan and A. V. Demchishin, *Phys. Met. Metallogr.* **28**, 83 (1969).
5. J. V. Sanders, in *Chemisorption and Reactions on Metal Films*, ed. J. R. Anderson. Academic Press, New York, 1971.
6. J. A. Thornton, *Ann. Rev. Mater. Sci.* **7**, 239 (1977).
7. H. T. G. Hentzell, C. R. M. Grovenor, and D. A. Smith, *J. Vac. Sci. Technol.* **A2**, 218 (1984).
8. R. Messier, A. P. Giri, and R. Roy, *J. Vac. Sci. Technol.* **A2**, 500 (1984).
9. Q. Tang, K. Kikuchi, S. Ogura, and A. Macleod, *J. Vac. Sci. Technol.* **A17**, 3379 (1999).

10. P. J. Kelly and R. D. Arnell, *J. Vac. Sci. Technol.* **A17**, 945 (1999).
11. P. J. Kelly and R. D. Arnell, *Vacuum* **56**, 159 (2000).
12. J. M. Nieuwenhuizen and H. B. Haanstra, *Philips Tech. Rev.* **27**, 8 (1966).
13. K. Hara, M. Komiya, T. Hashimoto, K. Okamoto, and H. Fujiwara, *J. Magn. Magn. Mater.* **73**, 161 (1988).
14. L. Abelmann and C. Lodder, *Thin Solid Films* **305**, 1 (1997).
15. N. Kuratani, A. Ebe, K. Ogata, I Shimizu, Y. Setsuhara, and S. Miyake, *J. Vac. Sci. Technol.* **A19**, 153 (2001).
16. H. Pulker, *Coatings on Glass*. Elsevier, Amsterdam, 1984.
17. S. Nakahara, *Thin Solid Films* **64**, 149 (1979).
18. B. W. Dodson, *Crit. Rev. Solid State Mater. Sci.* **16**, 115 (1990).
19. E. S. Machlin, *Materials Science in Microelectronics*, Vol. 1. Giro Press, Croton-on-Hudson, NY, 1995.
20. K.-H. Müller, *J. Appl. Phys.* **58**, 2573 (1985).
21. K.-H. Müller, *SPIE* **821**, 36 (1987).
22. K.-H. Guenther, *SPIE* **346**, 9 (1982).
23. K.-H. Müller, *Phys. Rev.* **35**, 7906 (1987).
24. C. C. Fang, V. Prasad, and F. Jones, *J. Vac. Sci. Technol.* **A11**, 2778 (1993).
25. K.-H. Müller, *J. Appl. Phys.* **61**, 2516 (1987).
26. R. Messier, *J. Vac. Sci. Technol.* **A4**, 490 (1986).
27. B. Yang, B. L. Waldon, R. Messier, and W. B. White, *SPIE* **821**, 68 (1987).
28. G. S. Bales, A. C. Redfield, and A. Zangwill, *Phys. Rev. Lett.* **62**, 776 (1989).
29. G. S. Bales and A. Zangwill, *J. Vac. Sci. Technol.* **A9**, 145 (1991).
30. C. V. Thompson, *Mat. Res. Soc. Symp. Proc.* **343**, 3 (1994), and R. Carel, C. V. Thompson, and H. J. Frost, *Mat. Res. Soc. Symp. Proc.* **343**, 49 (1994).
31. J. M. E. Harper, C. Cabral, P. C. Andricacos, L. Gignac. I. C. Noyan, K. P. Rodbell, and C. K. Hu, *J. Appl. Phys.* **86**, 2516 (1999).
32. A. E. Lita and J. E. Sanchez, *Phys. Rev. B* **61**(11), 7692 (2000).
33. J. M. E. Harper and K. P. Rodbell, *J. Vac. Sci. Technol.* **B15**, 763 (1997).
34. J. M. E. Harper, J. Gupta, D. A. Smith, J. W. Chang, K. L. Holloway, C. Cabral, D. P. Tracy, and D. B. Knorr, *Appl. Phys. Lett.* **65**, 177 (1994).
35. D. A. Smith, private communication.
36. B. D. Cullity, *Elements of X-Ray Diffraction*, 2nd ed. Addison Wesley, Reading, MA, 1978.
37. D. P. Field and D. J. Dingley, *Solid State Technol.* **38**(11), 91 (1995).
38. D. B. Knorr, *Mat. Res. Soc. Symp.* **309**, 75 (1993).
39. A. Gittis and D. Dobrev, *Thin Solid Films* **130**, 335 (1985).
40. R. W. Vook and F. Witt, *J. Vac. Sci. Technol.* **2**, 49 (1965).
41. K. P. Rodbell, V. Svilan, L. M. Gignac, P. W. Dehaven, R. J. Murphy, and T. J. Licata, *Mat. Res. Soc. Proc.* **428**, 261 (1996).
42. I. Petrov, F. Adibi, J. E. Greene, L. Hultman, and J.-E. Sundgren, *J. Appl. Phys.* **73**, 8580 (1993).
43. N. Sonnenberg, A. S. Longo, M. J. Cima, B. P. Chang, K. G. Ressler, P. C. McIntyre, and Y. P. Liu, *J. Appl. Phys.* **74**, 1027 (1993).
44. D. Barr, "Electron Backscatter Diffraction and Focused Ion Beam Studies of VLSI Interconnect Microstructure," Ph.D. thesis, Stevens Institute of Technology (1996).
45. J. E. Sanchez, private communication.
46. R. A. Powel and S. M. Rossnagel, *PVD for Microelectonics: Sputter Deposition Applied to Semiconductor Manufacturing, Physics of Thin Films*, Vol. 26. Academic Press, San Diego, 1999.
47. S. K. Dew, T. Smy, and M. J. Brett, *IEEE Trans. Electron Dev.* **39**, 1599 (1992).

48. L. J. Friedrich, D. S. Gardner, S. K. Dew, M. J. Brett, and T. Smy, *J. Vac. Sci. Technol.* **B15**, 1780 (1997).
49. L. J. Friedrich, S. K. Dew, and M. J. Brett, *J. Vac. Sci. Technol.* **B17**, 186 (1999).
50. M. J. C. van der Homberg, P. F. A. Alkemade, J. L. Hurd, G. J. Leusink, and S. Radelaar, *Mat. Res. Soc. Proc.* **391**, 397 (1995).
51. K. Robbie and M. J. Brett, *J. Vac. Sci. Technol.* **A15**, 1460 (1997).
52. K. Robbie, C. Shafai, and M. J. Brett, *J. Mater. Res.* **14**, 3158 (1999).
53. R. Messier, T. Gehrke, C. Frankel, V. C. Venugopal, W. Otano, and A. Lakhtakia, *J. Vac. Sci. Technol.* **A15**, 2148 (1997).
54. W. Buckel, *Z. Physik* **138**, 136 (1954).
55. Y.-T. Cheng and W. L. Johnson, *Science* **235**, 997 (1987).
56. L. L. Kazmerski and F. Abulfotuh, in *Thin Film Technology Handbook*, eds. A. Elshabini-Riad and F. D. Barlow. McGraw-Hill, New York, 1998.
57. H. S. Chen, H. J. Leamy, and C. E. Miller, *Ann. Rev. Mater. Sci.* **10**, 363 (1980).
58. S. R. Mader, in *The Use of Thin Films in Physical Investigations*, ed. J. C. Anderson. Academic Press, New York, 1966.
59. S. B. Newcomb and K. N. Tu, *Appl. Phys. Lett.* **48**, 1436 (1986).
60. S. Schneider, R. Busch, and K. Samwer, *Mat. Res. Soc. Symp. Proc.* **343**, 223 (1994).
61. A. S. Nowick and S. R. Mader, *IBM J. Res. Develop.* **9**, 358 (1965).
62. W. L. Bragg and J. F. Nye, *Proc. Roy. Soc.* **A190**, 474 (1947).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 10

Characterization of Thin Films and Surfaces

10.1 INTRODUCTION

Scientific disciplines are identified and differentiated by the experimental equipment and measurement techniques they employ. The same is true of thin-film science and technology. For the first half of this century interest in thin films centered on optical applications. The role played by films was largely a utilitarian one necessitating measurement of film thickness and optical properties. At first single films on thick substrates were involved. However, with the explosive growth of thin-film utilization in microelectronics there was an important need to understand the intrinsic nature of films in more complex materials environments. Increasingly, the benefits of multilayer film structures have been realized in an assortment of high-technology applications. Examples include multilayer metal and insulating films in microelectronics, compound semiconductor films in optoelectronics, dielectric-film stacks for optical coatings, and ceramic film layers in hard coatings.

With the increasingly interdisciplinary nature of applications, new demands for film characterization and property measurements in both individual films and multilayer coatings have arisen. It was this necessity that drove the creativity and inventiveness that culminated in the development of an impressive array of commercial analytical instruments. These are now ubiquitous in the thin-film, coating, and broader scientific communities. In many instances it was a question of borrowing and adapting existing

techniques employed in the study of bulk materials (e.g., X-ray diffraction, microscopy, mechanical testing) to meet the challenges posed by thin-film applications. In other cases well-known physical phenomena (e.g., electron spectroscopy, nuclear scattering, mass spectroscopy) were exploited.

A partial list of the modern techniques employed in the characterization of thin-film materials and devices appears in Table 10-1 (Ref. 1). Among their characteristics are the unprecedented structural resolution and chemical analysis capabilities over both small lateral and depth dimensions. Some techniques only sense and provide information on the first few atom layers of the surface. Others probe more deeply but in most cases depths of a micron or less are analyzed. Virtually all of these techniques require a high- or ultrahigh-vacuum ambient. Some are nondestructive while others are not. All of them utilize incident electron, or ion, or photon beams. These interact with the surface and excite it in such a way that some combination of secondary beams of electrons, ions, or photons are emitted, carrying off valuable structural and chemical information in the process. A rich collection of acronyms — a savory alphabet soup — has emerged to differentiate the various techniques. These abbreviations now widely pervade the thin-film and surface-science literature.

General testing and analysis of thin films is carried out with equipment and instruments that are wonderfully diverse in character. For example, consider the following extremes in their attributes:

1. *Size.* This varies from a portable desktop interferometer to the 50-foot long accelerator and beam line of a Rutherford backscattering (RBS) facility.
2. *Cost.* There is a wide variation in cost from the modest levels for test instruments required to measure electrical resistivity of films, to the near million-dollar price tag of an Auger spectrometer.
3. *Operating environment.* This varies from the ambient in the measurement of film thickness to the 10^{-10} torr vacuum required for the measurement of film-surface composition.
4. *Sophistication.* At one extreme is the manual Scotch-tape film peel-test to evaluate adhesion; at the other is an assortment of electron microscopes and surface analytical equipment where operation, data gathering, display, and analysis are computer controlled.

What is remarkable is that films can be characterized structurally, chemically, and with respect to various properties with almost the same ease and precision that we associate with bulk measurements. This despite the fact that there are many orders of magnitude fewer atoms available in films. To appreciate this, consider AES analysis of a Si wafer surface layer containing 1 at.% of an impurity. Only the top ~ 15 Å is sampled and since

Table 10-1
Analytical Techniques Employed in Thin-Film Science and Technology[a]

Primary beam	Energy range	Secondary signal	Acronym	Technique	Application
Electron	20–200 eV	Electron	LEED	Low-energy electron diffraction	Surface structure
	0.3–30 keV	Electron	SEM	Scanning electron microscopy	Surface morphology
	1–30 keV	X-ray	EMP (EDX)	Electron microprobe	Surface region composition
	500 eV–10 keV	Electron	AES	Auger electron spectroscopy	Surface layer composition
	100–400 keV	Electron	TEM	Transmission electron microscopy	High-resolution structure
	100–400 keV	Electron, X-ray	STEM	Scanning TEM	Imaging, X-ray analysis
	100–400 keV	Electron	EELS	Electron energy loss spectroscopy	Local small-area composition
Ion	0.5–2.0 keV	Ion	ISS	Ion-scattering spectroscopy	Surface composition
	1–15 keV	Ion	SIMS	Secondary ion mass spectroscopy	Trace composition vs depth
	1–15 eV	Atom	SNMS	Secondary neutral mass spectrometry	Trace composition vs depth
	1 keV and up	X-ray	PIXE	Particle-induced X-ray emission	Trace composition
	5–20 keV	Electron	SIM	Scanning ion microscopy	Surface characterization
	>1 MeV	Ion	RBS	Rutherford backscattering	Composition vs depth
Photon	>1 keV	X-ray	XRF	X-ray fluorescence	Composition (1 μm depth)
	>1 keV	X-ray	XRD	X-ray diffraction	Crystal structure
	>1 keV	Electron	ESCA, XPS	X-ray photoelectron spectroscopy	Surface composition
	Laser	Ion	—	Laser microprobe	Composition of irradiated area
	Laser	Light	LEM	Laser emission microprobe	Trace element analysis

[a] From Ref. 1.

state-of-the-art systems have a lateral resolution of 500 Å, the total measurement volume corresponds to $\frac{\pi}{4}$ $(500)^2$ $15 = 3 \times 10^6$ Å^3. In Si this corresponds to about 150,000 matrix atoms and therefore only 1500 impurity atoms are detected in the analysis! Such measurements pose challenges in handling and experimental techniques but the problems are usually not insurmountable.

This chapter will only address, with roughly equal coverage, the experimental techniques and applications associated with determination of:

1. Film thickness
2. Film and surface morphology and structure
3. Film and surface composition

These represent the common core of information required of all films and multilayer coatings irrespective of ultimate application. Within each of these three categories only the most important techniques will be discussed. Beyond these characteristics there are a host of individual properties (e.g., hardness, adhesion, stress, electrical conductivity) which are specific to the particular application. Measurement techniques for specific properties will therefore be addressed in the appropriate context within the book.

10.2 FILM THICKNESS

10.2.1 INTRODUCTION

Thickness of a film is among the first quoted attributes of its nature. The reason is that thin-film properties usually depend on thickness. Historically the use of films in optical applications spurred the development of techniques capable of measuring film thicknesses with high accuracy. In contrast, other important film characteristics, such as structure and chemical composition, were only characterized in the most rudimentary way until relatively recently. The actual film thickness, within broad limits, is not particularly crucial to function in some applications. Decorative, metallurgical, and protective films and coatings are examples where this is so. On the other hand, microelectronic applications generally require the maintenance of precise and reproducible film metrology, i.e., thickness as well as lateral dimensions. Even more stringent thickness requirements must be adhered to in multilayer optical-coating applications.

The varied types of films and their uses have generated a multitude of ways to measure film thickness. They are basically divided into optical and mechanical methods, and are usually nondestructive but sometimes destructive in nature. In the overwhelming majority of techniques convenience

requires that films be removed from the deposition chamber for measurement. However, a number of methods are either commercially available or can be experimentally adapted for real-time monitoring of film thickness during growth. We start with the theory and practice of optical techniques, subjects that are extensively treated in many books and references on thin films (Refs. 2–4).

10.2.2 Optical Methods for Measuring Film Thickness

Optical techniques for film-thickness determinations are widely used because they are applicable to both opaque and transparent films and generally yield thickness values of high accuracy. In addition, measurements are rapidly performed, frequently nondestructive, and utilize relatively inexpensive equipment. A list of optical techniques devised to measure the thickness and related optical properties of thin films is given in Table 10-2 together with relevant details. The first five are traditional methods that are essentially limited to measuring the thickness of a single film layer on a substrate. In recent years newer techniques such as spectral reflectometry and spectroscopic ellipsometry have supplanted these older methods. The principles of physical optics on which they are based have, of course, not changed in the interim. Capitalizing on advances in spectroscopy instrumentation, coupled with computer control and powerful computational analysis capabilities, these newer instruments have revolutionized the metrology of thin films. Importantly, they are now extensively used in integrated-circuit production environments to characterize multilayer film processing.

Interferometry and ellipsometry are the two basic optical methods for determining film thickness. In both, incident light impinges on the film surface and a reflected portion is measured. Interferometry relies on the interference of two or more beams of light, e.g., from the air/film surface and film/substrate interfaces, where the optical path difference is related to film thickness. Light generally impinges normal to the film surface and the instrumentation differs depending on whether opaque or transparent films are involved. Ellipsometry is not an interferometric technique. Rather, changes in the state of polarization of light reflected from surfaces at nonnormal incidence is measured.

The more data an experimental technique provides, the greater are the parameters it can measure with confidence. We shall see the truth of this in the progressive increase of information content in the optical techniques considered, first for thickness determinations in single films, and then in multifilm layers.

Table 10-2
Optical Techniques for Measuring Film Thickness

Method	Range (nm)	Characteristics[a]
1. Multiple beam interferometry (FET)	$3\text{–}2000 \pm 1\text{–}3$	A step and reflective coating required (1λ, $\theta = 90°$)
2. Multiple beam interferometry (FECO)	$1\text{–}2000 \pm 0.5$	A step, reflective coating, and spectrometer required; time consuming ($M\lambda$)
3. VAMFO	$80\text{–}1000 \pm 0.02\text{–}0.05\%$	Transparent films on reflective substrate (1λ, $M\theta$)
4. CARIS	$40\text{–}2000 \pm 1$ nm	Transparent films on reflective substrate ($M\lambda$, $\theta = 90°$)
5. Ellipsometry	<0.1–	Transparent films and multilayers, uses polarized light, measures d, n, and k (1λ, fixed θ)
6. Spectral reflectometry (unpolarized)	$\sim 30\text{–}2000 \pm 1$ nm	Transparent films and multilayers, fast, measures d, n, and k ($\lambda = \sim 200\text{–}1000$ nm, $\theta = 90°$) (polarized reflectometry is also performed at 1λ, $M\theta$)
7. Spectroscopic ellipsometry	<0.1–	Transparent films and multilayers, uses polarized light ($M\lambda$, fixed θ) (multiple-angle ellipsometry is also performed at 1λ)

[a] M = multiple, i.e., wavelengths (λ) or angles (θ).

10.2.3 INTERFEROMETRY

10.2.3.1 Opaque Films

For opaque films a sharp step down to the substrate plane must be first generated either during deposition through a mask or by subsequent etching. A neighboring pair of light rays reflected from the film/substrate will travel different lengths and interfere by an amount dependent on the step height. To capitalize on the effect, multiple beam interferometry, a technique developed by Tolansky (Ref. 5), is employed. This requires that the optical reflectance of both the film and substrate be very high as well as uniform. A metal such as Al or, better yet, Ag is evaporated over both film and substrate for this purpose. Interference fringes are generated by placing a highly reflective, but semitransparent, optically flat reference plate very

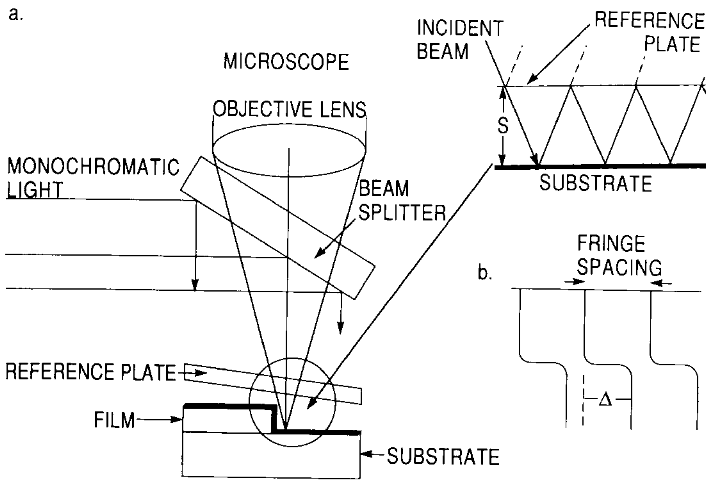


Figure 10-1 (a) Schematic view of experimental arrangement required to produce multiple-beam Fizeau fringes. (b) Fringe displacement at step.

close to the film-step-substrate region as shown in Fig. 10-1a. The two highly reflective surfaces are tilted lightly off-parallel enabling the light beam to be reflected in a zigzag fashion between them many times. A series of increasingly attenuated beams are now available for interference. This sharpens the resultant, so-called Fizeau fringes of equal thickness which can be viewed with a microscope.

The condition for constructive interference is that the optical path difference between successive beams be an integral number of wavelengths or

$$2S + 2\delta\lambda/2\pi = j\lambda. \tag{10-1}$$

Here S is the distance between film and flat, λ is the wavelength of the monochromatic radiation employed, and j is an integer. The phase change accompanying reflection, δ, is assumed to be the same at both surfaces and is taken to be π because of the high film reflectances. Therefore,

$$S = (j - 1)\lambda/2 \tag{10-2}$$

and the distance between maxima of successive fringes corresponds to $S = \lambda/2$. The existence of the step now displaces the fringe pattern abruptly by an amount Δ (Fig. 10-1b) proportional to the film thickness given by

$$d = \frac{\Delta}{\text{Fringe spacing}}\frac{\lambda}{2}. \tag{10-3}$$

For highly reflective surfaces, the fringe width is about $\sim 1/40$ of the fringe spacing. Displacements of about $\sim 1/5$ of a fringe width can be detected. For the Hg green line ($\lambda = 5640\,\text{Å}$) the resolution is therefore $(1/40)(1/5)(5640/2) = 15\,\text{Å}$. The resolution and ease of measurement are, respectively, influenced by the fraction of incident light reflected (R) and the fraction absorbed (A) by the film overlying the step. Raising R from 0.9 to 0.95 reduces the fringe width by half, whereas high A values reduce the fringe intensity.

The technique we have just described relies on fringes of equal thickness (FET). Fringes of equal chromatic order (FECO) form the basis of another optical method for measuring the thickness of opaque films (Ref. 2). A similar reflective surface and step are required, but white, rather than monochromatic, light is employed. Instead of a microscope the reflected light is spectrally analyzed by a spectrometer. Film thicknesses ranging from 1 to 2000 nm can be measured with an error of $\pm 5\,\text{Å}$.

10.2.3.2 Transparent Films

A perfectly suitable method for measuring the thickness of a transparent film is to first generate a step, metallize the film/substrate, and then proceed with either the FET or FECO techniques just discussed. However, transparent films are ideally suited for interferometry because interference of light occurs naturally between beams reflected from the two film surfaces. This means that a step is no longer required. Since different interfaces (the air/film and film/substrate) are involved for the beams that interfere, precautions must be taken to account for phase changes on reflection. Importantly, Fig. 10-2 summarizes what happens when monochromatic light of wavelength λ is normally incident on a transparent film/substrate combination. If n_1 and n_2 are the respective film and substrate indices of refraction, the intensity of the reflected light undergoes oscillations as a function of the optical film thickness or $n_1 d$. When $n_1 > n_2$, the maxima occur at film thicknesses equal to

$$d = \lambda/4n_1,\ 3\lambda/4n_1,\ 5\lambda/4n_1, \ldots \qquad (10\text{-}4)$$

For values of d halfway between these, the reflected intensity is minimum. When $n_1 < n_2$, a reversal in intensity occurs at the same optical film thickness. In Fig. 10-2 these results are shown for the case of a glass substrate ($n = 1.5$) and for films in which n_1 is either greater or less than 1.5. In order to exploit Fig. 10-2 for the measurement of film thickness, it is necessary to devise experimental arrangements so that the light-intensity oscillations can be revealed. Two of the techniques, VAMFO and CARIS, now primarily of historical interest (Ref. 2), have a useful pedagogical value.

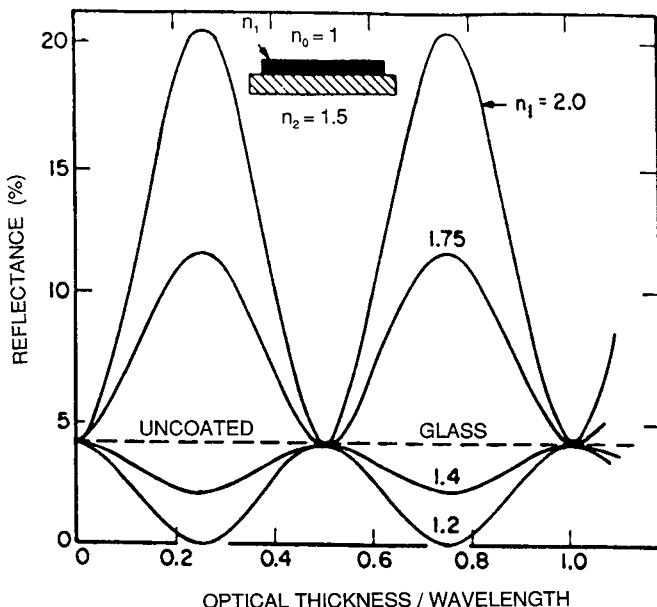


Figure 10-2 Calculated variation of reflectance (on air side) with normalized thickness ($n_1 d/\lambda$) for films of various refractive indices on a glass substrate of index 1.5. (From Ref. 3.)

In VAMFO (variable-angle monochromatic fringe observation) provision is made to vary the incident angle of monochromatic light on the film. As the stage and sample are rotated, bright (maxima) and dark (minima) fringes are observed on the film surface as the optical path changes. For the case of a transparent film on an absorbing reflecting substrate the film thickness is simply given by

$$d = \frac{N_0 \lambda}{2n_1 \cos\theta} \tag{10-5}$$

where θ is the angle of refraction in the film, and N_0 is the fringe order, which is measured by counting successive minima starting from perpendicular incidence. More accurate d values are obtained when the easier-to-detect intensity minima are measured ($N_0 = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, etc.). A disadvantage of VAMFO is that the refractive index of the film must be known at the wavelength of measurement. At best, film thicknesses ranging between 800 and 10,000 Å can be measured with a precision of 0.02 to 0.05%.

As FECO is to FET, so CARIS (constant-angle reflection interference spectroscopy) is to VAMFO. In CARIS, the wavelength of the incident light

rather than the angle of observation is systematically varied. Light is reflected from the film into a spectrometer with fringes being formed as a function of wavelength. For homogeneous films, the thickness (400–20,000 Å with 10 Å or 0.1% precision) is determined by measuring the number of fringes between two specified wavelengths.

10.2.3.3 Step Gauges

If there is a particular need to roughly estimate the thickness of one kind of film, it may be advantageous to construct a step gauge. Films of different but independently known thickness are deposited on the substrate of interest and are arrayed sequentially. Interference colors, observed when the specimen film is examined in reflected light, are matched to the color of the step gauge standard. For example, a step gauge for SiO_2 films on a Si wafer has proven to be useful in estimating film thickness to approximately 200 Å. To use such a gauge the spectral order must be known. In fluorescent light, for example, successive green-yellow colored oxides reappear at ~2000 Å intervals. An SiO_2 step gauge can be simply prepared by etching an oxidized Si wafer into a wedge shape by slowly lifting it from a dilute HF etchant in which it is immersed, at a constant velocity.

10.2.3.4 Spectral Reflectometry

Everything considered thus far with respect to interferometry has dealt with a single film on a thick substrate. Secondly, it has been assumed that the films are transparent so that light propagates without absorption. Therefore, the reflected intensity simply depends only on the index of refraction, n. But as noted earlier, multilayer film stacks are increasingly important. In the case of integrated circuits, dielectric and semiconductor layered stacks containing films such as SiO_2, silicon nitride, polysilicon, and photoresist on assorted substrates must be routinely characterized in production settings. Furthermore, all largely "transparent" films do absorb some light. Film-thickness measurement systems based on spectral reflectivity are currently employed to overcome the difficulties posed in measuring multiple, partially absorbing layers. In these systems a broad spectral range of wavelengths is employed to increase the information content required. This additional capability is needed because the films have wavelength-dependent optical properties.

To treat these complications we start by defining the *complex* index of refraction N, where

$$N = n - ik \qquad (i = \sqrt{-1}). \tag{10-6}$$

This form of N reflects an admixture of the ability of materials to both transmit (n) light at a speed of c/n (c = speed of light in vacuum) as well as partially absorb (k) it. The imaginary component k, known as the absorption or extinction coefficient, is a measure of the latter attribute. Its influence on reflectivity (R) of light from a surface can be appreciated if we recall from elementary physics that during normal incidence

$$R = (n - 1)^2/(n + 1)^2, \tag{10-7}$$

if there is no absorption. However, when k is nonzero,

$$R = \frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2}. \tag{10-8}$$

When films are present these formulas are more complex because of multiple reflections at internal interfaces, and the fact that n and k are both wavelength (λ) dependent. Unlike the simple case of Fig. 10-2, this causes the reflectivity of multilayer film stacks to undergo complex oscillations of variable periodicity and amplitude versus λ. An example of the reflectivity response is shown in Fig. 10-3a for a 442.8 nm thick film of Nb_2O_5 on fused silica.

What spectral reflectance methods attempt to do is extract values of film thickness (d), n, and k from the measurement of the normally reflected light intensity over a broad wavelength range. In working backward, this *normal*-incidence interferometry analysis first makes use of physical connections between n and k known as the Kramers–Kronig relations (Ref. 6); these enable n and k to be self-consistently determined. Then powerful desktop computer analysis uses these derived values of n and k to simulate a predicted R–λ dependence that is matched to the measured R–λ response. Values for d, n, and k are then reported for the iteration that yields the best "goodness of fit" to the data. The technique illustrated in Fig. 10-4 is quite simple experimentally and makes no use of microscopes, optical flats, or angular tilting of specimens, but rather fiber-optic spectrophotometry instead.

Key to the success of these methods is the assumptions made for n and k. Nonabsorbing films ($k = 0$) are usually modeled by the Cauchy approximation, i.e., $n(\lambda) = a + b/\lambda^2 + c/\lambda^4$ where a, b, and c are fitting parameters. Similarly, $k(\lambda) = a' \exp b'(1/\lambda - 1/c')$, with a', b', and c' fitting parameters, is sometimes used to model the absorption coefficient. For partially absorbing films, a set of energy ($E = hc/\lambda$)-dependent n and k values appropriate for a wide variety of semiconductors and dielectrics spanning the ultraviolet to the near infrared has the form (Refs. 7, 8)

$$n(E) = n(E = \infty) + \sum_i^q \frac{B_{0i}E + C_{0i}}{E^2 - B_iE + C_i} \tag{10-9}$$

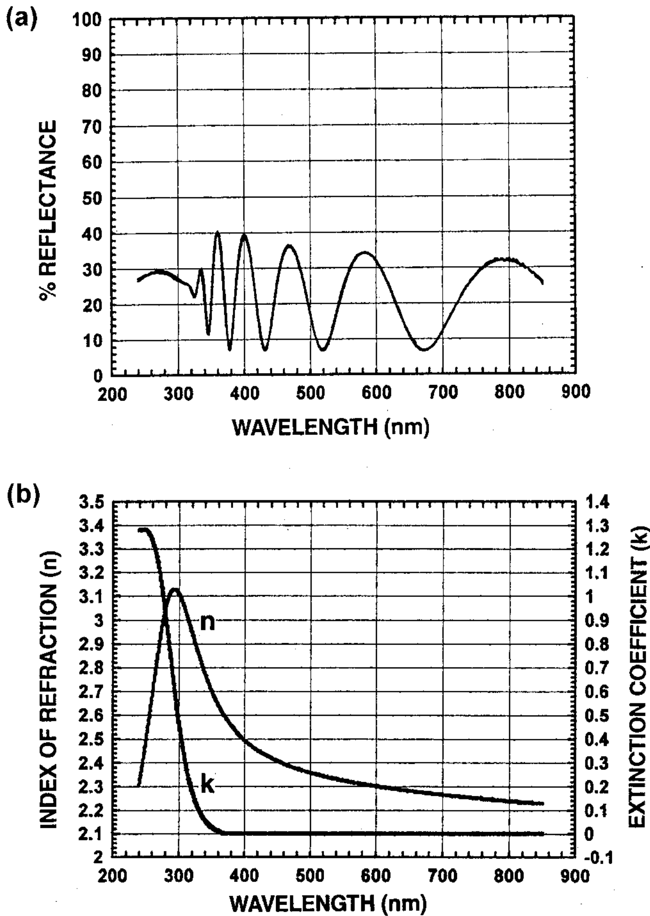


Figure 10-3 (a) Reflectivity vs wavelength for 442.8-nm thick Nb_2O_5 film on fused silica. (b) n and k vs wavelength for Nb_2O_5. (Reprinted with the permission of Scientific Computer International.)

and

$$k(E) = \sum_{i}^{q} \frac{A_i(E - E_g)^2}{E^2 - B_i E + C_i}. \tag{10-10}$$

Parameters A_i, B_i, and C_i are related to the electronic structure of the material; E_g is the energy bandgap; and B_{0i} and C_{0i} depend on these four previous quantities. The integer q specifies the number of terms to be used in the sum. Term $q = 1$ describes spectra from an amorphous structure while higher-order terms $q \geqslant 2$ arise from polycrystalline or crystalline material. Each term in the sum for $n(E)$ and $k(E)$ contributes either a peak or a

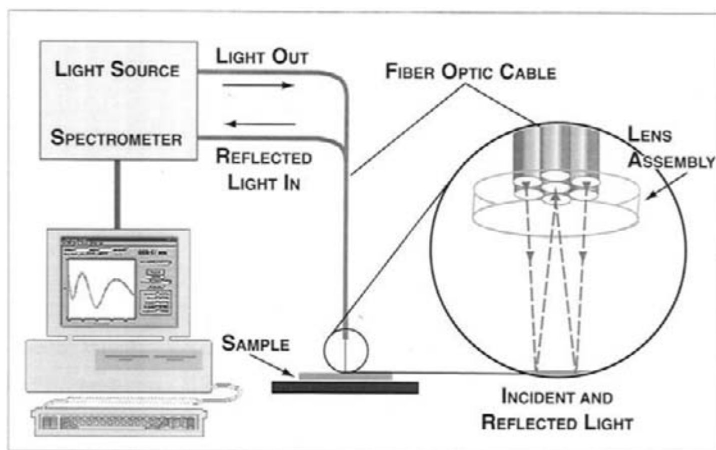


Figure 10-4 Experimental arrangement for spectral reflectivity. Light impinges on the film surface through the central optical fiber of the lens assembly. The reflected light signal is collected by the outer six fibers and analyzed by the spectrometer. A sophisticated program extracts film thickness and optical constants from the computed reflectivity that most closely matches the measured reflectivity data. (Courtesy of Filmetrics, Inc.)

shoulder to the respective spectra. The details of the calculation are complicated and well beyond the scope of this book. Results obtained for $n(\lambda)$ and $k(\lambda)$ in Nb_2O_5 are shown in Fig. 10-3b.

10.2.4 Ellipsometry

10.2.4.1 Introduction

Also known as polarimetry and polarization spectroscopy, the technique of ellipsometry is over a century old and associated with prominent scientists such as Brewster, Fresnel, Lord Rayleigh, and Drude (Ref. 9). Long used to obtain the thickness and optical constants of dielectric films, ellipsometry is now the preferred way to measure the optical properties of films and multilayer dielectric-film stacks. And as we shall see, it also possesses the capability of monitoring films *in situ* during deposition and processing (Refs. 10, 11). Because of its sensitivity in the submonolayer (sub-angstrom) thickness range and ability to function at high temperatures or pressures, through gases and liquids, etc., ellipsometry has been incorporated into plasma and CVD reactors, and MBE equipment, as well as electrochemical cells.

Ellipsometry consists of measuring and interpreting the change of polarization state that occurs when a polarized light beam is reflected at *non-normal* incidence from a film surface. Since ellipsometry does not rely on interference effects, film thickness determinations are not limited by the wavelength of light. Two variants of ellipsometry have gained great currency, namely multiple-angle-of-incidence and multiple-wavelength or spectroscopic ellipsometry. These will be briefly treated after some basics are presented and standard reflection ellipsometry is described.

10.2.4.2 Polarized Light and Its Interaction with Surfaces

Electromagnetic waves are transverse in nature consisting of mutually perpendicular electric and magnetic field vectors that are also perpendicular to the wave propagation direction. Only the electric-field vector $\mathscr{E}$ is pertinent to our discussion. Consider now a monochromatic light wave propagating in the plane of the page with two sinusoidally-oscillating components of $\mathscr{E}$, one parallel ($\mathscr{E}_{\mathrm{p}}$) to, and the other ($\mathscr{E}_{\mathrm{s}}$) perpendicular to the page plane. If these two sinusoidal field components of the same wavelength are in phase they add to yield plane-polarized light, but if they are out of phase elliptically polarized light emerges. As a special case, if the maximum of one coincides with the minimum of the other, the phase difference is 90° and we speak of circularly polarized light. Interestingly, when plane-polarized light reflects from a metal surface both components generally undergo different phase shifts and the emergent light is elliptically polarized.

In Fig. 10-5 plane-polarized light emanating from a medium whose refractive index is n_0 is incident on a bare surface of refractive index n_1. Angles of incidence (ϕ_1), reflection (ϕ_1), and refraction (ϕ_2), all measured relative to the surface normal, lie in and define the plane of incidence. Light rays in this plane consist of the transverse oscillating electric field components described above. Interaction with the surface causes both the phase and the amplitude of the reflected light (r) components to change, and relative to the incident (i) light it is both elliptically polarized and attenuated as shown. This contrasts with the case of light normally incident on a film/substrate combination where the distinction between parallel and perpendicular polarizations vanishes. Now, however, the two components of reflectivity, r_p and r_{s}, are distinct (Ref. 12). The ratio of their amplitudes, equal to a complex reflection coefficient (ρ), yields the fundamental equation of ellipsometry,

$$\rho = \frac{r_{\mathrm{p}}}{r_{\mathrm{s}}} = \frac{\mathscr{E}_{\mathrm{rp}}/\mathscr{E}_{\mathrm{ip}}}{\mathscr{E}_{\mathrm{rs}}/\mathscr{E}_{\mathrm{is}}} = \tan\psi\, e^{i\Delta}. \tag{10-11}$$

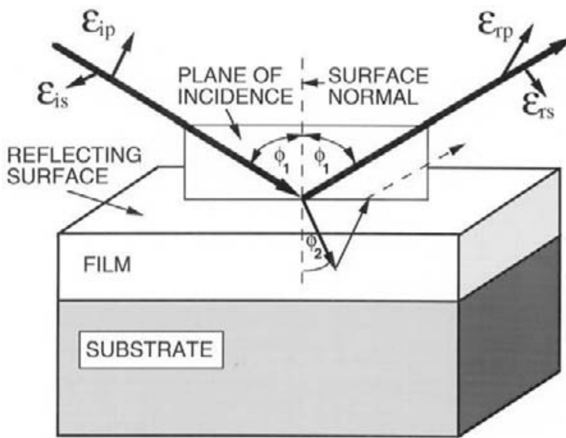


Figure 10-5 Reflection of a light beam from a surface. Refraction, reflection, and polarization effects are shown.

This means two quantities, $\tan\psi$ and Δ, must be measured; the former is related to the amplitude ratio, while Δ is the phase difference that develops between the s and p field components after reflection. A similar but more complex equation holds when a film is present on the substrate. Now Δ and ψ depend on λ, ϕ_1, and five material parameters, i.e., complex refractive indices (n and k) for both the film and substrate, and of course the film thickness. The objective is to extract either the film thickness, or alternatively, the film index of refraction, a task generally requiring computer analysis.

10.2.4.3 Reflection Ellipsometry

The experimental arrangement for standard single-wavelength ellipsometer measurements is shown in Fig. 10-6. First, the light source, which is normally a laser, is linearly polarized with a polarizer. Upon passing through a quarter-wave plate (compensator) the two field components emerge 90° out of phase so the light is circularly polarized. After impingement on the specimen surface the reflected light, now linearly polarized, is transmitted through a second polarizer which serves as the analyzer. The light intensity is monitored with a photomultiplier detector. Rotation of the polarizer and analyzer until a null or light extinction occurs enables Δ and ψ to be determined.

Data analysis is reasonably straightforward for smooth homogeneous films, but complications arise when the films are inhomogeneous, contain a phase mixture, or have rough surfaces and/or interfaces. For example,

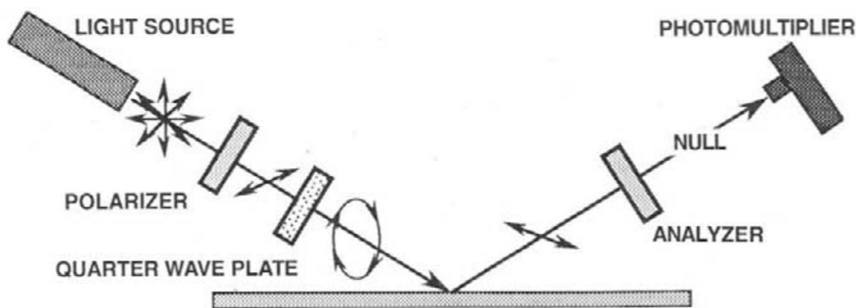


Figure 10-6 Experimental arrangement in ellipsometry. In null ellipsometry a quarter-wave plate is held in a fixed position prior to reflection. The polarizer and analyzer are rotated to obtain the null. (From Ref. 12. Copyright © 1993 by Academic Press, Inc. Reprinted with the permission of the publishers.)

surface roughness can be modeled as a film containing voids. In general, an optical response is computed from assumed models of the film structure and compared to experimental data using regression analysis to yield the desired parameters.

10.2.4.4 Multiple-Angle-of-Incidence Ellipsometry

When more independent measurements are required to determine the optical constants (n, k) of the film/substrate or multilayer-film/substrate and film thickness(es), one option is to employ multiple-angle-of-incidence (MAI) ellipsometry. This entails measuring Δ and ψ at a number of different ϕ_1 values. Among the advantages of MAI is the removal of order ambiguity when measuring the thickness of thick dielectric films. Order ambiguity arises because the phase difference is periodic and only a relative Δ within a particular period can be determined; an analogous situation arises when observing step gauges (Section 10.2.3.3). Although Δ and ψ vary at different ϕ_1 values, the optical constants remain the same. Jenkins (Ref. 13) has discussed a number of applications employing MAI ellipsometry, including studies of gold islands on GaAs, thick SiO_2 films on InP, and superlattices of $GaAs/GaP_xAs_{1-x}$.

10.2.4.5 Spectroscopic Ellipsometry

Of all the optical techniques for determining film thickness, spectroscopic ellipsometry (SE) is the most powerful providing both amplitude and phase change information across the wavelength spectrum. Just as the information content from a colored photograph exceeds that from a black-and-white one, by using many wavelengths a more extensive and precise determination

of optical constants is possible. The greater amount of raw data available in SE enables it to resolve the problem of *correlation* in thin-film optical systems. Correlation occurs because changes in both film thickness and index of refraction alter the measured spectra in similar ways. As a result it is difficult to disentangle the magnitudes of these two quantities without making assumptions. Both reflectometry and other ellipsometry methods which provide fewer data for analysis are subject to troublesome correlation effects. Having said this we note that a disadvantage of SE is the introduction of more variables because n and k vary with wavelength. Nevertheless, through the use of the Cauchy or similar approximations such problems are readily overcome. Finally, unlike other optical methods, SE requires neither reference samples nor a reference beam.

Commercial research grade spectroscopic ellipsometers operate over the UV–visible–IR range (200–1700 nm) and typically sample surface areas of $\sim 2000\ \mu$m in diameter. By improving instrument alignment and calibration, and overcoming nonlinearities in photodetectors, the accuracy of Δ and Ψ determinations has been increased to better than $\pm 0.01^\circ$. It is almost routine to obtain spectroellipsometric data over a wide wavelength range with high precision and accuracy in a time (seconds to minutes) which is very short compared to that for manual null instruments. In fact, real-time spectroscopic ellipsometers have been developed enabling millisecond determinations, a feature that is desirable when monitoring *in situ* thin-film deposition and etching.

A few examples will demonstrate the diversity and power of SE.

1. In the first (Ref. 14) a multilayer structure consisting of SiO_2 (25 Å)/c-Si + a-Si (120 ± 20 Å)/c-Si (550 ± 50 Å)/a-Si (250 ± 50 Å)/c-Si (substrate) was imaged in cross section within the transmission electron microscope, and the indicated film thicknesses were measured directly. The same structure was nondestructively probed by SE yielding SiO_2 (24 ± 3 Å)/c-Si + a-Si (119 ± 19 Å)/c-Si (511 ± 21 Å)/a-Si (270 ± 30 Å)/c-Si (substrate); note the capability of SE in distinguishing between amorphous (a) and crystalline (c) phases. With respect to characterization of multilayer structures, in addition to film thickness determinations SE has the capability of providing information on the composition of bulklike layers, the degree of film crystallinity, and the thicknesses of surface contaminant and microrough layers.

2. An example illustrating ellipsometric monitoring of real-time epitaxial growth is shown in Fig. 10-7 for the GaAs/GaAlAs system. The experiment was carried out in a CVD reactor at 600°C containing flows of AsH_3, $(CH_3)_3Ga$, and $(CH_3)_3Al$ gases past a GaAlAs film, illuminated by a He–Ne laser (6328 Å) at a 71° angle of incidence. After the Al species flow was stopped, growth of the GaAs film commenced. The resulting Δ and ψ

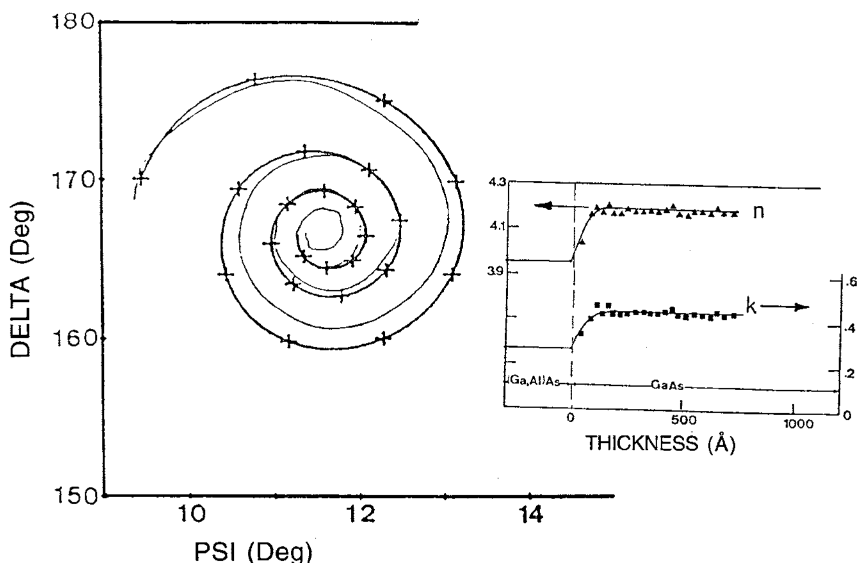


Figure 10-7 Ellipsometry measurement of AlGaAs–GaAs transition in the same organometallic CVD system. (a) Experimental trajectory of Δ and ψ, and comparison with theoretical growth model (solid line) with crosses corresponding to 100-Å increments in thickness. (b) Refractive index (n) and absorption coefficient (k) as a function of GaAs film thickness. (Reproduced with permission from Annual Reviews Inc. From Ref. 10.)

variation with film thickness (or time) is shown together with changes in the optical constants of GaAs. Within 150 Å the GaAlAs → GaAs growth transition is apparently complete.

3. As a final example, SE data for the early growth of a PECVD amorphous $Si_{0.65}Ge_{0.35}$:H film on a chromium substrate are depicted in Fig. 10-8. Of two possible growth modes that can account for the observed Δ–Ψ spiral trajectory, it appears the nucleation of hemispherical caps followed by coalescence rather than layer growth is the operative mechanism.

10.2.5 Mechanical Techniques for Measuring Film Thickness

10.2.5.1 Profilometry

Paralleling the operation of a phonograph needle on a record, profilometry consists of electromagnetically sensing the mechanical movement

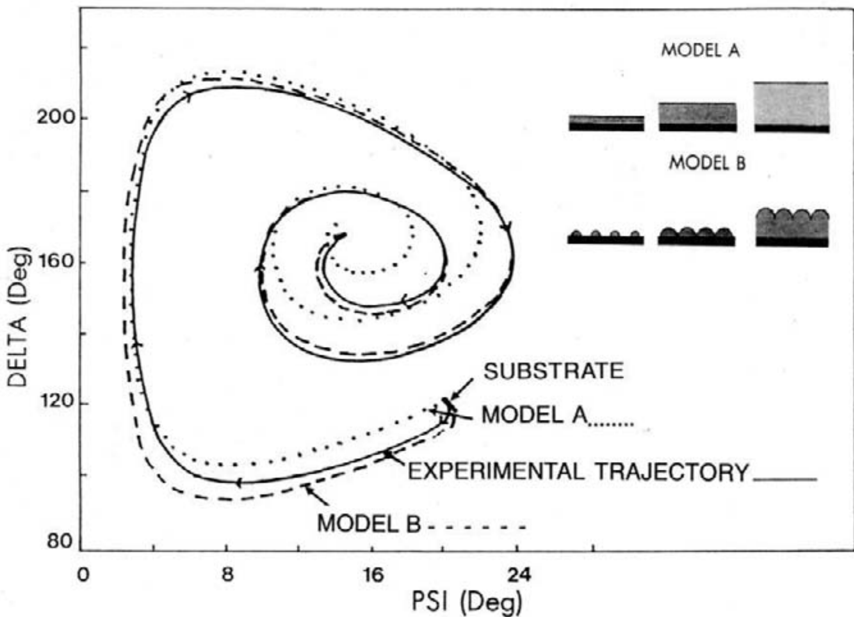


Figure 10-8 Experimental trajectory of Δ and ψ for the growth of an amorphous $Si_{0.65}Ge_{0.35}$:H film deposited by PECVD methods. Predicted trajectories for layer (Model A) and island growth (Model B) mechanisms are also indicated. (Courtesy of Instruments SA, Inc. Reprinted with permission.)

of a stylus as it traces the topography of a film–substrate step. The step can be made by masking or etching, during or after deposition, respectively. Diamond styli with cone angles of either 45° or 60° and tip radii ranging from 0.2 to 25 μm are commercially available for this purpose. The stylus force typically spans a range from 0.1 to 50 mg, enabling step heights from 50 Å to 800 μm to be measured; vertical magnifications of a few thousand up to ~1 million times are possible. Film thickness is directly read out as the height of the resulting step-contour trace. Vertical height resolutions of ~1 Å are possible, and the reproducibility of successive scans is within ±10 Å in a 1000-Å thick film. Several factors which limit the accuracy of stylus measurements are:

1. Stylus penetration and scratching of films. This is sometimes a problem in very soft films, e.g., In, Sn.
2. Substrate roughness. Excessive noise is introduced into the measurement as a result and this creates uncertainty in the position of the step.

3. Vibration of the equipment. Proper shock mounting and rigid supports are essential to minimize background vibrations.

In modern instruments the leveling and measurement functions are computer controlled. The vertical stylus movement is digitized and data can be processed to magnify areas of interest and yield best-profile fits. Calibration profiles are available for standardization of measurements.

Both thin films and thick coatings are amenable to profilometry. One of the interesting applications is to determine the flatness and depth of a sputtered crater during depth-profiling analysis by Auger electron spectroscopy (AES) (Section 10.4.4) or secondary ion mass spectrometry (SIMS) (Section 10.4.8). In this technique a circular region on the film surface is sputtered away and an electron (AES) or ion (SIMS) probe beam ideally analyzes the flat bottom of the crater formed. Such a crater profile generated during SIMS analysis is shown in Fig. 10-9 where the total depth sputtered exceeds 2.5 μm. Since the sputtering times are known, this information can be converted into an equivalent depth scale for use in determination of precise concentration profiles. If the crater walls are slanted rather than vertical, the analyzing beam may not sample the bottom flat surface but some portion of the side wall. This leads to errors in the depth profile which must be corrected for.

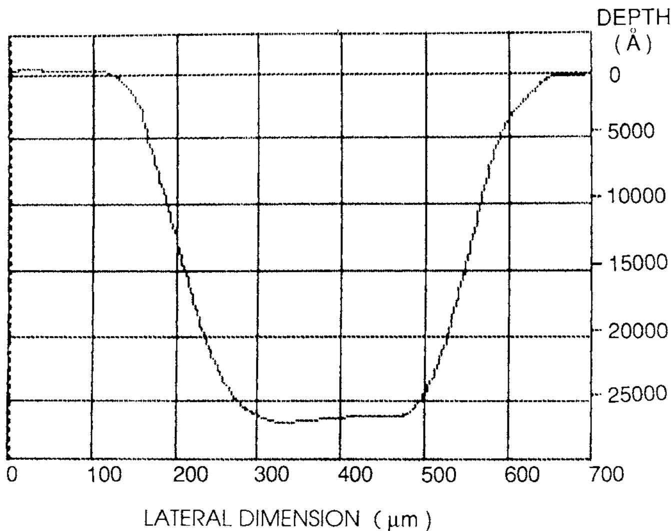


Figure 10-9 Profilometer trace of crater in InP produced during raster scanning of 12.5 keV Cs^+ ion beam across surface. (Courtesy of H. Luftman, AT&T Bell Laboratories.)

Related to probing sputtered craters is the demanding application of profiling trench depths in integrated circuits. Depending on trench width, the maximum depth that can be accessed depends on the stylus radius. For example, with a 0.2 μm tip radius traversing a 1 μm wide trench, a trench depth of 0.66 μm could be reached; with a 2.5 μm tip radius the maximum depth that can be probed is only 0.05 μm (Ref. 15).

10.2.5.2 Quartz Crystal Microbalances

The universally employed technique for *in situ* monitoring of the thickness of physical vapor-deposited films is based on the use of quartz crystal oscillators. Because the mass of the film is sensed the technique is also known as the quartz crystal microbalance method. Interestingly, very delicate and novel mechanical microbalances have been constructed and widely employed in the past to monitor film thickness during deposition. Microbalance designs have relied on such effects as the elongation of a thin quartz-fiber helix, the torsion of a wire, or the deflection of a pivot-mounted beam. Sensitive optical and, more commonly, electromechanical transducers and compensators for null measurements have been employed, enabling detection of film deposits weighing $\sim 10^{-8}$ g. If the film density is known, its thickness can be determined (Ref. 16).

To understand the operation of quartz crystal oscillators, consider homogeneous elastic plates set into mechanical vibration. These have resonant frequencies which depend on their dimensions, elastic moduli, and importantly on density. Additional mass in the form of a deposited thin film alters (lowers) the resonant frequency by effectively changing the properties of the composite vibrating plate. In practice, AT quartz crystals, i.e., cut $\sim 35^\circ$ with respect to the c axis, are employed. These same piezoelectric crystals are used for precise timing in watches and electronic instruments. They are powered through metal film electrodes on both wide faces of the crystal and are mounted within the deposition chamber close to the substrate. The fundamental frequency, f, of the shear vibrational mode is given by $f = v_q/2d_q$ where v_q is the elastic wave velocity and d_q is the plate thickness; d_q is also equal to half the wavelength of the transverse wave. If mass δm deposits on one of the crystal electrodes, the film/crystal thickness increases by an amount given by $\delta d = \delta m/A\rho_f$, where A and ρ_f are the film area and density, respectively. Combination of these two equations yields a frequency change given by

$$\delta f = -\frac{f^2 \delta m}{C \rho_f A} \tag{10-12}$$

where $C = v_q/2$ is defined as the frequency constant whose value is 1656 kHz-mm for AT cut quartz. In deriving Eq. 10-12 we have assumed that the addition of a small foreign film mass can be treated as an equivalent mass change of the quartz crystal. The formula is not rigorously correct because the elastic properties of the film are not the same as those of quartz and A is generally not equal to the total crystal face area. These effects greatly complicate the frequency-response analysis. Nevertheless, as long as the accumulated mass deposited on the crystal does not shift the resonant frequency by a few percent of its original value, δf varies linearly with δm.

To appreciate the kind of numbers involved, we note that a 6 MHz crystal is commonly employed. Since a frequency shift of 1 Hz is readily measurable, Eq. 10-12 reveals that this is equivalent to a mass of Al $= 1.24 \times 10^{-8}$ gms, and if $A = 1\ \text{cm}^2$, to a film thickness of 0.46 Å. This sensitivity is suitable for most applications. It can be enhanced, however, by more than an order of magnitude by employing thinner crystals with higher resonant frequencies, and by detecting smaller frequency shifts.

Since differences rather than absolute magnitudes are usually easier to detect with precision, the frequency shift is usually measured by beating the crystal signal against that from a reference (undeposited) crystal and subtracting. Quartz crystal oscillators are commonly employed for the measurement of deposition rate rather than film thickness. Therefore, commercial rate monitors contain circuitry to mathematically differentiate the frequency change with respect to time, display the rate, and provide feedback to control the power delivered to evaporation sources. In these functions it is essential to eliminate uncertainty in the frequency shift measurement. A potentially important source of error arises from the temperature increase of the crystal due to radiant-heat exposure from the evaporant source, and from the heat of condensation liberated by depositing atoms. Typically, temperature increases of a few degrees Celsius above that of the reference crystal result in frequency shifts of 10–100 Hz which are equivalent to a mass change of 10^{-7} to 10^{-6} g/cm^2. For this reason crystals are enclosed in water-cooled shrouds having a small entrance aperture to sample the evaporant stream. Lastly, precise work requires a correction due to the geometry of the monitor relative to the substrate.

10.2.5.3 Ultrasonic Multilayer-Film Metrology

Previously we noted that spectral reflectance (Section 10.2.3.4) and spectroscopic ellipsometry (Section 10.2.4.5) techniques enabled the accurate determination of the thickness of individual films within layered *dielectric* stacks. But this is not possible for opaque metal films because light cannot penetrate through them. In such cases film thickness is usually inferred from

sheet resistance measurements knowing the film resistivity. For a stack of metal films consisting of individual blanket layers this means tedious sequential determinations following successive depositions. However, because of exciting recent advances in instrumentation there appears to be a much better option. It is now possible to determine the thickness of each film in a multilayer stack in a single measurement that can be made in a few seconds!

To appreciate the significance of the technique, consider the following metrology requirements for metallizations during the processing of Si wafers (Ref. 17):

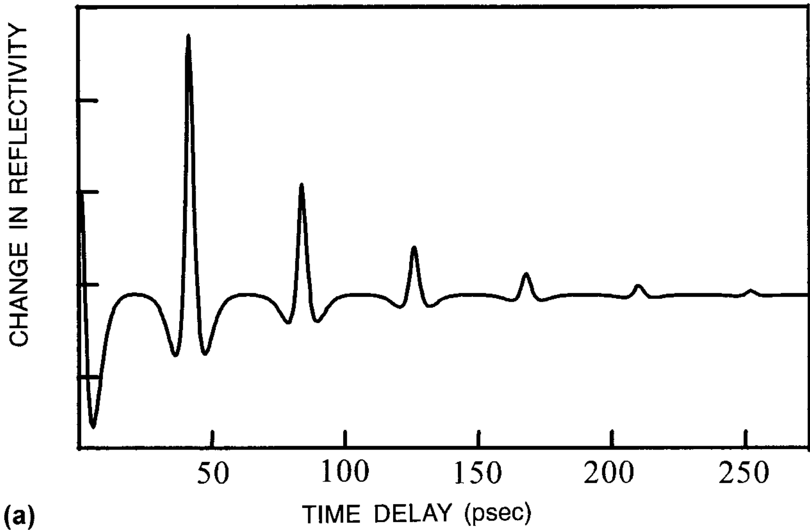
1. Capability of simultaneously measuring the thickness of up to six (or more) stacked optically opaque films
2. A noncontact and nondestructive technique
3. Probe spot sizes of 10 to 20 μm diameter which are compatible with the size of test structures
4. Measurement time no longer than the process deposition time
5. Film thickness determinations from ~ 2 nm to a few micrometers
6. A high degree of precision, accuracy (1–2%), and reproducibility (1%) in measurement
7. Robust operation and a high level of automation so that it can be used in manufacturing environments

Satisfying this wish list of requirements, particularly the first one, was unimaginable a few years ago. However, there are now practical commercial instruments capitalizing on picosecond ultrasonic laser-sonar methods that meet each of these demands (Ref. 18).

In this photoacoustic technique an ultrafast laser produces a short ($\sim 10^{-13}$ s) light pulse that impinges on the film surface. Absorption of this light causes a localized rapid thermal expansion that in turn launches an acoustic wave which propagates away from the surface with the speed of sound (v_s). The local temperature rise of a few degrees Celsius and subsequent heat dissipation over a few nanoseconds, based on a pulse energy of ~ 0.5 nJ, play no further role in the measurement. As the sound wave encounters subsurface film interfaces part of the wave reflects back while the remainder propagates further. When the echo from the first interface reaches the surface it changes its reflectivity by perhaps only one part in 10^5. A second light pulse, diverted from the initial laser pump light by a beam splitter, detects this change but the signal is delayed by the time (t) it takes sound to twice traverse the film thickness (d), i.e.,

$$t = 2d/v_s. \tag{10-13}$$

The measurement, with characteristics of both sonar and interferometry, is capable of measuring distance traveled to within an angstrom.



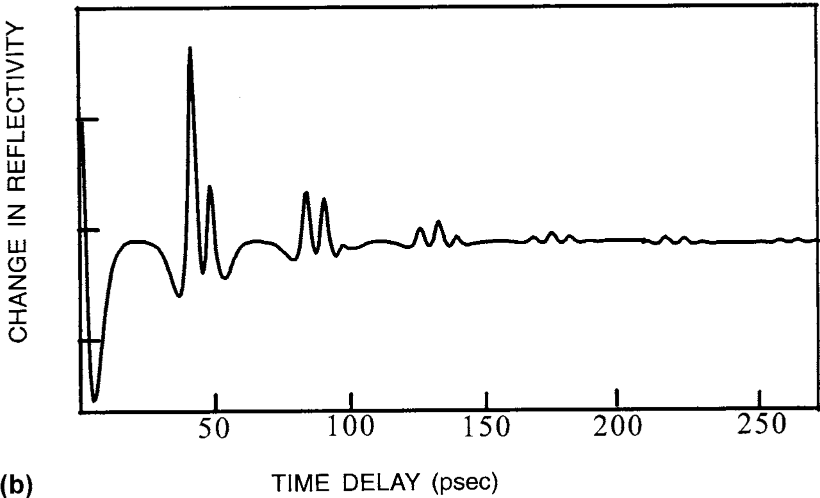


Figure 10-10 Time-dependent change in reflectivity for (a) 200 nm TiN/Si and (b) 200 nm TiN/20 nm Ti/Si. (Courtesy of G. J. Collins, Rudolph Technologies Inc. From Ref. 17.)

In Fig. 10-10a we see that a time of 42 ps elapses between a sequence of echoes obtained from a TiN film deposited on a Si substrate. If it is independently known that $v_s = 9.5$ nm/ps (9500 m/s), simple calculation yields a film thickness of 200 nm. Now consider a 20 nm Ti film sandwiched between the TiN and Si. A more complicated reflectivity response is

recorded in Fig. 10-10b because sound now reflects first from the TiN/Ti interface and then from the Ti/Si interface. The 6.5-ps interval between the first and second peaks corresponds to the time it takes sound to travel through the buried Ti film and echo back from the Si surface. From this information we deduce that v_s is equal to 2×20 nm/6.5 ps or 6150 m/s in Ti.

In addition to film thickness measurement, picosecond ultrasonics have the important capability of characterizing more subtle material and structural properties such as film density and interfacial roughness and adhesion. Such information is provided by the relative amplitudes of successive echoes and the shape and widths of the reflected signals. For example, the echo amplitude decay rate depends on the acoustic reflection coefficient (R_a), a quantity that is film-density dependent. In addition, R_a is also a sensitive measure of the strength of interfacial bonding. A film that is poorly bonded to another film or the substrate inefficiently transfers acoustic energy across the interface. Therefore, more energy is reflected and R_a would be larger than the value obtained for a clean interface. Images of R_a as a function of position across wafer surfaces can thus assess the quality of blanket films on wafers. Analogous ultrasonic imaging methods have been very successfully employed in exposing delamination of chips from substrates in plastic encapsulated packages (Ref. 19). Other applications to multilayer films have been reviewed by Stoner *et al.* (Ref. 20).

10.3 STRUCTURAL CHARACTERIZATION OF FILMS AND SURFACES

10.3.1 INTRODUCTION

There are several hierarchies of structural information which are of interest to thin-film scientists and technologists in research, process development, and reliability and failure analysis activities. These are listed next in roughly ascending order of structural resolution required.

1. The first broadly deals with metrology of patterned films where issues of lateral or depth dimensions and tolerances, uniformity of thickness and coverage, completeness of etching, etc., are of concern.
2. Secondly, film surface topography and microstructure in plan view including grain size and shape, existence of compounds, presence of hillocks or whiskers, evidence of film voids, microcracking or lack of adhesion, etc., are of interest.
3. Next are cross-sectional views of multilayer structures exposing interfacial regions, columnar grain morphology, and substrate interactions. Such

images are crucial in microelectronics and optical coating technologies, enabling direct verification of device dimensions and structural defects in devices.

4. Somewhat more challenging are the high-resolution lattice images of both plan-view and transverse film sections. Among the applications here are defect structures in films and devices, structure of grain boundaries, identification of phases, and a host of issues related to epitaxial structures, e.g., the crystallographic orientations, direct imaging of atoms at interfaces, interfacial quality and defects, and perfection of assorted thin film superlattices.

5. Finally, there is structural information required of substrate and film surfaces such as the details of film-nucleation and growth processes, crystallography of deposits, and reconstruction of surface layers.

In addressing all these needs there is a growing arsenal of techniques that includes optical, electron, and scanning probe microscopes. Thus the scanning electron microscope (SEM), and occasionally, the reflection metallurgical microscope would primarily be used to obtain the required information included under items 1, 2, and 3. The transmission electron microscope (TEM) addresses item 3 as well. Augmented perhaps by X-ray diffraction methods, the TEM is indispensable for item 4. And finally to confront item 5 there is the atomic force microscope (AFM) and the scanning tunneling microscope (STM).

This section discusses those techniques that have the common objective of structurally characterizing films and surfaces. Together with a description of their operation and an introduction to the interpretation of results they provide, the advantages and disadvantages of each will be noted.

10.3.2 Scanning Electron Microscopy

Because seeing is believing and understanding, the SEM is perhaps the most widely employed thin-film and coating characterization instrument (Refs. 21, 22). There is an interesting distinction between the SEM and TEM. The latter is a true microscope in that all image information is acquired simultaneously or in parallel. In the SEM, however, only a small portion of the total image is probed at any instant and the image builds up serially by scanning the probe. Strictly speaking, the SEM has more in common with the scanning Auger electron (Section 10.4.4) and SIMS (Section 4.4.8) microprobes than a traditional microscope.

A schematic of the typical SEM is shown in Fig. 10-11. Electrons thermionically emitted from a tungsten or LaB_6-cathode filament are drawn

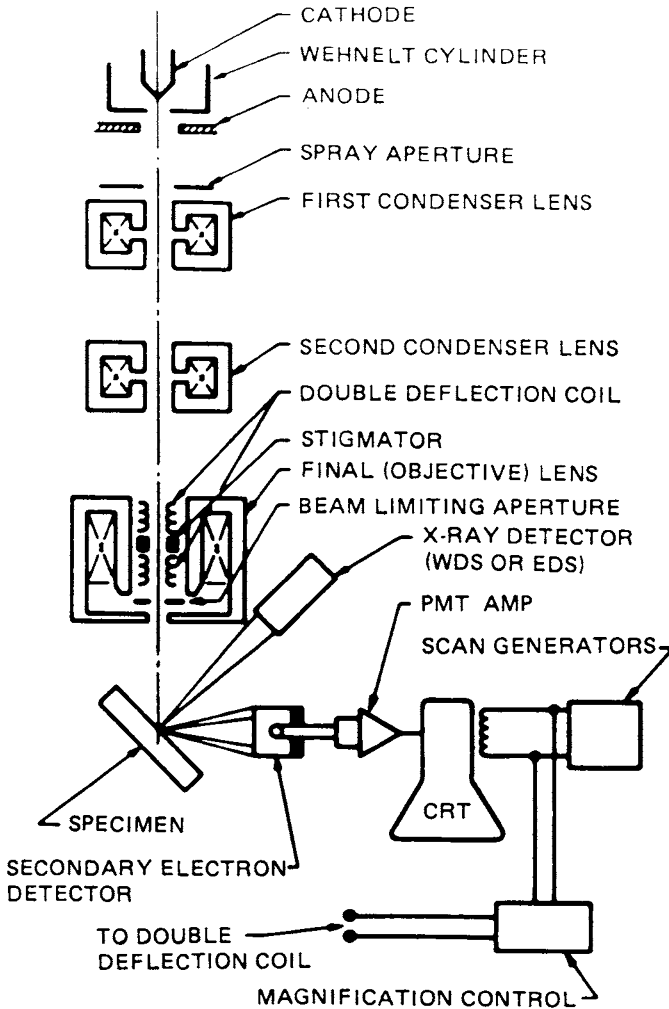


Figure 10-11 Schematic of the scanning electron microscope. (From Ref. 21 with permission from Plenum Publishing Corp.)

to an anode and focused by two successive condenser lenses into a beam with a very fine spot size that is typically 10 Å in diameter. Pairs of scanning coils located at the objective lens deflect the beam either linearly or in raster fashion over a rectangular area of the specimen surface. Electron beams having energies ranging from a few keV to 50 keV, with 30 keV a common

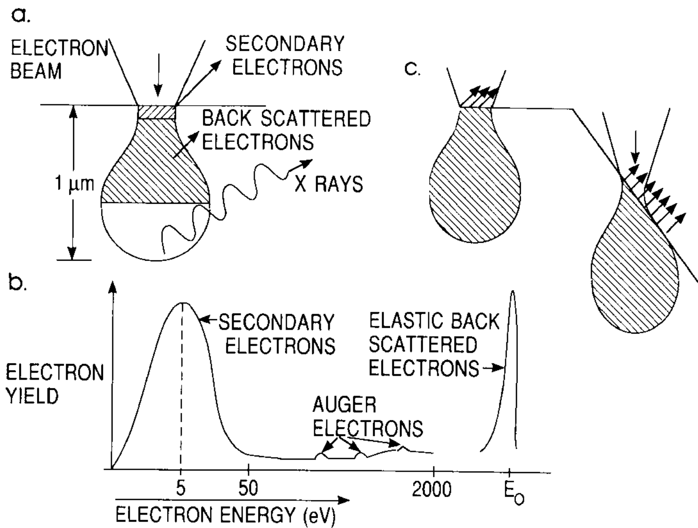


Figure 10-12 (a) Electron and photon signals emanating from tear-shaped interaction volume during electron-beam impingement on specimen surface. (b) Energy spectrum of electrons emitted from specimen surface. (c) Effect of surface topography on electron emission.

value, are utilized. Upon impinging on the specimen, the primary electrons decelerate and in losing energy, transfer it inelastically to other atomic electrons and to the lattice. Through continuous random scattering events the primary beam effectively spreads and fills a teardrop-shaped interaction volume (Fig. 10-12a) with a multitude of electronic excitations. The result is a distribution of electrons which manage to leave the specimen with an energy spectrum shown schematically in Fig. 10-12b. In addition, target X-rays are emitted and other signals such as light, heat, and specimen current are produced; the sources of their origin can be imaged with appropriate detectors. Various SEM techniques are differentiated on the basis of what is subsequently detected and imaged. Several of these are indicated next.

10.3.2.1 Secondary Electrons

The most common imaging mode relies on detection of this very lowest portion of the emitted energy distribution. Their very low energy means they originate from a subsurface depth of no larger than several angstroms. The signal is captured by a detector consisting of a scintillator/photomultiplier combination and the output serves to modulate the intensity of a cathode-ray tube (CRT) which is rastered in synchronism with the raster-scanned

primary beam. Image magnification is then simply given by the ratio of scan length on the CRT to that on the specimen. Resolution specifications quoted on research-quality SEMs are less than 2 nm. Great depth of focus enables images of beautiful three-dimensional quality to be obtained from non-planar surfaces. The contrast variation obtained can be understood with reference to Fig. 10-12c. Sloping surfaces produce a greater secondary-electron yield because the portion of the interaction volume projected on the emission region is larger than on a horizontal surface. Similarly, edges will appear even brighter. Several examples of secondary-electron SEM images have been reproduced in various places throughout the book.

10.3.2.2 Backscattered Electrons

These are the high-energy electrons which are elastically scattered and essentially possess the same energy as the incident electrons. The probability of backscattering increases with the atomic number Z of the sample material. Since the backscattered fraction is not a very strong function of Z (varying very roughly as $\sim 0.05Z^{1/2}$) elemental identification is not feasible from such information. Nevertheless, useful contrast can develop between specimen regions that differ widely in Z. Since the escape depth for high-energy backscattered electrons is much greater than for low-energy secondaries, there is much less topological contrast in the images,

An interesting phenomenon which makes use of backscattered electrons is electron backscatter diffraction (EBSD), also known as backscatter Kikuchi diffraction (BKD). Since an important application to the microtexture of polycrystalline films was previously described in Section 9.4.5, it is worth briefly noting the principle of the technique here. As the finely focused electron beam of an SEM (Fig. 10-13a) penetrates a crystalline specimen, i.e., a grain, electrons are inelastically scattered and lose angular correlation with the primary beam. In the process, a point source of electrons effectively forms and it Bragg diffracts from the sample lattice. Electrons, scattered at large angles in the shape of conelike beams, intercept a flat detector and produce an image consisting of intersecting hyperbolic sections as shown in Fig. 10-13b. Computer analysis of these geometrically complex EBSD patterns enables the crystallographic orientation of individual grains to be extracted. As a result, a microtexture pole figure can build up point by point as the beam scans from grain to grain.

10.3.2.3 X-Rays

An SEM is like a large X-ray vacuum tube used in conventional X-ray diffraction systems. Electrons emitted from the filament (cathode) are accelerated to high energies where they strike the specimen target (anode).

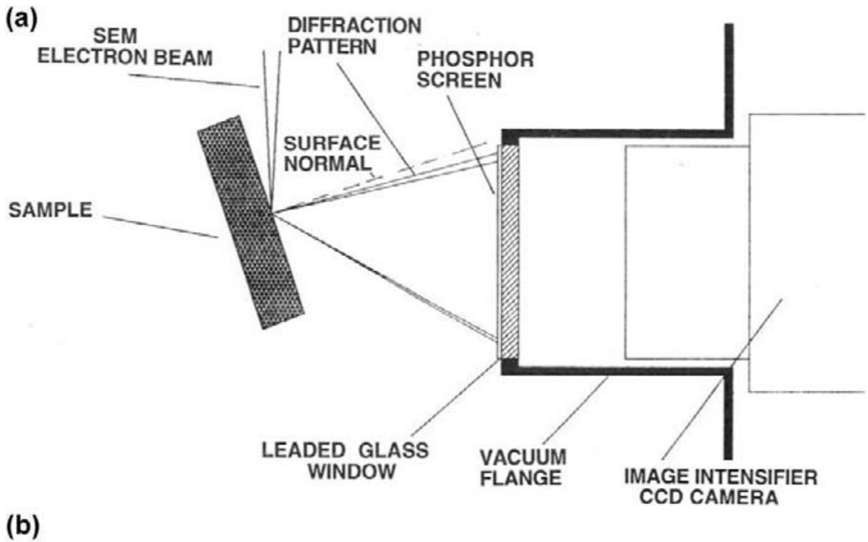


Figure 10-13 (a) Schematic of SEM based EBSD detection system. (b) Typical EBSD pattern of an aluminum grain taken with 17 keV electrons. The pole at lower right is (101). (From D. Barr, Doctoral Thesis, Stevens Institute. Reprinted with permission of the author.)

In the process, X-rays characteristic of atoms in the irradiated area are emitted. By analyzing their energies the atoms can be identified and by counting the numbers of X-rays emitted, the concentration of atoms in the specimen can be determined. Instrumentation for this major X-ray spectroscopy technique, known as X-ray energy dispersive analysis (EDX), is practically always attached to the SEM column. X-ray spectroscopy will be further discussed in Section 10.4.3.

10.3.3 Transmission Electron Microscopy

10.3.3.1 Introduction

Normally a research tool, the transmission electron microscope (TEM) is indispensable for the structural imaging of nanometer-sized features. In comparison, the resolution of the SEM is about an order of magnitude poorer. As its name implies, the TEM is used to obtain structural information from specimens thin enough to transmit electrons. Thin films are, therefore, ideal for study but they must be removed from electron-impenetrable substrates prior to insertion in the TEM.

As a gross simplification, the TEM may be compared to a slide projector with the slide (specimen) illuminated by light (electron beam) that first passes through the condenser lens (electromagnetic condenser lens). The transmitted light forms an image that is magnified by the projector lens (electromagnetic objective and projector lenses) and viewed on a screen (or photographed). In operation, electrons are thermionically emitted from the gun and typically accelerated to anywhere from 125 to 300 keV, or higher (e.g., 1 MeV) in some microscopes. High magnification in TEM methods is a result of the small effective wavelengths (λ) employed. According to the deBroglie relationship,

$$\lambda = h/(2mqV)^{1/2} \tag{10-14}$$

where m and q are the electron mass and charge, and V is the potential difference. Electrons of 100 keV energy have wavelengths of 0.037 Å and are capable of effectively transmitting through about 0.6 μm of Si. In operation, electrons are projected onto the specimen by the condenser lens system. The scattering processes they undergo during their passage through the specimen determine the kind of information obtained. Elastic scattering, involving no energy loss when electrons interact with the potential field of the ion cores, gives rise to diffraction patterns. Inelastic interactions between beam and matrix electrons at heterogeneities such as grain boundaries, dislocations, second-phase particles, defects, and density variations cause

complex absorption and scattering effects leading to a spatial variation in the intensity of the transmitted beam. Primary and diffracted electron beams that emerge from the specimen are now made to pass through a series of postspecimen lenses. The objective lens produces the first image of the object and is, therefore, required to be the most perfect of the lenses. Just how the beams reaching the back focal plane of the objective lens are subsequently processed distinguishes the different methods of operation.

10.3.3.2 Operational Modes

There are two very broad modes of TEM operation, namely imaging and analytical. In the former, structural images ranging from low magnification to atomic resolution are directly revealed, whereas in the latter structural information is indirectly revealed through analysis of diffracted beam geometries and energies. These modes are differentiated in Figs. 10-14a and 10-14b, where the lenses, beam path, and method of image formation are schematically indicated. Each of the techniques depicted is briefly described in turn below.

1. *Bright-field imaging* (*BF*). Also known as conventional TEM, bright-field imaging intentionally excludes all diffracted beams and only allows the central beam through. This is done by placing suitably sized apertures in the back focal plane of the objective lens. Intermediate and projection lenses then magnify this central beam to provide images of the microstructure and morphology of features.

2. *Dark-field imaging* (*DF*). Dark-field images are also formed by magnifying a single beam; this time one of the diffracted beams is chosen by means of an aperture which blocks both the central beam and the other diffracted beams. The micrograph of alternating 40 Å wide films in the GaAs–$Al_{0.5}Ga_{0.5}As$ superlattice structure shown in Fig. 10-15 is a dark-field image employing the 200 diffracted beam. In both bright- and dark-field imaging we speak of amplitude contrast because diffracted beams with their phase relationships are excluded from the imaging sequence.

3. *Lattice imaging*. In this now-popular method of imaging periodic structures with atomic resolution, the primary transmitted and one or more of the diffracted beams are made to recombine, thus preserving both beam amplitudes and phases. High-resolution lattice images of the interfacial region between epitaxial $CoSi_2$ and Si (Fig. 1-4) and the polysilicon/oxide/silicon structure (Fig. 11-10) are examples of this remarkable technique.

4. *Diffraction*. Diffraction yields crystallographic and orientation effect information on structural features, defects, and phases. Microdiffraction in localized regions is conveyed in patterns that look much like the spotted arrays in Fig. 8-33 or Fig. 9-24 when crystalline material is analyzed. For

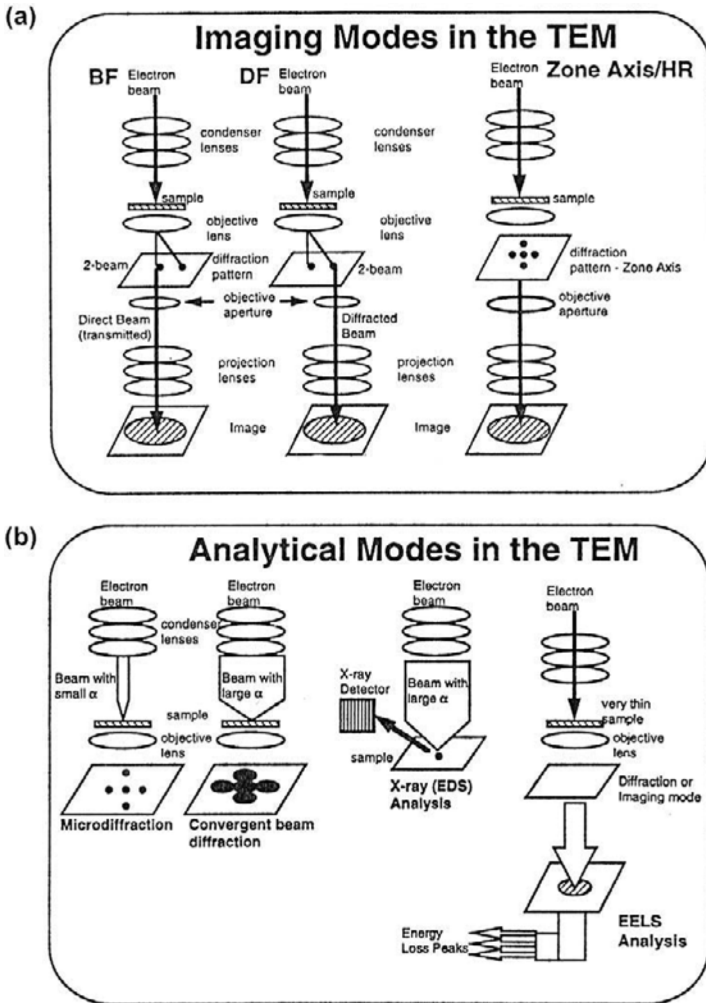


Figure 10-14 Two broad modes of TEM operation. (a) Imaging modes. (b) Analytical modes. (From V. S. Kaushik, IEEE International Reliability Physics Symposium, Tutorial Notes, p. 3.64, 1994. Reprinted with permission of the author.)

polycrystalline and amorphous materials the spots give way to diffraction rings of varying sharpness and width. Convergent-beam electron diffraction enables detection of less than $\sim \pm 0.001$ Å variations in lattice parameter with high precision. This sensitivity makes it possible, for example, to measure deviations in texture orientations and strains in individual grains of polycrystalline films.

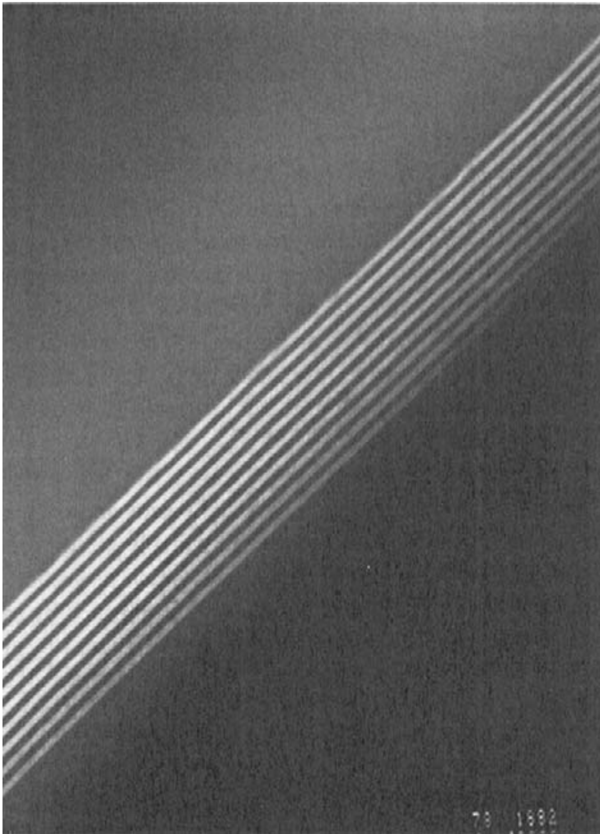


Figure 10-15 Dark-field TEM image of alternating 40 Å wide GaAs–$Ga_{0.5}Al_{0.5}As$ superlattice films. Light bands contain Al. (Courtesy of S. Nakahara, AT&T Bell Laboratories.)

5. *X-ray spectroscopy*. This analytical capability, otherwise known as X-ray energy dispersive analysis (EDX), allows elemental identification through measurement of characteristic X-ray energies. Core electron transitions which are discussed in Section 10.4.2 are the basis of the technique.

6. *Electron energy loss spectroscopy* (*EELS*). Through energy analysis of the transmitted electron beam, composition analysis is possible. EELS is particularly useful for detecting low-Z elements.

The theory and practice of electron optics and various TEM techniques are beyond our scope but have been detailed in several excellent books on the subject (Refs. 23, 24).

10.3.3.3 TEM Cross-Sections

Perhaps the most dramatic images of film structures and devices are those taken in cross section. In this important technique, specimens that are already thin in one dimension are now thinned in the transverse direction; it is like imaging this page in edge view rather than in the plane the words appear on. What is involved in the case of integrated circuits is cleaving a number of wafer specimens transversely, bonding these slivers in an epoxy button, and thinning them by grinding and polishing. Finally, the resulting disk is ion milled until a hole appears. By preparing many specimens simultaneously the probability of capturing images from desired device features is enhanced.

As an outstanding example of a TEM cross-sectional image, consider the nanotransistor shown in Fig. 10-16. This field-effect transistor structure has

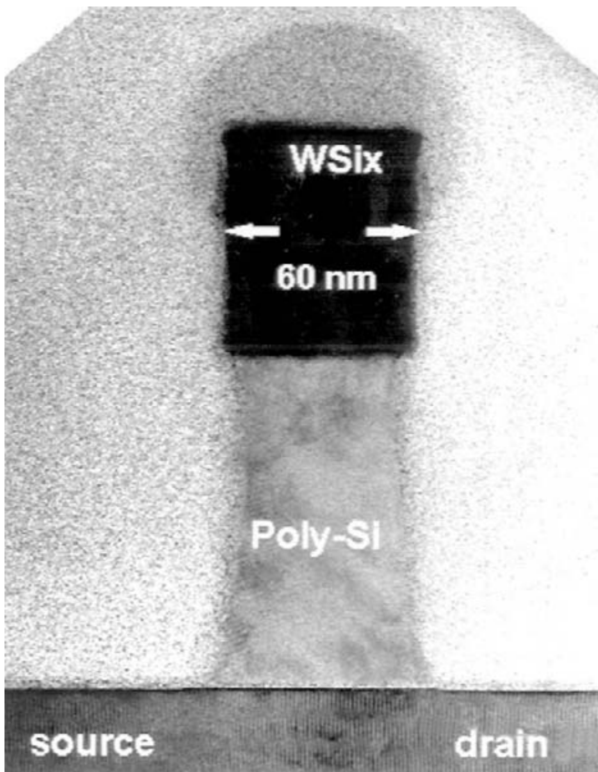


Figure 10-16 A TEM cross-sectional view of a nano-field-effect transistor thinned by conventional (not FIB) methods. (Courtesy of Frieder Baumann, Bell Laboratories, Lucent Technologies.)

a 61 nm channel (182 atoms long) and a gate oxide, barely visible beneath poly-Si, that is only 1.2 nm or about three atoms thick (see Fig. 11-10). Smaller features together with shallower source- and drain-junction depths, high-aspect-ratio polysilicon, and gate electrodes can be expected in the future. Lower operating voltages and higher transconductances of nanotransistors mean impressive power savings of 60–160 times relative to present-day devices.

10.3.3.4 Focused Ion Beam Microscopy

The use of focused ion beams (FIB) represents a recent major advance in TEM sample preparation as well as for microscopy in its own right. Briefly, the FIB resembles an SEM but instead of electrons, gallium ions generated in a liquid-metal ion source are employed as the imaging vehicle. In the ion-optical column these ions are electrostatically focused into a fine beam that raster scans the specimen surface releasing secondary electrons for structural imaging. When operated as a high-resolution microscope the FIB has some advantages as well as disadvantages relative to the SEM and TEM. The FIB image of an Al–0.5% Cu film shown in Fig. 10-17a reveals grain orientation contrast unlike secondary electron SEM imaging; also unlike TEM methods, thick specimens that require no sample preparation can be imaged by FIB microscopy. In Fig. 10-17b the SEM electron backscatter (EBSD) image (see Sections 9.4.5 and 10.3.2.2) demonstrates how well the techniques correlate with each other.

At the same time the Ga ions help form images, they contaminate specimens, introduce artifacts, and sputter away the material being observed. The latter undesirable feature for imaging is capitalized upon in preparing TEM cross sections, where it has eased the drudgery and eliminated the element of chance in locating features for examination (Ref. 25). Operating like a milling machine, the 10 to 25 keV Ga beam precisely etches material from submicron regions of specimens. Suitably thinned features can then be subsequently examined in the TEM. This sample preparation technique illustrated in Fig. 10-18 has greatly advanced structural characterization methodology in scientific and technological, e.g., failure analysis, applications.

10.3.4 X-Ray Diffraction

This very important experimental technique has long been used to address all issues related to the crystal structure of bulk solids including

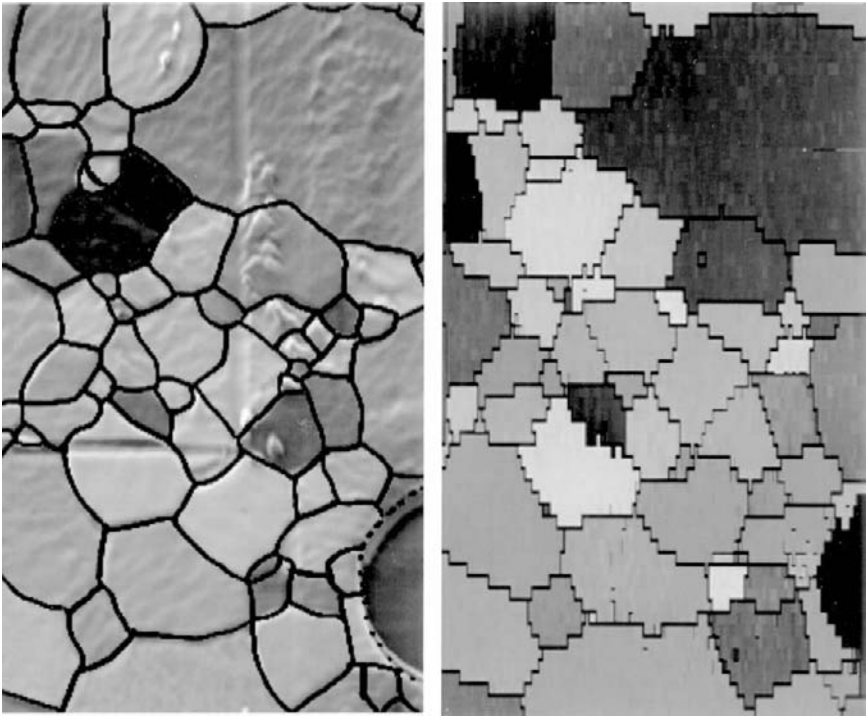


Figure 10-17 (Left) FIB secondary electron image of Al–0.5% Cu blanket film taken with a 30 keV Ga beam. *Note:* A grain tracing is overlayed. (Right) EBSD pattern map of same area. (Courtesy of L. Gignac, IBM, T. J. Watson Research Center.)

lattice constants and crystallography, identification of unknown materials, orientation of single crystals and preferred orientation of polycrystals, defects, stresses, etc. Extension of X-ray diffraction methods to thin films has not been pursued with vigor for two main reasons: First, the great penetrating power of X-rays means that with typical incident angles, their path length through films is too short to produce diffracted beams of sufficient intensity. Under such conditions the substrate, rather than the film, dominates the scattered X-ray signal; thus, diffraction peaks from films require long counting times. Second, the TEM provides similar diffraction information with the added capability of performing analysis over very small selected areas. Nevertheless X-ray methods have advantages because they are nondestructive and do not require elaborate sample preparation or film removal from the substrate. Two applications to thin films follow.

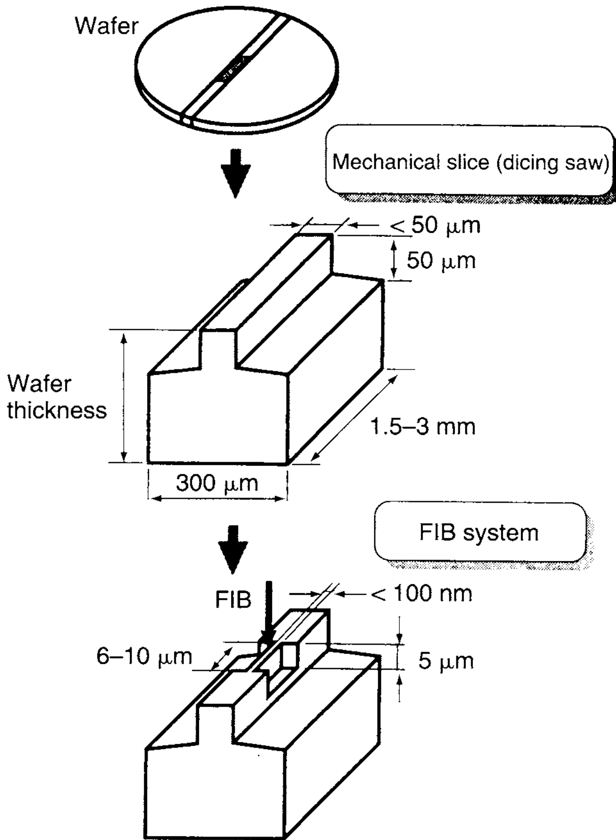


Figure 10-18 Cross-section sample preparation for TEM employing rough mechanical slicing followed by FIB milling techniques to remove submicron layers of material. (From Ref. 25. Reprinted by permission.)

10.3.4.1 Monitoring Interdiffusion in Thin Films

Consider a specimen consisting of consecutively evaporated polycrystalline films of Ag, Au, Ag, and Pb on fused quartz (Fig. 10-19b). After the composite structure is annealed at 200°C for 24 h it is desired to know the composition of the reaction products. To accomplish this by X-ray methods the films must appear to be thicker to the beam than they actually are. Employing a grazing angle of incidence $\gamma = 5^\circ$ makes the films effectively 12 times thicker. The Seeman–Bohlin diffraction geometry (Fig. 10-19a) is used with the focal point of the X-ray source, film specimen, and detector slit all

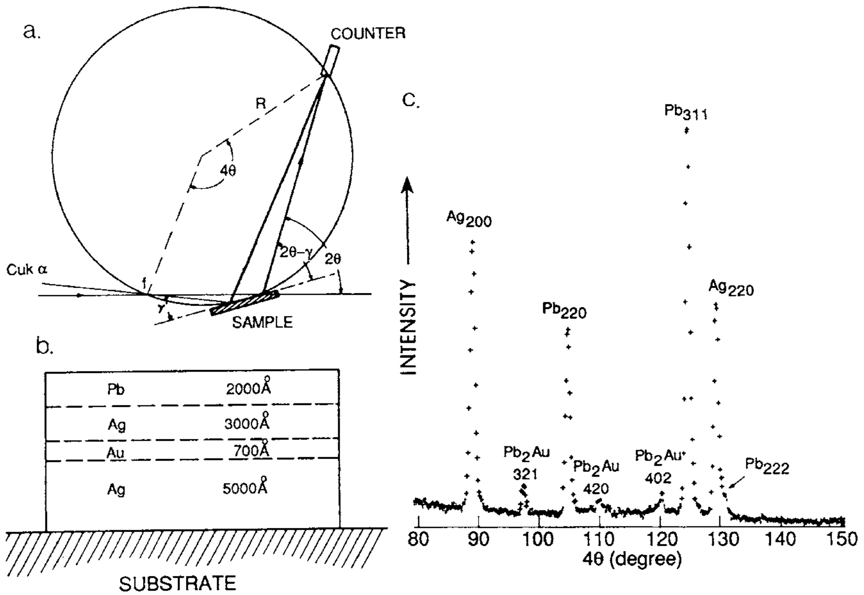


Figure 10-19 (a) Seeman–Bohlin diffraction geometry. f = effective location of X-ray source, γ = angle of incidence, θ = Bragg angle. (b) Schematic of layered film specimen. (c) A selected portion of the diffraction pattern with reflections from Ag, Pb, and Pb_2Au shown. (From K. N. Tu, *J. Appl. Phys.* **43**, 1303 (1972). (Courtesy of K. N. Tu, IBM, T. J. Watson Research Center.)

located on the circumference of one great circle. Each of the diffracted peaks (Fig. 10-19c) appear at different angles which are sequentially swept through as the X-ray detector moves along the circumference. All the while, a servomotor rotates the detector to keep it aimed at the specimen and preserve the overall focusing geometry.

The fact that Pb_2Au peaks emerged indicates that Au atoms diffused through the polycrystalline Ag layer and then reacted with the Pb. Grain boundaries in Ag were the likely diffusion path because no penetration of single-crystal Ag films by Au was observed.

10.3.4.2 Analysis of Epitaxial Films

Let us suppose we wish to nondestructively measure the composition of a ternary epitaxial film of $Al_xGa_{1-x}As$ on GaAs to an accuracy of 2%. One way to do this is to make use of the connection between the lattice parameter a_0 and x. Vegard's law then suggests that a_0 must be measured

to a precision of

$$\frac{\Delta a_0}{a_0} = \frac{2}{100} \frac{[a_0(\text{AlAs}) - a_0(\text{GaAs})]}{a_0(\text{GaAs})} = 2.8 \times 10^{-5}, \tag{10-15}$$

or 1 part in more than 35,000. This is quite a formidable challenge and neither LEED nor RHEED can even remotely approach such a capability. X-ray diffraction methods can, however, but not easily. By virtue of Bragg's law (Eq. 1-2), differentiation yields

$$\frac{\Delta a}{a} = \frac{\Delta \lambda}{\lambda} - \frac{\Delta \theta}{\tan \theta}. \tag{10-16}$$

This equation reveals the inadequacy of conventional X-ray diffraction methods in meeting the required measurement precision. For example, typical CuKα ($\lambda = 1.5406$ Å) radiation from an X-ray tube exhibits a so-called spectral dispersion of 0.00046 Å, so that $\Delta\lambda/\lambda = 0.0003$. This causes unacceptable diffraction peak broadening. In addition, the angular divergence of the beam must be several seconds of arc and it is not possible to achieve this with usual slit-type collimation.

For the foregoing reasons the high-resolution *double-crystal* diffractometer, shown schematically in Fig. 10-20, is indispensable. It has three special features:

1. Very high angular stepping accuracy on the θ axis (i.e., $\sim$1 arc second)
2. Excellent angular collimation of the incident X-ray beam (i.e., $<$10 arc seconds)
3. Elimination of peak broadening due to spectral dispersion

The diffractometer consists of a point-focus X-ray source of monochromatic radiation which falls on a first collimator crystal composed of the same material as the sample epitaxial film (second crystal). When the Bragg condition is satisfied both crystals are precisely parallel. The Bragg condition is simultaneously satisfied for all source wavelength components, i.e., no wavelength dispersion. During measurement, the sample is rotated or *rocked* through a very small angular range. The resulting rocking curve diffraction pattern contains the very intense substrate peak which serves as the internal standard against which the position of the low-intensity epitaxial film peak is measured.

In the following example the power and importance of the technique is illustrated. A rocking curve of an 1100 Å ZnSe film grown epitaxially on a (001) GaAs substrate is shown in Fig. 10-20 for the (004) reflection. In GaAs $a_0 = 5.6537$ Å and $a(004) = 1.4134$ Å. Since $\lambda = 1.5406$ Å, Bragg's law yields $\theta = 33.025^\circ$. For ZnSe with $a_0 = 5.6690$ Å, $a(004) = 1.4173$ Å and the expected

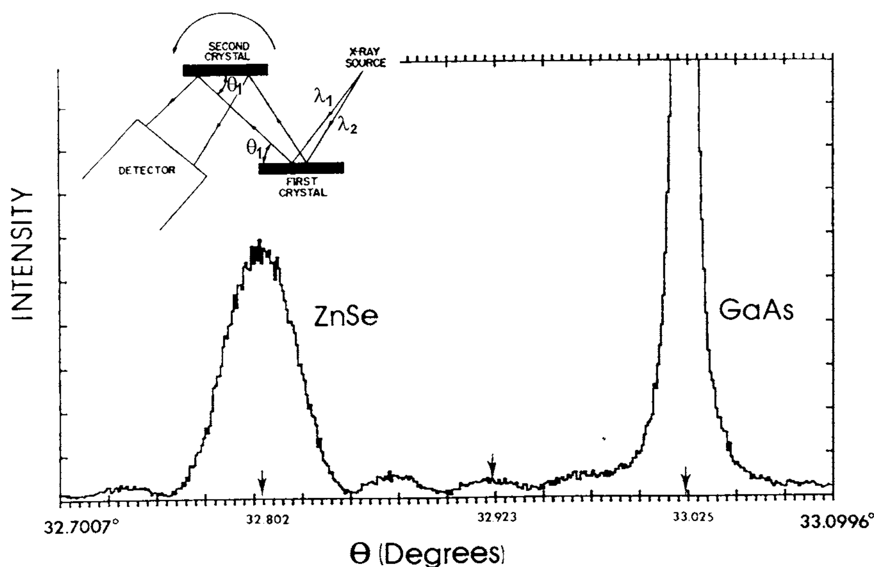


Figure 10-20 Rocking curve for (004) reflection of ZnSe on GaAs. (Courtesy of B. Greenberg Philips Laboratories, North American Philips Corp.) (Inset) Schematic of high-resolution double-crystal diffractometer. (From A. T. Macrander, *Ann. Rev. Mater. Sci.* **18**, 283, 1988.)

Bragg angle for unstrained ZnSe is 32.923°. But the actual (004) peak appears at 32.802° which corresponds to $a(004) = 1.4219$ Å. To interpret these findings it should be noted that the misfit (Eq. 8-1) in this system is -0.27% and hence ZnSe is biaxially compressed in the film plane. But since the film thickness is less than d_c (Eq. 8-5), it has grown pseudomorphically with GaAs; we can therefore assume that ZnSe has the same lattice constant as GaAs in the interfacial plane. However, normal to this plane the ZnSe lattice expands and assumes a tetragonal distortion. The measured increase in the (004) interplanar spacing of ZnSe is thus consistent with this explanation.

10.3.5 Scanning Probe Microscopies

10.3.5.1 Introduction

When Binnig and Rohrer invented the scanning tunneling microscope (STM) in 1981 they could scarcely have imagined that its impact would eventually encompass disciplines as diverse as surface science, thin-film

technology, and cellular biology. The STM and variants of it, most notably the atomic force microscope (AFM), are collectively known as scanning-probe microscopes (SPMs). These instruments are imaging tools that have a resolution range spanning the capabilities of the SEM at one extreme and TEM lattice imaging at the other. In addition, assorted physical properties such as surface conductivity, elastic moduli, static-charge distributions, and magnetic fields have been measured by SPM methods, suggesting a sense of the breadth of applications. Aspects of scanning-probe microscopies are reviewed in Refs. 26–28.

10.3.5.2 Scanning Tunneling Microscopy

The STM, the progenitor of all subsequent scanning-probe microscopes, won the 1986 Nobel prize in physics for its inventors. It was not the first microscope that could image atoms, however. Some three decades earlier E. Müller invented the field ion microscope (FIM). Though not as popular a surface-science tool as it once was, the FIM has been employed over the years in studies of defects, surface diffusion, crystallography, desorption, and alloying (Ref. 29). In this microscope (Fig. 10-21a) the surface of a *strong metal tip*, sharpened to a radius of curvature less than $\sim 0.1\,\mu$m, is the specimen being imaged. The *cryogenically cooled* tip is contained within a system pumped to *high vacuum* ($\sim 10^{-10}$ torr) and then backfilled with an imaging gas, typically helium. When the tip is raised to a *high positive potential* (~ 10 *kV*), *very high electric fields* (several hundred MV/cm) are generated in the vicinity of tip atom projections. This has the effect of stripping electrons from the gas atoms into the tip atoms through tunneling from the former to the latter. The resulting positive He ions then fly off the metal along the electric-field lines and light up the grounded fluorescent screen, providing a *projected image* of the atomic arrangement on the terraced terrain of the tip, shown in Fig. 10-21b for tungsten. Terms that are italicized may be viewed as disadvantages of FIM because they correspond to limitations in specimens that can be imaged, experimental difficulties, and complexities in interpreting images.

The STM eliminates many, but not all of these restrictions and shortcomings, i.e., all conductors are imageable in principle and there is less danger of distorting and altering surface features in high electric fields. Importantly, unlike other indirect techniques for determining surface crystallography that depend on the amplitude and phases of scattered electron, photon or ion waves, the STM enables direct imaging of atoms. Like the FIM, the STM also employs a metal needle or probe tip and it is etched to almost an atomic radius. But this tip serves as a conductive electrode probe and is not the specimen being observed. In operation the tip is normally brought to

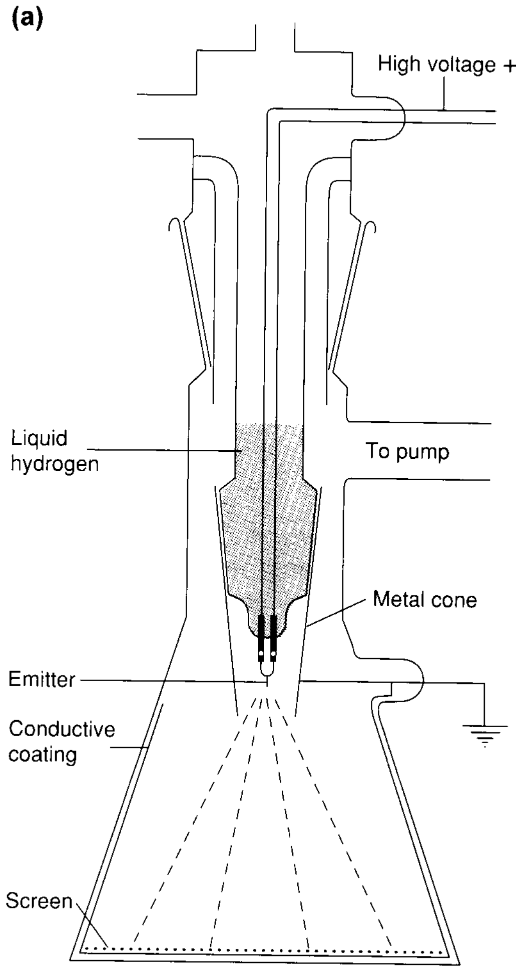


Figure 10-21 (a) Schematic of a field-ion microscope employing helium ions. (b) Field-ion microscope image of tungsten tip. (From E. W. Müller, *Advances in Electronics and Electron Physics*, Vol. 13. Reprinted with the permission of Academic Press, Inc.)

within about 1 nm of the specimen surface. There it is rastered in atomic-scale increments along both x and y surface directions by means of piezoelectric transducers, as shown in Fig. 10-22. An electron tunneling current, i, flows across the narrow vacuum-gap distance s between tip and specimen when a voltage V_T (typically millivolts to a few volts) is applied between

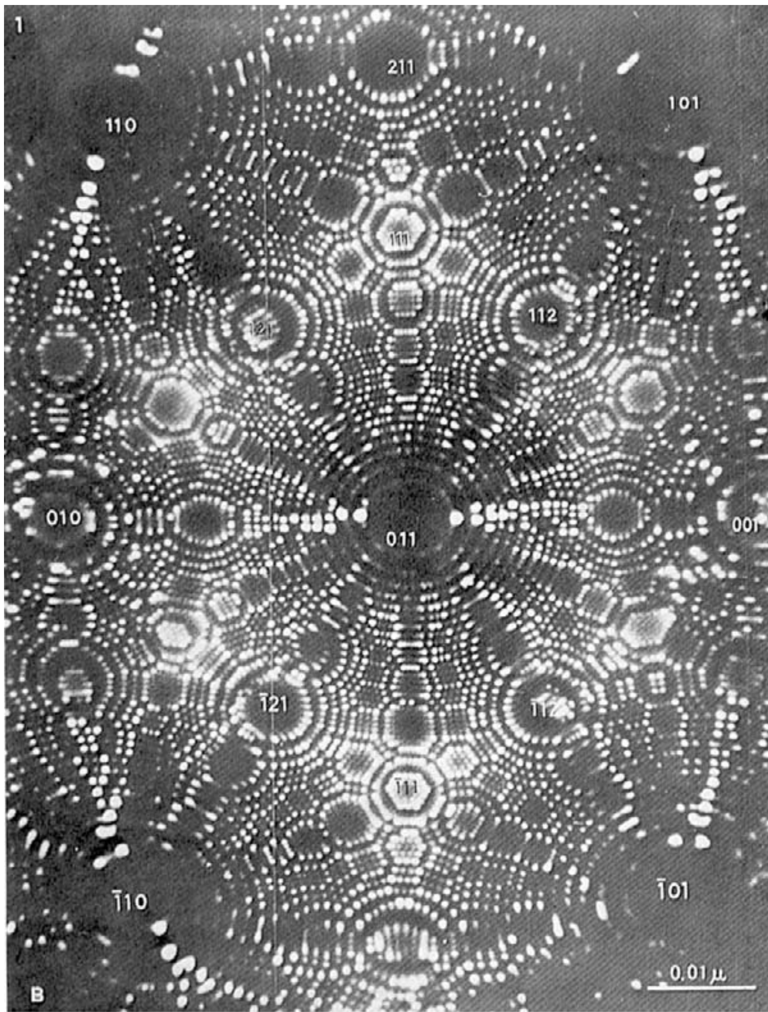


Figure 10-21 (Continued).

them. This current arises when electrons occupying filled electron states in a negatively biased tip tunnel into unfilled states in the specimen; electrons flow in the reverse direction when the tip–specimen polarity is reversed. Theory supported by measurement indicates that for small values of V_T the current is given by

$$i = aV_T \exp - (A\phi^{1/2}s), \tag{10-17}$$

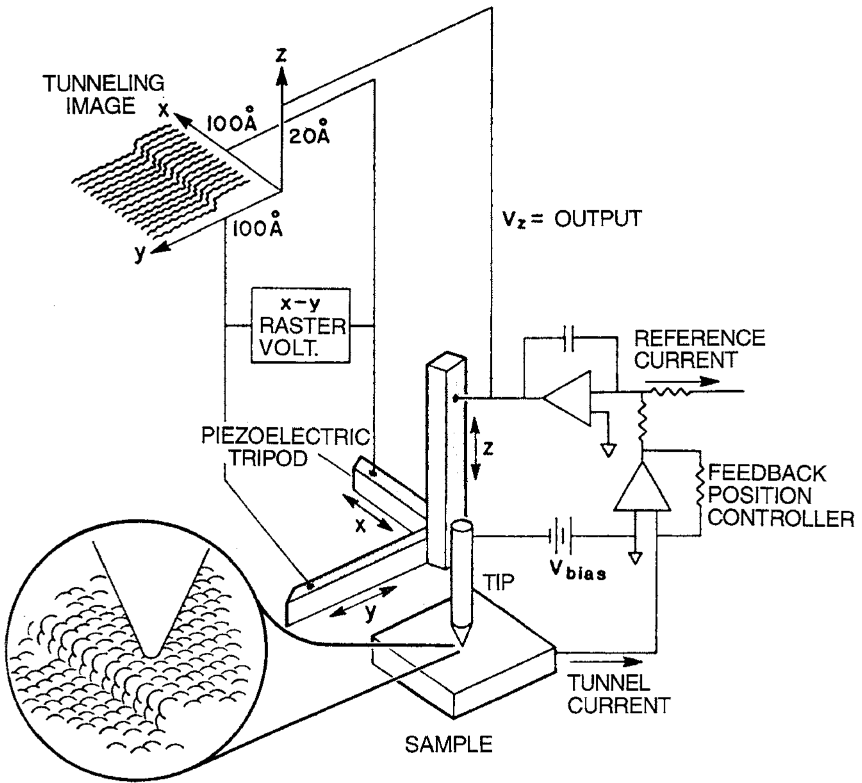


Figure 10-22 Schematic illustration of the scanning tunneling microscope. The tunneling current is maintained constant during surface scanning through electronic feedback. (From J. A. Golovchenko, *Science* **232**, 48, 1986. Copyright © 1986 by the American Association for the Advancement of Science. Reprinted with the permission of *Science* and J. A. Golovchenko, Harvard University.)

where a and A are constants and ϕ is the average work function. The interdependence of i, V_T, and s for a given ϕ tightly fixes experimental conditions such that the tunneling current is typically a very steep function of s.

There are two basic modes of operating an STM. In the first, the tunnel current is maintained constant during lateral scanning by keeping s constant as the tip rides up and down over the atomic terrain. This is accomplished through feedback electronics which continually adjusts the tip height. Alternatively, the tip scans at a preset height so that the tunnel current varies periodically in concert with the atomic-scale surface corrugations. In either case the electronic signal variations are converted into an image such as the one shown for a reconstructed (111) Si surface (Fig. 7-9).

10.3.5.3 An Atomic Corral

One of the most intriguing features of the STM is its ability to move individual adsorbed atoms to selective sites on a surface. Since the tip exerts both van der Waals and electrostatic forces on such atoms they can be extracted, dragged, and displaced along the surface by simply varying the tip position and bias voltage (Ref. 30). An early example of this was the well-publicized manipulation of 35 individual xenon atoms on a nickel surface to spell out the IBM logo. Even more fascinating from a scientific point of view is the 14.3-nm diameter *quantum corral* shown in Fig. 10-23 that was created by carefully positioning 48 iron atoms on a (111) copper surface (Ref. 31). Electrons near the surface essentially form a two-dimensional electron gas that is sandwiched between the Cu work function on the vacuum side and a bandgap in the energy levels of the bulk electrons. Free to roam within the corral, the electrons scatter off the confining Fe atom fence. As a result they establish electron-density standing waves corresponding to allowed quantum states in a two-dimensional circular box. This quantum roundup of electrons in the corral should be contrasted with trapping of electrons in quantum wells that arise in layered semiconductor

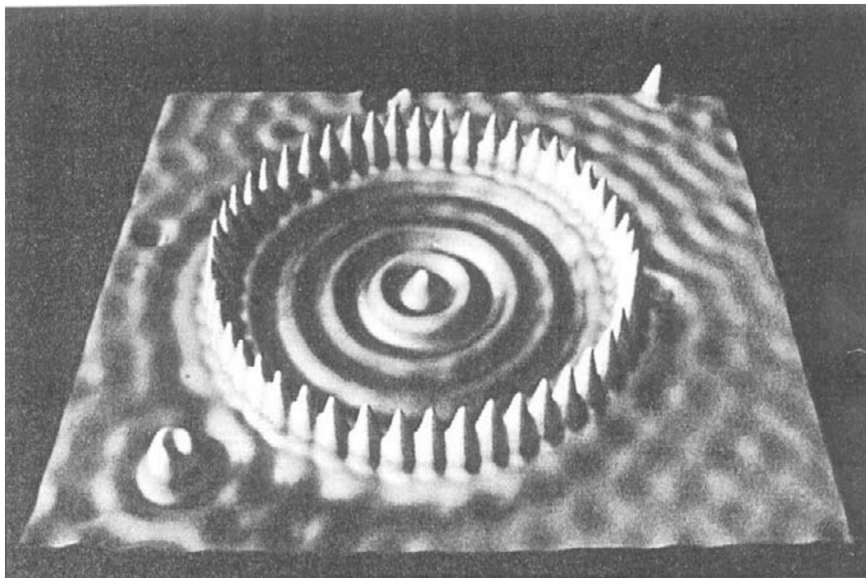


Figure 10-23 Quantum corral of iron atoms positioned on a (111) copper surface and then imaged by an STM at 4 K. (From cover of *Physics Today* **46**, Nov. (1993); also Ref. 31. Copyright © 1993 by the American Institute for Physics. Reprinted by permission.)

films (Ref. 32). Both experimental arrangements have corroborated very basic predictions of wave mechanics that were previously relegated to texts on theoretical physics.

10.3.5.4 Atomic Force Microscopy

Shortly after the STM appeared, the atomic force microscope (AFM) was conceived as a response to the question: if surfaces could be imaged by a current, why not by a force? A major advantage of detecting forces rather than current is that all kinds of material surfaces including metals, semiconductors, and insulators are imageable. The inventors of the AFM realized that in order to mechanically sense atomic-scale surface topography, the sharp tip should be mounted at the end of a soft cantilever spring (Ref. 33). Furthermore, the cantilever spring constant k must be smaller than the spring constant K_s(a) that effectively exists between atoms, and the applied force should not be large enough to displace surface atoms. Treating a vibrating atom as a mass on a spring, the simple theory of the harmonic oscillator indicates that $2\pi v = (K_s(a)/m)^{1/2}$, where v is the frequency and m is the atomic mass. Since $2\pi v$ is typically 10^{13} Hz and $m \simeq 10^{-25}$ kg, $K_s(a) \simeq 10$ N/m. In contrast, we know from solid mechanics that the spring constant or force per unit deflection of a cantilever of length l, width w, and depth d is given by $k = Ywd^3/4l^3$. Since Young's modulus (Y) for inorganic materials ranges from ~ 50 to 1000×10^9 N/m^2, large values of l coupled with small values of both w and d are required to reduce the magnitude of k. For example, if the cantilever were made of household aluminum foil with $Y = 72 \times 10^9$ N/m^2, $d = 25 \times 10^{-6}$ m, $l = 4 \times 10^{-3}$ m, and $w = 1 \times 10^{-3}$ m, $k =$ 4.4 N/m. Presently, AFM cantilevers are micromachined from silicon, silicon dioxide, or silicon nitride and have spring constants ranging from 0.1 to ~ 50 N/m. Many of the essential experimental features of AFM parallel those of STM. There is a mechanical raster-scanning system (usually piezoelectric), a method to sense the cantilever deflection (usually optical, e.g., interferometry), a feedback system to monitor and control the cantilever force, and a display to convert the force–position data into an image.

There are two basic modes of operation, and they can be understood in terms of Fig. 1-8b if we imagine this well-known curve represents the tip force as a function of tip–specimen separation. When the tip and specimen are widely separated, van der Waals forces cause them to weakly attract. But when they draw too closely together, their electron clouds overlap and electrostatic repulsive forces physically push them apart. In the former *non-contact* AFM (NC-AFM) mode the cantilever is located tens to hundreds of angstroms from the specimen surface. To prevent surface contact a stiff cantilever is used resulting in low tip–specimen forces of $\sim 10^{-12}$ N.

Because it is difficult to detect small forces the cantilever is vibrated, typically at 100 to 400 kHz. Changes in vibrational amplitude (due to topography) or resonant frequency (due to force gradients and variations in $K_s(a)$) are then detected by sensitive AC methods, converted to tip–sample spacings, and ultimately recorded as surface images. Among the advantages of NC-AFM are the ability to probe soft or elastic materials and minimization of both surface contamination and tip degradation. Normally, the spatial resolution of AFMs is poorer than that of STMs. But with sharp probe tips, a stiff cantilever, and operation in ultrahigh vacuum, atomic-resolution surface reconstructions of Si (100) 7×7 and Si (100) 2×1 have been readily imaged by NC-AFM means (Ref. 34).

In the latter or *contact* AFM mode, also known as the repulsive mode, the tip actually makes physical contact with the surface, and forces in the range of 10^{-6} to 10^{-8} N are typically generated.

10.3.5.5 Scanning Probe Applications

Spurred by the promise of nanometer-scale or even atomic resolution with relatively inexpensive instruments capable of imaging all classes of materials under different environmental ambients, there has been an unceasing flood of studies in the past decade having applications to physics, chemistry, and biology. To limit the ensuing discussion it must be appreciated that scanning-probe methods reveal *surface* properties that may or may not reflect the bulk of the underlying thin film. The list of representative applications to thin films and substrate materials in the scientific and technical areas noted in Table 10-3 is perhaps the most appropriate way to briefly summarize this body of information. For the MFM, CFM, and SThM techniques mentioned in the table special sensing tips are required. For example, most MFM force sensors are composed of etched ferromagnetic materials, e.g., Fe, Ni, while CFM and SThM rely on chemically modified tips and very tiny thermocouples, respectively.

10.4 CHEMICAL CHARACTERIZATION OF SURFACES AND FILMS

10.4.1 INTRODUCTION

In this section we turn our attention to the subject of chemical characterization of thin films. This includes identification of surface and near-surface atoms and compounds, as well as their lateral and depth spatial distributions. To meet these needs an important subset of the analytical techniques listed in Table 10-1 is employed. Space limitations will restrict the

Table 10-3
Scanning Probe Applications in Thin-Film Science and Technology

Technique	Subject, information	Materials, features	Reference[a]
1. STM, AFM	Metrology, feature dimensions, roughness	Surface topography of ICs, trenches, photoresists	a,b
2. STM	Surface structure, reconstruction, defects	Si, Ge, GaAs surfaces	c
	Dry etching mechanisms	Si (100) and (111) surfaces	d
	Silicide interface structure and defects, growth kinetics	$CoSi_2$/Si	e
3. MFM[b]	Magnetic domain structure, disk particle grain size, surface defects	Magnetic recording films, disks, heads, magnetooptical films	f
4. CFM[c] AFM	Adhesion and friction between chemical groups, nanotribology of surfaces	Organic monolayers Surface lubricants	g, h
	Friction, lubrication	Si, graphite	i
5. SThM[d]	Thermal probing of hot spots in devices	Si-MOSFET	j
6. STM, AFM	Nucleation of films, nucleus size, shape, density, epitaxy growth mechanisms	Si/Si, Si/SiO_2, metal films on assorted metal, oxide and semiconductor substrates	(see Chapter 7)

[a] References: a, J. E. Griffith and D. A. Grigg, *J. Appl. Phys.* **74**(9), R83 (1993); b, Y. Martin and H. K. Wickramasinge, *J. Vac. Sci. Technol.* **B13**(6), 2335 (1995); c, J. A. Kubby and J. J. Boland, *Surf. Sci. Rep.* **26**, 61 (1996); d, J. J. Boland and J. H. Weaver, *Physics Today* **8**, 34 (1998); e, P. A. Bennett and H. von Kanel, *J. Phys. D: Appl. Phys.* **32**, R71 (1999); f, U. Hartmann, *Ann. Rev. Mater. Sci.* **29**, 53 (1999); g, A. Noy, D. V. Vezenov, and C. M. Lieber, *Ann. Rev. Mater. Sci.* **27**, 381 (1997); h, M. B. Salmeron, *MRS Bull.* **5**, 20 (1993); i, C. M. Mate, *IBM J. Res. Develop.* **39**, 617 (1995); j, J. Lai, M. Chandrachood, A. Majumdar, and J. P. Carrejo, *IEEE Electron Dev. Lett.* **16**, 312 (1995).
[b] MFM, magnetic force microscopy.
[c] CFM, chemical force microscopy.
[d] SThM, scanning thermal microscopy.

discussion to include only the most popular methods (SEM/EDX, AES, XPS, RBS, and SIMS) and variants based on them. The justification for selecting these and not others is that they, together with the SEM and TEM, form the core of the diagnostic facilities associated with all phases of the research, development, processing, and reliability and failure analysis of

Table 10-4

Summary of Major Chemical Characterization Techniques

Method	Elemental sensitivity	Detection limit (at.%)	Lateral resolution	Effective probe depth
Scanning electron microscope/energy dispersive X-ray (SEM/EDX)	Na–U	~0.1	~1 μm	~1 μm
Auger electron spectroscopy (AES)	Li–U	~0.1–1	500 Å	15 Å
X-ray photoelectron spectroscopy (XPS)	Li–U	~0.1–1	~100 μm	15 Å
Rutherford backscattering (RBS)	He–U	~1	1 mm	~200 Å
Secondary-ion mass spectrometry (SIMS)	H–U	10^{-4}	~1 μm	15 Å

thin-film electronic and optoelectronic devices. In integrated-circuit technology some of these methods have gained wide acceptance as support tools for manufacturing lines. In addition, all of the associated equipment is now commercially available, albeit at high cost. All excellent film-characterization laboratories are outfitted with the total complement of this equipment. Table 10-4 will help the reader to distinguish among the various chemical analytical methods. Capabilities and limitations of each are indicated and the comparative strengths and weaknesses for particular analytical applications can, therefore, be assessed. The following remarks summarize several of these distinctions.

1. AES, XPS, and SIMS are true surface analytical techniques since the detected electrons and ions are emitted from surface layers less than ~15 Å deep. Provision is made to probe deeper, or depth profile, by sputter etching the film and analyzing the newly exposed surfaces.
2. EDX and RBS generally sample the total thickness of the thin film (~1 μm) and frequently some portion of the substrate as well. Unlike RBS with a depth resolution of ~200 Å, EDX has little depth resolution capability.
3. AES, XPS, and SIMS are broadly capable of detecting, with few exceptions, all of the elements in the periodic table.

4. EDX ordinarily detects elements with $Z > 11$ and RBS is restricted only to selected combinations of elements whose spectra do not overlap.

5. The detection limits for AES, XPS, EDX, and RBS are similar, ranging from about ~0.1 to 1 at.%. On the other hand, the sensitivity of SIMS is much higher and parts per million can be detected. Even lower concentration levels ($\sim 10^{-6}$ at.%) are detectable in certain instances.

6. Quantitative chemical analysis with AES and XPS is problematical with error bounds of a few atomic percent. EDX is better and SIMS significantly worse in this regard. Composition standards are essential for quantitative SIMS analysis.

7. Only RBS is quantitatively precise to within an atomic percent or so from first principles and without the use of composition standards. It is the only nondestructive technique which provides simultaneous depth and composition information.

8. The lateral spatial resolution of the region over which analyses can be performed is highest for AES (~500 Å) and poorest for RBS (~1 mm). In between are EDX (~1 μm), SIMS (several μm) and XPS (~0.1 mm). AES has the distinction of being able to sample the smallest volume for analysis.

9. Only XPS, and to a much lesser extent AES, is capable of readily providing information on the nature of chemical bonding and valence states.

The preceding characteristics earmark certain instruments for specific tasks. Suppose, for example, a film surface is locally discolored due to contamination, or contains a residue, and it is desired to identify the source of the unknown impurities. Assuming access to all instruments at equal cost, AES and EDX would initially be the techniques of choice. If only ultrathin surface layers are involved, EDX would probably be of little value. The presence of trace elements would necessitate the higher sensitivity of SIMS analysis. If preliminary examination pointed to the presence of Cl from an etching process, then evidence of the actual chemical compound formed would be obtained from XPS measurements. In a second example a broad-area, thin-film metal-bilayer structure is heated. Here we know which elements are initially present but wish to determine the stoichiometry of intermetallic compounds formed as well as their thicknesses. This information is without question most unambiguously provided by RBS methods.

In what follows, the various techniques are considered individually where details of instrumentation, particular capabilities and limitations, and applications will be discussed. First, however, it is essential to understand the scientific principles underlying each type of analysis, and this is treated first. More detailed discussions of these characterization techniques are given in Refs. 35–44.

10.4.2 FINGERPRINTING ATOMS THROUGH ELECTRON TRANSITIONS

We start with a discussion of atomic core-electron spectroscopy since it is the basis for identification of the elements by EDX, AES, and XPS techniques. Consider the electronic structure of an unexcited atom schematically depicted in Fig. 10-24a. Both the K, L, M, etc., shell notation and the corresponding 1s, 2s, 2p, 3s, etc., electron states are indicated. Through excitation by an incident electron or photon, a hole or electron vacancy is created in the K shell (Fig. 10-24b).

In EDX an electron from an outer shell lowers its energy by filling the hole and an X-ray is emitted in the process (Fig. 10-24c). If the electron transition occurs between L and K shells, $K\alpha$ X-rays are produced. Different X-rays are generated, e.g., $K\beta$ X-rays from $M \rightarrow K$, and $L\alpha$ X-rays from

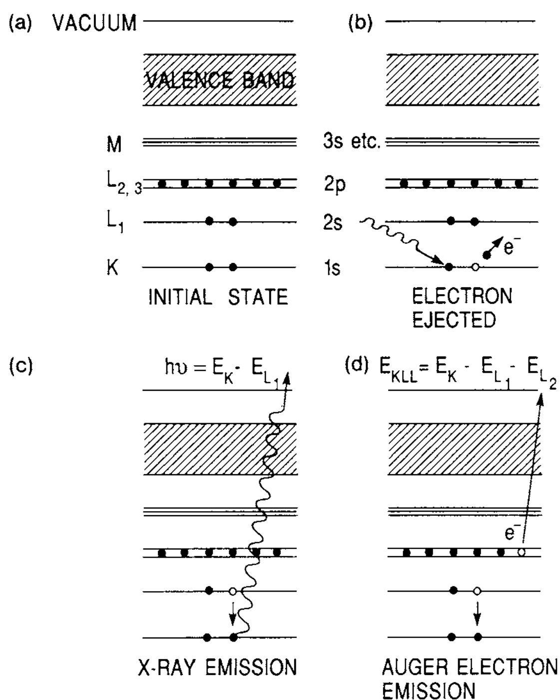


Figure 10-24 Schematic of electron energy transitions: (a) initial state; (b) incident photon (or electron) ejects K shell electron; (c) X-ray emission when 2s electron fills vacancy; (d) Auger electron emission. KLL transition shown.

$M \rightarrow L$ transitions. There are two facts worth remembering about these X-rays.

1. The difference in energy between the levels involved in the electron transition determines the energy (or wavelength) of the emitted X-ray. For example,

$$E_{K\alpha_1} \times hc/\lambda_{K\alpha_1} \times E_K - E_{L_3} \qquad (10\text{-}18)$$

where the symbols h, c, and λ have their usual meaning.

2. The emitted X-rays are *characteristic* of the particular atom undergoing emission. Thus, each atom in the periodic table exhibits a unique set of K, L, M, etc., X-ray spectral lines which serve to unambiguously identify it. These characteristic X-rays are also known as fluorescent X-rays when excited by incident photons, e.g., X-rays and gamma rays.

There is, however, an alternative process by which the electron hole in Fig. 10-24b can be filled. It involves a more complex transition in which three rather than two electron levels, as in EDX, participate. The Auger process, which is the basis of AES, first involves an electron transition from an outer level, e.g., L_1, to the K hole. This excess energy is not channeled into the creation of a photon but is expended in ejecting an electron from yet a third level, e.g., L_2. As shown in Fig. 10-24d, the atom finally contains two electron holes after starting with a single hole. The electron which leaves the atom is known as an Auger electron and it possesses an energy given by

$$E_{KLL} = E_K - E_{L_1} + E_{L_2} \times E_K - E_{L_2} + E_{L_1}. \qquad (10\text{-}19)$$

The last equality indicates that KL_1L_2 and KL_2L_1 transitions are indistinguishable. Similarly, other common transitions observed are denoted by LMM and MNN. Since the K, L, and M energy levels in a given atom are unique, the Auger spectral lines are characteristic of the element in question. By measuring the energy of the Auger electrons emitted by a material, its chemical identity can be ascertained.

To quantitatively illustrate these ideas, let us consider the X-ray and Auger excitation processes in titanium. The binding energies of each of the core electrons are indicated in Fig. 10-25, where electrons orbiting close to the nucleus are strongly bound with large binding energies. Electrons at the Fermi level are far from the pull of the nucleus and are taken to have zero binding energy, thus establishing a reference level. They would still have to acquire the work function energy to be totally free of the solid. Some notion of the rough magnitude of core energy levels can be appreciated from the well-known formula for hydrogen-like levels, i.e.,

$$E = 13.6Z^2/n^2 \text{ (eV)} \qquad (10\text{-}20)$$

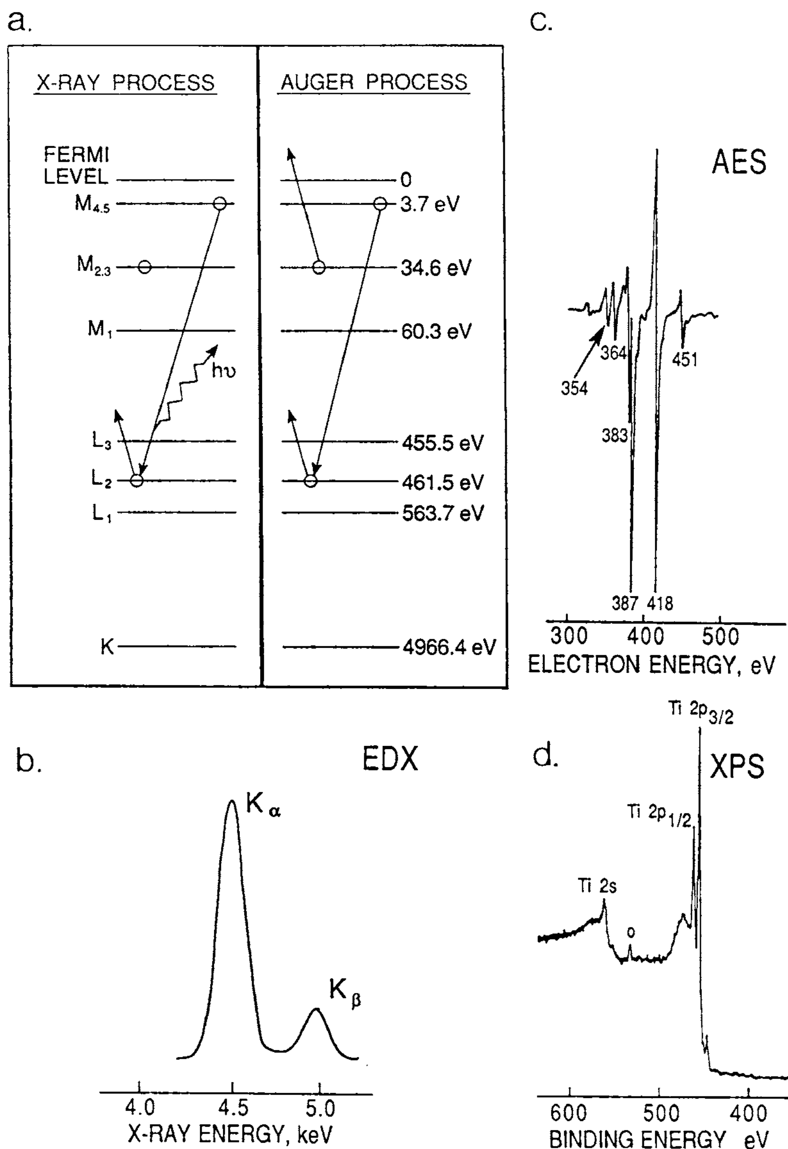


Figure 10-25 Electron excitation processes in Ti: (a) energy-level scheme; (b) EDX spectrum of Ti employing Si(Li) detector; (c) AES spectral lines for Ti ($dN(E)/dE$ vs E); (d) a portion of the XPS spectrum for Ti (MgKα radiation).

where Z is the atomic number and n is the principal quantum number. For Ti ($Z = 22$), the calculated energy of the K shell ($n = 1$) is 6582 eV. Complex electron–electron interactions and shielding of the nucleus makes this formula far too simplistic for multielectron atoms. Both effects reduce electron binding energies relative to Eq. 10-20. Several of the prominent characteristic X-ray energies and wavelengths for Ti are:

$$\mathbf{K}\alpha_1;\ E_K - E_{L_3} = 4966.4 - 455.5 = 4511\,\text{eV};\ \lambda(K\alpha_1) = 2.75\,\text{Å}$$

$$\mathbf{K}\beta_1;\ E_K - E_{M_3} = 4966.4 - 34.6 = 4932\,\text{eV};\ \lambda(K\beta_1) = 2.51\,\text{Å}$$

$$\mathbf{L}\alpha' = E_{L_3} - E_{M_{4,5}} = 455.5 - 3.7 = 452\,\text{eV};\ \lambda(L\alpha) = 27.4\,\text{Å}.$$

Similarly, a prominent Ti Auger spectral transition is LMM or

$$E_{LMM} = E_{L_3} - E_{M_3} - E_{M_4} = 461.5 - 34.6 - 3.7 = 423\,\text{eV}.$$

A question that may have occurred to the reader at this point is, when do atoms with electron holes undergo X-ray transitions and when do they execute Auger processes? The answer is that both processes go on simultaneously. In the low-Z elements the probability is greater than an Auger transition will occur, whereas X-ray emission is favored for high-Z elements. The fractional probabilities of characteristic X-ray and Auger yields for KLM transitions, as well as electron binding energies of the elements, can be found in Ref. 35.

The variation of the principal characteristic X-ray and Auger lines with atomic number is shown in Figs. 10-26 and 10-27, respectively. Commercial

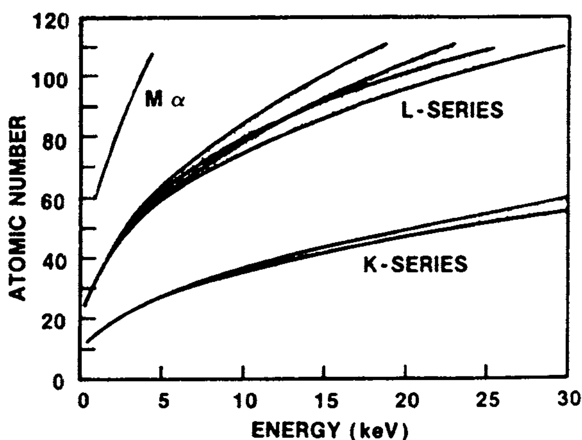


Figure 10-26 Characteristic X-ray emission energies of the elements. (Courtesy of Princeton Gamma Tech, Inc.)

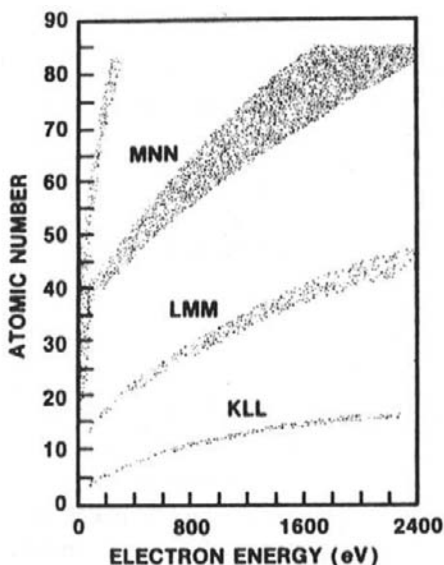


Figure 10-27 Principal Auger electron energies of elements. (Courtesy of Physical Electronics Industries, Inc.)

spectrometers typically operate within the energy range spanned in these figures. Therefore, in EDX, K X-ray transitions are conveniently measured in low-Z materials while L series X-rays appear when high-Z elements are involved. Similarly, in AES, KLL and LMM transitions are involved for low-Z elements while LMM and MNN lines appear for high-Z elements. Although keeping track of the particular shell involved is sometimes annoying, spectra from virtually all of the elements in the periodic table can be detected with a single excitation source and a single detector. Fortunately, the resolution of X-ray or electron detectors is such that prominent lines of neighboring elements do not seriously overlap. This facilitates spectral interpretation and atomic fingerprinting.

The basis for understanding XPS lies in the same atomic core-electron scheme that we have been considering. Rather than incident electrons in the case of EDX and AES, relatively low-energy X-rays impinge on the specimen in this technique. The absorption of the photon results in the ejection of electrons via the photoelectric effect. Governing this process is the well-known but modified Einstein equation for the photoelectric effect expressed as

$$E_{KE} = h\nu - E_B, \tag{10-21}$$

where E_{KE}, hv, and E_B are the energies of the ejected electron, the incident photon, and the involved bound electron state. Since values of the binding energy are element specific, atomic identification is possible through measurement of photoelectron energies.

10.4.3 X-RAY ENERGY DISPERSIVE ANALYSIS (EDX)

10.4.3.1 Equipment

Most X-ray energy dispersive analysis systems are interfaced to SEMs where the electron beam serves to excite characteristic X-rays from the probed area of the specimen. Attached to the SEM column is the liquid-nitrogen dewar with its cooled Si(Li) detector aimed to efficiently intercept emitted X-rays. The Si(Li) detector is a reverse biased Si diode doped with Li to create a wide depletion region. An incoming X-ray generates a photoelectron which eventually dissipates its energy by creating electron–hole pairs. The incident photon energy is linearly proportional to the number of pairs produced or equivalently proportional to the amplitude of the voltage pulse they generate when separated.

After the pulses are amplified they are sorted according to voltage amplitude by a multichannel analyzer which also counts and stores the number of pulses within given increments of the voltage (energy) range. The result is the characteristic X-ray spectrum shown for Ti in Fig. 10-25b. Si(Li) detectors typically have a resolution of about 150 eV so that overlap of peaks occurs when they are not separated in energy by more than this amount. Overlap sometimes occurs in multicomponent samples or when neighboring elements in the periodic table are present.

In an important variant known as X-ray wavelength dispersive analysis (WDX), *wavelength* rather than *energy* is dispersed, resulting in a factor of 20 or so improvement in X-ray linewidth resolution. In this case emitted X-rays, rather than entering a Si(Li) detector, are diffracted from single crystals with known interplanar spacings. From Bragg's law, each characteristic wavelength reflects at corresponding angles which can be measured with very high precision. As the goniometer/detector assembly rotates, the peak is swept through as a function of angle. The electron microprobe (EMP) is an instrument specially designed to perform WDX analysis. This capability can also be achieved by attaching a diffractometer to an SEM column. The high spectral resolution of WDX is offset by its relatively slow speed of analysis.

10.4.3.2 Quantification

Quantitative analysis of an element in a multicomponent matrix is a complicated matter. The expected X-ray yield, $Y_x(z)$, originating from some depth z below the surface depends on a number of factors: $I_0(z)$, the intensity of the electron beam at z; C, the atomic concentration; σ, the ionization cross section; ω_x, the X-ray yield; μ, the X-ray absorption coefficient; and ε, Ω, and θ, the detector efficiency, solid angle, and angle with respect to the beam, respectively. Therefore,

$$Y_x(z) \sim I_0(z) C \sigma \omega_x e^{-\mu z/\cos\theta} \varepsilon \, d\Omega/4\pi \tag{10-22}$$

and the total signal detected is the sum contributed by all atomic species present integrated over the depth range. It is sometimes simpler to calibrate the yields against known composition standards. Excellent computer programs, both standardless and employing standards, are available for analysis and compositions are typically computed to approximately 0.1 at.%

10.4.4 Auger Electron Spectroscopy (AES)

10.4.4.1 Equipment

The typical AES spectrometer, shown schematically in Fig. 10-28, is housed within an ultrahigh-vacuum chamber maintained at $\sim 10^{-10}$ torr. This level of cleanliness is required to prevent surface coverage by contaminants (e.g., C, O) in the system. The electron-gun source aims a finely focused beam of ~ 2 keV electrons at the specimen surface, where it is scanned over the region of interest. Emitted Auger electrons are then energy analyzed by a cylindrical (or hemispherical in some systems) analyzer. This consists of coaxial metal cylinders (or hemispheres) raised to different potentials. The electron pass energy, E, is proportional to the voltage on the outer cylinder while the incremental energy range, ΔE, of transmitted electrons determines the resolution ($\Delta E/E$), which is typically 0.2 to 0.5%. Electrons with higher or lower energies (velocities) than E hit either the outer or inner cylinders, respectively. They do not exit the analyzer and are not counted. By sweeping the bias potential on the analyzer, the entire electron spectrum is obtained. Auger electrons are but a part of the total electron yield, $N(E)$, intermediate between low-energy secondary and high-energy elastically scattered electrons. They are barely discernible as small bumps above the background signal. Therefore, to accentuate the energy and magnitude of the Auger peaks, the spectrum is electronically differentiated and this gives rise to the common AES spectrum, or $dN(E)/dE$ vs E response, shown in Fig. 10-25c for Ti. The reader should verify that

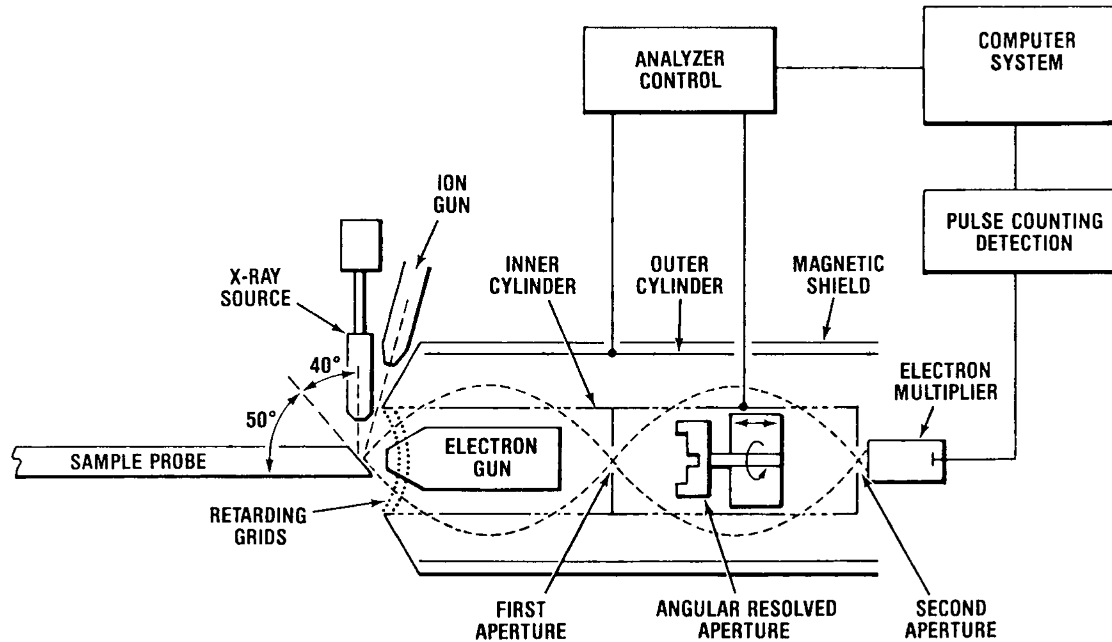


Figure 10-28 Schematic of spectrometer with combined AES and XPS capabilities. (Courtesy of Physical Electronics Industries, Inc.)

differentiation of a Gaussian-like peak yields the wiggly narrow double peak response. By convention, the Auger line energy is taken at the resulting peak minimum.

Two very useful capabilities for thin-film analysis are depth profiling and lateral scanning. The first of these is accomplished with incorporated ion guns that enable the specimen surface to be continuously sputtered away while Auger electrons are being simultaneously detected. Multielement composition depth profiles can thus be determined over total film thicknesses of several thousand angstroms by sequentially sampling and analyzing arbitrarily thin layers. While depth resolution is extremely high, the frequently unknown sputter rates make precise depth determinations problematical. Through raster or line scanning of the electron beam, the AES is converted into an SEM and images of the surface topography are obtained. By modulating the imaging beam with the Auger electron signal, lateral composition mapping of the surface distribution of particular elements is achieved. Unlike EDX/SEM composition mapping, only the upper few atom layers are probed in this case.

Complete AES spectrometers are commercially available and often combined with XPS capabilities.

10.4.4.2 Quantification

The determination of the Auger electron yield from which atomic concentrations can be extracted is expressed by a formula similar in form to Eq. 10-22 for the X-ray yield. External standards are important in quantifying elemental analysis since standardless computer programs for AES are more imprecise than those for EDX analysis. A formula which has been widely used to determine the atomic concentration of a given species A in a multielement matrix is

$$C_{\mathrm{A}} = \frac{I_{\mathrm{A}}/S_{\mathrm{A}}}{\Sigma_{i=\mathrm{A,B}\ldots} I_i/S_i}. \tag{10-23}$$

The quantity I_i represents the intensity of the Auger line and is taken as the peak-to-peak span of the spectral line. Relative Auger sensitivities, S_i, also enter Eq. 10-23. They have values ranging from 0.02 to 1 and depend on the element in question, the particular transition selected, and the electron-beam voltage. Uncertainties in C values so determined are perhaps a few at.%

10.4.5 X-RAY PHOTOELECTRON SPECTROSCOPY (XPS)

In order to capitalize on the X-ray-induced photoelectron effect, a spectrometer like the one used for AES (Fig. 10-28) is employed. The only difference is the excitation source, which is now a beam of either Mg or Al

Kα X-rays. These characteristic X-rays have relatively low energy, e.g., $h\nu_{\mathrm{Mg}} = 1254\,\mathrm{eV}$ and $h\nu_{\mathrm{Al}} = 1487\,\mathrm{eV}$, and set an upper bound to the kinetic energy of the detected photoelectrons.

A portion of the XPS spectrum for Ti is shown in Fig. 10-25d, where 2s, $2p_{1/2}$, and $2p_{3/2}$ peaks are evident. Interestingly, characteristic Auger electron transitions (not shown) frequently appear at precisely the energy locations indicated in Fig. 10-25c. The XPS peak positions, however, are shifted slightly by a few eV from those predicted by Eq. 10-21 because of the work function difference between specimen and detector.

It is beyond our scope to discuss spectral notation, the chemistry and physics of transitions, and the position and width of lines. What is significant is that linewidths are considerably narrower than those associated with Auger transitions. This fact makes it possible to gain useful chemical-bonding information which can also be obtained with less resolution by AES, but not by the other surface analytical techniques. It is for this reason that XPS is alternately known as ESCA (electron spectroscopy for chemical analysis).

Effects due to chemical bonding originate at the valence electrons and ripple beyond them to alter the energies of the core levels in inverse proportion to their proximity to the nucleus. As a result, energy shifts of a few eV occur and are resolvable. For example, in the case of pure Ti, the $2p_{3/2}$ line has a binding energy of 454 eV. In compounds this electron is more tightly bound to the Ti nucleus because electron charge clouds of the neighboring atoms "repel" it. In TiC, TiN, and TiO the same line is located at $E_B = 455\,\mathrm{eV}$. Similarly, for the compounds TiO_2, $BaTiO_3$, $PbTiO_3$, $SrTiO_3$, $CaTiO_3$, and $(C_2H_5)_2TiCl_2$ the transition occurs between $E_B = 458$ and 459 eV. Clearly the magnitude of the chemical shift alone is not sufficient to establish the nature of the compound.

Aside from chemical bonding information, XPS has an important advantage relative to AES. X-rays are less prone to damage surfaces than electrons. For example, electron beams can reduce hydrocarbon contaminants to carbon, destroying the sought-after evidence. For this reason, XPS tends to be preferred in assessing the cleanliness of substrates and semiconductor films during MBE growth.

10.4.6 A Couple of Applications in GaAs Films

10.4.6.1 AES

As an example of AES consider the depth profiles for Ga and Al shown in Fig. 10-29. The structure represents a single-crystal GaAs substrate onto which a 2000 Å thick, compositionally graded film of $Al_xGa_{1-x}As$ was

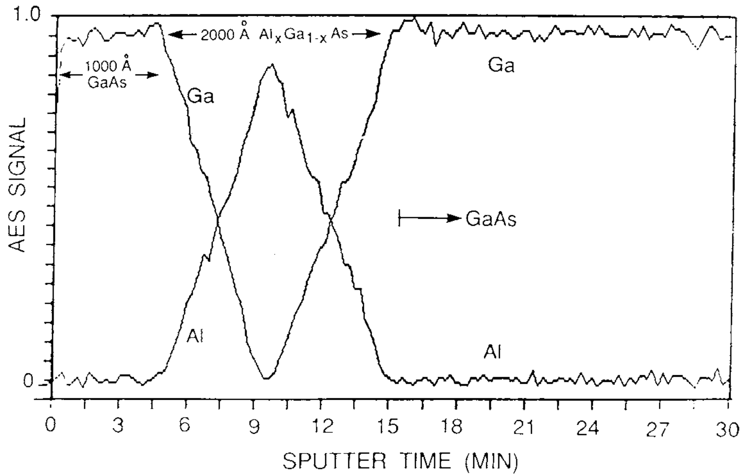


Figure 10-29 AES depth profiles of Al and Ga through GaAs and $Al_xGa_{1-x}As$ films. Signal for As not shown. (Courtesy of R. Kopf, AT&T Bell Laboratories.)

grown by molecular beam epitaxy. At first the deposition rate of Al was increased linearly while that of Ga correspondingly decreased until AlAs formed; then the deposition rates were reversed until the GaAs composition was attained after 2000 Å. Finally, a 1000-Å cap film of GaAs was deposited resulting in a 2000-Å wide, V-shaped quantum well sandwiched between GaAs layers. Because the lattice constants of GaAs, AlAs, and the intermediate $Al_xGa_{1-x}As$ compositions differ by only 0.16% at most, the entire "lattice-matched" structure is free of crystallographic defects. During AES depth profiling, the film was sputtered away exposing surface compositions in reverse order to those which were initially deposited. The linearly graded walls of the quantum well reflect the precise control of deposition that is possible. In this case the values of S_{Al} and S_{Ga} are apparently independent of composition.

10.4.6.2 XPS

Here we consider an AlGaAs film that has been etched in a $CF_2Cl_2 + O_2$ plasma in order to fabricate devices. It is desired to determine the composition of the film. But both electron-beam impingement and ion bombardment during sputter depth-profiling would alter or even remove surface compounds. What is needed is a technique to probe surface layers nondestructively. Angle-resolved XPS, based on altering the take-off angle for electron detection, is such a method. If the angle (θ) exiting photo-

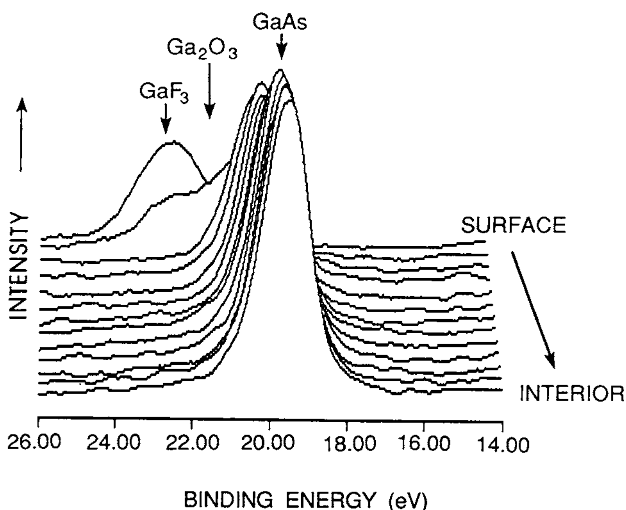


Figure 10-30 Angle-resolved XPS spectra of Ga 3d line as a function of electron detection angle. (Courtesy of M. Vasile, AT&T Bell Laboratories.)

electrons make with the surface plane is large, then chemical information on deep surface layers is sampled. However, if electrons exiting at a small grazing angle are detected, then only the top surface layers are probed; photoelectrons generated within deeper layers simply never emerge because their effective range (l_e) is smaller than the now-longer geometric escape-path length. In fact, the signal intensity from atoms at depth z varies as

$$I = I_0 \exp - \frac{z}{l_e \cos\theta}.$$

In the analysis shown in Fig. 10-30, θ was varied from 90° to 20° while the Ga 3d peak was scanned. At the surface, bonding associated with GaF_3 and Ga_2O_3 compounds was detected. Deeper within the film only Ga bound within GaAs or AlGaAs is evident.

10.4.7 RUTHERFORD BACKSCATTERING (RBS)

10.4.7.1 Physical Principles

This popular thin-film characterization technique (Ref. 41) relies on the use of very high energy (MeV) beams of low-mass ions. These have the property of penetrating thousands of angstroms or even microns deep into

films or film/substrate combinations. Interestingly, such beams cause negligible sputtering of surface atoms. Rather, the projectile ions lose their energy through electronic excitation and ionization of target atoms. These electronic collisions are so numerous that the energy loss can be considered to be continuous with depth. Sometimes the fast-moving light ions (usually $^4He^+$) penetrate the atomic-electron-cloud shield and undergo close-impact collisions with the nuclei of the much heavier stationary target atoms. A resulting scattering from the Coulomb repulsion between ion and nucleus has been long known in nuclear physics as Rutherford scattering. A primary reason that this phenomenon has been so successfully capitalized upon for film analysis is that classical two-body elastic scattering is operative. This, perhaps, makes RBS the easiest of the analytical techniques to understand.

We start by first considering an ion of mass (atomic weight) M_0 and energy E_0 incident on a surface as shown in Fig. 10-31. An elastic collision between this ion and a surface atom of mass M occurs, such that afterward the ion energy is E_1. The collision is insensitive to the electronic configuration or chemical bonding of target atoms, but depends solely on the masses and energies involved. As a consequence of conserving energy and momentum it is readily shown that

$$E_1 = \left\{ \frac{(M^2 - M_0^2 \sin^2\theta)^{1/2} + M_0 \cos\theta}{M_0 + M} \right\}^2 E_0 \qquad (10\text{-}24)$$

where θ is the scattering angle. For a particular combination of M_0, M, and θ the simple formula

$$E_1 = K_M E_0 \qquad (10\text{-}25)$$

relates the energy of emergent ions to that of the incident ions. K_M is known as the kinematic factor and can be calculated from Eq. 10-24. Once the incident ion, e.g., $^4He^+ (M_0 = 4)$ at $E_0 = 2$ MeV, and angular position of the ion detector are selected (θ is typically 170°), K_M just depends on the atomic weight of the target atom. For example, under the conditions just given, $K_{Pt} = 0.922$ and $K_{Si} = 0.565$. This means that if a 2 MeV He ion collides with a Pt atom ($M = 195$) located on the outer surface of a PtSi film, it will backscatter into the detector with an energy of 1.844 MeV; similarly, He ions scattering from surface Si atoms will have their energy reduced to 1.130 MeV.

Consider now what happens when a 2 MeV, $^4He^+$ ion beam impinges on a 900 Å platinum silicide film on a Si wafer substrate. In Fig. 10-31a the film/substrate geometry is shown together with a schematic of the energy changes the ions undergo during nuclear scattering and the corresponding

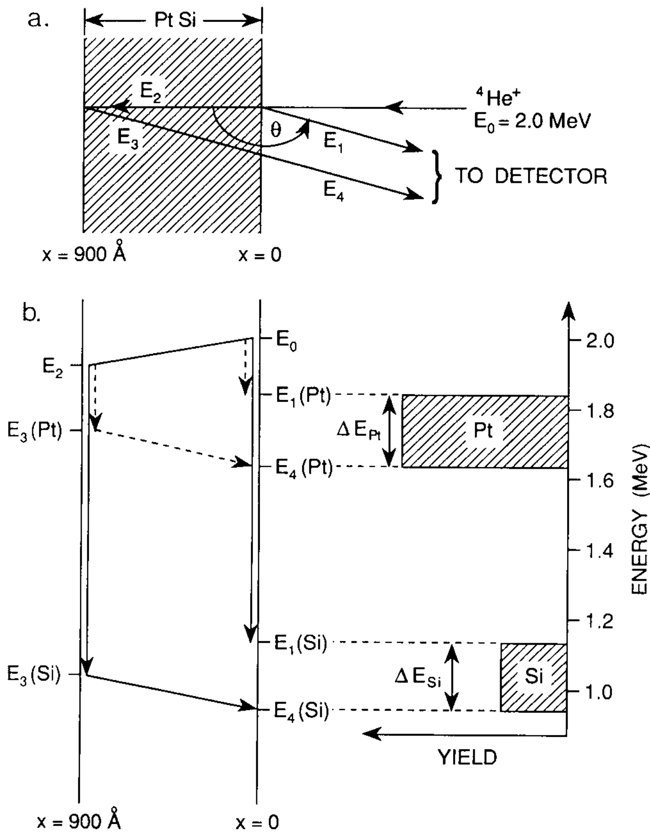


Figure 10-31 (a) Geometry of scattering and notation of energies at the front and back surfaces of a 900 Å thick PtSi film. (b) $^4He^+$ ion energy as a function of film depth due to scattering from Pt and Si. Schematic RBS spectrum shown rotated by 90°. (From W. K. Chu, J. W. Mayer, M. A. Nicolet, T. M. Buck, G. Amsel, and F. Eisen, *Thin Solid Films* **17**, 1 (1963), with permission from Elsevier Sequoia S.A.)

RBS spectrum (Fig. 10-31b). Energy changes ($E_0 \rightarrow E_1$) for scattering from Pt and Si surface atoms are represented on the $x = 0$ axis. The majority of the He ions penetrate below the film surface, however, where they continuously lose energy at a linear rate ($E_0 \rightarrow E_2$) with distance traversed. At any film depth they can suffer an atomic collision. The scattered ion energy is still given by Eq. 10-25, but E_0 is now the incident He^+ energy at that point in the matrix. Some of the energy-attenuated He^+ ions can reach the PtSi/Si interface where they may backscatter ($E_3 = K_M E_2$) and again lose energy

($E_3 \rightarrow E_4$) in traversing the film backward until they finally exit. Other He^+ ions can even penetrate the Si substrate where they eventually backscatter. It is important to realize that in the course of passage through the film, the ion beam can be thought of as splitting into two separate elemental components, each spanning a different range of energies. For each broad elemental peak detected, the highest and lowest energies correspond to atoms on the front and back film surfaces, respectively. Statistics largely govern the depth at which scattering occurs and whether Pt or Si atoms participate in the collisions.

Thus, by measuring the number and energy of backscattered He ions, information on the nature of the elements present, their concentration, and their depth distribution can all be simultaneously acquired without appreciably damaging the specimen.

10.4.7.2 Equipment

A schematic of the experimental arrangement employed for RBS is shown in Fig. 10-32. Lest the reader be deceived by the figure, it should be appreciated that the actual facility shown in the photograph of Fig. 10-33 is some 15 meters long and occupies more than 100 m^2 of floor space. Ions for analysis, e.g., $^4He^+$, $^{12}C^+$, or $^{14}N^+$, are accelerated at high voltages generated by the Van de Graaff accelerator. After entering the evacuated ($\sim 10^{-6}$–10^{-7} torr) beam line, the ions are then collimated and focused. Ion mass selection occurs in the bending magnet which geometrically disperses ions according to their mass. The resultant ion beam is then raster-scanned across the surface of the specimen. Backscattered ions are analyzed with respect to their energy by a silicon surface-barrier detector capable of an energy resolution of ~ 15 keV (peak width at half maximum amplitude). The

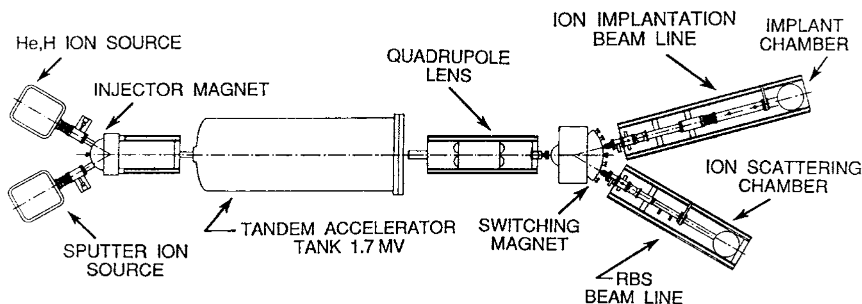


Figure 10-32 Schematic of the 1.7 MV tandem accelerator, RBS facility at AT&T Bell Laboratories, Murray Hill, NJ.

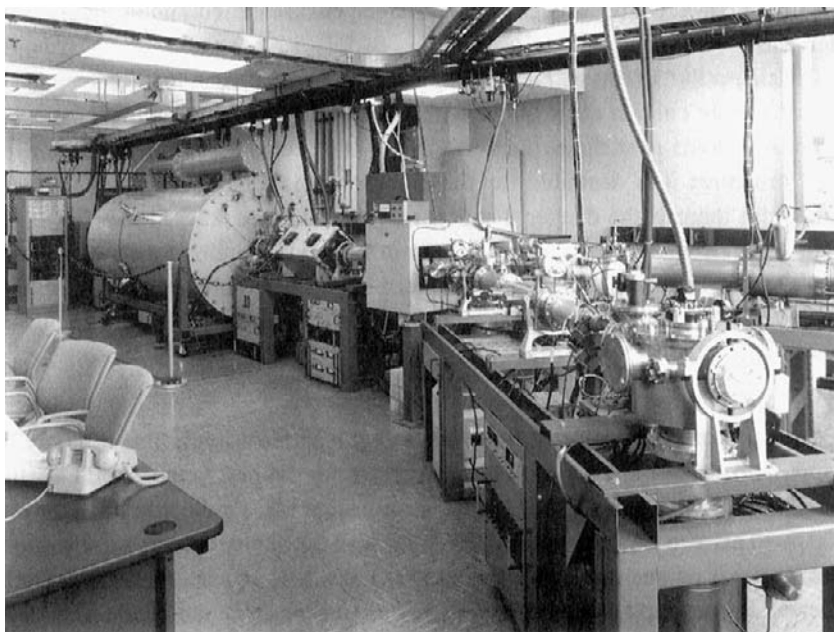


Figure 10-33 Photograph of 1.7-MV tandem accelerator/RBS facility. (Courtesy of D. C. Jacobson, AT&T Bell Laboratories.)

electronic pulses are then amplified and sorted according to voltage amplitude, or energy, by a multichannel analyzer to yield the resulting RBS spectrum.

10.4.7.3 Elemental Information

All elements and their isotopes including Li and those above it in the periodic table are, in principle, detectable with $^4He^+$ ions. The critical test is how well neighboring elements are resolved and this ultimately depends on the detector resolution. With 2 MeV $^4He^+$ ions, isotopes with $\Delta M = 1$ can generally be separated for M below approximately 40. At values of $M \simeq 200$ only atoms for which $\Delta M > 20$ can be resolved. Thus, ^{209}Bi and ^{190}Os would be indistinguishable. The apparent advantage in separating low-Z elements is offset by their low cross sections (σ_i) for scattering. The σ_i are a measure of how efficiently target atoms scatter incoming ions and depend on Z_i, particle energy E, the masses involved, and the angle of scattering. To a good approximation their dependence on these quantities

for the two ions—(M_0, Z_0) and (M, Z)—involved varies as

$$\sigma = (Z_0 Z q^2/4E)^2[(\sin\theta/2)^{-4} - 2(M_0/M)^2], \qquad (10\text{-}26)$$

where the term in square brackets is an important correction for low-mass targets. Clearly, high-Z elements produce a stronger backscattered signal than low-Z elements.

Consideration of these factors suggest that specimens for RBS analysis should ideally contain elements of widely different atomic weight stacked with the heavy atoms near the surface and the lighter atoms below them. In bilayer film couples it is desirable for the high-M film to be at the outer surface. Otherwise there is the danger that elemental peaks will overlap. In general, when applicable, RBS can detect concentration levels of about 1 at.%. The technique is unmatched in determining the stoichiometry of thin-film binary compounds such as metal silicides where accuracies of $\sim \pm 1\%$ or so are achieved.

10.4.7.4 Spatial and Depth Resolution

Since MeV ion beams can only be focused to a spot size of a millimeter or so in diameter, the lateral spatial resolution of RBS is not great. The depth resolution is commonly quoted to be 200 Å but can be improved to 20 Å by altering the geometry of detection. Grazing exit angles are employed to make the film appear to be effectively thicker. For example, when $\theta = 95^\circ$, corresponding to an exit angle of 5°, ion scattering at a given film depth means that the energy loss path length is an order of magnitude longer. Implicit in the use of RBS is the desirability that the specimen surface and underlying layered structures be planar. Fortunately, polished Si wafers are extraordinarily flat. Films grown or deposited on Si maintain this planarity and are thus excellently suited for RBS analysis. Films with rough surfaces yield broadened RBS peaks.

The maximum film depth that can be probed depends on the ion used, its energy, and the nature of the matrix. Typically, $\sim 1\ \mu$m is an upper limit for 2 MeV $^4He^+$ on Si. On the other hand, $^3H^+$ beams of 2 MeV penetrate $\sim 5\ \mu$m deep.

10.4.7.5 Quantification

Despite the limitations noted earlier, RBS enjoys the status of being the preferred method of analysis in situations where it is applicable. The basic reasons are that it is quantitative from first principles, does not require elemental standards, and yields simultaneous depth and film thickness information. Clearly the area (A) under a spectral peak represents the total

number of atoms of a given element present within the probed region or layer. The peak height (H) is directly proportional to the atomic concentration. Since the peak width (ΔE) is dependent on the maximum length traversed by projectile ions in the layer, ΔE is directly proportional to the layer or film thickness if the ion energy attenuation with distance is known. Particle range equations derived for nuclear physics applications yield this information to a high degree of accuracy. Hence, RBS provides a way to determine film thickness.

As an exercise in chemical analysis let us consider the experimental PtSi spectrum (Fig. 10-34) and calculate the relative proportions of Pt and Si present. By analogy to Eq. 10-23 the concentration ratio is given by

$$\frac{C_{\mathrm{Pt}}}{C_{\mathrm{Si}}} = \frac{A_{\mathrm{Pt}}/\sigma_{\mathrm{Pt}}}{A_{\mathrm{Si}}/\sigma_{\mathrm{Si}}} \tag{10-27}$$

where A_i are the peak areas, and the scattering cross sections are, in effect, sensitivity factors. For PtSi, $A_{\mathrm{Pt}}/A_{\mathrm{Si}}$ is measured to be 32, and $\sigma_{\mathrm{Pt}}/\sigma_{\mathrm{Si}}$ is estimated to be equal to $(Z_{\mathrm{Pt}}/Z_{\mathrm{Si}})^2 = 31.0$, so that $C_{\mathrm{Pt}}/C_{\mathrm{Si}} = 1.03$. Alterna-

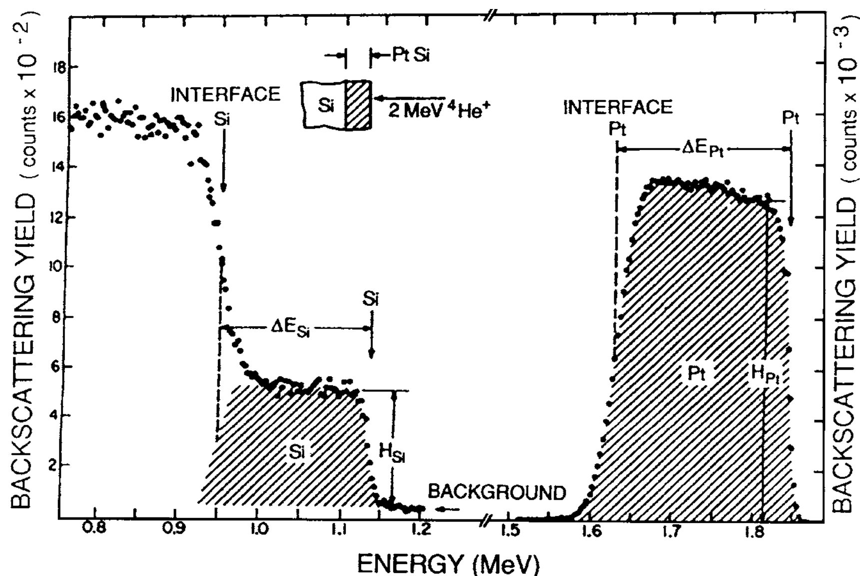


Figure 10-34 Energy spectrum for 2 MeV $^4He^+$ ions backscattered from 900 Å of PtSi. (From W. K. Chu, J. W. Mayer, M. A. Nicolet, T. M. Buck, G. Amsel, and F. Eisen, *Thin Solid Films* **17**, 1 (1963), with permission from Elsevier Sequoia S.A.)

tively, since $A_i = H_i \Delta E_i$,

$$\frac{C_{\text{Pt}}}{C_{\text{Si}}} = \frac{\sigma_{\text{Si}} H_{\text{Pt}} \Delta E_{\text{Pt}}}{\sigma_{\text{Pt}} H_{\text{Si}} \Delta E_{\text{Si}}}. \tag{10-28}$$

Experimental measurement reveals that $H_{\text{Pt}}/H_{\text{Si}} = 27.5$ and $\Delta E_{\text{Pt}}/\Delta E_{\text{Si}} = 1.15$, and therefore $C_{\text{Pt}}/C_{\text{Si}}$ is calculated to be 1.02.

Although film thickness can be determined to within 5%, the procedure is somewhat involved. The required ion energy loss with distance data, expressed as stopping cross sections, are tabulated and can be calculated for any matrix (Ref. 35). Film thicknesses can, in principle, also be determined either from known total incident charge and detector solid angle, or through the use of thickness standards. Experimental difficulties, however, limit the utility of these latter methods in practice.

10.4.7.6 Channeling

The role of channeling was already discussed in Section 4.6.4.2 in connection with development of preferred orientations during film growth under ion bombardment. A review of this section reveals how channeling effects might be advantageously used to distinguish RBS spectral features when crystalline (c), polycrystalline, and amorphous (a) film layers coexist. If, for example, an a-Si film covers c-Si, the higher $^4He^+$ yield from a-Si will abruptly drop to the lower c-Si value at the a–c interface because of channeling in c-Si along certain crystallographic directions. During ion-implantation doping of semiconductors, channeling effects are unwelcome because dopant ions travel further, modifying their resulting profile in not easily predictable ways.

10.4.8 Secondary-Ion Mass Spectrometry (SIMS)

10.4.8.1 Physical Principles

The mass spectrometer, a common tool in the chemistry laboratory for the analysis of gases, has been dramatically transformed in recent years to create SIMS apparatus capable of analyzing the chemical composition of solid surface layers. A critical need to measure thermally diffused and ion-implanted depth profiles of dopants in semiconductor devices spurred the development of SIMS. In typical devices, peak dopant levels are about $10^{20}/\text{cm}^3$ while background levels are $\sim 10^{15}/\text{cm}^3$. These correspond to atomic concentrations in Si of 0.2% to 2×10^{-6}%, respectively. None of the analytical techniques considered thus far has the capability of detecting such

low concentration levels. The price paid for such high sensitivity is a complex spectrum of peaks that resembles Fig. 2-15, corresponding to the masses of detected ions and ion fragments. This necessitates the use of standards, composed of the specific elements and matrices in question, for quantitative determinations of composition.

In SIMS, a source of ions bombards the surface and sputters from its outermost film-layer neutral atoms, for the most part, but also positive and negative ions. Once in the gas phase, the ions are mass analyzed in order to identify the species present as well as determine their abundance. Since it is the secondary ion emission current which is detected in SIMS, high-sensitivity analysis requires methods for enhancing sputtered ion yields. Secondary-ion emission may be viewed as a special case of (neutral atom) sputtering. However, a comprehensive theory to quantitatively explain all aspects of secondary-ion emission (e.g., ion yields S^+ or S^-, escape velocities and angles, dependence on ion projectile and target material) does not yet exist. Reliable experiments to test proposed theories are difficult to perform. Experimentally, it has been found that different ion beams interact with the specimen surface in profoundly different ways. For example, the positive metal ion yield of an oxidized surface is typically enhanced tenfold and frequently more, relative to a clean surface. This accounts for the common practice of using O_2^- beams to flood the surface when analyzing positive ions. Similarly, the negative ion signals can be enhanced by using Cs^+ primary ion beams.

One of the theories which attempts to explain the opposing effects of O_2^- and Cs^+ beams involves charge transfer by electron tunneling between the target and ions leaving the target surface. Negative ion (O_2^-) bombardment repels charge from the surface, in effect lowering its Fermi energy and raising its effective work function (ϕ). Tunneling is now favored from the surface atom (soon to be an ejected positive ion) into the now-empty electron states of the target. Similarly, positive-ion (Cs^+) bombardment lowers the target work function. Now electrons tunnel from the target into empty levels of surface atoms, enhancing the creation of negative ions. Since these charge transfer processes depend exponentially on ϕ, very large changes in ion yields with small shifts in ϕ are possible.

10.4.8.2 Instrumentation

A schematic depicting the basic elements of a conventional double-focusing SIMS spectrometer is shown in Fig. 10-35. The primary ions most frequently employed are Ar^+, O_2^-, and Cs^+, and these are focused into a beam with energy ranging from 2 to 15 keV. Sputtered charged atoms and compound fragments that are extracted enter an electrostatic energy

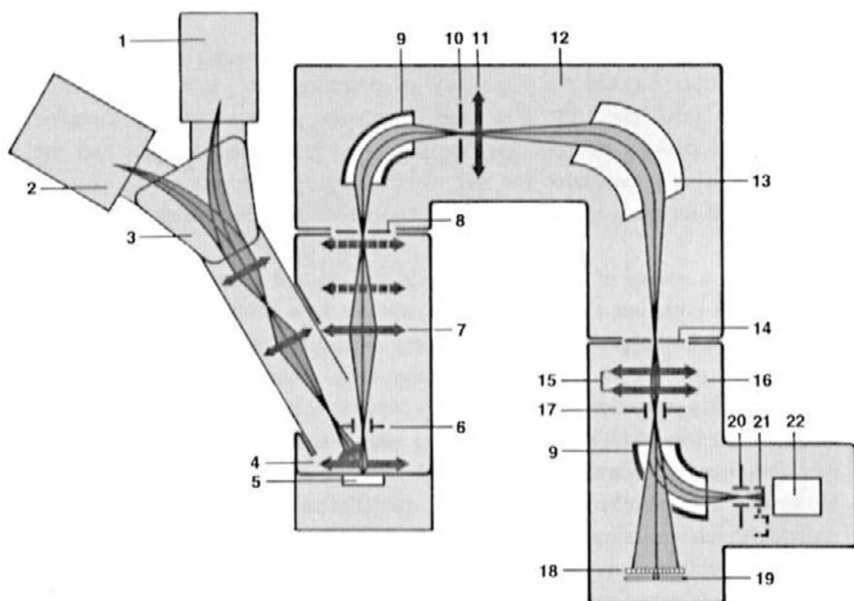


Figure 10-35 Schematic of the ion optical system in the Cameca double-focusing mass spectrometer. 1, Cs ion source; 2, duoplasmatron source; 3, primary beam mass filter; 4, immersion lens; 5, specimen; 6, dynamic transfer system; 7, transfer optical system; 8, entrance slit; 9, electrostatic sector; 10, energy slit; 11, spectrometer lens; 12, spectrometer; 13, electromagnet; 14, exit slit; 15, projection lens; 16, projection display and detection system; 17, deflector; 18, channel plate; 19, fluorescent screen; 20, deflector; 21, Faraday cup; 22, electron multiplier. (Courtesy Cameca Inc., Stamford, Connecticut.)

analyzer similar to the cylindrical and spherical electron energy analyzers employed in AES and XPS work. Those secondary ions which pass now enter a magnetic-sector mass filter whose function is to select a particular mass for detection. The desired ion of mass M, charge q, and velocity v traces an arc of radius r in the magnetic field (B) of the electromagnet, given by

$$r = Mv/qB. \tag{10-29}$$

Two modes of operation are possible. In the imaging mode, B is fixed and simultaneous projection of ions of different mass on the channel plate yields mass spectra maps in the form of an ion micrograph (spectrography). In the mass spectrometry mode, B is scanned and each ion mass is sequentially detected by the electron-multiplier detector. Magnetic sector mass filters possess high resolution capability ($M/\Delta M \sim 3000$–$20{,}000$). Another com-

Table 10-5
Comparison between Static and Dynamic SIMS

Variable	Static SIMS	Dynamic SIMS
Residual gas pressure (torr)	10^{-9}–10^{-10}	10^{-7}
Primary ion energy (keV)	0.5–3	3–20
Current density (A/cm^2)	10^{-9}–10^{-6}	10^{-5}–10^{-2}
Area of analysis (cm^2)	0.1	10^{-4}
Sputter rate (s/monolayer)	10^4	1

mon version of SIMS employs a quadrupole lens of lower resolution, i.e., $M/\Delta M \sim 500$.

A further distinction is made between what is known as "static" and "dynamic" SIMS. The issue which distinguishes them is the rate of specimen erosion relative to the time necessary to acquire data. Static SIMS requires that data be collected before the surface is appreciably modified by ion bombardment. It is well suited to surface analysis and the detection of contaminants such as hydrocarbons. Dynamic SIMS, on the other hand, implies that high sputtering rates are operative during measurement. This, of course, enables depth profiling of surface layers. Typical operating parameters for both static and dynamic SIMS are given in Table 10-5.

10.4.8.3 Two Extraordinary Capabilities

Among the unique features of SIMS are the ability to discriminate between ions of different mass, and to quantitatively measure very shallow dopant profiles. The first capability often provides interesting processing information. For example, consider a bipolar transistor which contains both ^{121}Sb and ^{123}Sb in the emitter but only ^{121}Sb in the collector. The emitter doped with both naturally occurring isotopes of antimony must have been diffused while the collector was obviously ion implanted. Extreme discrimination is required when two species of very similar mass must be distinguished. For example, to separate the ^{31}P (mass = 30.974) signal from the interfering ^{30}SiH (mass = 30.982) background level, the resolution required is $M/\Delta M = 3872$. With high-resolution mass spectrometers these species have been separated and phosphorous doping profiles obtained as shown in Fig. 10-36.

The second and more important capability is the critical detection and monitoring of dopant-ion implantation profiles in field-effect transistors.

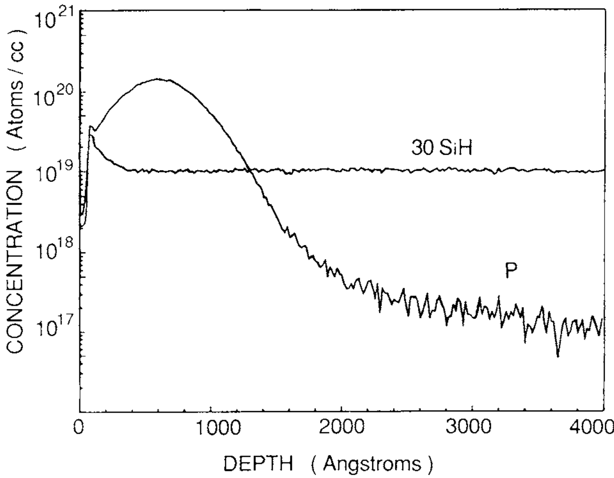


Figure 10-36 SIMS depth profile of P implanted in Si obtained with Cs^+ ions. (Courtesy of H. Luftman, AT&T Bell Laboratories.)

For this purpose primary O_2^+ ion beams are essential to enhance ionization efficiency. It is also important that the implant profile lie beneath or deeper than a surface transition region, a damaged layer where ion yields change continuously and quantification is problematical. However, once this layer is sputtered through and a steady state between incident and emitted primary ions occurs, constant secondary yields and accurate quantification are possible. When minimum feature dimensions were 0.25 μm or greater, i.e., prior to 1997, SIMS analysis encountered little difficulty because the implant profile lay deeper than the ~10 to 30-nm thick surface transition region. In current 0.18 μm technology, a typical ultralow-energy dopant-ion implantation employs 500 eV boron with a projected range of only 3 nm. This means the transition region must be reduced to ~0.1 to 0.4 nm, which is achievable with correspondingly low primary ion-beam energies. With such techniques the extremely shallow boron implant profile shown in Fig. 10-37 was characterized (Ref. 45).

10.4.9 Applications

The chapter closes with the self-explanatory Table 10-6 in which some applications of the surface analytical techniques just considered are listed. Varied phenomena, materials, and structures included extend into all facets

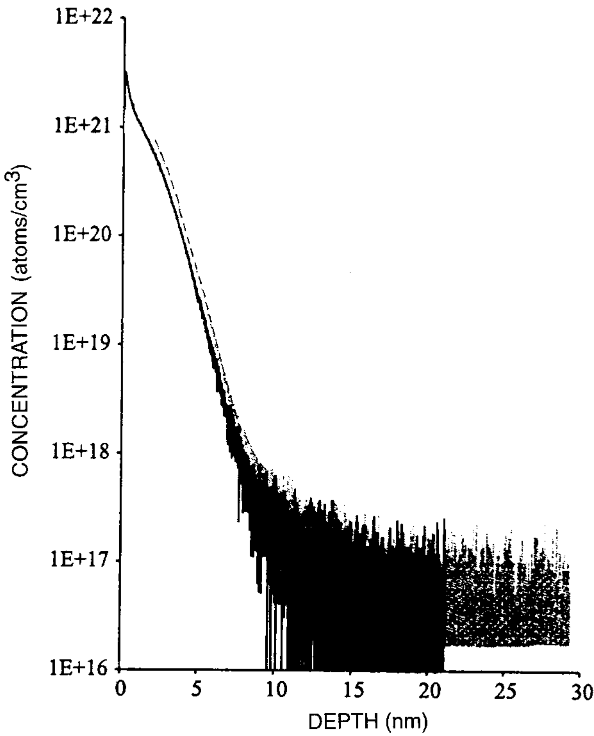


Figure 10-37 SIMS profiles of 1 keV BF_2 in silicon with dose of 3.5×10^{14} atoms/cm^2. Analytical conditions: (dotted gray) 310 eV O_2^+ primary ions, oblique incidence at 50° with oxygen flood; (solid black) 220 eV O_2^+ primary ions, normal incidence without oxygen flood. (From Ref. 45. Reprinted with the permission of Charles W. Magee.)

of thin-film science and technology. Ever-expanding applications will ensure future growth of the list.

10.5 CONCLUSION

The emergence of characterization instrumentation with unprecedented measurement capabilities has helped fuel the remarkable advances we have witnessed in thin-film science and technology. This chapter has attempted to convey how the insatiable and seemingly insurmountable demands for precise monitoring, control, and measurement of film thickness, structure, and composition are accomplished. It is remarkable that such

Table 10-6

Applications of Surface Analytical Techniques in Thin Films

Application	Information obtained, comments	Technique
1. Film nucleation, growth	Distinctions among island, layer, and S-K growth	AES
2. Diffusion in films	Diffusion coefficients are obtained by sputter sectioning and surface accumulation methods	AES, SIMS, RBS
3. Doped semiconductors	Diffused and ion-implanted depth profiles in Si and GaAs	SIMS
4. Compound formation	Stoichiometry of intermetallic and silicide compounds, growth kinetics	RBS, AES
5. Compound chemistry	Identification of valence and bonding state	XPS
6. Investigation of surface residues, stains, haze, and discoloration	Identification of elemental contaminants, reliability and failure studies	AES, SIMS, XPS
7. Contamination of surfaces by organic materials	Identification of elements Identification of compounds	AES, SIMS, XPS XPS
8. Interfacial analysis	Cause of adhesion failure, segregation of impurities at grain boundaries and interfaces	AES, SIMS, XPS
9. Multilayer films and coatings, superlattices	Stoichiometry, layer thickness, interfacial impurities	AES, RBS, SIMS
10. Determination of crystalline perfection	Channeling spectra distinguish between single-crystal and amorphous Si films	RBS
11. Fracture of coatings	Segregation of impurities at fracture surfaces	AES, SIMS
12. Metal–semiconductor contacts	Adhesion, contact reactions in Si and GaAs	AES, XPS, SIMS, RBS
13. Dielectric films on metals and semiconductor	Surface contamination, impurity diffusion, and segregation at interfaces (e.g., SiO_2–Si)	AES, SIMS, XPS
14. Molecular beam epitaxy	Assessment of surface cleanliness prior to deposition, detection of C and O contaminants	AES, XPS

characterization has been made against a backdrop of shrinking film dimensions, the need to analyze multilayered film structures, and the necessity to make real-time measurements in film deposition and etching environments. Capitalizing on an infrastructure of advances in electronic and optical instrumentation, computer technology and software development, all film characterization capabilities have dramatically expanded in the past decade. For example, prior thin-film thickness determinations by both interferometry and ellipsometry were practically limited to a single film; extraction of n and k as a function of wavelength was very time consuming. Today commercial spectral reflectometers and ellipsometers enable measurements over a broad spectral range to greatly expand the optical information content; besides n and k values, dielectric multilayer stacks can be analyzed, and all in a relatively short time. Furthermore, a totally new photoacoustic technique based on picosecond ultrasonics enables the thickness of individual opaque (metal) films within stacks to be rapidly measured with good accuracy. In summary, it is now possible to nondestructively measure the thickness, characterize the surface roughness, and model subsurface defects in both individual dielectric and metal films and multilayers.

Microstructural and morphological characterization of films by TEM and SEM methods have been augmented in recent years by a host of scanning probe microscopies. Few structural secrets remain now that atomic-resolution imaging can be attained on film surfaces by the STM and AFM, as well as in the film interior and at interfaces with TEM plan and cross-sectional views. Compositional characterization of films relies either on the unique electronic structure or the mass of the elements. Fingerprinting atoms by AES and XPS methods capitalizes on the former, whereas SIMS and RBS are based on the latter. It is a desideratum in all of these techniques to be able to unambiguously identify elemental species, detect fewer atoms, reduce the lateral and depth dimensions being probed, and perform the resulting analysis in shorter time periods. Prime objectives are real-time compositional monitoring and control of thin-film processing where possible, and the extension of these capabilities to multilayers. These challenges are daunting and require the most sophisticated and costly instrumentation employed in film characterization.

EXERCISES

1. The reflectance plotted in Fig. 10-2 has the mathematical form

$$R = \frac{r_1^2 + r_2^2 + 2r_1 r_2 \cos\delta}{1 + r_1^2 r_2^2 + 2r_1 r_2 \cos\delta},$$

where $\delta = 4\pi n_1 d/\lambda$, $r_1 = (n_0 - n_1)/(n_0 + n_1)$ and $r_2 = (n_1 - n_2)/(n_1 + n_2)$.

(a) Verify this equation by plotting R vs optical thickness/wavelength, i.e., $(n_1 d/\lambda)$, for the case $n_0 = 1.0$, $n_1 = 1.7$, and $n_2 = 1.4$.

(b) Replot for the case where $n_0 = 1.0$, $n_1 = 1.3$, and $n_2 = 1.4$.

(c) Physically distinguish between reflection of light from the surface of a 1000 Å thick SiO_2 film on a Si substrate, and reflection from the surface of bulk SiO_2.

2. Consider a 4428 Å thick Nb_2O_5 film on a quartz substrate ($n = 1.45$). Plot the reflectivity of this combination using $n(\lambda)$ values shown in Fig. 10-3b and the formula for R given in exercise 1. Note that this form of R is valid only when $k = 0$. How does the R response compare with that shown in Fig. 10-3a over the wavelength range 400–800 nm?

3. (a) A three-layer interconnect metallization has the following film thicknesses: TiN (40.83 nm) top; Ti (31.91 nm); Al–Cu (417.91 nm) bottom. If v_s for Al–Cu is 6350 m/s, sketch to scale the expected reflectivity–time response for the first two sets of echoes if this metallization is probed by ultrasonic methods.

(b) Will ultrasonic metrology be effective in determining layer thicknesses in dielectric film stacks? Explain.

4. Model the vibrating quartz-crystal thickness monitor as a mass (m)–spring combination, where k is the spring constant.

(a) What is the resonant frequency?

(b) Show that as additional mass δm deposits the difference in resonant frequency or frequency shift is given by $\Delta f \simeq -f_0 \delta m/2m$ for $\delta m/m \ll 1$.

5. After monitoring the thickness of a deposited Au film with a 6.0 MHz quartz (AT cut) crystal monitor, a researcher decides to confirm his results employing interferometry. A frequency shift of 10.22 Hz was recorded for the film measuring 1.00 cm^2 in area. Interferometry with the Hg green line revealed a displacement of 1.75 fringes across the film step. Are these measurements consistent? If not, suggest plausible reasons why not. *Note:* Density of Au is 19.3 g/cm^3.

6. Based on the diffraction pattern of Fig. 10-19, what are the lattice parameters for Ag and Pb? *Note:* $\lambda_{CuK\alpha} = 1.54$ Å.

7. Explain how an SEM/EDX facility could be used to measure the thickness of a film without detaching it from substrate B (film thickness standards may be necessary). Can the film thickness be measured this way if film and substrate have the same composition?

8. Suppose the specimen–tip current in an STM is given by

$$i = q^2 V_T/2\pi h \exp - [(4\pi/h)(2m)^{1/2}\phi^{1/2}s],$$

where q is the electronic charge, h is Planck's constant, ϕ is the work function of the tip metal, V_T is the tip voltage, and m is the mass of the electron.

(a) Derive an expression for the fractional change in tip current as a function of fractional change in tip spacing.

(b) If $\phi = 4$ V, select a reasonable set of V_T values so that a 1 Å increase in s will cause a factor of 10 decrease in i.

9. X-ray rocking curves for an epitaxial GaInAsP film lattice-matched to InP were recorded for the (002), (006), and (117) reflections using CuKα radiation.
   (a) What are the Bragg angles for these reflections in InP?
   (b) Suppose the lattice mismatch between the epilayer and substrate is 1.3×10^{-4}. What Bragg angles correspond to the InGaAsP peaks?
   (c) For which reflection is the peak separation greatest?

10. The surface of a film contained the following elements with the indicated sensitivities, S, and Auger intensities, I (in arbitrary units)

Element	S	I
Ga	0.68	6,950
As	0.68	5,100
Al	0.23	3,040
O	0.71	26,900
C	0.30	5,000
F	1.0	40,500
Cl	0.89	3,700

   (a) What is the composition of the surface in atom percent?
   (b) Are the binding energies of Ga in the compounds GaF_3, Ga_2O_3, and GaAs (Fig. 10-30) consistent with what you know about the chemistry of these materials?

11. Auger electrons emanating from A atoms located a depth z below the surface give rise to a signal of intensity

$$I_{\mathrm{AES}} = K \int_0^{\infty} C(z) \exp - (z/\lambda_e)\, dz,$$

where λ_e is the electron escape depth, $C(z)$ is the concentration of A atoms, and K is a constant.

   (a) What is the intensity from a pure A material?
   (b) A pure A substrate is covered by a film of thickness d so that $C_A(z) = 0$ for $d \geqslant z$, and $C_A(z) = 1$ for $\infty \geqslant z \geqslant d$. Show that $I_A = I_A(0) \exp - (z/\lambda_e)$, where $I_A(0)$ is the signal intensity from a pure A surface. Does this suggest a way to determine d?

12. A 2000-Å tungsten film is sputtered on a flat polycrystalline diamond substrate.
    (a) Determine the kinematic factors for tungsten and carbon if 2.0 MeV $^4He^+$ ions are employed and the detector angle is 170°.
    (b) Draw the expected RBS spectrum.
    (c) After annealing at a temperature of 1200°C some W_2C formed. Sketch the RBS spectrum assuming only half of the original tungsten reacted to form carbide.
    (d) What would the RBS spectrum look like if all of the original tungsten film reacted with diamond to form W_2C?

13. (a) Sketch the RBS spectrum for a 900 Å thick Pt film on a Si wafer. How does it differ from the spectrum of 900 Å of PtSi on Si? In both cases 2.0 MeV $^4He^+$ ions are employed.
    (b) Sketch the RBS spectrum for the case of a 900 Å Si film on a thick Pt substrate.

14. The implanted P dopant distribution shown in the SIMS spectrum of Fig. 10-36 can be described by Eq. 4-43.
    (a) Show that the distribution appears to be Gaussian by replotting the data as $\ln C$ vs $(z - R_p)^2$.
    (b) What is the value of R_p, the projected range? What is the value for ΔR_p, the longitudinal straggle?
    (c) What is the dose ϕ?

15. Recommend specific structural (e.g., film thickness, topography) and chemical characterization (i.e., composition) techniques for analyzing the following samples. In each case indicate potential difficulties in the analysis.
    (a) The surface of a glass slide.
    (b) A 1000 Å thick layer of crystalline SiO_2 buried 3000 Å below the surface of a crystalline silicon wafer.
    (c) A thin film of a Bi–Th alloy. (Thorium is radioactive, emitting high-energy α, β, and γ radiation.)
    (d) A thin carbon deposit on a Be substrate.
    (e) A "black" Al film surface consisting of a deeply creviced, rough, moundlike topography. (Surface features are typically 1 μm in size.)
    (f) A liquid Ga film surface at 35°C.
    (g) WC nanopowder, 40 nm in size.
    (h) The purple film coating a glass camera lens.
    (i) A steel drill coated with a gold-colored metallic layer.
    (j) A plastic potato chip bag coated with a thin film.

REFERENCES

1. J. B. Bindell, in *VLSI Technology*, 2nd ed., ed. S. M. Sze. McGraw-Hill, New York, 1988.
2. W. A. Pliskin and S. J. Zanin, in *Handbook of Thin Films Technology*, eds. L. I. Maissel and R. Glang. McGraw-Hill, New York, 1970.
3. K. L. Chopra, *Thin Film Phenomena*. McGraw-Hill, New York, 1969.
4. H. K. Pulker, *Coatings on Glass*, 2nd ed. Elsevier, Amsterdam, 1999.
5. S. Tolansky, *Multiple Beam Interference Microscopy of Metals*. Academic Press, London, 1970.
6. J. D. Jackson, *Classical Electrodynamics*, 3rd ed. John Wiley & Sons, New York, 1999.
7. A. R. Forouhi and I. Bloomer, *Phys. Rev.* **B38**, 1865 (1988).
8. A. Callegari and K. Babich, *SPIE* **3050**, 507 (1997).
9. K. Vedam, *Thin Solid Films* **313–314**, 1 (1998).
10. J. B. Theeten and D. E. Aspenes, *Ann. Rev. Mater. Sci.* **11**, 97 (1981).
11. E. A. Irene and J. A. Woollam, *MRS Bull.* **XX**, 24 (1995).
12. H. G. Tompkins, *A User's Guide to Ellipsometry*. Academic Press, Boston, 1993.
13. T. E. Jenkins, *J. Phys. D: Appl. Phys.* **32**, R45 (1999).
14. P. J. McMarr, K. Vedam, and J. Narayan, *J. Appl. Phys.* **59**, 694 (1986).
15. C. T. Ballinger and R. Olds, *Vacuum Coating Technol.*, Jan/Feb, p. 33 (2000).
16. R. Glang, in *Handbook of Thin-Film Technology*, eds. L. I. Maissel and R. Glang. McGraw-Hill, New York, 1970.
17. C. J. Morath, G. J. Collins, R. G. Wolf, and R. J. Stoner, *Solid State Technology*, June (1997).
18. R. DeJule, *Semicond. Int.* **21**(5), 52 (1998). Also see p. 48, same issue.
19. M. Ohring, *Reliability and Failure of Electronic Materials and Devices*. Academic Press, Boston, 1998.
20. R. J. Stoner, C. J. Morath, G. Tas, S. Sengupta, S. M. Merchant, J. B. Bindell, and H. J. Maris, *Proc. SPIE* **3269**, 104 (1998).
21. J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, A. D. Ronig, C. E. Lyman, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*, 2nd ed. Plenum, New York, 1992.
22. D. E. Newbury, D. C. Joy, P. Echlin, and C. E. Fiori, *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum, New York, 1986.
23. G. Thomas and M. J. Goringe, *Transmission Electron Microscopy of Materials*. J. Wiley & Sons, New York, 1979.
24. D. B. Williams and C. B. Carter, *Transmission Electron Microscopy — A Textbook for Materials Science*. Plenum, New York, 1996.
25. T. Ishitani, T. Yaguchi, and H. Koike, *Hitachi Rev.* **45**(1), 19 (1996).
26. R. Wiesendanger and H.-J. Guntherodt, eds., *Scanning Tunneling Microscopy II*, 2nd ed. Springer-Verlag, Heidelberg, 1995.
27. J. A. Kubby and J. J. Boland, *Surf. Sci. Rep.* **26**, 61 (1996).
28. R. Howland and L. Benatar, *A Practical Guide to Scanning Probe Microscopy*. Park Scientific Instruments, 1996.
29. D. P. Woodruff and T. A. Delchar, *Modern Techniques of Surface Science*. Cambridge University Press, Cambridge, 1994.
30. T. Nakayama, D.-H. Huang, and M. Aono, *Microelectron. Eng.* **32**, 191 (1996).
31. M. F. Crommie, C. P. Lutz, and D. M. Eigler, *Science* **262**, 218 (1993).
32. M. Ohring, *The Materials Science of Thin Films*. Academic Press, Boston, 1991.
33. G. Binnig, C. F. Quate, and C. Gerber, *Phys. Rev. Lett.* **56**, 930 (1986).

34. S. Kitamura and M. Iwatsuki, *Jpn. J. Appl. Phys.* **34**, L145 (1995). Also S. Kitamura, K. Suzuki, and M. Iwatsuki, *JEOL News* **32E**(1), 42 (1996).
35. L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and Thin-Film Analysis*. North-Holland, New York, 1986.
36. J. C. Riviere, *Surface Analysis Techniques*. Oxford University Press, London, 1990.
37. J. B. Wachtman, *Characterization of Materials*. Butterworth–Heinemann, Boston, 1993.
38. D. Briggs and M. P. Seah, eds., *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. J. Wiley & Sons, Chichester, U.K., 1983.
39. C. R. Brundle and A. J. Baker, eds., *Electron Spectroscopy: Theory, Techniques, and Applications*, Vol. I–X. Academic Press, New York, 1979–present.
40. J. F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*. Oxford University Press, London, 1990.
41. J. R. Bird and J. S. Williams, eds., *Ion Beams for Materials Analysis*. Academic Press, Sydney, 1989.
42. R. G. Wilson, F. A. Stevie, and C. W. Magee, *Secondary Ion Mass Spectrometry*. J. Wiley & Sons, New York, 1989.
43. A. W. Benninghoven, F. G. Rudenauer, and H. W. Werner, *Secondary Ion Mass-Spectrometry — Basic Concepts, Instrumental Aspects, Applications and Trends*. J. Wiley & Sons, New York, 1987.
44. H. W. Werner and P. P. H. Garten, *Rep. Prog. Phys.* **47**, 221 (1984).
45. V. K. F. Chia, G. R. Mount, M. J. Edgell, and C. W. Magee, *J. Vac. Sci. Technol.* **B17**, 2345 (1999).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 11

Interdiffusion, Reactions, and Transformations in Thin Films

11.1 INTRODUCTION

There is hardly an area related to thin-film formation, properties, and performance that is uninfluenced by mass-transport phenomena. This is especially true of microelectronic applications where very small lateral as well as depth dimensions of device features and film structures are involved. When these characteristic dimensions (d) become comparable in magnitude to atomic diffusion lengths, then compositional change can be expected. New phases in the form of precipitates, layered compounds, or even voids may form from ensuing reactions and alter the initial film integrity through, for example, generation of stress or decrease of adhesion. This, in turn, frequently leads to component and device malfunction and electrical instabilities manifested by decreases in conductivity as well as short- or even open-circuiting of conductors. The time it takes for such effects to evolve can be roughly gauged by noting that the diffusion length is given by $d \simeq 2(Dt)^{1/2}$ (Eq. 1-29), where D and t are the appropriate diffusivity and time, respectively. Therefore, $t = d^2/4D$. As we shall later see, D values in films are relatively high even at low temperatures so that small film dimensions serve to make these characteristic times uncomfortably short. These effects frequently surface when neighboring combinations of materials are chemically reactive.

To illustrate the difference between diffusion in bulk and thin-film solids, consider the pitfalls involved in designing a Cu–Ni film couple as part of the contact structure for solar cells (Ref. 1). Readily available high-temperature data in bulk metals extrapolated to 300°C yield a value of $3.8 \times 10^{-24}\ \mathrm{cm^2/s}$ for the diffusion coefficient of Cu in Ni. For a 1000 Å thick Ni film, the interdiffusion time is thus predicted to be $(10^{-5})^2/4(3.8 \times 10^{-24})$ s or more than 200,000 years! Experiment, however, revealed that these metals intermixed in less than an hour as observed through film color or reflectivity change. What caused the rapid interdiffusion? As we shall see, the high density of defects, e.g., grain boundaries and vacancies, causes deposited films to behave differently from bulk metals, and it is a purpose of this chapter to quantitatively define the distinctions. Indeed, a more realistic estimate of the Cu–Ni reaction time will be made utilizing concepts developed in Section 11.2.3.2. Other examples will be cited later involving interdiffusion effects between and among various metal–film layer combinations employed in Si chip packaging applications. In fact, practical problems associated with making both stable contacts to semiconductor surfaces and reliable interconnections between devices have generated the bulk of the mass-transport-related concerns including interdiffusion, compound formation, and phase transformations.

Whereas interdiffusion phenomena are driven by chemical concentration gradients, there are other mass-transport effects which take place even in homogeneous films. These rely on other driving forces such as electric fields, thermal gradients, and stress fields which give rise to respective electromigration, thermomigration, and creep effects that can similarly threaten film integrity. The Nernst–Einstein equation provides an estimate of the characteristic times required for such transport effects to occur. Consider a narrow film stripe that is as wide as it is thick. If it can be assumed that the volume of film affected is $\sim d^3$ and that mass flows through a cross-sectional area $\sim d^2$, then the appropriate velocity is d/t. By utilizing Eq. 1-35 it is concluded that $t = RTd/DF$. Large driving forces (F), which sometimes exist in films, can conspire with both small d and high D values to reduce t to an unacceptably short time. As circuit dimensions continue to inexorably shrink in the drive toward higher device densities and faster operating speeds, diffusion lengths will decrease and the surface-area-to-volume ratio will increase. Despite these tendencies, processing temperatures and heat generated during operation are not being proportionately reduced. Therefore, interdiffusion problems are projected to persist and even worsen in the future.

In addition to what may be termed thin-film reliability concerns, there are beneficial mass-transport effects that are operative during film deposition and processing heat treatments. The reliance on surface diffusion during film deposition, production of semiconductor contacts, and growth of gate

oxides are examples. Aspects of both of these broad applications will be discussed in this chapter in a fundamental way.

11.2 FUNDAMENTALS OF DIFFUSION

11.2.1 SCOPE

This section expands on the survey of diffusion phenomena given in Chapter 1, which the reader may wish to review. The emphasis is on analytical modeling of mass transport mechanisms and effects, particularly grain-boundary diffusion and compound formation. Applications of the theories presented to actual film systems begin in Section 11.3.

11.2.2 COMPARATIVE DIFFUSION MECHANISMS

Diffusion mechanisms attempt to describe the details of atomic migration associated with mass transport through a medium. The resulting atom movements reflect the marginal properties of materials in that only a very small fraction of the total number of atomic sites, namely, those which are unoccupied, interstitial, or on surfaces, is involved. An illustration of this was the vacancy mechanism for diffusion given earlier in Section 1.6.2.1. By countless random repetitions of the fundamental atom-vacancy exchange step shown in Fig. 1-17, this mechanism gives rise to macroscopic transport effects in bulk solids. Although vacancy diffusion occurs in both polycrystalline thin films and bulk solids, the former usually contain a greater defect density as a consequence of low-temperature growth from the vapor rather than melt phase. Less tightly bound atoms at these nonlattice grain-boundary (GB), dislocation, and surface sites are expected to attract different point-defect populations and be more mobile than lattice atoms. Although the detailed environment may be complex and even varied, the time-averaged atomic transport is characterized by a similar effective diffusion coefficient

$$D_i = D_{0i} \exp - (E_i/RT), \tag{11-1}$$

where i represents lattice (L), grain boundary (GB), dislocation (d), or surface (s) diffusion mechanisms. Most importantly, the activation energies for grain-boundary (E_{GB}), dislocation (E_d), and surface diffusion (E_S) are smaller than E_L, leading to higher diffusivities. Therefore, such heterogeneities and defects serve as diffusion paths which short-circuit the lattice.

In order to better appreciate the consequences of allowing a number of uncoupled transport mechanisms to freely compete, it is instructive to

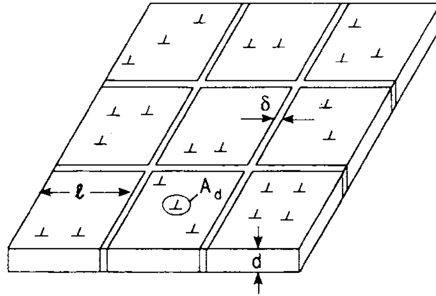


Figure 11-1 Highly idealized polycrystalline film containing square grains, grain boundaries, and dislocations.

consider the highly idealized polycrystalline film matrix shown in Fig. 11-1. Grain-boundary slabs of width δ serve as short-circuit diffusion paths even though they may only be 5–10 Å wide. They separate square grains of side l. Within the grains are dislocations oriented normal to the film surface. They thread the latter with a density ρ_d per cm^2, and diffusion is assumed to occur through the dislocation core whose cross-sectional area is A_d. Parallel transport processes normal to the film plane are assumed to occur for each mechanism. Under these conditions, the number of atoms ($\dot{n}_i$) which flow per unit time is essentially equal to the product of the appropriate diffusivity (D_i), concentration gradient $(dc/dx)_i$, and transport area involved. Therefore, for diffusion through a single grain,

$$\text{Lattice:} \qquad \dot{n}_L = D_L l^2 (dC/dx)_L, \tag{11-2a}$$

$$\text{Grain boundary:} \qquad \dot{n}_{GB} = \delta D_{GB} l (dC/dx)_{GB}, \tag{11-2b}$$

$$\text{Dislocation:} \qquad \dot{n}_d = A_d D_d l^2 \rho_d (dC/dx)_d. \tag{11-2c}$$

The importance of short-circuit mass flow relative to lattice diffusion can be quantitatively understood in the case of face-centered cubic metals where data for the individual mechanisms are available. A convenient summary of resulting diffusion parameters is given by (Ref. 2)

$$\text{Lattice:} \qquad D_L = 0.5 \exp - \frac{17.0 T_M}{T} \quad \frac{\text{cm}^2}{\text{s}}, \tag{11-3a}$$

$$\text{Grain boundary:} \qquad \delta D_{GB} = 1.5 \times 10^{-8} \exp - \frac{8.9 T_M}{T} \quad \frac{\text{cm}^3}{\text{s}}, \tag{11-3b}$$

$$\text{Dislocation:} \qquad A_d D_d = 5.3 \times 10^{-15} \exp - \frac{12.5 T_M}{T} \quad \frac{\text{cm}^4}{\text{s}}. \tag{11-3c}$$

These approximate expressions represent average data for a variety of FCC metals normalized to the reduced or homologous temperature T/T_M, where T_M is the melting point. As an example, the activation energy for lattice self-diffusion in Au is easily estimated through comparison of Eqs. 11-1 and 11-3a, which gives $E_L/RT = 17.0T_M/T$. Therefore, $E_L = 17.0RT_M$ or $(17.0)(1.99 \text{ cal/mol-K})(1336 \text{ K}) = 45{,}200$ cal/mol. As a first approximation, the foregoing equations can be assumed to be valid for both self- and dilute impurity diffusion.

Regimes of dominant diffusion behavior, normalized to the same concentration gradient, can be mapped as a function of l and ρ_d by equating the various $\dot{n}_i$ in Eq. 11-2. The equations of the boundary lines separating the operative transport mechanisms are thus

$$\frac{1}{l} = \frac{D_L}{\delta D_{GB}}, \qquad \rho_d = \frac{D_L}{A_d D_d}, \qquad \text{and} \qquad \frac{1}{l} = \frac{A_d D_d \rho_d}{\delta D_b}.$$

These are plotted as in $1/l$ vs ρ_d in Fig. 11-2 at four different levels of T/T_M employing Eqs. 11-3a,b,c. The dotted rectangles represent the range of thin

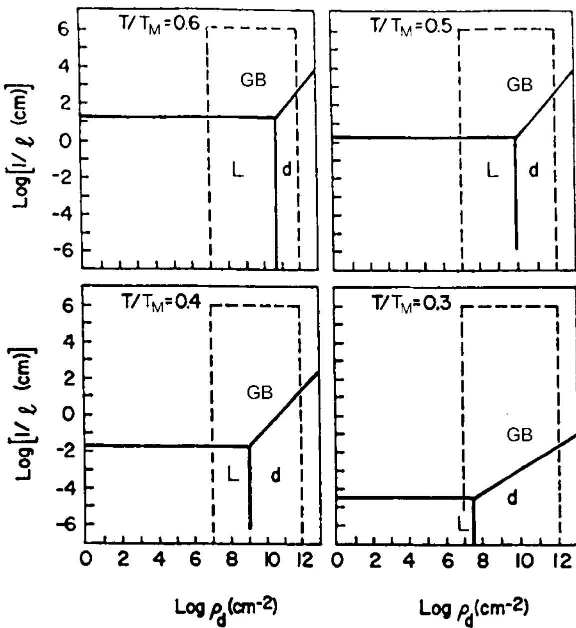


Figure 11-2 Regimes of dominant diffusion mechanism in FCC metal films as a function of temperature. (From Ref. 2.)

film values for l and ρ_d which occur in practice. For typical metal films with a grain size of 1 μm or less, grain-boundary diffusion dominates at all practical temperatures. Similarly, for dislocation-free epitaxial films where $1/l \sim 0$, lattice diffusion dominates. Transport at these extremes is intuitively obvious. Not so apparent is what happens when the film structure is such that admixtures of mechanisms are operative in different temperature regimes. Generally, lower temperatures will favor GB and dislocation short circuiting relative to lattice diffusion.

Surface diffusion is another transport mechanism of importance in thin films because of their large surface-to-bulk ratio. As noted in Chapters 7 and 9 this mechanism plays an important role in film nucleation and growth processes. Reduced parameters describing measured surface transport in FCC metals have been suggested (Ref. 3), e.g.,

$$D_S = 0.014 \exp - \frac{6.54 T_M}{T} \quad \frac{\text{cm}^2}{\text{s}} \qquad \text{for} \quad \frac{T_M}{T} > 1.3. \tag{11-4}$$

It is well known, however, that surface diffusion varies strongly with ambient conditions, surface crystallography, and the nature and composition of surface atoms.

Generalized Arrhenius plots for D_L as a function of T_M/T have already been introduced for metal, semiconductor, and alkali-halide matrices (Fig. 7-15). Therefore, systematics similar to those depicted in Fig. 11-2 also govern mass transport in these materials when grain boundaries and dislocations act as short-circuit paths. In semiconductors a great amount of impurity diffusion data exists, and these are used in designing and analyzing doping treatments for devices. This is a specialized field and complex modeling (Refs. 4, 5) is required to accurately describe diffusional doping profiles. Because of the importance of Si and GaAs films, preferred lattice-dopant diffusion data are presented in Fig. 11-3. Data on diffusion of noble metals Au, Ag, and Cu in amorphous Si films interestingly reveal that activation energies for diffusion in disordered matrices are very similar to values obtained for lattice diffusion in crystalline Si (Ref. 6).

11.2.3 Grain-Boundary Diffusion

Of all the mass-transport mechanisms in polycrystalline films, grain-boundary (GB) diffusion has probably received the greatest attention. Rapid diffusion within individual GBs, coupled with their great profusion due to small grain size, make them pathways through which the major amount of mass is transported. Low diffusional activation energies foster low-temperature transport creating serious reliability problems whose origins can

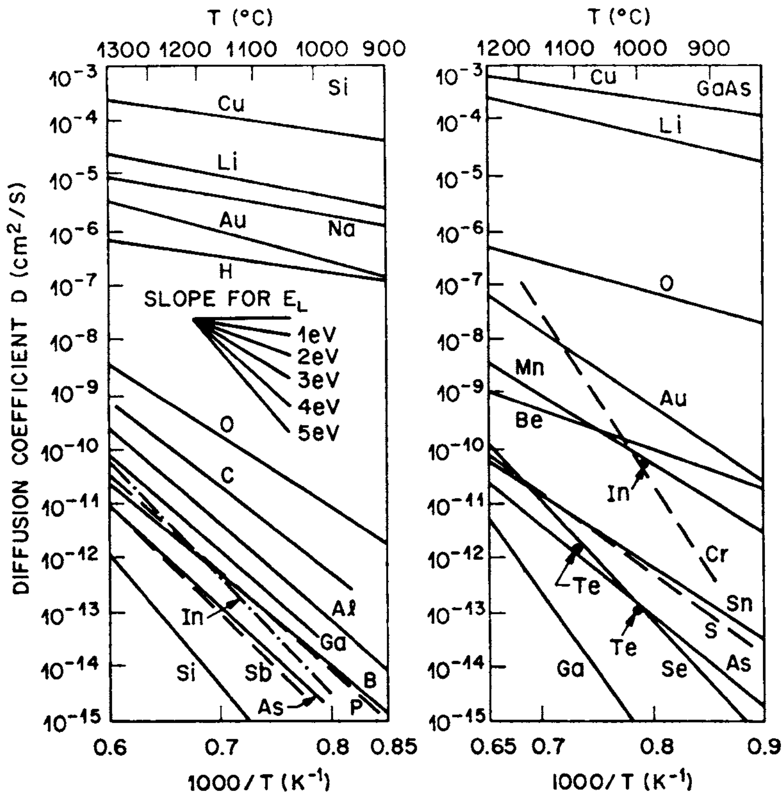


Figure 11-3 Diffusion coefficients of various elements in Si and GaAs as a function of temperature. (Reprinted with permission from John Wiley & Sons. From S. M. Sze, *Semiconductor Devices — Physics and Technology*, copyright © 1985, John Wiley & Sons.)

frequently be traced to GB involvement. This has motivated the modeling of both GB diffusion and phenomena related to film degradation processes.

11.2.3.1 Modeling Grain Boundary Diffusion Profiles

The first treatment of GB diffusion was due to Fisher (Ref. 7) and appeared a half century ago. This model considers transport within a semi-infinite bicrystal film that is initially free of diffusant as shown in Fig. 11-4. A species of concentration C_0 is permanently maintained at plane $y = 0$ and diffuses into both the GB as well as the two adjacent grains. At low temperatures in typical polycrystalline films, it is easily shown that there is

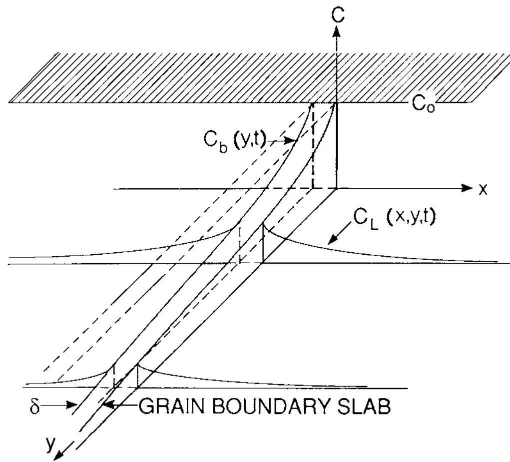


Figure 11-4 Representation of diffusional penetration down a grain boundary (y direction) with simultaneous lateral diffusion into adjoining grains (x direction).

far more transport down the GB than there is into the bulk lattice of the grains. The ratio of these two fluxes can be estimated through the use of Eqs. 11-2 and 11-3 for FCC metals, i.e.,

$$\frac{\dot{n}_{\mathrm{GB}}}{\dot{n}_{\mathrm{L}}} = \frac{dD_{\mathrm{GB}}}{lD_{\mathrm{L}}} = \frac{3 \times 10^{-8}}{l} \exp 8.1 T_{\mathrm{M}}/T. \tag{11-5}$$

Assuming $l = 10^{-4}$ cm and $T/T_{\mathrm{M}} = \frac{1}{3}$, $\dot{n}_{\mathrm{GB}}/\dot{n}_{\mathrm{L}} = 1.1 \times 10^{7}$. For this reason we may initially envision transport as consisting primarily of a deep, rapid penetration down the GB. Subsequently, diffusant spreads laterally into the adjoining grains building up the concentration level there. This is shown schematically in Fig. 11-4 and described mathematically by

$$C_{\mathrm{L}}(x, y, t) = C_0 \exp -\left(\frac{2\sqrt{D_{\mathrm{L}}}}{\delta D_{\mathrm{GB}}\sqrt{\pi t}}\right)^{1/2} y \cdot \operatorname{erfc}\left(\frac{x}{2\sqrt{D_{\mathrm{L}} t}}\right) \tag{11-6}$$

where $C_{\mathrm{L}}(x, y, t)$ is the diffusant concentration at any position and time. The Fisher analysis of the complex, coupled GB-lattice diffusion process yields simplified decoupled solutions—an exponential diffusant profile in the GB and an error function profile within adjoining grains. Experimental verification of Eq. 11-6 is accomplished by measurement of the integrated concentration $\bar{C}$ within incremental slices Δy thick (e.g., by sputter etching) normal

to the $y = 0$ surface, i.e.,

$$\bar{C} = \int_{-\infty}^{\infty} C_L(x, y, t)\, dx\Delta y = \text{const} \exp - \left(\frac{2\sqrt{D_L}}{\delta D_{GB}\sqrt{\pi t}}\right)^{1/2} y. \quad (11\text{-}7)$$

The last equation suggests that a plot of $\ln \bar{C}$ vs y is linear. Therefore, the useful result

$$\delta D_{GB} = 1.0/(\pi)^{1/2}(d \ln \bar{C}/dy)^{-2}(4D_L/t)^{1/2} \quad (11\text{-}8)$$

emerges. However, in order to obtain δD_{GB}, the value of D_L in the same system must be independently known. This usually poses no problem since lattice diffusivity data are relatively plentiful. Exact, but far more complicated, integral solutions, which are free of the simplifications of the Fisher analysis, have been obtained by Whipple (Ref. 8) and Suzuoka (Ref. 9). A conclusion based on these analyses which has been extensively used is

$$\delta D_{GB} = -0.66(d \ln \bar{C}/dy^{6/5})^{-5/3}(4D_L/t)^{1/2}. \quad (11\text{-}9)$$

Apart from overriding questions of correctness, the difference between Eqs. 11-8 and 11-9 is that $\ln \bar{C}$ is plotted vs y in the former and vs $y^{6/5}$ in the latter. Frequently, however, the experimental concentration profiles are not sufficiently precise to distinguish between these two spatial dependencies. It does not matter that actual films are not composed of bicrystals, but rather polycrystals with GBs of varying type and orientation; the physical character of the solutions is preserved despite the geometric complexity. A schematic representation of equiconcentration profiles in a polycrystalline film containing an array of parallel GBs is shown in Fig. 11-5. At elevated temperatures the extensive amount of lattice diffusion masks the penetration through GBs, whereas at the lowest temperatures virtually all of the diffusant is partitioned to GBs. In between, an admixture of diffusion mechanisms results in initial rapid migration down the short-circuit GB

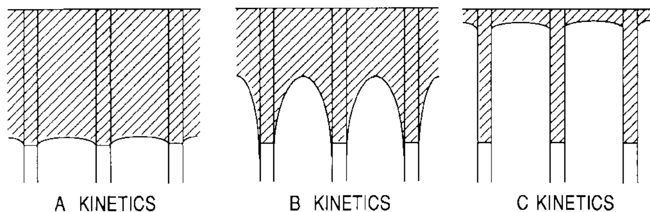


Figure 11-5 Schematic representation of type A (highest temperature), B, and C (lowest temperature) diffusion kinetics. (From Ref. 10.)

network which slows down as atoms leak into the lattice. The behaviors indicated in Fig. 11-5 represent the so called A, B, and C type kinetics (Ref. 10). Diffusion phenomena in polycrystalline films have been studied in the B to C range for the most part. Excellent reviews of the mathematical theories of GB diffusion, including discussions of transport in these different temperature regimes and applications to thin-film data, are available (Refs. 11, 12). A readable account of solid-state diffusion theory and practice in thin films is due to Tu, Mayer, and Feldman (Ref. 13). Their textbook also serves as an auxiliary reference for much of the material discussed in this chapter.

11.2.3.2 Estimating Grain-Boundary Diffusivities

What does one do when diffusion parameters in polycrystalline thin films must be estimated but experimental data are lacking? In such a case the master curves derived by Gupta (Ref. 14) and shown in Fig. 11-6 can provide considerable help. These yield the ratio of D_{GB}/D_L given the grain size, time, and homologous temperature. As an example let us reconsider the problem of interdiffusion in the Cu–Ni solar cell contact posed at the beginning of the chapter and try to estimate the ratio of D_{GB}/D_L. In the Cu–Ni system the average T_M is 1268°C (1541 K). Thus, at $T =$ 300°C (573 K), $T/T_M = 0.37$. Assuming the grain size is 1000 Å (10^{-5} cm), $\log 2(D_L t)^{1/2} = -5$. Following the recipe given in the caption for Fig. 11-6, we find that $\log(D_{GB}/D_L) \sim 15$. Therefore, $D_{GB} = 10^{15} D_L$ (300°C) or $10^{15} \times 3.8 \times 10^{-24} = 3.8 \times 10^{-9}$ cm^2/s. As a result, the diffusion length after a 1 h diffusion treatment is $2(D_{GB}t)^{1/2} = 2(3.8 \times 10^{-9} \times 3600)^{1/2}$ or 7.4×10^{-3} cm. Despite the approximate nature of the calculation, there is little wonder that interdiffusion occurred in less than 1 h.

11.2.3.3 Grain-Boundary Self-Diffusion in Gold Films

Experimental measurements (Refs. 15, 16) of the penetration of radioactive ^{195}Au into epitaxial (Fig. 11-7a) and polycrystalline (Fig. 11-7b) Au films provide a test of the foregoing theories. They also importantly illustrate how the spectrum of diffusion behavior can be decomposed into the individual component mechanisms through judicious choice of film temperature and grain size. These data were obtained by incrementally sputter-sectioning the film, collecting the removed material in each section, and then counting its activity level. Very low concentration levels can be detected because radiation-counting equipment is quite sensitive and highly selective. This makes it possible to measure shallow profiles and detect

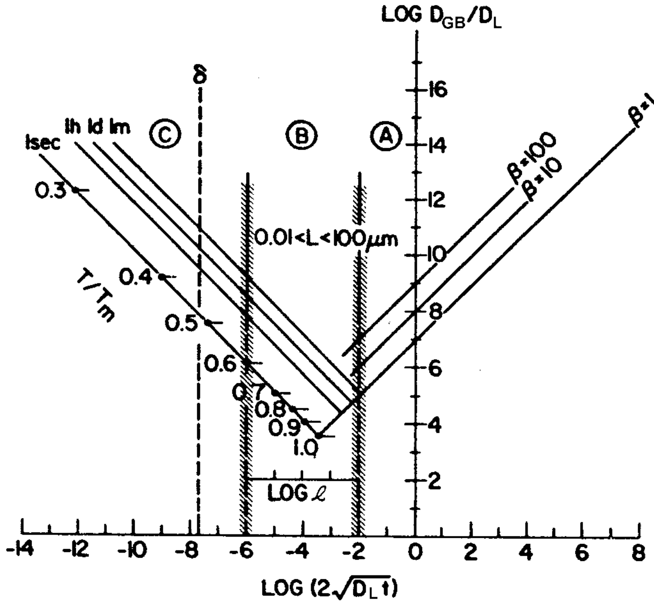
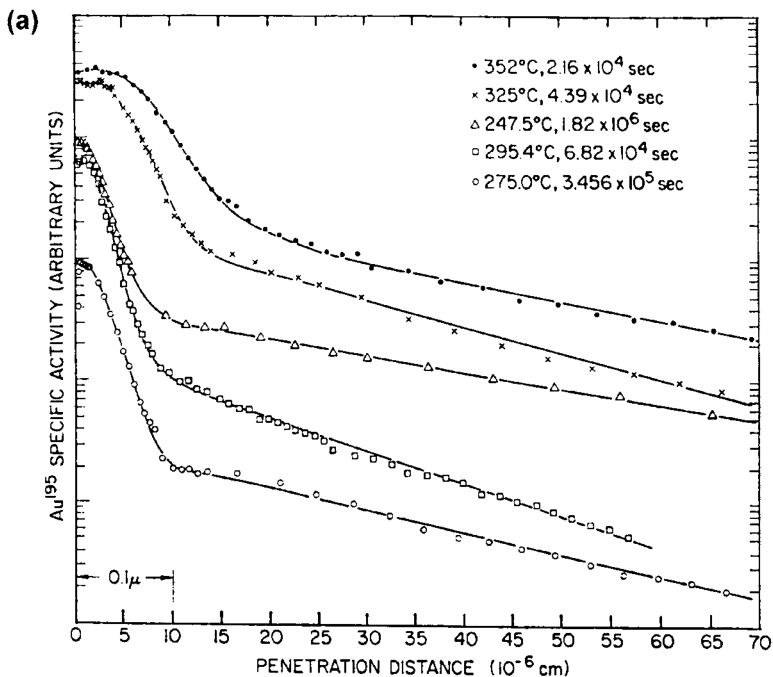


Figure 11-6 Master curves for estimating D_{GB}/D_L. Steps: (1) First project T/T_M on the appropriate time line (i.e., 1 s, 1 h, 1 day, 1 month) and draw a horizontal line. (2) Draw a vertical line upward from the horizontal grain-size axis. (3) At the point of intersection draw a line parallel to the slanted β axis on the right-hand side; where it intersects the vertical log D_{GB}/D_L axis, the value for D_{GB}/D_L may be read off. Also, the value of $2(D_L t)^{1/2}$ is read off the horizontal axis. Note that the quantity β, a dimensionless parameter, is defined as $[(D_{GB}/D_L)-1]\delta/2(D_L t)^{1/2}$ and increases as the type of kinetics goes from A to C. (From Ref. 14. Reprinted with permission.)

penetration at very low temperatures. The epitaxial film data display Gaussian-type lattice diffusion for the first 1000 to 1500 Å, followed by a transition to apparent dislocation short-circuit transport beyond this depth. Rather than high-angle boundaries, these films contained a density of some 10^{10} to 10^{11} dissociated dislocations per cm^2. On the other hand, extensive low-temperature GB penetration is evident in Fig. 11-7b without much lattice diffusion. The large differences in diffusional penetration between these two sets of data, which are consistent with the systematics illustrated in Fig. 11-2, should be noted. For epitaxial Au a mixture of lattice and dislocation diffusion is expected for $\rho_d \sim 10^{10}$/cm^2 at temperatures of $\sim 0.4T_M$. However, only GB diffusion is expected at temperatures of $\sim 0.3T_M$ for a grain size of 5×10^{-5} cm, and this is precisely what was observed.



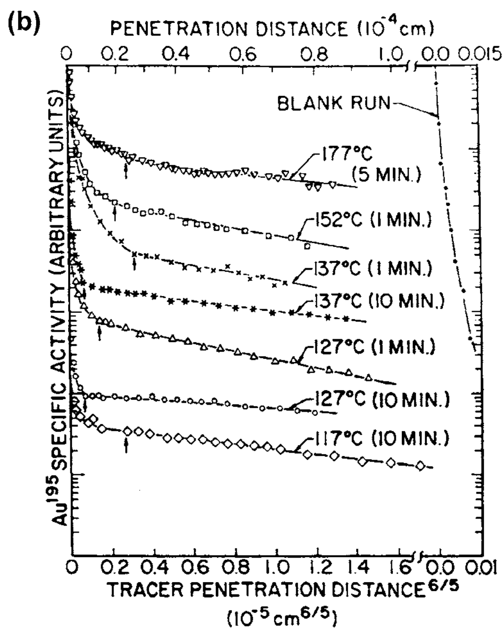


Figure 11-7 (a) Diffusional penetration profiles of ^{195}Au in (001) epitaxial Au films at indicated temperatures and times. Lattice and dislocation diffusion dominate. (From Ref. 15.) (b) Diffusional penetration profiles of ^{195}Au in polycrystalline films at indicated temperatures and times. Only GB diffusion is evident. (Reprinted with permission from Elsevier Sequoia S.A. From Ref. 16.)

11.2.4 Thin-Film Diffusion Couples

We now treat the important cases of interdiffusion between atoms of a film and substrate or between layered film diffusion couples. Consider a film of thickness d consisting of A atoms that lies above a much thicker (semiinfinite) substrate containing B atoms. The diffusing species is A and complete mutual solid solubility of A in B is assumed. If the top surface is $x = 0$, and no matter flows through it, $dC(0, t)/dx = 0$. Further, if we initially assume $C(x, 0) = C_0$ (or 1) for $d \geqslant x > 0$, and $C(x, 0) = 0$ for $x > d$, the solution is given by

$$C(x, t) = \frac{C_0}{2}\left\{\operatorname{erf}\left(\frac{x + d}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{x - d}{\sqrt{4Dt}}\right)\right\}. \qquad (11\text{-}10)$$

When d becomes sufficiently large relative to the size of the diffusion zone so that the A–B couple is effectively semiinfinite in length on either side, and $C = C_0$ for $x < 0$ and $C = 0$ for $x > 0$, it is not difficult to show that

$$C(x, t) = \frac{C_0}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{4Dt}}\right). \qquad (11\text{-}11)$$

Four features of these solutions, noted below, mask the actual details of mass transport in thin-film couples and lead to errors in quantitative analysis and interpretation of experimental concentration profiles. In ideal, homogeneous media the first two shortcomings can be overcome mathematically, whereas in real materials the last two raise important issues relating to the physical diffusion mechanisms.

1. No provision is made for the limited solubility of A in B or B in A. Limited solubilities lead to compound layer formation, a subject dealt with shortly.

2. It is usually incorrect to assume that thin films are of infinite extent with regard to diffusional effects. This questions the validity of the boundary conditions and solutions based on them. Treating the case of an A–B film couple of finite layer thickness is a relatively simple matter and is left as an exercise (see Exercise 11-7).

3. A single concentration-independent value of D is used in Eqs. 11-10 and 11-11. However, in an A–B couple we note that D_A and D_B can differ because there is no reason to expect A atoms to exchange with vacancies at the same rate that B atoms do. As a result of this and the fact that the diffusion coefficient is composition-dependent in alloys, the concentration profiles are altered in a complex way. The Boltzmann–Matano analysis

(Ref. 14) can be used to obtain a value for the interdiffusion or chemical coefficient $\tilde{D}$ for such a case. It turns out that $\tilde{D}$ at a particular concentration is composed of an admixture of diffusivities for components A and B, i.e.,

$$\tilde{D} = C_A D_B + C_B D_A \tag{11-12}$$

where C_A and C_B are the respective concentrations. Furthermore, an important consequence of the fact that $D_A \neq D_B$ is a vacancy accumulation or depletion near the couple interface, and the possibility of void formation. The Kirkendall effect which includes these phenomena will be discussed again in Section 11.5.4.2.

4. The A and B film matrices contain complex distributions of grain boundaries, dislocations, and point defects that establish rapid mass-transport pathways, each with a characteristic diffusivity. If both members of the diffusion couple were single crystals, the only avenue for atom transport would be through the bulk lattice. In contrast, atoms will preferentially migrate along the short-circuit GB paths if polycrystalline films are involved.

11.2.4.1 Compound Formation

Many of the interesting binary combinations employed in thin-film technology react to form compounds. They include all sorts of semiconductor–metal contacts, metal–metal interconnections, and solder–base metal couples at different electronic packaging levels. Since the usual configuration is a planar composite structure composed of elemental films on a flat substrate, layered compound growth occurs. The concentration–distance profile in such systems is schematically indicated in Fig. 11-8. Each of the terminal phases is assumed to be in equilibrium with the intermediate compound. The compound shown is stable over a narrow rather than broad concentration range. With time, the compound thickens as it consumes the α phase at one interface and the β phase at the other interface.

It is instructive to begin with the simplified analysis of the kinetics of compound growth based on Fig. 11-8. Only the γ phase (compound) interface in equilibrium with α is considered. Since A atoms lost from α are incorporated into γ, the shaded areas shown are equal. With respect to the interface moving with velocity v, the following mass fluxes of A must be considered:

$$\text{Flux into interface} = C_\alpha v - D_\alpha (dC_\alpha/dx)_{\text{int}}$$

$$\text{Flux away from interface} = C_\gamma v.$$

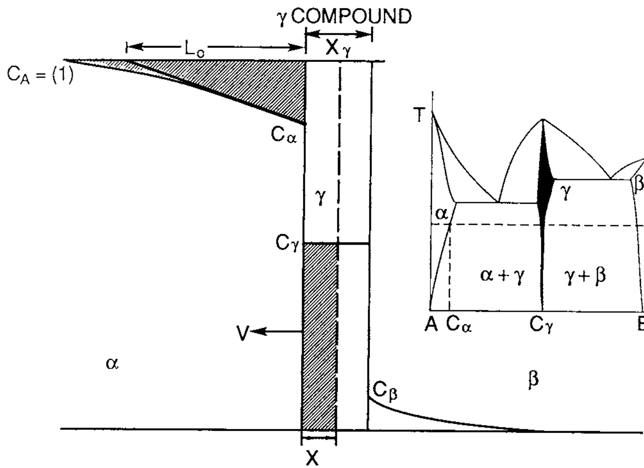


Figure 11-8 Depiction of intermediate compound formation in an A–B diffusion couple. Reaction temperature is dotted in on phase diagram.

These fluxes remain balanced for all times, so that by equating them we have

$$v = \frac{dX}{dt} = \frac{D_\alpha \dfrac{dC_\alpha}{dx}}{C_\alpha - C_\gamma}, \tag{11-13}$$

where X is the compound layer thickness. From the simple geometric construction shown, dC_α/dx can be approximated by $(C_A - C_\alpha)/L_0$. Therefore, as growth proceeds, L_0 increases while v decreases. If the shaded area within the α phase can be approximated by $\frac{1}{2}L_0(C_A - C_\alpha)$, and this is set equal to $C_\gamma X$, then $L_0 = 2C_\gamma X/(C_A - C_\alpha)$. Substituting for dC_α/dx in Eq. 11-13 leads to

$$\frac{dX}{dt} = \frac{D_\alpha(C_A - C_\alpha)^2}{2X(C_\alpha - C_\gamma)C_\gamma}. \tag{11-14}$$

Upon integrating, the α–γ portion of the compound layer thickness is obtained as

$$X = \frac{[D_\alpha(C_A - C_\alpha)^2]^{1/2} t^{1/2}}{[(C_\alpha - C_\gamma)C_\gamma]^{1/2}}. \tag{11-15}$$

A similar expression holds for the β–γ interface, and both solutions can be

added together to yield the final compound layer thickness X_γ, i.e.,

$$X_\gamma = \text{const}\, t^{1/2} \qquad (\text{const} \sim \exp - (E_p/RT)). \tag{11-16}$$

The important feature to note is that parabolic growth kinetics with activation energy E_p is predicted, and as we shall see, often observed. When diffusion is sufficiently rapid, however, growth may be limited by the speed of chemical reaction at the compound interface. Linear kinetics, with activation energy E_l varying simply as t, then ensue, i.e.,

$$X_\gamma = \text{const}\ t \qquad (\text{const} \sim \exp - (E_l/RT)), \tag{11-17}$$

but only for short times. At longer times, linear growth gives way to diffusion-controlled parabolic growth. In either case thermally activated growth is anticipated, but with an effective activation energy dependent on an admixture of diffusion parameters from both α and β phases. Combined linear–parabolic (interface reaction–diffusion) compound growth kinetics have been modeled in the literature (Ref. 17); a similar version for these dual mechanisms will be derived next for the case of silicon oxidation.

11.2.4.2 Growth of SiO_2 Films

As an example of dual-mechanism kinetics for compound formation consider an infinite diffusion couple consisting of oxygen gas on one side and a silicon wafer substrate on the other. During reaction between these elements a film of SiO_2 grows on the Si and thickens with time. In addition to the diffusional exchange between gas atoms and growing film, and the redistribution of atoms between film and substrate, there is the added complexity of transport across a moving oxide interface. Such effects occur during high-temperature oxidation of Si, one of the most studied film growth processes. The resulting amorphous SiO_2 film serves the critical function of the gate oxide dielectric in field-effect transistors. For this demanding application SiO_2 forms a remarkable chemical and electronic interface; it passivates the Si, allows very low current leakage, and leaves a minimum of dangling bonds to alter doping behavior in the underlying Si. In contrast to film deposition where all atoms in the deposit originate from the vapor phase (as in CVD of SiO_2), oxidation relies on a gas–solid reaction to sustain oxide-film growth. This means that substrate atoms participate in the process such that for every 1000 Å of SiO_2 growth, 440 Å (i.e., $1000\rho_{SiO_2}M_{Si}/\rho_{Si}M_{SiO_2}$, where ρ and M are respective densities and atomic weights) of Si substrate is consumed.

The now classic analysis of oxidation due to Grove (Ref. 18) has a simple elegance and yet accurately predicts the kinetics of thermal oxidation. In this treatment of the model, we assume a flow of gas containing oxygen parallel to the plane of the Si surface. For oxide to form at the Si/SiO_2 interface, the

following sequential steps are assumed to occur:

1. Oxygen is transported from the bulk of the gas phase to the gas–oxide interface.
2. Next oxygen diffuses through the growing solid oxide film.
3. When oxygen reaches the Si/SiO_2 interface, it chemically reacts with Si and forms oxide.

The respective mass fluxes corresponding to these steps can be expressed by

$$J_1 = h_G(C_G - C_0) \tag{11-18}$$

$$J_2 = \frac{D(C_0 - C_i)}{d_0} \tag{11-19}$$

$$J_3 = k_S C_i \tag{11-20}$$

where the concentrations of oxygen in the bulk of the gas, at the gas/SiO_2 interface, and at the SiO_2/Si interface are respectively C_G, C_0, and C_i. Quantities h_G, D, and k_S represent the gas mass-transport coefficient, the diffusion coefficient of oxygen in SiO_2, and the chemical reaction rate constant k_S, respectively. Constants D and k_S display the usual Boltzmann behavior but with different activation energies, while h_G has a weak temperature dependence.

By assuming steady-state growth implying $J_1 = J_2 = J_3$, it is a simple matter to solve for C_i and C_0 in terms of C_G or

$$C_0 = \frac{C_G(1 + k_S d_0/D)}{(1 + k_S/h_G + k_S d_0/D)} \tag{11-21a}$$

$$C_i = \frac{C_G}{(1 + k_S/h_G + k_S d_0/D)}. \tag{11-21b}$$

Clearly, the grown SiO_2 has a well-fixed stoichiometry so that C_0 and C_i differ only slightly in magnitude, but sufficiently to establish the concentration gradient required for diffusion. In fact, $C_0 = C_i = C_G/(1 + k_S/h_G)$ in the so-called reaction limited case where $D \gg k_S d_0$. Here diffusion is assumed to be very rapid through the SiO_2, but the bottleneck for growth is the interfacial chemical reaction. On the other hand, under diffusion control, D is small so that $C_0 = C_G$ and $C_i \simeq 0$. In this case the chemical reaction is sufficiently rapid but the supply of oxygen limits the rate of oxide formation.

The actual oxide growth rate is related to the flux, e.g., J_3, and therefore the thickness of oxide at any time (d_0) is expressed by

$$\frac{d(d_0)}{dt} = \frac{k_S C_i}{N_0}, \tag{11-22}$$

where N_0 is the number of oxidant molecules incorporated into a unit volume of film. For oxidation in dry O_2 gas, $N_0 = 2.2 \times 10^{22}\,\text{cm}^{-3}$ whereas for steam (wet) oxidation, $N_0 = 4.4 \times 10^{22}\,\text{cm}^{-3}$ because half as much oxygen is contained per molecule. Substitution of Eq. 11-21b into 11-22 followed by direct integration of the resulting differential equation yields

$$d_0^2 + Ad_0 = B(t + \tau), \tag{11-23}$$

where $A = 2D(1/h_G + 1/k_S)$, $B = 2DC_G/N_0$, and $\tau = (d_i^2 + Ad_i)/B$. The constant of integration τ arises only if there is an initial oxide film of thickness d_i present prior to oxidation and therefore Eq. 11-23 is useful in describing sequential oxidations. A solution to this quadratic equation is

$$\frac{d_0}{A/2} = \sqrt{1 + \frac{t + \tau}{A^2/4B}} - 1 \tag{11-24}$$

from which the limiting long- as well as short-time growth-kinetics relationships are easily shown to be

$$d_0^2 \simeq Bt \qquad \text{for} \qquad t \gg A^2/4B \tag{11-25}$$

$$d_0 \simeq B(t + \tau)/A \qquad \text{for} \qquad t + \tau \ll A^2/4B. \tag{11-26}$$

As we shall see, parabolic and linear growth in SiO_2 parallels that for intermetallic compound films such as metal silicides. All film growth in such cases is probably linear to begin with because parabolic growth implies an infinite initial thickening rate. A single dimensionless thickness–time plot shown in Fig. 11-9 very neatly summarizes Si oxidation behavior. The limiting linear and parabolic growth kinetics regimes are clearly identified. Not all oxidation processes, however, display linear or parabolic growth rates; cubic, logarithmic, and inverse logarithmic oxidation kinetics have been observed (Ref. 19).

Gate oxide technology has greatly changed in recent years in response to continual device miniaturization. Increasing MOS transistor speeds resulting from decreasing gate widths mean that gate-dielectric films must be correspondingly scaled to thinner dimensions. Gate oxides, the thinnest fabricated films in integrated circuit technology, are already less than 40 Å. At these thicknesses, critical properties gate dielectrics must possess can no longer be met by pure grown SiO_2. For example, SiO_2 allows boron originating in the polysilicon gate electrode to penetrate it; furthermore, leakage currents are excessive, causing unreliable transistor operation (Ref. 20). To circumvent these problems, ultrathin oxide films are now thermally nitrided employing NO, N_2O, or NH_3 sources which typically introduce about 10 at.% nitrogen into the SiO_xN_y structure. The resultant silicon oxynitrides serve as an effective diffusion barrier to boron and also limit current leakage.

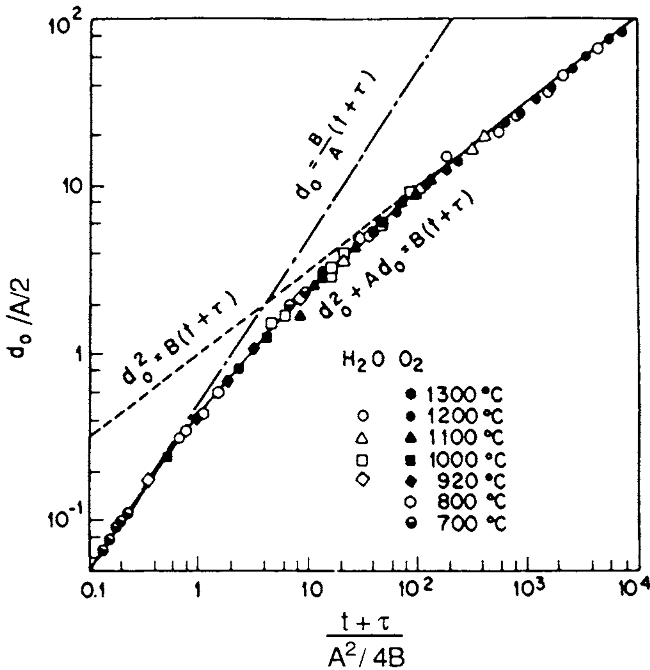


Figure 11-9 Oxidation kinetics behavior of Si in terms of dimensionless oxide thickness and time. The two limiting forms of the kinetics are shown. (From A. S. Grove, *Physics and Technology of Semiconductor Devices*, copyright © 1967, John Wiley & Sons. Reprinted with permission.)

Since so few atoms are involved in the growth of these ultrathin oxides, the macroscopic mass transport equations developed earlier have limited applicability. Such an oxide film only 1.2 nm thick and shown in Fig. 11-10 is essentially the gate dielectric utilized in the experimental nano-field-effect transistor of Fig. 10-16.

11.3 INTERDIFFUSION IN THIN METAL FILMS

11.3.1 Scope

The mass transport effects considered in this and the next section have a number of common features: (1) they primarily deal with binary metal systems, (2) thin-film couples are usually involved, (3) diffusion and reaction

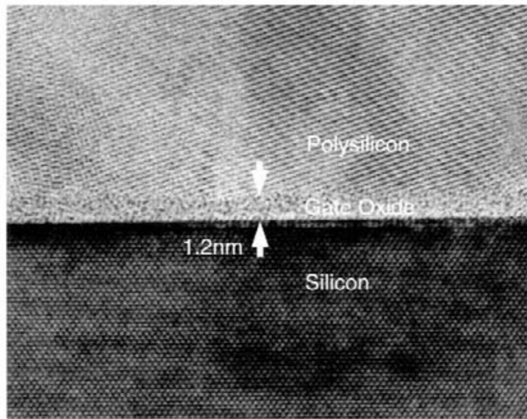


Figure 11-10 Transmission electron micrograph cross section of a 1.2 nm gate-oxide film sandwiched between silicon and polysilicon. (Courtesy of T. W. Sorsch, Agere Systems.)

occurs through the film thickness rather than along the film surface, (4) intermetallic compounds are often the reaction products, and (5) issues within the microelectronics industry have been the chief drivers for interest in the subject matter treated.

With respect to the last item, a good way to appreciate the importance of interdiffusion effects in metal films is to consider the interfacial region between a fabricated silicon chip and the solder joint which connects it to the outside world. This is shown in Fig. 11-11, and following Tu (Ref. 21), we note that the aluminum metallization levels that contact the Si devices above must also be bonded to the solder ball below at the contact pad. Anyone who has tried to solder Al is acquainted with the difficulties involved. In this case, they are overcome by using an evaporated Cr–Cu–Au thin-film structure. Since the surface of Al is easily oxidized, it is difficult to solder with the Pb–Sn alloy and so a Cu layer is introduced. The intention is to utilize the fast Cu–Sn reaction to form intermetallic compounds. However, Cu adheres poorly to oxidized Al and SiO_2 surfaces. Moreover, when molten, the relatively massive Pb–Sn consumes the Cu and then tends to dewet on the Al surface. Therefore, Cr is introduced as a glue layer between the Al and SiO_2 and to prevent the molten solder from dewetting. The Cu surface needs to be protected against atmospheric corrosion because corroded Cu surfaces do not solder well. That is why a thin film of Au is introduced to passivate the Cu surface. Since Au dissolves in both Pb

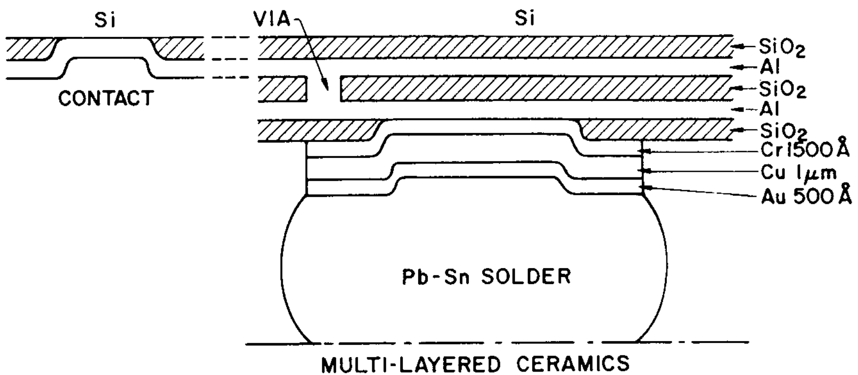


Figure 11-11 Schematic diagram of solder contact to Al by means of the trimetal Cr–Cu–Au film metallization. *Note:* The chip is flipped 180° (flip chip) in this interconnection scheme. (Reprinted with permission. From Ref. 21.)

and Sn and forms intermetallic compounds with Sn, a solder richer in Pb than the eutectic composition is used. This allows for enough Pb to dissolve Au and sufficient Sn to react with Cu. The reaction between the excess Pb and Cu is limited because these elements do not form extended solid solutions or intermetallic compounds. Thus, in order to fulfill the functions of adhesion, soldering, and passivation, this elaborate trimetal thin-film structure is required.

There are additional solid-state diffusion effects between metal layers to contend with. At temperatures close to 200°C, Cu can diffuse rapidly through GBs of Cr even though these metals are basically immiscible. When this happens, the interfacial adhesion at the Cr–SiO_2 interface is adversely affected. Moreover, Cu can diffuse outward through the Au in which it is miscible. At the Au surface it forms an oxide which interferes with soldering. This one method of joining has spawned a host of mass-transport phenomena. For this reason there have been many thin-film interdiffusion studies of the involved metals, e.g., Cr–Cu, Au–Cu, Cu–Sn. Other binary combinations, such as Al–Au and Al–Cu, have received even more attention. Miscible, immiscible, and compound-forming systems are all represented. For the most part vacancy diffusion is the accepted transport mechanism. However, the metals Cu and Au are believed to undergo anomalously rapidly migration through interstitial sites in the group IV matrices of Sn and Pb.

The remainder of this section treats miscible metals; Section 11.4 deals with compound formation issues.

11.3.2 Interdiffusion in Miscible Alloy Systems

11.3.2.1 A Macroscopic View

It is helpful to initiate the discussion of diffusion in miscible systems by excluding the complicating effects of grain boundaries. Bulk materials contain large enough grains that the influence of GBs is frequently minimal. In thin films a couple where both layers are single crystals, e.g., a hetero-epitaxial system, must be imagined. Under such conditions, the well-established macroscopic diffusion analyses hold. Upon interdiffusion in miscible systems, there is no crystallographic change, for this would imply new phases. Rather, each composition will be accessed at some point or depth within the film as a continuous range of solid solutions is formed. When the intrinsic atomic diffusivities are equal, i.e., $D_A = D_B$, the profile is symmetric and Eq. 11-11 governs the resultant diffusion. More often, however, Eq. 11-12 applies and a Boltzmann–Matano analysis must be performed to extract $\tilde{D}(C)$.

As an example of a miscible system consider the much studied Au–Pd polycrystalline thin-film couple shown in Fig. 11-12 (Ref. 22). Both AES sputter sectioning and RBS methods were employed to obtain the indicated profiles whose apparent symmetry probably reflects the lack of a strong diffusivity dependence on concentration. It is very tempting to analyze these data by fitting them to Eq. 11-11. Although effective diffusivity values can be

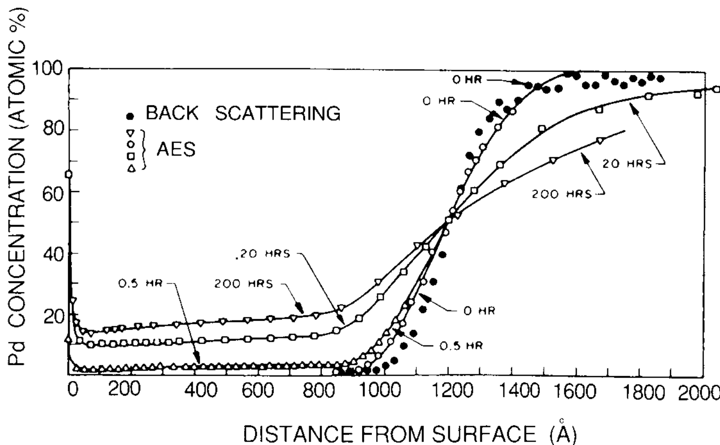


Figure 11-12 Palladium concentration profiles in a Au–Pd thin-film diffusion couple determined by RBS and AES techniques. (Reprinted with permission from Elsevier Sequoia, S.A. From Ref. 22.)

extracted, they would tend to have limited value because of the heterogeneous character of the film matrix. Since grain boundary diffusion is the dominant mechanism in this couple, an appropriate GB analysis revealed that a defect-enhanced admixture of GB and lattice diffusion was operative. Diffusional activation energies obtained for this film system are typically 0.4 times that for bulk diffusion, in accord with the systematics for GB diffusion.

11.3.2.2 Grain-Boundary Diffusion in Alloy Films

There have been a considerable number of fundamental interdiffusion studies in metal alloy films which have been interpreted in terms of GB transport models. In order to exclude the complicating effects of compounds and precipitates, the systems considered here are primarily limited to those displaying solid solubility. Results from a representative group of investigations are entered in Table 11-1 where the diffusivities are expressed

Table 11-1
Grain-Boundary Diffusion in Thin Films[a]

Diffusant	Matrix	δD_0 (cm^3/s)	E_{GB} (eV)	Remarks
Al	Cu	5.1×10^{-7}	0.94	Polycrystalline films
^{195}Au	Au	9.0×10^{-10}	1.00	Polycrystalline films
^{195}Au	Au	1.9×10^{-10}	1.16	(100) epitaxial
Cu	Al	4.5×10^{-8} 1.8×10^{-10}	1.00 0.87	Polycrystalline films
Ag	Au	4.5×10^{-12}	1.20	(111) epitaxial
^{195}Au	Ni–0.5% Co	1.4×10^{-10}	1.60	Polycrystalline films
Sn	Pb	6.9×10^{-10}	0.62	Polycrystalline films
Cr	Au	5.0×10^{-11}	1.09	Textured film
Ag	Cu	1.5×10^{-13}	0.75	Polycrystalline films
Pt	Cr	5.0×10^{-10}	1.69	Polycrystalline films
Cu	Al–0.2% Co	2.0×10^{-7}	0.56	Polycrystalline films
Ag	Au	5.0×10^{-14}	1.10	(100) epitaxial
^{119}Sn	Sn	5.0×10^{-8}	0.42	Polycrystalline films
Co	Au–0.19% Co	6.5×10^{-10}	1.2	Electrodeposited films
^{195}Au	Au–1.2% Ta	5.0×10^{-7}	1.2	Low-angle GB
^{63}Ni	Ni	2.2×10^{-8}	1.77	Low-angle GB

[a] From Ref. 14.

in terms of the preexponential factor (δD_0) and activation energy (Refs. 14, 23).

Two broad categories of experimental techniques were employed in gathering these data. Sputter sectioning through the film is the basis of the first technique. In the commonly employed AES depth-profiling method, the film is analyzed continuously as it is simultaneously thinned by sputter etching. Profiles appear similar to those shown in Fig. 11-7a, b.

The second category is based on permeation and surface accumulation techniques. An example shown in Fig. 11-13 utilizes AES signal-sensing methods to detect Ag penetration through a Au film and subsequent spreading along the exit surface (Ref. 24). The signal reflects this by building slowly at first after an incubation period, then changing more rapidly and finally saturating at long times. Diffusivities are then unfolded by fitting data to the assumed kinetics shown in Fig. 11-13 for surface concentration (C_S) buildup with time. When type C kinetics (Fig. 11-5) prevails, D_{GB} values may be simply estimated by equating the film thickness to the GB diffusion length $\sim 2(D_{GB}t)^{1/2}$, where t is the time required for the AES signal to appear. Interestingly, a grain-boundary Kirkendall effect displayed in Fig. 11-13 has been proposed to account for the unequal GB diffusion rates of Ag and Au.

The high sensitivity inherent in detecting a signal rise from the background enables very small atomic fluxes to be measured, making it possible to monitor transport at quite low temperatures. On the other hand, disadvantages include the electron beam heating of films during measurement, and the need to maintain ultrahigh-vacuum conditions because of the sensitivity of surface diffusion to ambient contamination.

Results for the sectioning methods suggest two GB-diffusion regimes:

1. $D_{GB} = 0.3 \exp - (7.5T_M/T)\ \text{cm}^2/\text{s}$ for large angle boundaries where $E_{GB} = 0.5E_L$
2. $D_{GB} = 2 \exp - (12.5T_M/T)\ \text{cm}^2/\text{s}$ for epitaxial films containing subboundaries or dissociated dislocations where $E_{GB} = 0.7E_L$

For the permeation and surface accumulation methods, the activation energies appear to be somewhat lower. Data derived from these techniques are more susceptible to GB structure since diffusant is transported through the most highly disordered boundaries first.

Studies on magnetic films used as sensors in recording heads have also confirmed the importance of GB diffusion. Thin tantalum films are often used as diffusion barriers to prevent atom transport into Permalloy ($Ni_{80}Fe_{20}$) recording heads. Indiffusion of Ta into $Ni_{80}Fe_{20}$ can thermally

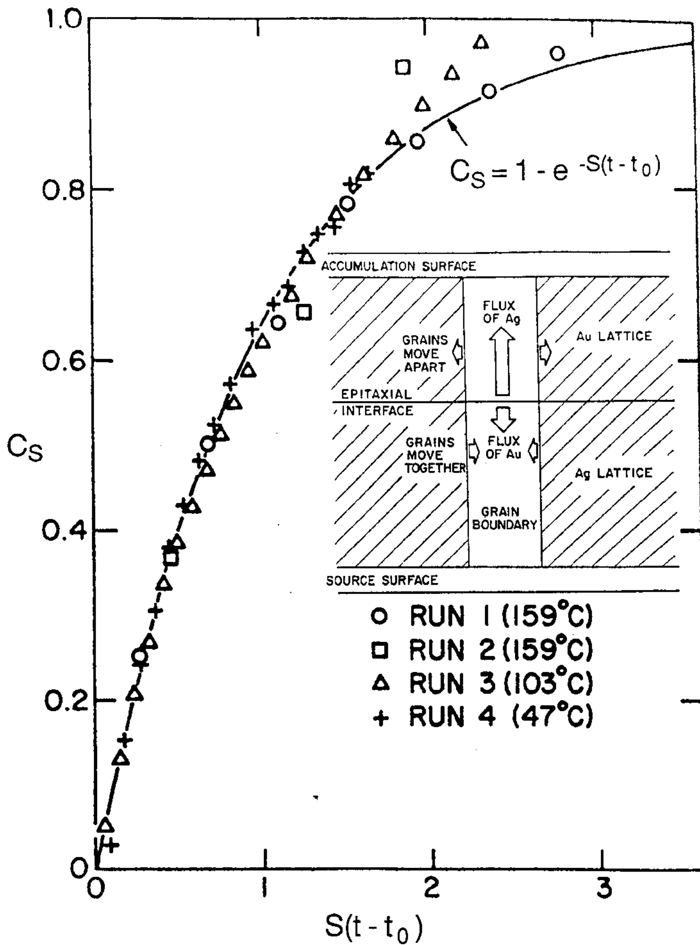


Figure 11-13 Accumulation (C_S) of Ag on Au film surfaces as a function of normalized (reduced) time for different diffusional annealing temperatures. Constants S and t_0 (incubation time) depend on the run involved. Inset shows a model for a grain-boundary Kirkendall effect. (From Ref. 24.)

roughen the interface and compromise the recording fidelity. To study the effect, thin-film bilayers of Ta–$Ni_{80}Fe_{20}$ as well as Ta–Ni and Ta–Fe were annealed in the range of 300 to 600°C and analyzed by RBS methods. Among the observations were fast GB diffusion of Ta through NiFe, Ni, and Fe, amorphous or microcrystalline alloy phase formation, compound growth, and Fe segregation in Ta–NiFe at higher temperatures (Ref. 25).

A related study of the interdiffusion of silver in Permalloy films by the permeation/surface accumulation technique likewise demonstrated the importance of Ag transport through NiFe GBs (Ref. 26). This work is interesting because giant magnetoresistance effects are observed in the Ag/NiFe system. Changes in the electrical resistance of the head material due to magnetic fields is the basis of sensing stored information, and maximizing the effect by minimizing interdiffusion is a critical concern.

11.3.2.3 A Strategy to Reduce Grain Boundary Diffusion

Reliability implications of undesirable low temperature interdiffusion effects in films (e.g., alloying, compound growth, and micro- as well as macrodefect formation) often stem from *solvent* diffusion through grain boundaries. Thus preventive strategies invariably attempt to minimize this mass-transport mechanism. Gupta (Ref. 14) has drawn together several phenomena related to GB behavior, and crafted a model to predict conditions for either enhanced or reduced solvent diffusion in metal films. Key equations will be presented in turn without derivation.

11.3.2.3.1 Relation of Grain Boundary Energy to Grain Boundary Diffusion

This interesting connection relates vacancy diffusivities in grain boundaries (D_{GB}) and the lattice (D_L) to the GB interfacial energy (γ_{GB}) and takes the form

$$\gamma_{GB} = \rho RT \ln(D_{GB}/D_L), \tag{11-27a}$$

where ρ is a conversion factor (Ref. 27).

11.3.2.3.2 Gibbs Adsorption

Atomic enrichment at surfaces or interfaces is related to the well-known thermodynamic Gibbs adsorption effect (Ref. 28). Basically, when atoms adsorb or accumulate at the surface of a phase, the surface energy decreases. Mathematically,

$$\Gamma = -1/RT(d\gamma_{GB}/d \ln X), \tag{11-27b}$$

where X is the solute mole fraction and Γ is the density of solute adsorbed. After integrating Eq. 11-27b and eliminating γ_{GB} (Eq. 11-27a) we obtain, with subscripts a and p referring to alloy and pure solvents, respectively,

$$\rho \ln\left[\frac{(D_{GB})_p(D_L)_a}{(D_{GB})_a(D_L)_p}\right] = \Gamma \ln(X_{GB}/X_L). \tag{11-27c}$$

11.3.2.3.3 Solute Segregation at Grain Boundaries

Using a statistical mechanical approach for alloys McLean (Ref. 29) derived an equation to describe the process of solute segregation at grain boundaries given by

$$X_{GB}/X_L = \exp(E_b/RT)/[1 + X_L \exp(E_b/RT)]. \qquad (11\text{-}27d)$$

X_{GB} and X_L are the solute mole fractions in GBs and in the lattice, respectively and E_b is the solute–GB binding energy. Furthermore, it has been observed (Refs. 28, 30) that the GB solute enrichment factor (X_{GB}/X_L) correlates inversely with atomic solid solubility. Thus in systems with limited solubility whatever solute there is in solution, it is magnified at grain boundaries.

Finally, through combination and simplification of the preceding equations, Gupta has suggested the workable form

$$\left[\frac{(D_{GB})_p(D_L)_a}{(D_{GB})_a(D_L)_p}\right] = 1 + X_0 \exp(E_b/RT) \qquad (11\text{-}27e)$$

where X_0, the solute mole fraction in the lattice, is generally small.

At elevated temperatures the second term on the right may be neglected. Therefore, for alloys where $(D_L)_a > (D_L)_p$, $(D_{GB})_p < (D_{GB})_a$, and an undesirable enhanced GB diffusivity occurs. At lower temperatures, even for $(D_L)_a > (D_L)_p$, it may happen that $(D_{GB})_p > (D_{GB})_a$, resulting in a desirable decreased GB solvent diffusivity due to alloying addition.

The magnitude of E_b is critical to predicting these diffusion effects and it appears to scale with relative atomic size difference between the solute and solvent atoms. This implies that elastic strain energy is the primary contributor to E_b. Strong correlations have been found between GB segregation of transition metal solutes and suppression of GB electromigration in aluminum base films. In addition, the model has proven useful in explaining diffusion effects in Pb–Sn solders, Au–1.2 at.% Ta films, and assorted dilute Cu alloys.

11.3.3 Effects of Ambients on Thin-Film Diffusion and Reaction

Annealing of thin films is a widely employed processing step in many technologies and therefore the effects of the ambient gas on film interactions

are of interest. Furthermore, by extrapolation to lower temperature, stability of film structures during use and storage are also dependent on the ambient atmosphere. Two basic effects, namely enhancement or reduction of diffusion effects between atoms in surface films and in underlying layers, are observed (Ref. 31). For example, in cases where gold ($\sim$200 nm thick) is the coating film, outdiffusion of the underlying species is typically enhanced in steam, air, and oxygen. This has been observed for Si, Ge, GaAs, Cr, Cu, Ni, Fe, and Ti and results in the formation of oxide layers above the gold, which correspondingly sinks toward the substrate. In all of these cases the overlying metal is polycrystalline and so we may expect diffusion through grain boundaries. The use of vacuum or hydrogen ambients reduces these effects, however, e.g., Au/Co and Au/Al. Reduced diffusional interactions have been observed in oxygen for palladium films on top of Si, Ti, and Al. In these cases interfacial rather than surface oxide layers form.

Reasons for transport enhancement (i.e., surface oxide formation) or retardation (i.e., interface oxide formation) in air are not known with a certainty that enables prediction for a given metal couple. Diffusion in a direction that reduces the overall film surface energy is one suggested reason. A more workable explanation offered by Chang (Ref. 31) considers the relative electronegativity (χ) or tendency to accept electrons, between the overcoat (χ_0) and underlying layer (χ_u), and the ambient-induced work function (ϕ) changes in the overcoat. Thus, there are four possibilities:

1. If $\chi_0 > \chi_u$, enhanced diffusion occurs when ϕ is decreased by the ambient.
2. If $\chi_0 > \chi_u$, reduced diffusion occurs when ϕ is increased by the ambient.
3. If $\chi_0 < \chi_u$, enhanced diffusion occurs when ϕ is increased by the ambient.
4. If $\chi_0 < \chi_u$, reduced diffusion occurs when ϕ is decreased by the ambient.

In case 1, for example, which applies to Au on Si, $\chi_{Au} = 2.4$ and $\chi_{Si} = 1.8$. A partial ionic bond between Au and Si is thus favored and results in formation of Au^-–Si^+. The bond is further stabilized by a negative potential on Au which occurs when its work function is decreased by the ambient, e.g., oxygen. Apparently once Si bonds to Au it is then free to outdiffuse and eventually react with oxygen. Hydrogen, on the other hand, raises ϕ_{Au}, inhibiting outdiffusion of Si, as observed. Similar arguments appear to apply in other cases.

11.4 COMPOUND FORMATION AND PHASE TRANSFORMATIONS IN THIN FILMS

11.4.1 INTERMETALLIC COMPOUND FORMATION

11.4.1.1 The Al–Au System

One of the best known examples of intermetallic compound formation between metal films occurs in the aluminum–gold system. For some 40 years the combination of Al films and Au wires or balls served to satisfy bonding requirements in the semiconductor industry, but not without a significant number of reliability problems. When devices are heated to 250–300°C for a few days, an Al–Au reaction proceeds to form a porous intermetallic phase around the Au ball bond accompanied by a lacy network of missing Al. One of the alloy phases, Al_2Au, is purple, which accounts for the appelation "purple plague." Even though the evidence is not definitive that the presence of Al_2Au correlates with bond embrittlement, lack of strength, or degradation, formation of this compound is viewed with alarm.

An extensive RBS study (Ref. 32) of compound formation in the Al–Au thin-film system is worth reviewing since it clarifies the role of film thickness, temperature, and time in influencing the reaction. Film couples with Al on top of Au were annealed, yielding the RBS spectra shown in Fig. 11-14. Because there is such a large mass difference between Al and Au, there is no peak overlap despite the unfavorable spatial ordering of the metal layers. The stoichiometry of the peak shoulders in the annealed films corresponds to the compounds Au_2Al and $AuAl_2$, and the high-energy tail is indicative of Au transport through the Al film to the surface. From the fact that the $AuAl_2$ always appears at a higher energy than the Au_2Al, we know that it is closer to the air surface of the couple. By the methods outlined in Section 10.4.7.5, the compound thickness can be determined and Fig. 11-15a reveals that both compounds grow with parabolic kinetics. The slopes of the compound thickness vs time$^{1/2}$ curves are proportional to the ubiquitous Boltzmann factor. Therefore, by plotting these slopes (actually the logs of the square of the slope in this case) vs $1/T$ K in the usual Arrhenius manner (Fig. 11-15b), activation energies for compound growth were obtained. The values of 1.03 and 1.2 eV can be roughly compared with the systematics given for FCC metals to elicit some clue as to the mass-transport mechanism for compound formation. Based on Au, these energies translate into equivalent Boltzmann factors of $\exp-(9.0T_M/T)$ and $\exp-(10.4T_M/T)$, respectively, suggesting a GB-assisted diffusion mechanism (see Eq. 11-3b). Lastly, it is interesting to note how

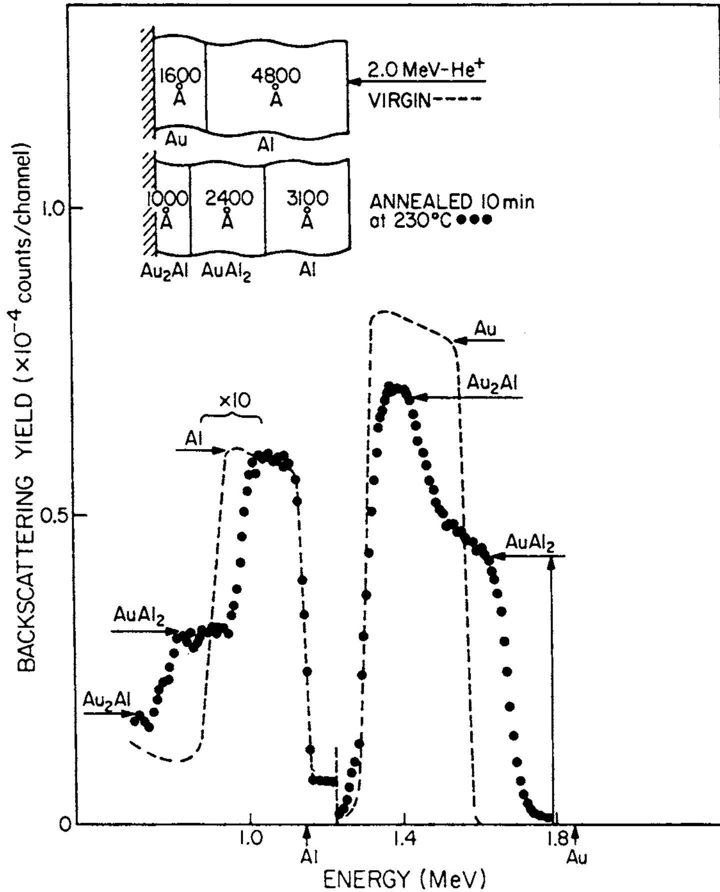


Figure 11-14 RBS spectrum showing formation of $AuAl_2$ and Au_2Al phases at 230°C. (From Ref. 32.)

the sequence of compound formation (Fig. 11-16) correlates with the equilibrium phase-diagram which is shown in Fig. 11-18. When the film thickness of Al exceeds that of Au, then the latter will be totally consumed leaving excess Al. The observed equilibrium between Al and $AuAl_2$ layers is consistent with the phase diagram. Similarly, excess Au is predicted to finally equilibrate with the Au_4Al phase as observed. Another complementary way to view the sequence of compound formation is treated in Section 11.4.1.4.

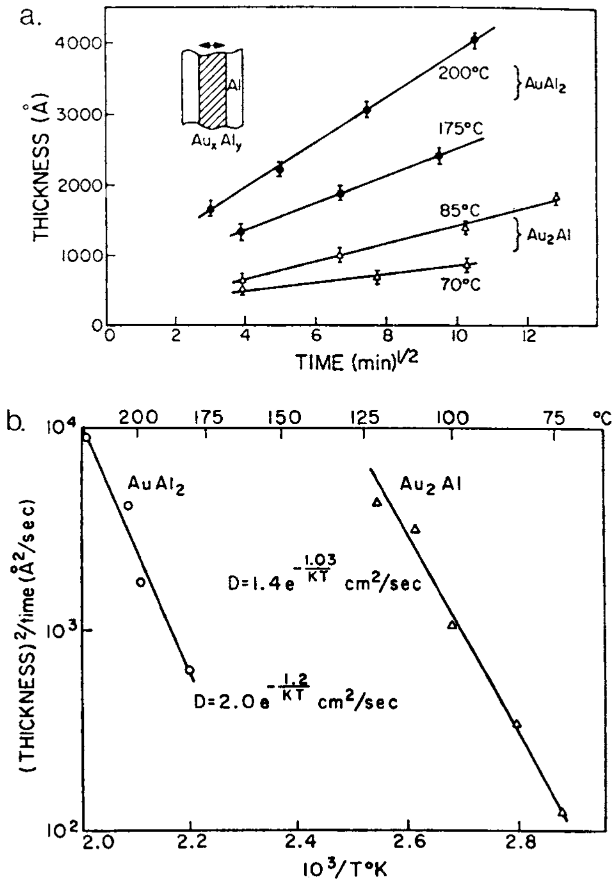


Figure 11-15 (a) Compound film thickness vs square root of time for $AuAl_2$ and Au_2Al at different annealing temperatures. (From Ref. 32.) (b) Arrhenius plots for the kinetics of formation of $AuAl_2$ and Au_2Al compounds. (From Ref. 32.)

11.4.1.2 Room-Temperature Compound Formation in Films

There is a worrisome prospect of compound formation in thin metal–film diffusion couples when widely used metals such as gold, silver, copper, and palladium are involved. Each of these metals was observed (Ref. 33) to react at room temperature with many relatively low-melting metals (Me) that include Al, Bi, Cd, Ga, In, Pb, Sb, Sn, Te, and Zn. X-ray analysis revealed the presence of assorted compounds, e.g., $AuGa_2$, $AuIn_2$, AuIn, Au_4In, $AuSn_2$, $AuCd_3$, AuCd, $AuPb_3$, $AuPb_2$, Au_2Pb, $AuZn_3$, AuZn, Au_3Zn,

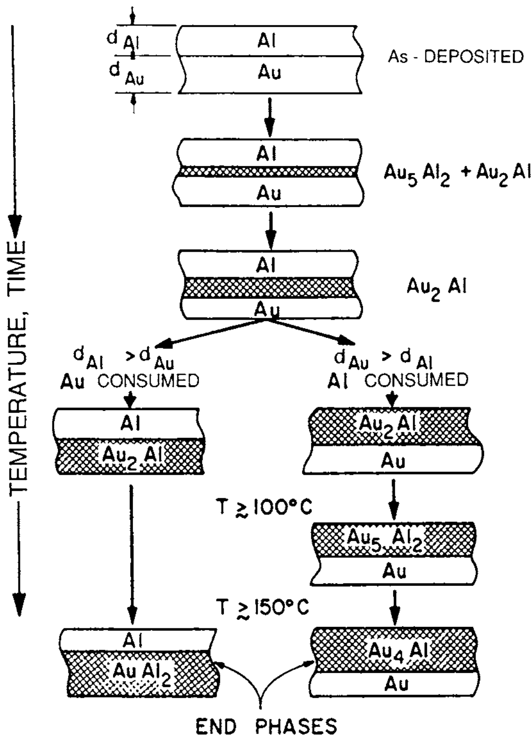


Figure 11-16 Schematic diagrams illustrating compound formation sequence in Al–Au thin film couples. End phases depend on whether $d_{Al} > d_{Au}$ or $d_{Au} > d_{Al}$.

$AuSb_2$, Au_2Al, and $AuAl_2$ in the case of Au–Me couples. Values for the diffusion coefficients were obtained using $D = X^2/4t$ (Eq. 1-29), with X the measured compound thickness. The results are summarized as follows, where T_{Me} is the melting temperature of metal Me:

1. For Au–Me, $D = 2.36 \times 10^{-11} \exp - (0.016T_{Me})$ cm²/s (11-28a)
2. For Ag–Me, $D = 3.20 \times 10^{-13} \exp - (0.0085T_{Me})$ cm²/s (11-28b)
3. For Cu–Me, $D = 7.34 \times 10^{-13} \exp - (0.00149T_{Me})$ cm²/s (11-28c)
4. For Pd–Me, $D = 5.59 \times 10^{-16} \exp - (0.0048T_{Me})$ cm²/s (11-28d)

11.4.1.3 Thermodynamics of Thin-Film Compound Formation

The use of differential scanning calorimetry (DSC) methods has provided an interesting way to study thermodynamic and kinetic aspects of

compound formation in thin films (Ref. 34). In DSC, both the sample (s) and reference (r) material, thermally insulated from one another, are independently heated at specific rates (dT/dt), e.g., 40 K/min, while their temperatures (T) are sensed. As a result, interdiffusion, reactions, and transformations occur. Heats that are evolved or absorbed in the sample for these processes are compensated for by respectively subtracting or adding an equivalent amount of electrical energy to the heater. In the null-balance measurement this power is automatically adjusted to maintain $T(s) = T(r)$, and a plot of the heat flow (dH/dt) vs temperature consisting of one or more peaks is recorded as shown in Fig. 11-17a. To have sufficient sensitivity the specimen is a multilayer stack several micrometers thick consisting of alternating film layers. The molar enthalpy of compound formation (ΔH_f) in the specimen is extracted by essentially integrating $(dH/dt)_s$ (relative to some baseline) over the range of temperature, normalizing the result to the heating rate, and accounting for the number of moles involved.

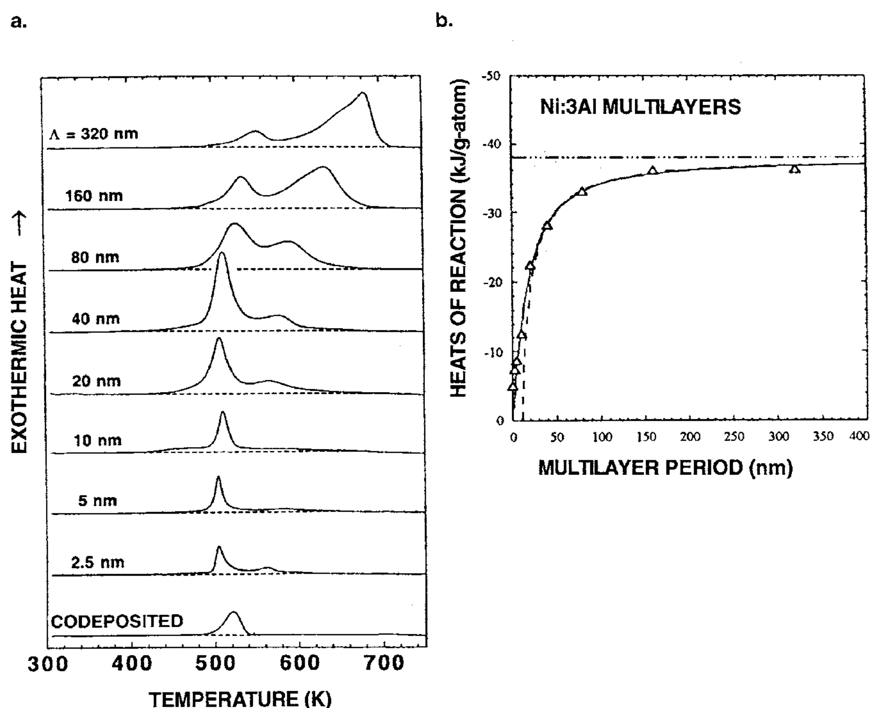


Figure 11-17 (a) Differential scanning calorimetry traces for 1 Ni:3 Al multilayer films with various periods, measured at a heating rate of 40 K/min. (b) Plot of the reaction heats for the transformation of 1 Ni:3 Al multilayers to form $NiAl_3$, vs multilayer period. The DSC heat at large values of Λ extrapolates to −36.2 kJ/mol. *Note:* Δ measured, ______ smooth composition profile, __...___..._ step composition profile. (From Ref. 34. Reprinted with permission.)

Values of reaction heats for nickel–aluminum multilayers having varying period spacings (Λ), but with an overall stoichiometry of 1Ni:3Al, are displayed in Fig. 11-17b. When the period, e.g., Al to Al film distance, is small the measured reaction heat is low, reflecting, perhaps, intermixing of film layers during deposition. However, for large layer periods the reaction heat extrapolates to $\Delta H_f = -36.2$ kJ/mol, which compares favorably with the accepted bulk value of -37.7 kJ/mol. Proceeding in this way ΔH_f values have been determined for other systems, e.g., metastable amorphous and crystalline Cu–Zr (-14.3 kJ/mol for $Cu_{51}Zr_{14}$) and assorted intermetallic Al–Zr compounds. Similarly, values for the heats of crystallization in amorphous films of $CoSi_2$ (-11.4 kJ/mol) and Nb–Cu alloys (ranging from -4.5 to 7.6 kJ/mol) have been measured.

In addition to thermodynamics, kinetic features of compound growth can also be revealed by DSC measurements. For example, under diffusion control, one-dimensional growth is described by

$$dX/dt = (\text{const}/X)\exp -(E_D/RT)$$

where X is the compound thickness and E_D is the diffusional activation energy. For a constant heating rate ($dT/dt = \beta$), integration approximately yields

$$X^2 = (2\,\text{const})\frac{k_B T^2}{\beta E_D}\cdot \exp - \frac{E_D}{k_B T}. \qquad (11\text{-}29)$$

This equation makes use of the fact that the integral of $\exp - (E_D/RT)$ with respect to temperature is equal to $k_B T^2/E_D \exp - (E_D/k_B T)$ within an error less than $\sim 10\%$. Since it is observed that the peak temperature (T_{peak}) for reactant consumption rises with heating rate, a plot of $\ln[\beta/(T_{peak})^2]$ vs $1/T_{peak}$ yields a straight line with a slope of $\sim E_D/k_B$. In this way E_D for Al/Ni was found to be 1.64 eV.

11.4.1.4 The First Thin-Film Compound to Form

Look at any binary phase diagram containing a few intermetallic compounds and try to predict the first compound phase to form when the two elements react. This thorny problem arises during thin-film processing, e.g., metal silicide formation, or whenever film bilayers react. Over the years the question of the first compound phase to form has been addressed by various investigators. Three samples of rules that have been formulated in response follow:

1. The most often quoted rule for predicting the transformation sequence in thin film metal–Si couples, for example, states that "the first phase to nucleate is the most stable congruently melting compound adjacent to the lowest-temperature eutectic on the bulk-equilibrium phase diagram" (Ref. 35). This rule is related to the fact that bulk amorphous phases are readily formed by quenching metal–Si eutectic melts (Section 9.6.5). Although the rule agrees with experiment in many silicides, refined analytical techniques reveal that thin amorphous film layers rather than crystalline phases appear first in many systems. Furthermore, several phases, rather than only one stable phase, form.
2. In their generalized kinetic model Zhang and Ivey (Ref. 36) addressed the issue of reaction sequence in metal silicide films in a more quantitative fashion. They suggested that when the choice is between two phases, the first to appear is the one which suffers the largest free energy degradation rate.
3. In a review of the subject by Pretorius *et al.* (Ref. 37), an effective heat of formation (EHF) model was proposed to predict first phase formation in thin film structures. The rule formulated for metal–semiconductor systems is "the first compound to form during metal–silicon or metal–germanium interaction is the congruently melting phase with the *most negative* effective heat of formation at the concentration of the liquidus minimum of the binary system." Similarly, for metal–metal interactions the same rule holds. A simple graphical construction shown in Fig. 11-18 will illustrate how the rule is translated into practice.

Shown is the previously considered Al–Au system where, in addition to the phase diagram, there is a corresponding plot of the heat of formation (ΔH_f) for each of the intermetallic compounds. From thermodynamics it is axiomatic that $\Delta H_f = 0$ for elements. Between the pure compound ΔH_f value[1] at the peak and the terminal element on each side, the EHF model assumes a linear concentration dependence to complete the triangulation. In the case under consideration the liquidus minimum occurs at $\sim$78 at.% Au. This dotted vertical line closely intercepts $AlAu_2$ and Al_2Au_5, the compounds with the most negative ΔH_f value, and these are the intermetallics predicted to appear first. Interestingly, they are the same ones experimentally reported in Fig. 11-16. In a similar vein the model has correctly predicted the first phase formed in a number of silicides, germanides, and aluminides.

[1]Values of ΔH_f for the pure intermetallic compounds reported in Fig. 11-18 are in units of kJ/mol-at, or the energy released when the total number of Al + Au atoms that react is equal to Avogadro's number. As a caveat, there are at least three other ways that ΔH_f values are reported in the literature (Ref. 37).

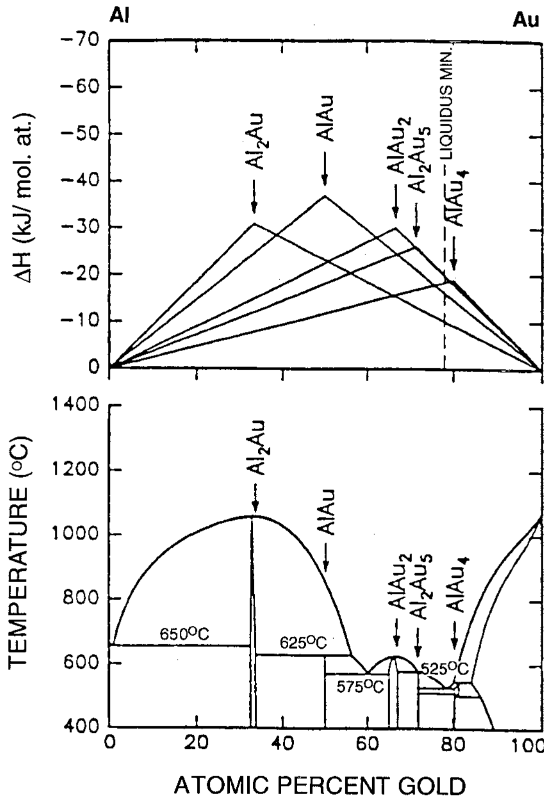


Figure 11-18 Illustration of EHF model applied to Al–Au system. (Top) Heats of formation of Al_2Au, AlAu, $AlAu_2$, Al_2Au_5 and $AlAu_4$ compounds. (Bottom) Equilibrium Al–Au phase diagram. (From Ref. 37. Reprinted with the permission of CRC Press, LLC.)

11.4.2 Introduction to Phase Transformations in Thin Films

By phase transformation we normally mean changing the state of a material at a fixed temperature (on a phase diagram) to another state while keeping the composition constant. Thermodynamic driving forces are the agents of this change which transform unstable materials into metastable or stable phases of lower free energy. And as noted in Chapter 1, a one-, two- or multicomponent system may be involved. Although there is a great body of literature dealing with bulk solid-state phase transformations in all classes

of materials, considerably less has been published on this topic in thin films. Other than the fact that thin specimens are involved, issues of scale rather than mechanistic differences often distinguish transformations in thin films from those in their bulk counterparts. For the most part solid-state phase transformations initiate by nucleation and proceed to completion by growth and impingement of new phase grains. The greater abundance of heterogeneous grain boundary and interfacial nucleation sites in polycrystalline films means higher diffusivities and reaction rates so that transformations generally proceed more rapidly and at lower temperatures.

As in bulk solids, phase transformations in films can occur isothermally (at a single temperature) or continuously (over a range of temperatures). The transformation of an amorphous film matrix to equilibrium crystalline phases can be accomplished both continuously, as we have seen for Au–Co in Section 9.6.5.1, and isothermally in the case of Co–Si, which we now consider.

11.4.2.1 Amorphous-to-Crystalline Transformation in $CoSi_2$

The amorphous-to-crystalline transformation of cobalt disilicide ($CoSi_2$), a metal used in integrated circuits, serves as a particularly instructive example of isothermal reaction (Ref. 38). First, thin amorphous films were deposited and subsequently heated to several elevated temperatures to crystallize them. Transformed spherical compound crystals form, surrounded by the amorphous matrix as shown in Fig. 11-19. The crystallization process was continually monitored through resistivity measurements, and the fractional amount of transformation (f) was determined at each temperature and time. Plots of the type shown in Fig. 11-20a are obtained at each annealing temperature. These S-shaped curves reflect an incubation period prior to transformation. Then nucleation of the crystalline phase is rapid and simultaneous growth results in a quickening of the reaction. Finally, as the crystal grains butt up against each other, f rises very slowly and saturates as the transformation draws to completion.

The experimental curves are excellent fits to the Avrami equation (see Section 7.4.5.4) which may be written in the form

$$f = 1 - \exp - (Kt^n) \tag{11-30}$$

where K and n are constants and t is the time. Furthermore,

$$K = K_0 \exp - (E_t/RT),$$

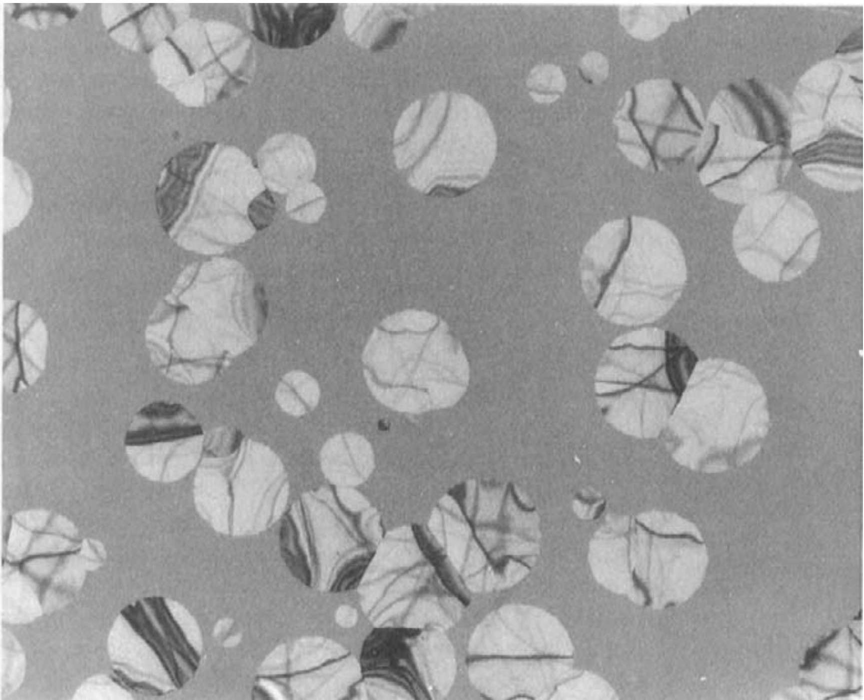


Figure 11-19 TEM image of a partially transformed $CoSi_2$ amorphous film. The circular crystalline $CoSi_2$ particles are immersed in the prior uniform amorphous matrix. (After Refs. 13, 38. Reprinted with the permission of Professor K. N. Tu.)

where K_0 is a constant, E_t is the apparent activation energy for transformation, and RT has the usual meaning. To analyze the data at each annealing temperature they are first plotted as $\ln(-\ln(1-f)$ vs $\ln t$. After this is done the slope is found to have a value of $n \sim 4$. When the times to reach $f = \frac{1}{2}$ are plotted on an Arrhenius plot (Fig. 11-20b), a value of $E_t = 296$ kJ/mol, or 3.07 eV/atom, is obtained.

Further analysis shows that constant $K = (\pi/3)\dot{N}\dot{G}^3$, reflecting a nucleation rate $\dot{N}$ ($\dot{N} \sim \exp - (E_N/RT)$) followed by a linear growth rate $\dot{G}$ ($\dot{G} \sim \exp - (E_G/RT)$) in each of three directions. Assuming both of these processes are thermally activated with respective activation energies for nucleation and growth, E_N and E_G, it is apparent that $E_t = E_N + 3E_G$. Knowing one energy enables the other to be determined. For $CoSi_2$, it was found that $E_N = 0.3$ eV and $E_G = 0.92$ eV. Often, E_N is taken to be zero so that $E_t = 3E_G$.

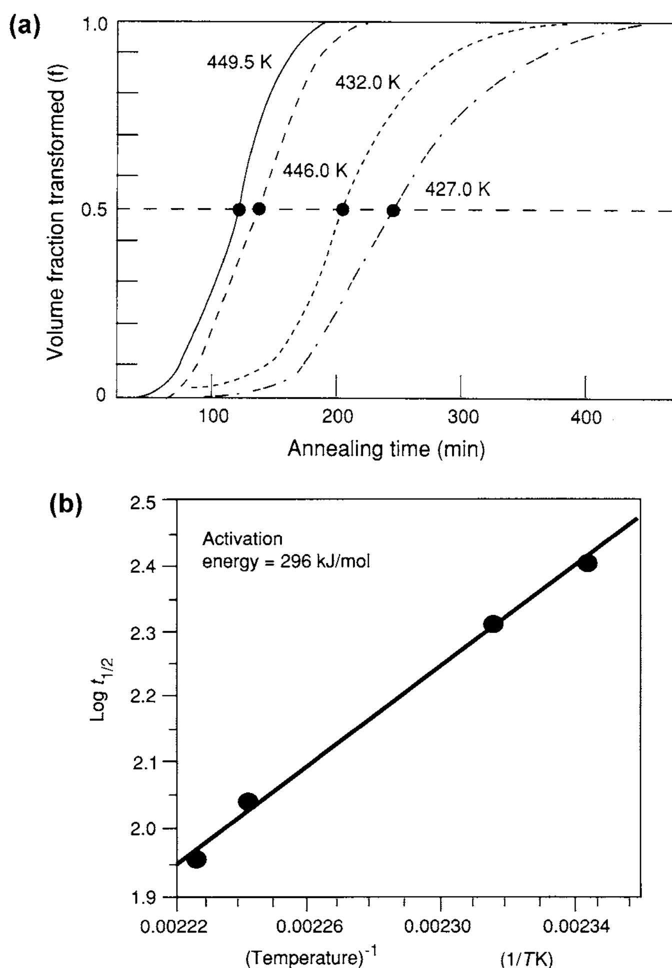


Figure 11-20 (a) Transformed fraction of $CoSi_2$ as a function of time as measured by change in resistivity. (b) Arrhenius plot of log $t_{1/2}$ vs $1/TK$. (After Refs. 13, 38. Reprinted with the permission of Professor K. N. Tu.)

In a more recent study (Ref. 39) of the same amorphous–crystalline transformation in $CoSi_2$ films, values of $n = 3$ and $E_G = 1.2$ eV were obtained. Differential scanning calorimetry results suggested that formation of $CoSi_2$ quickly consumed preferred nucleation sites, and this was followed by three-dimensional growth of crystallites. For constant-heating-rate ($dT/dt = \beta$) conditions, the Avrami equation is transformed so that the fractional change

as a function of temperature is given by

$$f(T) = 1 - \exp[-(kT^2/\beta E)^n \exp(-nE/k_B T)] \qquad (11\text{-}31)$$

where k is a constant. A study (Ref. 40) of phase changes in GeTe films used for optical storage applications is similarly instructive because data and analysis for both isothermal and continuous crystallization follow the foregoing systematics.

11.4.2.2 Epitaxial Regrowth of Amorphous Silicon

As another example of phase transformation in films let us consider the solid-state regrowth of epitaxial silicon from prior amorphous (a-Si) films. This amorphous-to-crystalline transformation differs from that of $CoSi_2$ in at least three ways:

1. Transformation is to a single component rather than a compound phase.
2. Unlike $CoSi_2$ films, which are essentially independent of the substrate, the a-Si films transform with the active participation of the single-crystal Si wafer.
3. One-dimensional consumption of the a-Si layer by a migrating planar epitaxial front is preordained by the film-growth geometry. In contrast, transformation in $CoSi_2$ films proceeds at random locations in three dimensions.

In an elegant study by Olsen and Roth (Ref. 41), surfaces of (100) Si wafers were amorphized either through Si^+ ion implantation, or by coating them with a thin amorphous film of electron-beam-evaporated Si. The specimen surface was heated with a laser which induced unidirectional epitaxial regrowth. Simultaneously, the thickness of the ever-shrinking a-Si film was continually monitored by time-resolved optical reflectance measurements. Recorded solid-phase epitaxial (SPE) growth rates (v) are shown in Fig. 11-21 as a function of temperature. The data were fit to the equation $v(\text{cm/s}) = v_0 \exp - (E_{SPE}/k_B T)$. For implanted and deposited films, the respective values for E_{SPE} were 2.68 and 2.71 eV, with corresponding v_0 values of 3.1×10^8 cm/s and 2.3×10^8 cm/s. Significantly, only a single intrinsic mechanism, probably interfacial-bond breaking or rearrangement, is operative over the entire temperature range. Within experimental error (± 0.05 eV), the common value for E_{SPE} suggests the same SPE mechanism irrespective of different a-Si microstructures. Apparently, bond breaking and rearrangement is facilitated by defect propagation at crystalline edges. In a

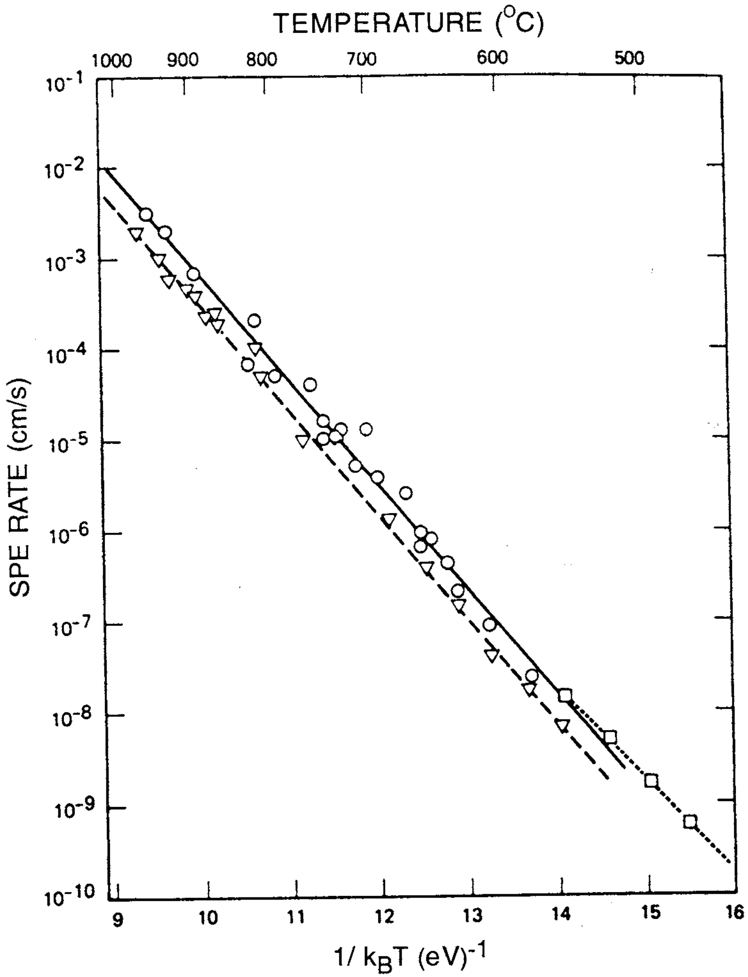


Figure 11-21 Temperature dependence of intrinsic solid-phase epitaxial growth velocity in Si^+ implanted and e-beam deposited amorphous silicon. ____ ○ and ...... ◇, ion implanted; _ _ _ ∇, e-beam deposited. (From Ref. 41. Reprinted with the permission of Elsevier Science Publishers and Dr. G. L. Olson.)

parallel study on Si films doped with implanted As^+, values obtained for E_{SPE} and v_0 were 2.75 eV and 3.68×10^8 cm/s. Significantly, data from this study reflected some 10 decades of SPE growth rate over temperatures spanning 450–1350°C. This writer is unaware of any study that addresses a comparably wide temperature range for mass transport in thin films.

11.5 METAL–SEMICONDUCTOR REACTIONS

11.5.1 INTRODUCTION

It was noted in the Introduction that problems associated with contacting semiconductor surfaces have generated the bulk of studies related to interdiffusion, compound formation, and phase transformations in thin films. This section is essentially devoted to discussing strategies that either prevent or, conversely, promote reactions between metals and semiconductors. In the former case the intent is to preserve the electrical performance of contacts; in the latter the purpose is to create useful metal–semiconductor compounds for contact purposes.

All semiconductor devices and integrated circuits require contacts to connect them to other devices and components. When a metal contacts a semiconductor surface, two types of electrical behavior can be distinguished in response to an applied voltage. In the first, the contact behaves like a p–n junction (Fig. 1-10b) so that current (i)–voltage (V) characteristics are nonlinear. The ohmic contact, on the other hand, passes current equally independent of voltage polarity so that linear i–V characteristics are obtained. There is an imperative to understand mass-transport-induced effects during both metal–semiconductor contact fabrication and subsequent service because they may influence the reliability of the i–V response.

Contact technology has dramatically evolved since the first practical semiconductor device, the point contact rectifier, which employed a metal whisker that was physically pressed into the semiconductor surface. Evaporated aluminum films on top of silicon served as the first contacts in integrated circuits. Reasons for the degradation of Al–Si contacts are interesting metallurgically and provide a good pedagogical vehicle for applying previously developed concepts of mass transport. Historically, minimizing Al–Si reactions led to both the imposition of diffusion barriers and the use of metal silicides. Therefore, it is instructive to start by treating this important contact system.

11.5.2 Al–Si REACTIONS

Nature has endowed us with two remarkable elements—Al and Si. Together with oxygen, they are the most abundant elements on earth. It was their destiny to be brought together in the minutest of quantities to make the computer age possible. Individually, each element is uniquely suited to perform its intended function in a device, but together they combine to form

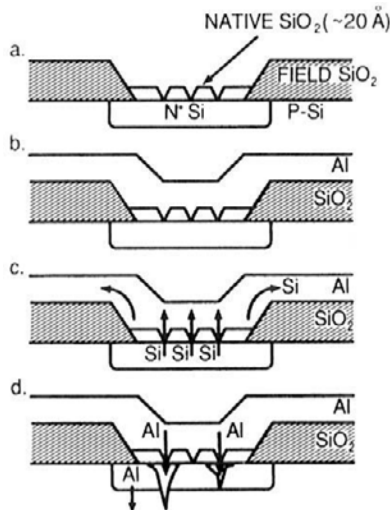


Figure 11-22 Schematic sequence of Al–Si interdiffusion reactions leading to junction spiking.

unstable contacts. In addition to providing the required electrical behavior, they form a diffusion couple where the extent of reaction is determined by the phase diagram and mass-transport kinetics. The processing of deposited Al films for contacts typically includes a 400°C heat treatment. This enables the Al to reduce the very thin native insulating SiO_2 film and sinter to the Si, thereby lowering the contact resistance. Reference to the Al–Si phase diagram (Fig. 1-14) shows that at this temperature Si dissolves in Al to the extent of about 0.3 at.%. During sintering, Si from the substrate diffuses into the Al via grain-boundary paths in order to satisfy the solubility requirement. Simultaneously, Al migrates into the Si by diffusion in the opposite direction. As shown by the sequence of events in Fig. 11-22, local diffusion couples are first activated at several sites within the contact area. When enough Al penetrates at one point, the underlying p–n junction is shorted by a conducting metal filament and junction "spiking" or spearing is said to occur.

The remedy for the problem seems simple enough. By presaturating the Al with Si the driving force for interdiffusion disappears. Usually a 1 wt% Si–Al alloy film is sputtered for this purpose. However, during processing another complication arises. During the heating and cooling cycle Si is first held in solid solution but then precipitates out into the Al grain boundaries as the latter become supersaturated with Si at low temperatures. The irregularly shaped Si precipitate particles, saturated with Al, grow

epitaxially on the Si substrate. These particles are p-type and alter the intended electrical characteristics of the contact. Dimensions of the spikes and precipitate phases were large (i.e., $\sim$100 nm) compared to the shallow junction depth dimensions of prior VLSI (very large scale integration) devices and led to unreliable contact performance. For these reasons the more stable noble metal silicides such as Pd_2Si largely replaced Al contacts. Nowadays in the ULSI (ultralarge scale integration) regime, junction depths are less than 100 nm deep and contact stability demands are even more stringent. Reference to Fig. 6-1 illustrates that these needs are being met by $TiSi_2$ in CMOS devices.

There is yet another example of Al–Si reaction which occurred in early field-effect transistors. In this case, however, the contact to the gate oxide (SiO_2) rather than to the semiconductor source and drain regions is involved. Historically Al films were first used as gate electrodes, but as suggested by Eq. 1-16a, they tend to reduce SiO_2. Other metals are also problematical because of the potential reaction to form a silicide as well as oxide, e.g., $3Ti + 2SiO_2 \rightarrow TiSi_2 + 2TiO_2$. For reliable device performance the preceding considerations led to the adoption of poly-Si films as the gate electrode. Although there was now no driving force promoting reaction between Si and SiO_2, the chronic problem of Si–Al interdiffusion reemerged since Al interconnections still had to contact the gate electrode. To make matters worse, reaction of Al with poly-Si is even more rapid than with single-crystal Si because of the presence of GBs. Reactions similar to those previously described for the Al/Si contact occur and resultant changes are sensitive to the ratio of the film thicknesses. It is easy to see why electrical properties would also be affected. Therefore, intervening silicide films and diffusion barriers were once again relied upon to eliminate contact between Al and Si.

11.5.3 Diffusion Barriers

In the broadest sense of the term, diffusion barriers are thin films used to separate materials from coming into direct contact in order to prevent them from reacting. Paint and electrodeposited layers are everyday examples of practical barriers employed to protect underlying materials from atmospheric attack. In a similar vein, diffusion barriers are used in thin-film metallization systems, an example being silicides to prevent direct Al–Si contact. The subject of diffusion barriers has been reviewed by Kattelus and Nicolet (Ref. 42) and Pauleau (Ref. 43). Following their treatments the discussion will be limited to microelectronic applications.

11.5.3.1 Barrier Requirements

Ideally, a barrier layer X sandwiched between A and B materials should:

1. Constitute a kinetic barrier to the traffic of A and B across it. In other words, the diffusivity of A and B in X should be small.
2. Be thermodynamically stable with respect to A and B at the highest temperature of use; ideally, compounds AX and BX should not form. Further, the solubility of X in A and B should be small.
3. Have low contact resistance with A and B and possess high electrical and thermal conductivity.
4. Be easy to deposit, adhere to the involved films, possess low stress, and be compatible with other processing.

Some of these requirements are difficult to achieve and even mutually exclusive so that it is necessary to make compromises. With regard to item 3, it is essential to maintain low resistance in the contact structure. Just how low can be estimated by considering a barrier layer of resistivity ρ and thickness d, sandwiched between the metallization above and the semiconductor of area A_0 below. It is assumed that the voltage drop (V_B) across the barrier has a value of 0.025 V (equivalent to $k_B T$ at 300 K) and the maximum current density is $j = 10^6$ A/cm^2. By Ohm's law, $V_B = jA_0\rho d/A_0$ or $j\rho d$, and for the conditions specified $\rho d = 2.5 \times 10^{-8}$ Ω-cm^2. For a typical barrier thickness of 100 nm the bulk resistivity must therefore be below 2500 $\mu\Omega$-cm. For less extreme current densities, higher barrier resistivities can be tolerated. This generally means that many materials other than metals can qualify as diffusion barriers from an electrical standpoint.

Another barrier requirement is thickness. If the reaction-rate kinetics of compounds, i.e., AX, BX, are known, then either the effective lifetime or the minimum thickness of the barrier can be predicted. The following example is instructive in this regard (Ref. 44). Suppose we consider a Ti diffusion barrier between Si and Al. Without imposition of Ti, the Al/Si combination is unstable. The question is, how much Ti should be deposited to withstand a thermal anneal at 500°C for 15 min? At the Al interface $TiAl_3$ forms with parabolic kinetics given by

$$d^2(\mathrm{TiAl_3}) = 1.5 \times 10^{15} \exp - \frac{1.85\ \mathrm{eV}}{k_B T} \cdot t\ (\text{Å}^2) \qquad (11\text{-}32)$$

where $d(\mathrm{TiAl_3})$ is the thickness of the $TiAl_3$ layer and t is the time in seconds. Similarly, the reaction of Ti with Si results in the formation of $TiSi_2$

with a kinetics governed by

$$d^2(\mathrm{TiSi_2}) = 5.74 \times 10^9 \exp - \frac{1.3\ \mathrm{eV}}{k_\mathrm{B}T} \cdot t\ (\text{Å}^2) \qquad (11\text{-}33)$$

For the specified annealing conditions, $d(\mathrm{TiAl_3}) = 1100$ Å and $d(\mathrm{TiSi_2}) =$ 130 Å. An insignificant amount of Ti is consumed under ambient operating conditions. Therefore, the minimum thickness of Ti required is the sum of these two values or 1230 Å.

11.5.3.2 Barrier Materials

A considerable number of materials have been investigated for use as barrier layers between semiconductor contacts and interconnects, as well as to prevent reaction between interconnects and surrounding dielectrics. At contacts there is usually either a three-layer, e.g., semiconductor–diffusion barrier film–interconnect line, or four-layer, e.g., semiconductor–contact layer–diffusion barrier film–interconnect line, structure. With one fewer layer, the first of these contact schemes is simpler. In preventing Si–Al interconnect reactions the following diffusion barrier materials have been tried (Refs. 42, 45): refractory metals (Cr, V, Ti); transition-metal alloys (TiW); transition-metal compounds (TiN, TiC, TaN); amorphous alloys (MoNi, NiW, WZr); silicides ($TiSi_2$, $CoSi_2$, PtSi, Pd_2Si); dual-layer films such as silicide/transition metal (Pd_2Si/Cr, $CoSi_2$/W); and silicide/transition-metal compound combinations (PtSi/TiN, PtSi/TiC, NiSi/TaN). Of these only a few have seen wide use in silicon technology, e.g., $TiSi_2$ and PtSi/TiW (Ref. 46). In another noteworthy use of diffusion barriers, reaction between Al interconnects and SiO_2 is prevented by TiN liners.

Similar strategies exist to create stable contacts to compound–semiconductor devices. For example, metals such as Au, Ni, Pd, Au/Ge, Ni/Ge, and AuGeNi have all been used to contact GaAs (Ref. 47).

To assess the effectiveness of contacts and diffusion barrier materials, analytical techniques are employed to reveal both metallurgical interactions and their effect on device properties. For this purpose RBS measurements, and SIMS and AES depth profiling to a lesser extent, have been complemented by various methods for determining interfacial electronic barrier heights between metals and semiconductors (Fig. 1-10b), a sensitive indicator of low-temperature contact reactions.

11.5.3.3 Barrier Behavior

To aid in the selection of diffusion barrier materials, it is useful to distinguish among three mechanisms that have been proposed to explain their behavior (Ref. 42).

1. *Stuffed barriers.* Stuffed barriers rely on the segregation of impurities along otherwise rapid diffusion paths such as GBs to block further two-way atomic traffic there. The marked improvement of sputtered Ti, Mo, and Ti–W alloys as diffusion barriers when they contain small quantities of intentionally added nitrogen or oxygen is apparently due to this mechanism. Such interstitial impurity concentrations of $\sim 10^{-1}$ to 10^{-3} at.% are typically required to decorate GBs and induce stuffed-barrier protection. In extending the electromigration life of Al, Cu may in effect "stuff" the conductor GBs.

2. *Passive compound barriers.* Ideal barrier behavior exhibiting chemical inertness and negligible mutual solubility and diffusivity is sometimes approximated by compounds. Although there are numerous possibilities among the carbides, nitrides, borides and even the more conductive oxides, only the transition-metal nitrides have been extensively exploited for device applications.

3. *Sacrificial barriers.* A sacrificial barrier maintains the separation of A and B but only for a limited duration. As shown in Fig. 11-23, sacrificial barriers exploit the fact that reactions between adjacent films in turn produce uniform layered compounds AX and BX which continue to be separated by a narrowing X barrier film. So long as some X remains and compounds AX and BX possess adequate conductivity, this barrier is effective. The first recognized application of a sacrificial barrier involved Ti which reacted with Si to form $TiSi_2$ and with Al to form $TiAl_3$; this was the barrier system modeled earlier. Judging by the many metal aluminide- and metal-silicide compounds, the possibility for sacrifical barrier reactions appears to be great when employing single metals.

In conclusion, we note that semiconductor contacts are usually thermodynamically unstable. The imposition of a diffusion barrier slows down the equilibration process but the instability is never actually removed. This has necessitated the use of "diffusion barriers" to protect diffusion barriers. Enhanced reliability is bought with diffusion barriers, but at the cost of increasing structural complexity and added processing expense.

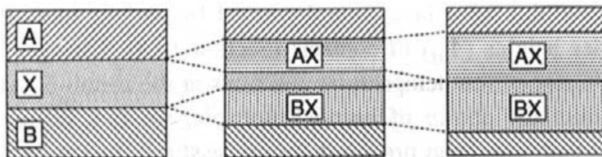


Figure 11-23 Model of sacrificial barrier behavior. A and B films react with barrier film X to form AX and BX compounds. Protection is afforded as long as X is not consumed. (Reprinted with permission from Elsevier Sequoia, S.A. From Ref. 44.)

11.5.4 SILICIDES

Over the years, silicon has been doped, contacted by, or reacted with virtually every element in the periodic table in an unceasing effort to advance semiconductor technology. Much of this research and development effort has dealt with reactions between thin metal films and single-crystal Si in order to develop silicides for contact applications. Among the issues and questions addressed by these investigations are:

1. Which silicide compounds form?
2. What is the time and temperature dependence of metal silicide formation?
3. Which atomic mass-transport mechanisms are operative during silicide formation? Does the metal or silicon species migrate more rapidly?
4. When the phase diagram indicates a number of different stable silicide compounds, which form preferentially and in what reaction sequence?

Virtually all thin-film characterization and measurement tools have been employed at one time or another in studying these aspects of silicide formation. In particular, RBS methods have played an important role in shaping our understanding of metal–silicon reactions by revealing compound stoichiometries, layer thicknesses, and the moving species. As an example, the RBS spectra shown in Fig. 11-24 depict the growth of Ru_2Si_3. Interpretation of these data is left to the reader (see Section 10.4.7.5).

11.5.4.1 Metal–Silicon Reactions

A summary of kinetic data obtained for a large body of near-noble, transition and refractory metal silicides is summarized as follows:

Silicide	Formation temp. (°C)	Growth kinetics	Activation energy (eV)
M_2Si	200	$t^{1/2}$	1.5
MSi	400	$t^{1/2}$	1.6–2.5
MSi_2	600	$t^{1/2}$	1.7–3.2

Nucleation (Ref. 48) and kinetic (Ref. 49) aspects of silicide formation have been reviewed by d'Heurle and Gas. Classical thermodynamics was found to be particularly fruitful in describing nucleation, a step normally ignored in the description of interdiffusion reactions. As a rough rule of thumb, silicide formation temperatures range from one-third to one-half of the melting point (in kelvins) of the corresponding silicide. Since fine-grained metal films are involved, it is not surprising that this rule is consistent with

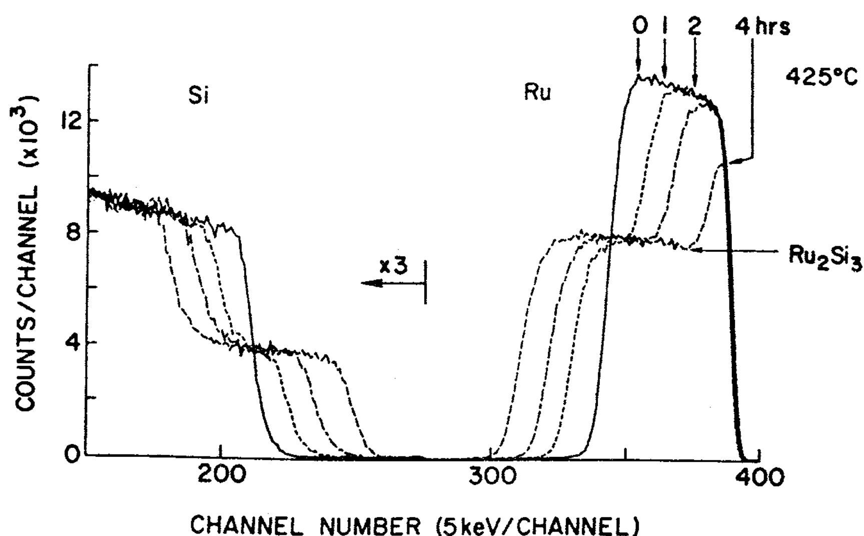


Figure 11-24 Rutherford backscattering spectra showing the progressive formation of Ru_2Si_3 through reaction between ruthenium and silicon at 425°C. (From Ref. 49. Reprinted with the permission of Dr. F. M. d'Heurle.)

the GB-diffusion regime. The activation energies correlate with the melting point of the silicide in accord with systematic diffusion trends noted earlier. Parabolic diffusional-growth kinetics are usually observed, but linear growth reflecting reaction limitations has also been reported, e.g., in NiSi, $CrSi_2$, and $MoSi_2$.

11.5.4.2 What Is the Migrating Species? The Kirkendall Effect

The Kirkendall effect has clarified a number of issues in solid-state diffusion. One of its great successes is the unambiguous identification of vacancy motion as the operative atomic transport mechanism during interdiffusion in binary alloy systems. The Kirkendall experiment requires a diffusion couple with a small inert marker located within the diffusion zone between the two involved migrating atomic species. An illustration of what happens to the marker during thin-film silicide formation is shown in Fig. 11-25. Assuming that metal (M) atoms exchange sites more readily with vacancies than do Si atoms, more M than Si atoms will sweep past the marker. In effect the lattice is transported toward the left! To avoid lattice stress or void generation, the marker responds by shifting as a whole toward the right. The reverse is true if Si is the dominant migrating species. Such

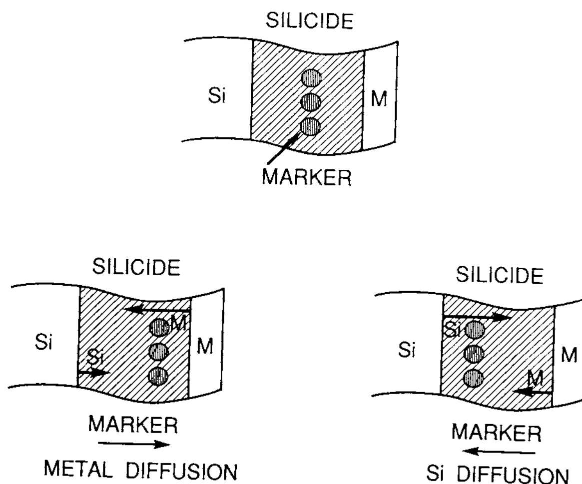


Figure 11-25 Schematic of Kirkendall marker motion during silicide formation. (Reprinted with permission of John Wiley & Sons. From Ref. 11.)

marker motion has indeed been observed in an elegant experiment (Ref. 50) employing RBS methods to analyze the reaction between a thin Ni film and a Si wafer. Implanted Xe, which served as the inert marker, moved toward the surface of the couple during formation of Ni_2Si. The interpretation, therefore, is that Ni is the dominant diffusing species.

In the metal-rich silicides the metal is observed to be the dominant mobile species, whereas in the mono- and disilicides Si is the diffusing species. The crucial step in silicide formation requires the continual supply of Si atoms through the breaking of bonds in the substrate. In the case of disilicides, high temperatures are available to liberate Si for reaction. But at lower temperatures there is insufficient thermal energy to cause breaking of Si bonds, and the metal-rich silicides probably form by a different mechanism; bond breaking through rapid interstitial migration of metal through the Si lattice has been suggested.

11.5.4.3 Sequence of Compound Formation

The question of the first silicide phase to form has already been addressed in Section 11.4.1.4. But the sequence of phase formation after the initial reaction has only been established in a few silicide systems (Ref. 21). One of the most extensively studied of these is the Ni–Si system for which the phase diagram and compound formation map are provided in Fig. 11-26. The map

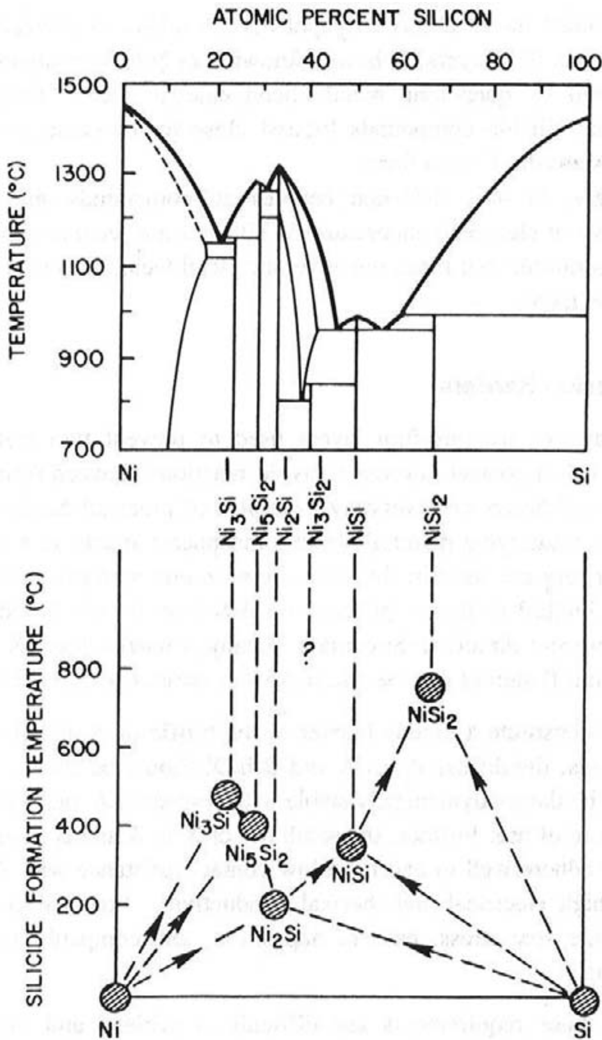
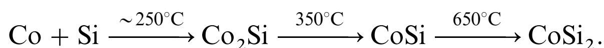


Figure 11-26 Map of thin-film Ni silicide formation sequence. Phase diagram of Ni–Si system shown on top. (Reprinted with permission of Annual Reviews Inc. From Ref. 21.)

shows that Ni_2Si is always the first phase to form during low-temperature annealing. Clearly Ni_2Si is not in thermodynamic equilibrium with either Ni or Si according to the phase diagram. What happens next depends on whether Si or Ni is present in excess. In the usual former case where a Ni thin film is deposited on a massive Si wafer, the sequence proceeds first to

NiSi and then to $NiSi_2$ at elevated temperatures. However, when a film of Si is deposited on a thicker Ni substrate, then the second and third compounds become Ni_5Si_2 and Ni_3Si. At elevated temperatures the resultant two-phase equilibrium, i.e., $Si/NiSi_2$ or Ni/Ni_3Si, conforms to the phase diagram. As with $NiSi_2$ the reaction of Co and Si to form $CoSi_2$ goes through the following stages:



$$\mathrm{Co} + \mathrm{Si} \xrightarrow{\sim 250^\circ\mathrm{C}} \mathrm{Co_2Si} \xrightarrow{350^\circ\mathrm{C}} \mathrm{CoSi} \xrightarrow{650^\circ\mathrm{C}} \mathrm{CoSi_2}.$$

Interestingly, in bulk diffusion couples all compounds appear to grow simultaneously at elevated temperatures. This does not seem to happen in films at low temperature, but more sensitive analytical techniques may be required to clarify this issue.

11.5.4.4 Silicide Properties

In addition to being diffusion barriers, metal silicides are specifically used in integrated circuit technology to counter the increasing resistance of shrinking polysilicon gate contacts to field-effect transistors. Although electrical resistivity is important, process integration and property requirements (i.e., thermal stability, oxidation resistance, chemical reactivity, diffusivity in Si) are critical factors in the selection of silicides for this demanding application.

The most exhaustive compilation of the crystal and epitaxial structure of silicides as well as their electrical, mechanical, thermodynamic, and interfacial properties can be found in the book by Maex and Rossum (Ref. 51). Culled from this as well as other references (Refs. 52, 53) are the properties for a number of monosilicides (MSi) as well as metal-rich (M_2Si) and silicon-rich (MSi_2) silicides that appear in Table 11-2. Of the silicides the electrical resistivity of $TiSi_2$ is an issue of considerable importance. When formed at $\sim 700^\circ$C, $TiSi_2$ exists as C49, a base-centered orthorhombic phase with a resistivity of 60–90 $\mu\Omega$-cm. But through a polymorphic transformation at about 800°C, the face-centered orthorhombic C54 phase with a more desirable lower resistivity ($\sim$12–25 $\mu\Omega$-cm) forms.

11.5.5 Technology of Silicides, Salicides, and Polycides

Two basic types of silicide processes are employed in the semiconductor industry, and they are schematically illustrated in Fig. 11-27. The *salicide* process is a way to self-align silicide to all exposed regions, i.e., source, drain,

Table 11-2
Properties of Metal Silicides[a]

Silicide	Resistivity ($\mu\Omega$-cm)	TCE (ppm/°C)	T_M (°C)	Formation energy[b] (kcal/mol)	Elastic modulus (GPa)
NiSi	20		992	−20.6	
PtSi	28–35		1229	−15.7	
Ni_2Si	25	16.5	1318	−33.0	161
Pd_2Si	25		1330	−20.7	
$TiSi_2$ (C54)	12–25	12–13	1540	−32.1	260
$CoSi_2$	18–25	~10	1326	−24.6	160
$MoSi_2$	20–100	~8	1980	−31.4	435
$TaSi_2$	10–50	8–11	~2200	−27.0	
WSi_2	50–120	6–8	2165	−22.2	~300–530

[a] Data taken from several sources (Refs. 51–53).
[b] Based on Avogadro's number of metal atoms consumed in forming silicide.

and gate. In this process, the polycrystalline silicon gate is patterned and SiO_2 sidewall spacers are formed prior to deposition of metal. Then metal is deposited and reacted to produce metal silicides at the exposed silicon; in contrast, no silicide forms over isolation regions. A selective etch removes unwanted reaction products and the silicide is often annealed to reduce its resistivity.

In the *polycide* process, the stoichiometric metal silicide is deposited as an amorphous film over the polysilicon gate before patterning. After the gate is etched a so-called polycide stack forms which is often capped by an insulator. Annealing at a relatively high temperature is essential in order to transform the amorphous silicide to a crystalline state with lower resistivity (see Section 11.4.2.1). Once the sidewall spacers are formed, silicide contacts to the source and drain regions are fabricated as in the salicide process.

Currently, the most popular metal silicide in CMOS technology is $TiSi_2$. A number of interesting reactions occur simultaneously during the formation of $TiSi_2$ by the salicide process. As Ti and Si react at the gate, source, and drain to form low-resistance contacts, Si can diffuse laterally over isolation regions and form thin filaments of $TiSi_2$. These tend to electrically short the gate to sources or drains and compromise the reliability of CMOS devices. This defect, known as bridging, is largely prevented by annealing in nitrogen. Apparently nitrogen diffuses into the Ti grain boundaries where it forms TiN and reduces any long-range Si transport into the Ti layer;

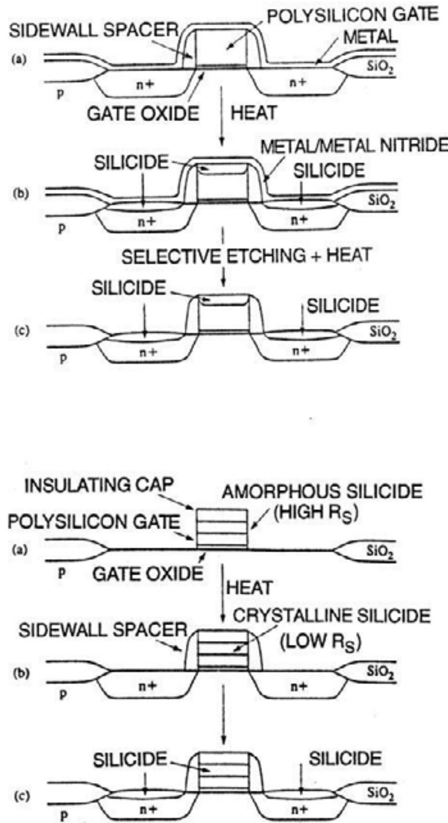
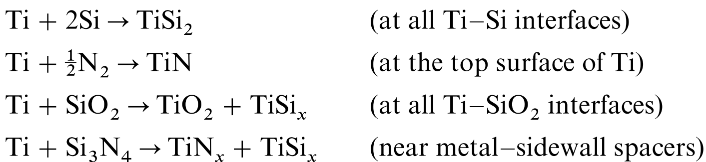


Figure 11-27 (Top) Schematic of the salicide process. (a) After metal deposition. (b) After annealing to form silicide. (c) After removal of excess metal and metal nitride by etching, and transformation annealing. (Bottom) Schematic of the polycide process showing change in the stack structure. (a) After patterning of the multilayer structure. (b) After spacer formation and annealing. (c) After formation of the source–drain contacts. (From Ref. 53. Reprinted with the permission of Dr. L. A. Clevinger.)

nitrogen also causes some Si_3N_4 formation. These reactions and the locations where they occur are summarized as



$$\mathrm{Ti} + 2\mathrm{Si} \rightarrow \mathrm{TiSi_2} \qquad \text{(at all Ti–Si interfaces)}$$

$$\mathrm{Ti} + \tfrac{1}{2}\mathrm{N_2} \rightarrow \mathrm{TiN} \qquad \text{(at the top surface of Ti)}$$

$$\mathrm{Ti} + \mathrm{SiO_2} \rightarrow \mathrm{TiO_2} + \mathrm{TiSi}_x \qquad \text{(at all Ti–SiO}_2\text{ interfaces)}$$

$$\mathrm{Ti} + \mathrm{Si_3N_4} \rightarrow \mathrm{TiN}_x + \mathrm{TiSi}_x \qquad \text{(near metal–sidewall spacers)}$$

Although the first reaction occurs below 700°C, higher processing temperatures enable the remaining three reactions to proceed readily. Because TiO_2, $TiSi_x$, and TiN form over the isolation regions and are difficult to etch away, there is an imperative to use low heat treatment temperatures.

Once formed, C49 $TiSi_2$ must be transformed to the low-resistivity C54 phase. In achieving this, temperatures above ~800°C must be avoided to prevent thin $TiSi_2$ film agglomeration and a corresponding increase in resistivity. The activation energy for the C49-to-C54 transformation on lightly doped (100) Si using the salicide process is a rather high 5.7 eV. Because the transformation is nucleation limited, elevated temperatures are effective in creating many nuclei and a fine-grained C54 phase. The disadvantages of high temperatures noted earlier mean a narrow processing window and compromises. Other variables that influence the transformation temperature are metal line width and the type and concentration of Si dopants. Thus the geometric constraints to reaction in patterned films cause them to transform to C54 at higher temperatures than blanket $TiSi_2$ films (Ref. 54). In general n-type dopants impede silicide formation, while troublesome Ti–B reactions raise the contact resistance of $TiSi_2$ on Si.

Returning to the parallel polycide process, the intially amorphous $TiSi_2$ film sequentially transforms to C49 and then to C54 as the temperature is ramped up, usually by rapid thermal annealing (RTA). During RTA, radiant heating occurs through the use of high-intensity tungsten-halogen flash lamps (Section 8.5.3.3). As a result the absorbing substrate heats very rapidly (typically several hundred degrees Celsius per second) to effect the required transformation without causing excessive unwanted diffusional transport and reaction. Problems with bridging and agglomeration in C54 have prompted experimentation with the other silicides.

11.6 MASS TRANSPORT IN THIN FILMS UNDER LARGE DRIVING FORCES

11.6.1 INTRODUCTION

There are a number of mass-transport phenomena observed with difficulty in bulk materials that are greatly magnified in thin films because of their small dimensions. By controlling deposition and patterning it is possible, for example, to shrink film thickness dimensions such that diffusional driving forces are very large either perpendicular to parallel to the film plane. The former occurs in superlattices and other artificially modulated structures that contain steep concentration gradients. On the other hand, metal interconnections having a small cross section and short

length can be subjected to very large mass-transport driving forces in the plane of the film due to the passage of high current densities. As a result, nonlinear diffusion and electromigration effects, respectively, may be manifested. In this section we shall explore a few of these mass-transport size effects in terms of their theoretical and practical implications.

11.6.2 Nonlinear Mass Transport Effects

11.6.2.1 Kinetic Nonlinear Effects

A linear diffusion theory has been utilized to describe the various transport effects considered to this point, and except for this section, is assumed for the remainder of the thin-film applications in this chapter and book. The macroscopic Fick diffusion equations defined by Eqs. 1-22 and 1-24 suffice as an operating definition of what is meant by linear diffusion theory. To understand nonlinear effects, let us reconsider atomic diffusion in the presence of a free-energy gradient driving force. Two pertinent equations (1-33 and 1-35) describing this motion are reproduced here:

$$r_{\mathrm{N}} = 2\nu \exp\left(-\frac{G_{\mathrm{D}}}{RT}\right) \sinh\left(\frac{\Delta G}{RT}\right) \qquad \text{and} \qquad v = DF/RT.$$

The latter follows when $\sinh(\Delta G/RT)$ is small, a condition met when $(\Delta G/RT) \ll 1$. But, since $\sinh(\Delta G/RT) = \Delta G/RT + 1/3!(\Delta G/RT)^3 + \ldots$, if $(\Delta G/RT) \simeq 1$ or greater, then higher order terms in the expansion must be considered. This has the effect of nonlinearizing all of the diffusion equations. Physically such a situation arises when there are large compositional differences over very small dimensions. It has then been shown by Tu (Ref. 21) that instead of the Nernst–Einstein equation,

$$v = \frac{D}{RT}\left\{F + \frac{a_0^2 F^3}{24(RT)^2}\right\} \tag{11-34}$$

holds, and rather than the linear nonsteady-state diffusion equation,

$$\frac{\partial C}{\partial t} = \frac{D\partial^2 C}{\partial x^2} + \frac{D a_0^2}{8C^2}\left(\frac{\partial C}{\partial x}\right)^2 \frac{\partial^2 C}{\partial x^2} \tag{11-35}$$

is appropriate. This fifth-degree nonlinear differential equation has been termed the “kinetic-nonlinear” equation. The author is unaware of its use in any applications.

11.6.2.2 Thermodynamic Nonlinear Effects

Equation 11-35 should not be confused with an earlier and outwardly similar "thermodynamic nonlinear" equation (Eq. 11-36) derived from an analysis of atomic diffusion in binary alloys possessing large compositional free-energy gradients. Assuming G is a function of C, dC/dx, $d^2C/dx^2, \ldots$, a complex thermodynamic–kinetic equation of the form

$$\frac{\partial C}{\partial t} = \frac{\tilde{D}\partial^2 C}{\partial x^2} - \frac{2\tilde{D}K}{G''_0(C)}\left[\frac{\partial^4 C}{\partial x^4}\right] \quad (11\text{-}36)$$

has been derived. Term $G''_0(C)$ is defined as d^2G_0/d^2C where $G_0(C)$ is the free energy of the homogeneous alloy, $\tilde{D}$ is the interdiffusion coefficient (Eq. 11-12), and K is a constant known as the gradient-energy coefficient. This equation has primarily been used to describe diffusion in superlattices or in composite layers with very small film thickness. As an example consider a harmonic composition modulation of wavelength λ with an average concentration level of C_0. Simple substitution reveals that the solution to Eq. 11-36 is

$$C(t, x)/C_0 = A_0 \exp\{-\tilde{D}h^2[1 + 2Kh^2/G''_0(C)]t\} \cos(hx), \quad (11\text{-}37)$$

where A_0 is a constant and $h = 2\pi/\lambda$. Depending on the sign of the term within the $\{\}$ brackets of the exponent, the composition wave can grow $\{+\}$ or decay $\{-\}$ in amplitude. Wave growth by unnatural "uphill" diffusion of atoms occurs for alloy compositions such that $G''_0(C)$ is negative, the condition for the existence of so-called "spinodal" regions (Ref. 55) on phase diagrams.

A review of diffusion in artificially modulated structures has been given by Greer (Ref. 56); multilayer systems discussed include crystalline Au/Ag, Cu/Pd, Au/Ni, Au/Cu, Cu/Ni, Ag/Pd, and GaAs/AlAs, as well as assorted amorphous metal alloy and semiconductor films. For example, in Ge/Si amorphous multilayers having a repeat length of 5.83 nm, X-ray diffraction revealed a decay in amplitude of (superlattice sideband) peaks during interdiffusion (Ref. 57). Assuming the intensity (I) simply varies as $I = I_0 \exp[-8\pi^2\tilde{D}/\lambda^2 t]$, values of $\tilde{D} = 1.07 \times 10^{-6} \exp-(1.6\,\text{eV}/k_B T)\ \text{cm}^2/\text{s}$ were obtained. Interestingly, in this system interdiffusion of Si and Ge atoms was sufficiently rapid that it could be measured without crystallization occurring.

11.6.3 Electromigration in Thin Films

11.6.3.1 Manifestations of Electromigration

The phenomenon of electromigration is not unlike electrolysis and involves the migration of metal atoms through metallic conductors carrying

large direct-current densities. It was observed in liquid metal alloys well over a century ago and is a mechanism responsible for the failure of tungsten lightbulb filaments. Bulk metal conductors melt when powered with current densities (j) approaching 5×10^4 A/cm^2 (Ref. 58). On the other hand, thin-metal films can tolerate densities of $\sim 10^7$ A/cm^2 without immediately melting or open circuiting because the Joule heat is effectively conducted away by the substrate, which behaves as a massive heat sink. In present integrated circuit chips, if all of the polycrystalline Al alloy interconnect stripes were laid end to end they would typically extend several kilometers or so. Typical cross sections of these interconnects, which are usually surrounded by SiO_2, are less than 0.3 μm wide and ~ 1 μm thick. When powered at high current densities, mass transport is manifested by such effects as void formation, mass pileups and hillocks, extruded metal, cracked dielectric-film overlayers, grain-boundary grooving, whisker bridging, localized heating, and thinning both along the conductor stripe and near contacts. Several examples of such film-degradation processes are shown in Fig. 11-28. In bootstrap fashion the damage accelerates to the point where open-circuiting terminates the life of the conductor. It is for these reasons that electromigration has been recognized as a major reliability problem in integrated circuit metallizations for the past 35 years. Indeed there is some truth to a corollary of one of Murphy's laws: "A million-dollar computer will protect a 25-cent fuse by blowing first."

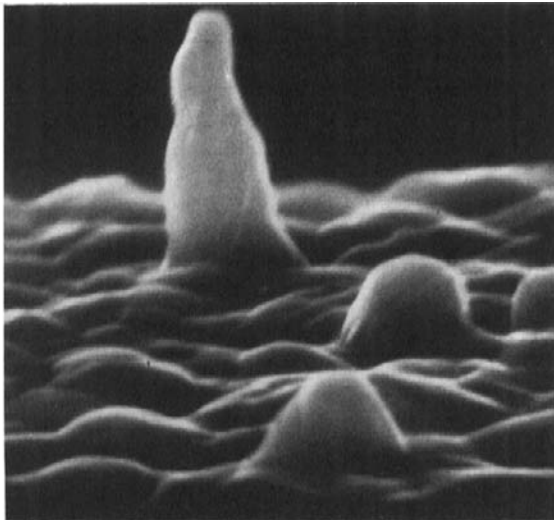
The subject of electromigration and its reliability implications has been reviewed by this author (Ref. 58). For this reason the discussion of electromigration will be largely limited to fundamental mass-transport phenomena rather than current technological issues.

11.6.3.2 Modeling Electromigration

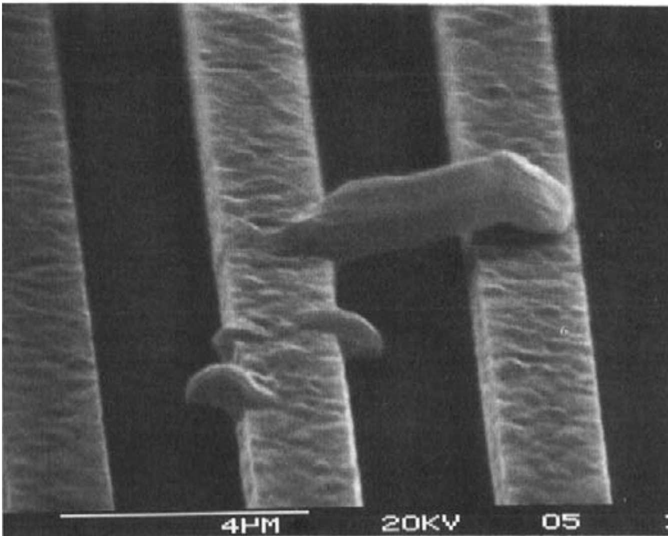
The mechanism of the interaction between the current carriers and migrating atoms is not entirely understood, but it is generally accepted that electrons streaming through the conductor are continuously scattered by lattice defects. At high enough current densities sufficient electron momentum is imparted to atoms to physically propel them into activated configurations and then toward the anode as shown in Fig. 11-29a. This so-called electron "wind" force is oppositely directed to and normally exceeds the well-shielded, cathode-directed electrostatic force on atom cores arising from the applied electric field, $\mathscr{E}$. Therefore, a net force given by

$$F = Z^* q \mathscr{E} = Z^* q \rho j \tag{11-38}$$

acts on the migrating ions, where q is the electronic charge and the electric field is given by the product of the electrical resistivity of the metal, ρ, and j.

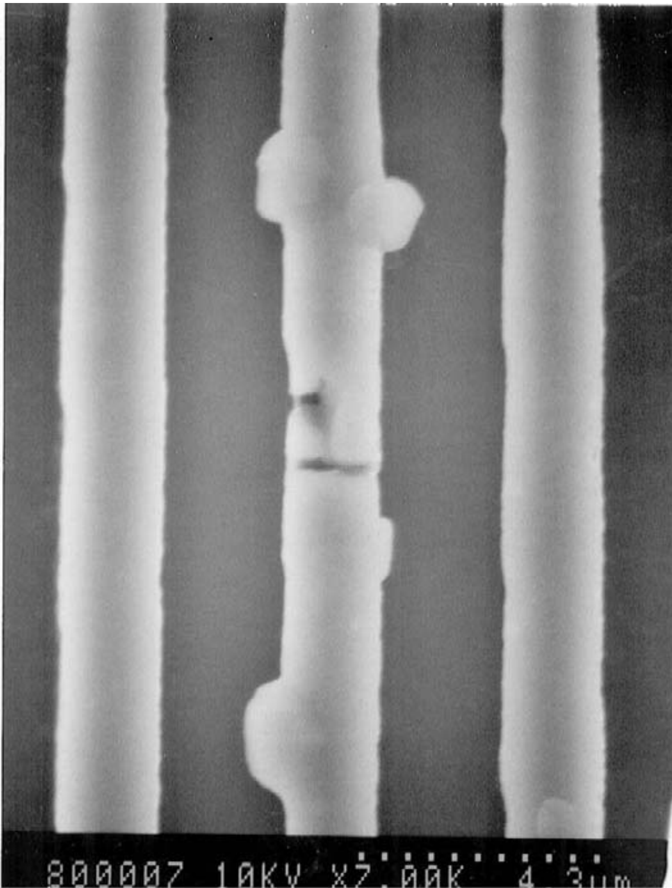


(a)



(b)

Figure 11-28 Manifestations of electromigration damage in Al films: (a) Hillock growth. (From M. Ohring and R. Rosenberg, *J. Appl. Phys.* **42**, 5671 (1971). Courtesy of L. Berenbaum.) (b) Whisker bridging two conductors. (Courtesy of R. Knoell, AT&T Bell Laboratories.) (c) Lateral mass growths and grain-boundary slit fracture in a conducting stripe. (Courtesy of K. P. Rodbell, IBM, T. J. Watson Research Center.)



(c)

Figure 11-28 Continued.

An "effective" ion valence Z^* may be defined, which for electron conductors is negative in sign because positive ions migrate to the anode. Furthermore, the measured magnitude of Z^* is usually far in excess of typical chemical valences. On a macroscopic level, the observed mass-transport flux for an element of concentration C is given by

$$J = Cv = \frac{CDZ^*q\rho j}{RT} \tag{11-39}$$

where use has been made of the Nernst–Einstein relation. Electromigration is thus characterized at a fundamental level by the terms Z^* and D.

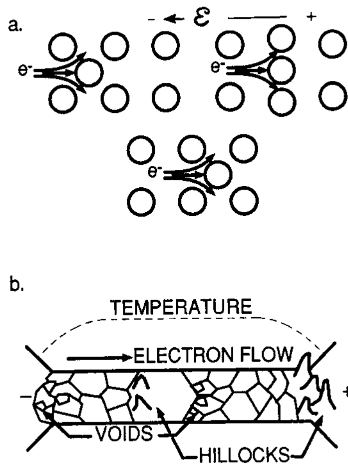


Figure 11-29 (a) Atomic model of electromigration involving electron momentum transfer to metal ion cores during current flow. (b) Model of electromigration damage in a powered film stripe. Mass flux divergences arise from nonuniform grain structure and temperature gradients.

Although considerable variation in Z^* exists, values of the activation energy for electrotransport in films reflect a grain-boundary diffusion mechanism.

Film damage is caused by a depletion or accumulation of atoms which is defined by either a negative or positive value of dC/dt, respectively. But dC/dt is equal to $-\nabla \cdot J$ and therefore by expanding Eq. 11-39,

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}\left\{\frac{CDZ^*q\rho j}{RT}\right\} - \frac{\partial}{\partial T}\left\{\frac{CDZ^*q\rho j}{RT}\right\}\frac{\partial T}{\partial x}. \qquad (11\text{-}40)$$

The first term on the right-hand side reflects the isothermal, structurally induced mass-flux divergence, while the second term represents mass transport in the presence of a temperature gradient. Transport under these distinct conditions can be qualitatively understood with reference to Fig. 11-29b, assuming that atom migration is solely confined to GBs and directed toward the anode.

11.6.3.3 Structural and Thermal Effects

Let us first consider electromigration under isothermal conditions. Because of varying grain size and orientation distributions, local mass-flux divergences may exist throughout the film interconnect structure. In addition, divergences in mass flux necessarily exist at the interfaces between

interconnects and semiconductor contacts, as well as at junctions between multilevel metallizations and vias or diffusion barriers, i.e., wherever the continuity of mass flow is interrupted. In fact, electromigration damage has been observed to occur at each of these sites. For the case of interconnects, each cross section contains a lesser or greater number of effective GB transport channels. If more atoms enter a region such as a junction of grains than leave it, a mass pileup or growth can be expected. A void develops when the reverse is true. At highly heterogeneous sites where, for example, a single grain extends across the stripe width and abuts numerous smaller grains, the mass accumulations and depletions are exaggerated. This is why a uniform distribution of grain size is desirable. Of course, single-crystal films would make ideal interconnections because the source of damage sites is eliminated, but it is not practical to deposit them. Nevertheless, bamboo-like interconnect structures that statistically emerge when large-grained metallizations are lithographically patterned are resistant to electromigration damage because their grain boundaries lie perpendicular to the current flow.

Electromigration also frequently occurs in the presence of nonuniform temperature distributions which develop at various sites within device structures, e.g., at locations of poor film adhesion, in regions of different thermal conductivity such as metal–semiconductor contacts or interconnect–dielectric crossovers, at nonuniformly covered steps, and at terminals of increased cross section. In addition to the influence of microstructure noted earlier, there is the added complication of thermal gradients. The resulting damage pattern can be understood by considering the second term on the right-hand side of Eq. 11-40. For the polarity shown, all terms in parentheses are positive and $Cq\rho j/RT$ is roughly temperature independent, whereas DZ^* increases with temperature. Therefore, $\partial C/\partial t$ varies as $-\partial T/\partial x$. Voids will thus form ($\partial C/\partial t < 0$) at the negative electrode where $\partial T/\partial x > 0$, and hillocks will grow at the positive electrode because $\partial T/\partial x < 0$. Physically, the drift velocity of atoms at the cathode increases as they experience a rising temperature. More atoms then exit the region than flow into it. At the anode the atoms decelerate in experiencing lower temperatures and thus pile up there. An analogy to this situation is a narrow strip of road leading into a wide highway (at the cathode). The bottleneck is relieved and the intercar spacing increases. If further on down the highway it again narrows to a road, a new bottleneck reforms and cars will pile up (at the anode).

11.6.3.4 Effect of Film Stress on Electromigration

Film stress has an interesting effect on electromigration. As already noted, when long Al conductor stripes are powered mass that is depleted at

the cathode accumulates at the anode. However, this manifestation of mass transport is reduced when the conductor length (L) is shortened and completely halted when L is less than a critical value L_c, known as the Blech length. Two opposing mass-transport driving forces are at work here, one due to electromigration (Eq. 11-38) and one due to a stress (σ) gradient along the conductor which physically biases atomic motion from compressive to tensile regions. The latter force is given by $F = \Omega\, d\sigma/dx$, where Ω is the atomic volume. When the two forces balance,

$$L_c = \Omega\sigma_c/Z^*q\rho j(c) \tag{11-41}$$

where critical values are denoted by c. Thus shorter interconnects can withstand higher current densities.

11.6.3.5 Electromigration Failure and Materials Issues

Analysis of extensive accelerated testing that has been performed to characterize electromigration damage in interconnections has led to a general relationship between film mean time to failure (MTTF) and j given by

$$(\mathrm{MTTF})^{-1} = K \exp(-E_e/RT) j^n. \tag{11-42}$$

As with virtually all mass-transport-related reliability problems, damage is thermally activated. For pure Al conductors n is typically 2, and E_e, the activation energy for electromigration failure, ranges from 0.5 to 0.8 eV depending on grain size. In contrast, an energy of 1.4 eV is associated with lattice diffusion. Therefore, low-temperature electromigration damage in films is clearly dominated by GB transport. The constant K is dependent on film structure and processing. In order to improve the life of Al metallizations more electromigration-resistant alloys have been developed over the years. In this regard, Al alloyed with a few percent Cu can extend conductor life by perhaps an order of magnitude relative to pure Al. The reasons for this are not completely understood, but it appears that Cu reduces the GB migration of the solvent Al (see Section 11.3.2.3). The higher observed values for E_e are consistent with such an interpretation.

The latest development has been the introduction of copper metallizations to replace aluminum. Higher conductivity and shorter signal delay times are the primary reasons for the change. Besides, Cu also promises to be less prone to electromigration damage simply because its higher melting point means reduced atomic diffusivity. Adherence to design rules (e.g., $j < 10^5$ A/cm^2) has largely eliminated electromigration damage.

11.7 CONCLUSION

Mass-transport phenomena in thin films can be largely understood by extrapolating bulk material behavior into regimes of small grain size and low temperature. With appropriate modifications for the diffusional paths involved, $x^2 \simeq 4D_{0i} \exp -(E_i/RT)t$ roughly connects the variables of interest. Grain boundaries exert the overriding influence on atom movements in polycrystalline films by making available abundant defects and short-circuit diffusion paths to inhabit. As a result of reduced activation energies, barriers to transport can be readily surmounted at low temperatures. In virtually all applications, such diffusional effects are viewed with alarm because they threaten film adhesion, promote compositional change, alter stress, and generally lead to film degradation. In these areas metals exhibit the worst behavior. Metal films are reactive, generally polycrystalline, and often present in contacting layers that function at relatively high ratios of use to melting-point temperature (T/T_M). That is why most of the research on mass transport and reactions in films has focused on metals. In contrast, single-crystal semiconductors do not ordinarily suffer low-temperature diffusional effects. As for amorphous insulators (oxides, nitrides, etc.), T/T_M is usually too low for extensive atomic mobility, despite extremely defective matrices.

The overwhelming majority of thin-film metal–metal and metal–semiconductor systems of technological interest have phase diagrams that often contain several intermetallic compounds. When such thin-film couples are reacted, compound layers grow at generally low temperatures with low activation energies, and usually with parabolic kinetics. Unlike bulk couples where all compound phases appear simultaneously and in a generally predictable order, this is not true of film systems. Like bulk materials, thermodynamically unstable homogeneous thin films undergo phase transformations. The transformation of amorphous to crystalline $CoSi_2$ illustrates how our traditional understanding of such processes applies.

By their nature thin films enable mass transport phenomena to occur under very large driving forces that are normally difficult to apply in bulk. Diffusion between multilayers containing large concentration gradients, and electromigration in metals exposed to large electric fields, are examples.

EXERCISES

1. At what temperature will the number of Au atoms transported through grain boundaries equal that which diffuses through the lattice if the grain size is 2 μm? 20 μm?

2. A p–n junction is produced by diffusing phosphorus from a continuous source ($C_0 = 10^{19}\ \mathrm{cm}^{-3}$) into an epitaxial Si film with a background boron level of $10^{15}\ \mathrm{cm}^{-3}$. Diffusion is carried out at 1100°C for 30 min.
   (a) How far beneath the Si surface is the junction (i.e., where $C(n) = C(p)$)?
   (b) If there is a 1% error in the diffusion temperature, what is the percent change in junction depth?
   (c) By what percent will the junction depth change for a 1% error in diffusion time?

3. The equation due to Fisher for transport of atoms down a single grain boundary where there is simultaneous diffusion into the adjoining grains is

$$\frac{\partial C_{\mathrm{GB}}(y,t)}{\partial t} = D_{\mathrm{GB}}\frac{\partial C_{\mathrm{GB}}(y,t)}{\partial y^2} - \frac{2D_{\mathrm{L}}}{\delta}\frac{dC_{\mathrm{L}}(x,y,t)}{dx}\bigg|_{x=\delta/2},$$

   where GB and L refer to grain boundary and lattice quantities. Derive this equation by considering diffusional transport into and out of an element of grain boundary δ wide and dy long. *Note:* $C_{\mathrm{L}}(x, y, t)$ refers to the lattice concentration.

4. (a) Calculate the activation energy for dislocation-pipe diffusion of Au in epitaxial Au films from the data of Fig. 11-7a.
   (b) Calculate the activation energy for grain-boundary diffusion of Au in polycrystalline Au films from the data of Fig. 11-7b.
   In both cases make Arrhenius plots of the diffusivity data. Assume

$$D_{\mathrm{L}} = 0.091 \exp -\left[\frac{41.7\ (\mathrm{kcal/mol})}{RT}\right]\ \mathrm{cm^2/s}.$$

5. It is desired to make polycrystalline-silicon solar cells by doping p-type wafers composed of columnar grains. Consider a single grain boundary oriented normal to the wafer surface, through which n-type dopant diffuses from a continuous source of concentration C_0. If the background dopant level in the wafer is C_{p}, write an expression for the resulting p–n junction profile (y vs x) after diffusion.

6. In the Au–Pd thin-film diffusion couple of Fig. 11-12, an approximate fit to the data can be made using the equation

$$C(x,t) = C_0/2\ \mathrm{erfc}[x/2(Dt)^{1/2}].$$

   (a) Plot $C(x, t)$ vs x.
   (b) From values of dC/dx at $x = 0$, estimate the values of D for the 0 h, 20 h, and 200 h data. Are these D values the same?

(c) What accounts for the apparent interdiffusion between Au and Pd at 0 h?

7. Film A of thickness a lies atop a substrate, while film B of thickness b lies above film A. During interdiffusion in this thin-film couple atoms originating from film A permeate the B film and increase their concentration $C(\mathrm{A})$ on the outer surface $(x = a + b)$ with time (t). The following conditions hold: (1) Initially both A and B films are pure elements. (2) The diffusivities of A in B and B in A are equal and independent of composition. (3) No mass flows normal to the bounding film surfaces, i.e., at the substrate interface, $dC(x = 0, t)/dx = 0$, and at the top surface, $dC(x = a + b, t)/dx = 0$. Show that if $C(\mathrm{A})_0(=1)$ is the initial concentration of A in the A layer, the time-dependent solution for $C(\mathrm{B})$ is

$$C(\mathrm{B}) \text{ at } (x=a+b) = \frac{aC(\mathrm{A})_0}{a+b} + \frac{2C(\mathrm{A})_0}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}(-1)^n \sin\left(\frac{n\pi a}{a+b}\right)\exp-\left(\frac{n^2\pi^2 Dt}{(a+b)^2}\right).$$

8. It is known that the energy of low-angle grain boundaries depends on the angle of misorientation (θ) as $\gamma_{\mathrm{GB}} = A\theta(B - \ln\theta)$ where A and B are constants.
   (a) Derive a relation between grain boundary diffusivity and θ.
   (b) Schematically plot D_{GB} vs θ.

9. Values for the parabolic (B) and linear (B/A) rate constants for both dry and wet SiO_2 growth reveal thermally-activated temperature dependencies, i.e.,

$$\text{Dry: } B = 950 \exp-(1.24\,\mathrm{eV}/k_{\mathrm{B}}T)\ \mu\mathrm{m}^2/\mathrm{h},$$

$$B/A = 5.89 \times 10^6 \exp-(2.0\,\mathrm{eV}/k_{\mathrm{B}}T)\ \mu\mathrm{m/h}$$

$$\text{Wet: } B = 186 \exp-(0.71\ \mathrm{eV}/k_{\mathrm{B}}T)\ \mu\mathrm{m}^2/\mathrm{h},$$

$$B/A = 7.31 \times 10^7 \exp-(1.96\,\mathrm{eV}/k_{\mathrm{B}}T)\ \mu\mathrm{m/h}$$

where all constants are normalized to 760 torr.
   (a) Why do the activation energies for B differ while activation energies for B/A are essentially the same?
   (b) Wet oxidation is more rapid than dry oxidation. Why?
   (c) An oxide-free Si wafer is oxidized at 1200°C for an hour in dry oxygen. How long would a wafer have to be oxidized in a wet atmosphere to produce an SiO_2 film that is as thick as the dry oxide film? What is the oxide thickness?

10. Thermodynamic data in the Ni–Si system reveals the following ΔH_f values:

Compound	ΔH_f (kJ/mol-at)	Compound	ΔH_f (kJ/mol-at)
Ni_2Si	−47.1	Ni_3Si	−37.5
NiSi	−42.6	Ni_5Si_2	−42.5
$NiSi_2$	−29.2	Ni_3Si_2	−45

Predict the first intermetallic compound to form when Ni and Si are reacted.

11. A 100 nm thick film of cobalt is deposited on a silicon wafer and totally reacts to form assorted silicides.
    (a) If Co_2Si forms, what thickness of silicon is consumed? What is the height of this silicide above (below) the original wafer surface?
    (b) Suppose $CoSi_2$ forms. What thickness of silicon is consumed? Determine the height of this silicide above (below) the original wafer surface.
    (c) Similarly, 100 nm of titanium is deposited on a silicon wafer and reacted to form TiSi and $TiSi_2$. For each silicide determine the thickness of silicon consumed and the height of the silicide above (below) the original wafer surface. Note the following formula densities in units of 10^{22} cm^{-3}: Si = 4.99, Co = 8.95, Co_2Si = 3.06, $CoSi_2$ = 2.59, Ti = 5.5, TiSi = 3.44, $TiSi_2$ = 2.32.

12. Isothermal transformation curves for the fraction of $TiSi_2$ (C49) converted to $TiSi_2$(C54) at 720°C followed Avrami kinetics. It was reported that:
    1. Blanket 32 nm thick films transformed more rapidly than either patterned 2.0 μm and 0.7 μm wide lines fabricated from the same film.
    2. n values for the blanket, 2.0 μm wide and 0.7 μm wide films were 2.43, 1.67, and 1.33, respectively.
    3. The activation energy for nucleation was ~4 eV, independent of line width.
    4. Narrower lines had to be heated to higher temperatures to transform them.

    Give plausible explanations for as many of these findings as you can. *Hint:* Consult Ref. 54.

13. A 2500 Å thick amorphous (a) silicon film is deposited on a crystalline (c) silicon wafer.
    (a) When the surface is analyzed with 2.0 MeV $^4He^+$ ions, schematically depict the features of the resultant RBS spectrum.

(b) The a-Si film surface is recrystallized by laser light. Sketch the resultant RBS spectrum after $\frac{1}{4}$ of the a-Si film has epitaxially regrown. Repeat for the case where $\frac{3}{4}$ of the a-Si film has regrown.

14. By replotting the data from Fig. 11-24 determine the time dependence of Ru_2Si_3 film growth. Make any appropriate assumptions.

15. For electromigration in Al stripes, assume $E_e = 0.7$ eV and $n = 2.5$ in Eq. 11-42. By what factor is MTTF shortened (or extended) at 40°C by
 (a) A change in E_e to 0.6 eV?
 (b) No change in E_e but a temperature increase to 85°C?
 (c) A decrease in stripe thickness at a step from 1.0 to 0.75 μm?
 (d) An increase in current from 1 to 1.5 mA?

16. (a) When there are simultaneous electromigration and diffusional fluxes of atoms, show that if the velocity of ion migration is v,

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - \frac{v\partial C}{\partial x}.$$

 (b) For a diffusion couple ($C = C_0$ for $x < 0$, and $C = 0$ for $x > 0$) show that

$$C(x, t) = \frac{C_0}{2}\left(\exp\frac{vx}{D}\,\text{erfc}\frac{x + vt}{\sqrt{4Dt}} + \text{erfc}\frac{x - vt}{\sqrt{4Dt}}\right)$$

 is a solution to the equation in part (a) and satisfies the boundary conditions.

 (c) A homogeneous Al film stripe is alloyed with a cross stripe of Cu creating two interfaces: (+) Al–Cu/Al and (−) Al/Al–Cu. Using the preceding solution, show that the concentration profiles that develop at these interfaces obey the relation

$$\ln\{C_+(x, t)/C_-(x, t)\} = vx/D.$$

REFERENCES

1. M.-A. Nicolet and M. Bartur, *J. Vac. Sci. Technol.* **19**, 786 (1981).
2. R. W. Balluffi and J. M. Blakely, *Thin Solid Films* **25**, 363 (1975).
3. N. A. Gjostein, *Diffusion*. American Society for Metals, Metals Park, OH, 1973.
4. J. C. C. Tsai, in *VLSI Technology*, 2nd ed., ed. S. M. Sze. McGraw-Hill, New York, 1988.
5. S. M. Sze, *Semiconductor Devices—Physics and Technology*. Wiley, New York, 1985.
6. D. C. Jacobson, "Ion-Beam Studies of Noble Metal Diffusion in Amorphous Silicon Layers," Ph.D Thesis, Stevens Institute of Technology (1989).
7. J. C. Fisher, *J. Appl. Phys.* **22**, 74 (1951).
8. R. T. Whipple, *Phil. Mag.* **45**, 1225 (1954).
9. T. Suzuoka, *Trans. Jap. Inst. Met.* **2**, 25 (1961).

10. L. G. Harrison, *Trans. Faraday. Soc.* **57**, 1191 (1961).
11. J. M. Poate, K. N. Tu, and J. W. Mayer, eds., *Thin Films — Interdiffusion and Reactions*, John Wiley & Sons, New York, 1978.
12. D. Gupta and P. S. Ho, eds., *Diffusion Phenomena in Thin Films and Microelectronic Materials*. Noyes, Park Ridge, NJ, 1988.
13. K.-N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science for Electrical Engineers and Materials Scientists*. Macmillan, New York, 1992.
14. D. Gupta, in *Diffusion Phenomena in Thin Films and Microelectronic Materials*, eds. D. Gupta and P. S. Ho. Noyes, Park Ridge, NJ, 1988.
15. D. Gupta, *Phys. Rev.* **7**, 586 (1973).
16. D. Gupta and K. W. Asai, *Thin Solid Films* **22**, 121 (1974).
17. U. Gösele and K. N. Tu, *J. Appl. Phys.* **66**, 2619 (1989).
18. A. S. Grove, *Physics and Technology of Semiconductor Devices*. John Wiley & Sons, New York, 1967.
19. F. P. Fehlner, *Low-Temperature Oxidation: The Role of Vitreous Oxides*. Wiley-Interscience, New York, 1986.
20. D. A. Buchanan, *IBM J. Res. Develop.* **43**(3), 243 (1999).
21. K. N. Tu, *Ann. Rev. Mater. Sci.* **15**, 147 (1985).
22. P. M. Hall, J. M. Morabito, and J. M. Poate, *Thin Solid Films* **33**, 107 (1976).
23. D. Gupta and P. S. Ho, *Thin Solid Films* **72**, 399 (1985).
24. J. C. M. Hwang, J. D. Pan, and R. W. Balluffi, *J. Appl. Phys.* **50**, 1349 (1979).
25. A. J. Kellock, J. E. E. Baglin, K. R. Coffey, J. K. Howard, M. A. Parker, and D. L. Neiman, *Mat. Res. Soc. Symp. Proc.* **343**, 211 (1994).
26. M. Gall, J. G. Pellerin, P. S. Ho, K. R. Coffey, and J. K. Howard, *Mat. Res. Soc. Symp. Proc.* **343**, 217 (1994).
27. V. T. Borisov, V. M. Golikov, and G. V. Scherbedinskiy, *Phys. Metals. Metallogr.* **17**, 80 (1964).
28. P. Lejcek and S. Hofmann, *Crit. Rev. Solid State Mater. Sci.* **20**(1), 1 (1995).
29. D. McLean, *Grain Boundaries in Metals*. Oxford, London, 1957.
30. E. D. Hondros, *J. de Phys.* **36**, Coll-C4, 117 (1975).
31. C.-A. Chang, in *Diffusion Phenomena in Thin Films and Microelectronic Materials*, eds. D. Gupta and P. S. Ho. Noyes, Park Ridge, N.J., 1988.
32. S. U. Campisano, G. Foti, R. Rimini, and J. W. Mayer, *Phil. Mag.* **31**, 903 (1975).
33. Z. Marinkovic and V. Simic, *Thin Solid Films* **217**, 26 (1992).
34. C. Michaelsen, K. Barmak, and T. P. Weihs, *J. Phys. D. Appl. Phys.* **30**, 3167 (1997).
35. R. M. Walser and R. W. Bene, *J. Appl. Phys. Lett.* **28**, 624 (1976).
36. L. Zhang and D. G. Ivey, *Mat. Res. Soc. Symp. Proc.* **311**, 299 (1993).
37. R. Pretorius, C. C. Theron, A. Vantomme, and J. W. Mayer, *Crit. Rev. Solid State Mater. Sci.* **24**(1), 1 (1999).
38. K. N. Tu, D. A. Smith, and B. Z. Weiss, *Phys. Rev. B* **36**, 8948 (1987).
39. L. A. Clevinger, Q. Z. Hong, R. Mann, J. M. E. Harper, K. Barmak, C. Cabral, Jr., C. Nobili, and G. Ottaviani, *Mat. Res. Soc. Symp. Proc.* **311**, 253 (1993).
40. M. Libera, *Mat. Res. Soc. Symp. Proc.* **343**, 89 (1994).
41. G. L. Olsen and J. A. Roth, *Mater. Sci. Rep.* **3**, 1 (1988).
42. H. P. Kattelus and M.-A. Nicolet, in *Diffusion Phenomena in Thin Films and Microelectronic Materials*, eds. D. Gupta and P. S. Ho. Noyes, Park Ridge, NJ, 1988.
43. Y. Pauleau, in *Multicomponent and Multilayered Thin Films for Advanced Microtechnologies: Techniques, Fundamentals and Devices*, eds. O. Auciello and J. Engemann. Kluwer, Dordrecht, The Netherlands, 1993.
44. M.-A. Nicolet, *Thin Solid Films* **52**, 415 (1978).

45. M. Wittmer, *J. Vac. Sci. Technol.* **A2**, 273 (1984).
46. T. Castan, J-L. Peyre, and K. Beschorner, *Solid State Technol.* **41**(6), 127 (1998).
47. T.-J. Kim and P. H. Holloway, *Crit. Rev. Solid State Mater. Sci.* **22**(3), 239 (1997).
48. F. M. d'Heurle, *J. Mater. Res.* **3**, 167 (1988).
49. F. M. d'Heurle and P. Gas, *J. Mater. Res.* **1**, 205 (1986).
50. K. N. Tu, W. K. Chu, and J. W. Mayer, *Thin Solid Films* **25**, 403 (1975).
51. K. Maex and M. V. Rossum, *Properties of Metal Silicides.* INSPEC IEEE, London, 1996.
52. S. P. Murarka, *Metallization Theory and Practice for VLSI and ULSI.* Butterworth–Heinemann, Boston, 1993.
53. R. W. Mann, L. A. Clevenger, P. D. Agnello, and F. R. White, *IBM J. Res. Develop.* **39**, 403 (1995).
54. J. L. Jordan-Sweet, *IBM J. Res. Develop.* **44**, 457 (2000).
55. D. A. Porter and K. E. Easterling, *Phase Transformations in Metals and Alloys*, 2nd ed. Chapman and Hall, London, 1992.
56. A. L. Greer, in *Diffusion Phenomena in Thin Films and Microelectronic Materials*, eds. D. Gupta and P. S. Ho. Noyes, Park Ridge, NJ, 1988.
57. S. Prokes and F. Spaepen, *Appl. Phys. Lett.* **47**, 234 (1985).
58. M. Ohring, *Failure and Reliability of Electronic Materials and Devices*, Academic Press, Boston, 1998.

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

Chapter 12

Mechanical Properties of Thin Films

12.1 INTRODUCTION

Interest in mechanical-property effects in thin films has traditionally focused on two broad issues. The primary one has been the deleterious effects that stresses cause in films. This has prompted much research to determine the type, magnitude, and origin of stress as well as means of minimizing or controlling stresses. A second major issue concerns enhancing the hardness and wear resistance of assorted coatings. Of the two issues, the first has historically generated the greatest scientific attention and is the one primarily addressed in this chapter. On the other hand, the mechanical properties of metallurgical and protective coatings are important technologically and are the subject of an extensive literature.

It is virtually always the case that stresses are present in thin films. What must be appreciated is that stresses exist even though films are not externally loaded. Such residual or internal stresses directly affect a variety of phenomena including adhesion, generation of crystalline defects, perfection of epitaxial layers and formation of film surface growths such as hillocks and whiskers. Since they tend to increase with thickness, promoting film peeling, stresses are a prime limitation to the growth of very thick films. Substrate deformation and distortion also necessarily arise from stresses in the overlying films. For most applications this is not troublesome because

substrates are usually relatively massive compared to films. However, in integrated circuit technology, even slight bowing of silicon wafers presents significant problems with regard to maintaining precise tolerances in the lithographic definition of device features. And as devices continue to shrink in size, stress effects in increasingly smaller regions must be understood, measured, and controlled. These significant challenges have spurred much thin film stress research. Film stresses additionally influence electronic properties, e.g., semiconductor bandgap shifts, superconductor transition temperatures, and magnetic anisotropy.

In the packaging and attachment of semiconductor chips another important collection of structural-mechanics applications has emerged (Ref. 1). The involved components are small. Although they are larger and thicker than thin films, they are vastly smaller than conventional engineering structures. When these tiny components are subjected to thermal stresses during processing and service, creep, fatigue, and cracking are issues of primary concern. Whatever understanding we possess of the mechanical behavior and structural mechanics of electronic-packaging materials, i.e., metals (e.g., Pb–Sn solder, Al), ceramics (e.g., Al_2O_3, SiO_2), semiconductors (e.g., Si, GaAs), and polymers (e.g., circuit boards, epoxies), and combinations of involved components, e.g., solder bumps and joints, chip-bonding pads, die supports, and encapsulants, has evolved from well-established theories and models of macroscopic elastic and plastic phenomena.

Similarly, the varied mechanical phenomena thin films and multilayered structures display also span both the elastic and plastic realms of behavior. The chapter starts by noting how the important mechanical properties of films such as Young's modulus (Y), yield stress (σ_0), ultimate tensile stress (σ_{UTS}), and hardness (H) are derived from the response to *applied* stresses. Next the discussion will shift to *internal* stress, its origins, its magnitudes, and the methods employed to measure it. Internal stress is a subject largely rooted in concepts of elasticity. However, when internal stresses are relaxed, plastic deformation effects are often manifested. The resulting stress relaxation has important implications with regard to processing, the quality of epitaxy, and the properties of films and coatings that undergo large temperature excursions. Reserved for the last topic in the book is film adhesion, a very important subject since films cannot be practically exploited unless they adhere to substrates.

Many of the topics treated in this chapter are addressed more fully in a number of reviews (Refs. 2–6) dealing with mechanical behavior of thin films. Although the reader will find them all useful, the one by Nix is particularly recommended.

12.2 MECHANICAL TESTING AND STRENGTH OF THIN FILMS

12.2.1 SCOPE

In an effort to determine their mechanical properties, in particular their strength, a number of methods (Ref. 7) to test thin films have evolved over the years. The more popular of these are schematically depicted in Fig. 12-1. In all cases, mechanical forces are applied to films that are either free-standing (no substrate) or attached to a more massive substrate. Obviously, the interpretation of mechanical properties in free-standing films is less ambiguous than when the influence of the attached substrate and questions of adhesion complicate matters. Each test method is capable of yielding values for Young's modulus (Y) and strength but sometimes, as we shall see,

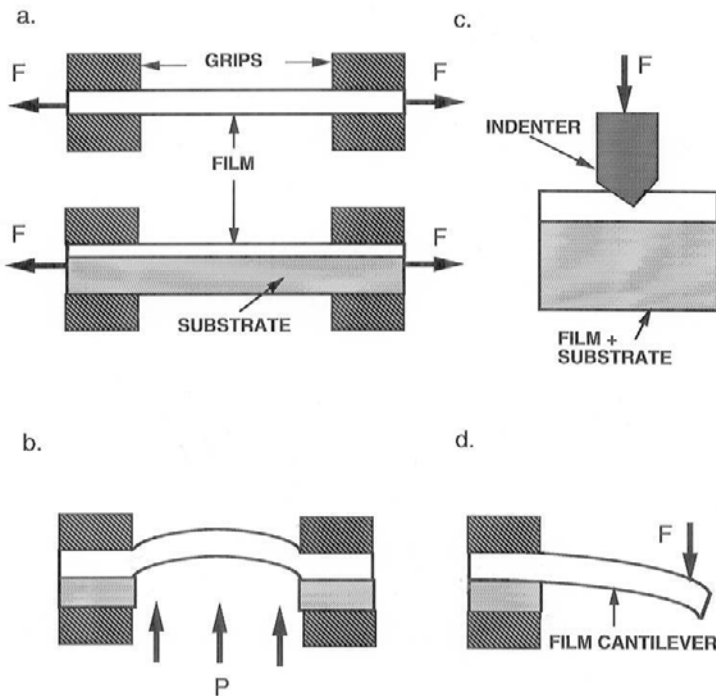


Figure 12-1 Methods for mechanical testing of thin films: (a) tensile testing, (b) bulge testing, (c) indentation (micro or nano) hardness testing, (d) deflection of microbeams. (Adapted from Ref. 7.)

with considerable scatter and questions as to range of validity. In this section the testing methods are individually described to compare their relative merits and assess the information gained. Note that the testing methods described next are quite distinct from those used to determine residual stress in films.

12.2.2 TENSILE TESTING

The tensile test is very widely used to evaluate both the elastic and plastic response of bulk materials. Although direct tensile tests of free-standing metal films are not conducted with any frequency today, past measurements are interesting because of the basic information they conveyed on the nature of deformation processes. Many results and interpretations of this earlier work can be found in the review by Hoffman (Ref. 8). Unlike its bulk counterpart, tensile testing of thin films is far from routine. The extreme delicacy required in handling and detaching thin films from substrates, gripping and aligning them, applying load, and measuring the mechanical response are some of the experimental challenges. Loading is commonly achieved by electromagnetic force transducers, and strains are typically measured by optical methods. Novel microtensile testing devices have been incorporated within electron microscopes, enabling direct observation of defects and recording of diffraction patterns during straining.

Tensile data obtained in thin films deposited on thicker substrates which are then gripped and loaded are only interpretable if the strains are fully transferred from substrate to film. X-ray and optical methods are available to determine whether the film strain continuously matches that of the substrate. Since each set of crystal planes in a film essentially acts like an *in situ* strain gauge, tensile data can be unfolded by X-ray diffraction methods using Bragg's law. However, X-ray-strain determinations have limited utility because only the elastic response (a sum of applied plus residual strains) is measurable. Tensile testing employing such X-ray methods has been reviewed by Noyan *et al.* (Ref. 9).

A typical stress–strain curve for Au is shown in Fig. 12-2a (Ref. 10). The maximum tensile strength, or stress at fracture in this case, is only somewhat higher than typical intrinsic or residual stress levels about which we will speak later. At fracture the strain is only 0.8%, a value well over an order of magnitude smaller than that observed in bulk Au. Loading and unloading curves of varying slope (or modulus of elasticity) have raised questions as to whether E differs from the bulk value and whether it is film-thickness dependent. On both accounts experimental data indicate no abnormal effects.

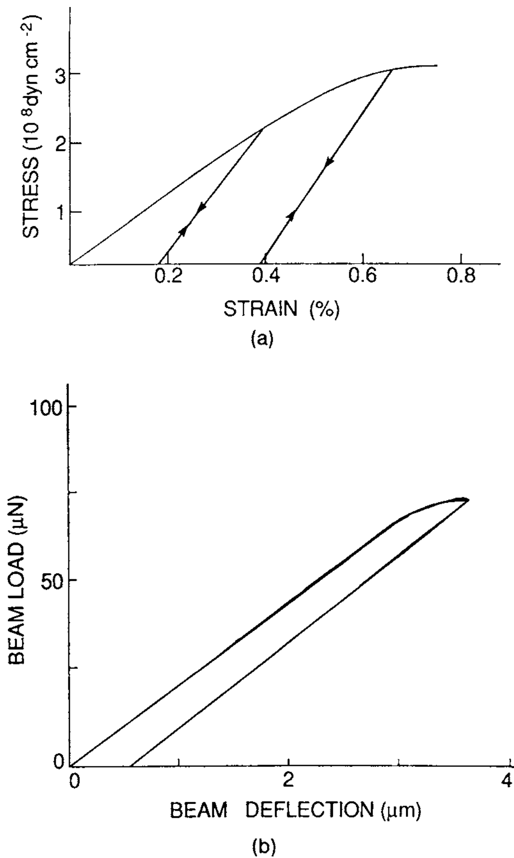


Figure 12-2 (a) Stress–strain behavior for an Au film. (From Ref. 10.) (b) Load-deflection behavior for a 0.87 μm thick Au cantilever film. (From Ref. 2.)

Unlike bulk metals, there is a limited regime of easy dislocation glide in films because deposited polycrystalline metal films already contain initial dislocation densities of 10^{10} to $10^{12}\,\text{cm}^{-2}$. Such values are comparable to those found in heavily worked and strain-hardened bulk metals. That is why maximum tensile strengths, often reported in terms of shear modulus μ, are very high, e.g., σ_{UTS} for FCC metal films span a range from $\mu/40$ to $\mu/120$. These surprisingly high strengths exceed those for hard-drawn metals and may be almost 100 times that for annealed bulk metals. Polycrystal metal films are stronger than single-crystal films, reflecting the role of grain boundaries as obstacles to dislocation motion. The highest strengths obtained are close to those theoretically predicted assuming deformation by

rigid lattice displacements. This means that the high initial dislocation densities leave few avenues available for either generation or motion of new dislocations.

12.2.3 Bulge Testing

Bulge testing is widely used to determine the mechanical properties of thin films and membranes. In this test the film/substrate assembly is sealed to the end of a hollow cylindrical tube and pressurized with gas. Then the maximum height of the resulting hemispherical bulge in the film is then measured optically with a microscope or interferometer and converted to strain. The relationship between the dome height (h) and applied differential pressure (P) has been determined to be (Ref. 8)

$$P = 4dh/r^2[\sigma_i + 2Yh^2/3(1 - v)r^2], \qquad (12\text{-}1)$$

where the film thickness and specimen radius are d and r, respectively; σ_i is the residual stress in the film under a zero pressure differential. To illustrate bulge testing and how data are analyzed, consider the results for epitaxial and polycrystalline Si membranes (Ref. 12). These materials are candidates for X-ray lithography mask substrates, an application requiring high fracture strength and excellent dimensional stability. Free-standing membranes 1 μm thick and 38 mm in diameter, supported by a thicker substrate ring of either Si or SiO_2, were prepared by selective masking and etching methods.

The pressure–deflection response for the polycrystalline membrane is shown in Fig. 12-3a. An excellent fit of the data to Eq. 12-1 is evident, enabling values of σ_i and $Y/1 - v$ to be determined. It should be noted that despite the nonlinear membrane deflection with pressure, the response is actually elastic and not plastic. Repeated pressurization cycles did not result in any appreciable deterioration of reproducibility in mechanical response. In Fig. 12-3b the membrane is stressed to failure at higher pressure levels. The fracture stress is given by $\sigma_f = P_f r^2/4dh_f$, where P_f and h_f are the values at fracture.

For epitaxial Si it was found that $Y = 181$ GPa, $\sigma_f = 190$ MPa, $\sigma_i = 52$ MPa; on the other hand for polycrystalline Si, $Y = 250$ GPa, $\sigma_f = 420$ MPa, $\sigma_i = 170$ MPa. In general, accurate and reproducible property values derived from bulge testing require the avoidance of film pinholes and other defects.

12.2.4 Nanoindentation Testing

The testing techniques discussed below were largely developed in response to a number of challenges posed by assorted technologies including

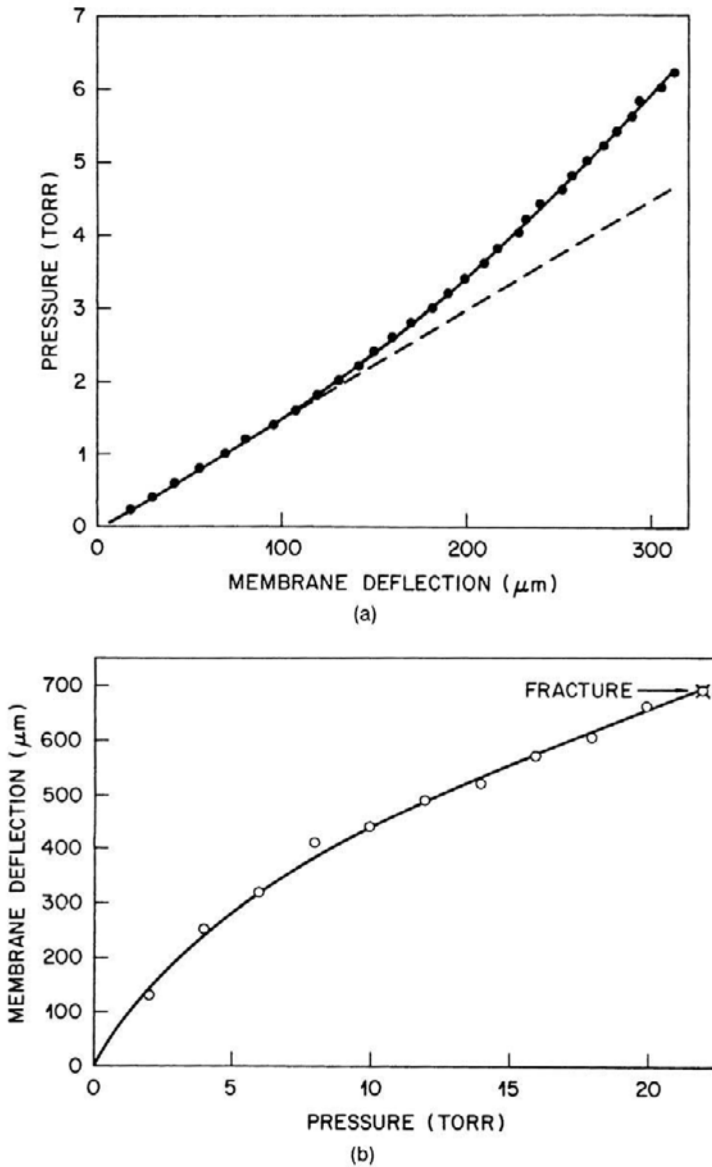


Figure 12-3 (a) Pressure–membrane deflection characteristics of a 1 μm thick poly-Si membrane. (Solid curve is Eq. 12-1. Dashed curve represents $P = 4dh\sigma_i/r_0^2$.) (b) Membrane deflection vs differential pressure measured to the point of fracture. (From Ref. 11.)

microelectronics, micromechanical device fabrication, and hard as well as optical coating production (Ref. 7). In microelectronics, for example, mechanical-property information is often required at varied locations on the surface of patterned films. Not only are the films thin and feature sizes small, it is impractical to detach the films from their substrates. Nanoindentation testing allows mechanical probing of film specimens having characteristic dimensions that are typically 1 μm in size. Testing is accomplished with the Nanoindentor, a commercially available, load-controlled submicron indentation instrument. The Berkovich diamond indenter possessing a three-sided pyramidal geometry is most frequently employed. Load is applied with a coil and magnet assembly while displacement is usually measured by capacitance sensors. Resolutions better than 0.01 μN for load and 0.01 nm for displacement have been claimed in the better instruments. Because of the very small volume sampled ($\sim 1\ \mu m^3$), the technique may be regarded as a mechanical properties microprobe by analogy to chemical microprobes, e.g., AES, SIMS. Nanoindentation testing has been primarily applied to the determination of nanohardness and secondarily to the measurement of microbeam deflections. These applications are discussed in turn.

12.2.4.1 Nanohardness

Through simultaneous measurement of load (F) and displacement (h), the very local response of a film to indentation shown in Fig. 12-4 is typically obtained for both loading and unloading. Two quantities can be extracted from these curves. Once the contact area (A) is measured, the hardness (H) is defined as $H = F_{max}/A$ where F_{max} is the maximum force applied. (See Eqs. 1-46, 1-47 for comparison). The initial slope of the unloading curve yields a quantity defined as the contact stiffness S ($S = dF/dh$). S is generally related in complex ways to the elastic moduli of both the film and indenter as well as indentation area and film thickness. For an axisymmetric indenter, $S = (2/\pi^{1/2})\langle Y \rangle A^{1/2}$, where $\langle Y \rangle$ is a weighted average of the film and indenter moduli.

An issue that always arises in the hardness testing of thin films and coatings is the influence of the substrate. As a rule of thumb, when the depth of the indentation exceeds 10 to 25% of the film thickness, substrate effects must be considered. One set of analytic formulas proposed for the apparent hardness (H) in terms of the film (f) and substrate (s) hardnesses, yield stresses, and elastic moduli takes the form (Ref. 7)

$$H/H(\mathrm{s}) = 1 + \{H(\mathrm{f})/H(\mathrm{s}) - 1\} \exp[-\xi]. \qquad (12\text{-}2)$$

For soft films on hard substrates, $\xi = [\sigma_0(\mathrm{f})Y(\mathrm{s})/\sigma_0(\mathrm{s})Y(\mathrm{f})](h/d)^2$, while for hard films on soft substrates $\xi = H(\mathrm{f})/H(\mathrm{s})[\sigma_0(\mathrm{s})Y(\mathrm{s})^{1/2}/\sigma_0(\mathrm{f})Y(\mathrm{f})^{1/2}](h/d)$,

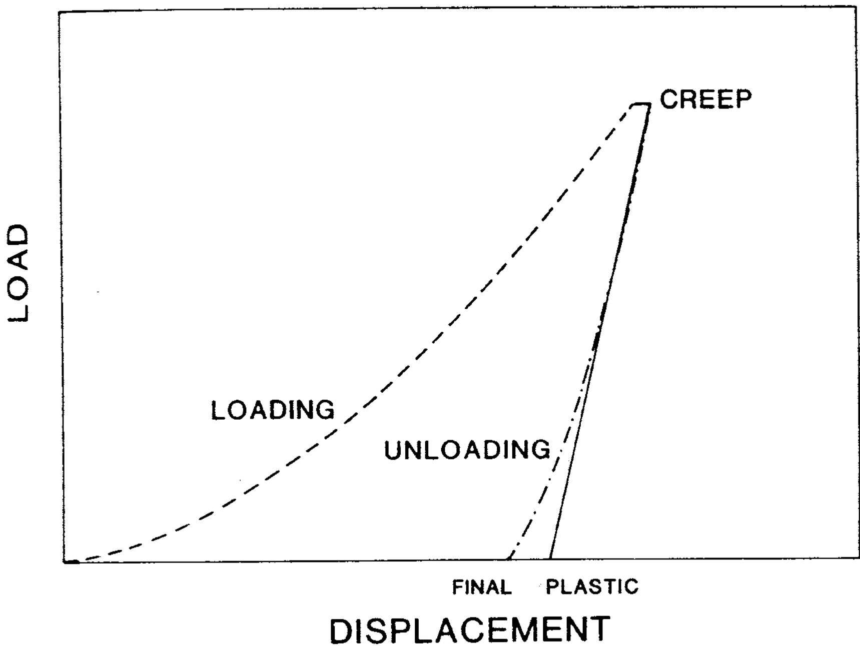


Figure 12-4 Typical indentation load–displacement curve during loading and unloading showing difference between the initial plastic and final depths. (After Ref. 2. Reprinted with the permission of Professor W. D. Nix.)

where d is the film thickness. These relations have been applied with mixed success; clearly more theoretical and experimental work is required to firmly establish their applicability.

12.2.4.2 Microbeam Testing

In a very popular variant of nanoindentation, indenter induced deflection of cantilever microbeams is measured. The microbeams are fabricated from Si wafers employing micromachining techniques involving standard photolithographic and etching processes borrowed from microelectronics technology. Beam dimensions, typically 1 μm thick, 20 μm wide and 100 μm long, are larger than microelectronic features, but significantly smaller than what can be machined or fabricated by traditional methods. By evaporating or growing films on the Si substrate, then etching the latter away, the technique can be extended to fabricating microbeams of assorted film materials. Because of their tiny size, microbeams have also been micromachined at different locations of wafers to locally monitor stress variations in subse-

quently deposited thin films (Ref. 12).

A typical beam–load deflection curve for a Au microbeam (Ref. 13) shown in Fig. 12-2b resembles the stress–strain curve of free-standing films. During loading, the linear elastic, as well as nonlinear, strain-hardening plastic-regimes are observed. Formulae for the elastic deflection (δ), i.e.,

$$\delta = 4F(1 - v^2)L^3/wd^3 Y, \tag{12-3}$$

and yield stress, i.e.,

$$\sigma_0 = 6LF_0/wd^2, \tag{12-4}$$

of the beam enable Y and σ_0 to be extracted from the data. Here L, w, and d are the beam length, width, and thickness, while F_0 is the load marking deviation from linearity in the loading response.

In a clever extension of the foregoing testing, the cantilever microbeams were electrostatically excited into transverse vibration and resonant frequencies (ω) were measured by detecting the movement of a reflected laser beam (Ref. 14). From such dynamic testing elastic moduli were extracted according to the relation

$$\omega = 0.162\, d/L^2(Y/\rho)^{1/2}, \tag{12-5}$$

where ρ is the mass density of the beam.

To shed some perspective on Young's modulus determinations in thin films it is interesting to compare results of testing aluminum and silicon films. Free-standing Al films were tested in uniaxial tension and by bulge testing; in addition, an Al film was deposited on a Si beam and tested. Values for Y, all determined by the same investigators (Ref. 15), were surprisingly close, i.e., 68.8, 73, and 69.3 GPa, respectively, and compared well with the bulk value of 70 GPa. In contrast, Schweitz (Ref. 16) lists six measurements of Y for crystalline (100) silicon by different experimenters that vary from 122 GPa to 220 GPa. Similarly, seven other determinations of Y for polycrystalline silicon ranged from 90 to 190 GPa. Micromechanical testing methods were employed in all cases. In view of the scatter, caution must be exercised when applying these values.

12.2.5 Dislocation Models for Film Strength

In bulk crystalline solids, dislocation theory is the basis for explaining all sorts of mechanical phenomena including strength properties. As a summary of topics aired in this section and a prelude to those raised in the next, it is instructive to once again model the role of dislocations in films particu-

larly as they affect strength. In a simplified version of a treatment by Freund (Ref. 17) and Nix (Ref. 2), we consider a film of thickness d, situated atop a thick substrate, containing threading and misfit dislocations (Section 8.3). Furthermore, the film is subjected to a biaxial state of internal stress of magnitude σ_i. Dislocation theory states that the force F (per unit length) on a dislocation having Burgers vector b is simply $F = \sigma_i b$. The energy per unit length (E_m) expended in moving threading dislocation segments along the slip plane a length equal to the film thickness is then approximately $E_m = \sigma_i bd$. For such dislocation migration the work done by the internal film stresses must be equal to or greater than the energy per unit length of the interfacial misfit dislocation (E_d); by Section 8.3.1, $E_d = [\mu b^2/4\pi(1-\nu)]\ln(\beta d/b)$. In this equation ν is Poisson's ratio, $\beta = 2.6$, and it is assumed that the shear modulus for the film and substrate are equal. By equating E_m and E_d the minimum biaxial yield stress is estimated to be

$$\sigma_0(\approx \sigma_i) = [\mu b/4\pi(1-\nu)d]\ln(\beta d/b). \tag{12-6}$$

This expression applies to a single crystal film having no obstacles to dislocation motion such as other dislocations, grain boundaries, precipitates, or point defects. Presence of an oxide layer would, for example, raise σ_0 because dislocations are necessarily present at the film/oxide interface. Importantly, Eq. 12-6 predicts that the film yield strength varies inversely with film thickness. The omission of the deformation crystallography is an important simplification in this treatment that limits its applicability to particular crystal structures. It is well known that dislocations glide on specific planes and in certain directions, and the analysis by Nix (Ref. 2) includes these effects.

Additional insight into the thickness dependence of film strength can be gained by considering two well-known relationships that describe mechanical behavior in bulk materials. The first simply estimates the stress (shear) required to cause dislocations to effectively bypass obstacles situated a distance l apart and thereby produce plastic deformation. Thus

$$\sigma_0 = \mu b/l \tag{12-7}$$

where b is the Burgers vector. The distance l is usually taken as the spacing between dislocation pinning points such as precipitates, grain boundaries, or other dislocations. If l is imagined to be the film thickness (d), then it is clear that film strength is predicted to vary inversely with d in basic agreement with Eq. 12-6.

The second relationship is the Hall–Petch equation. It describes grain boundary strengthening by connecting the yield stress to grain size l_g.

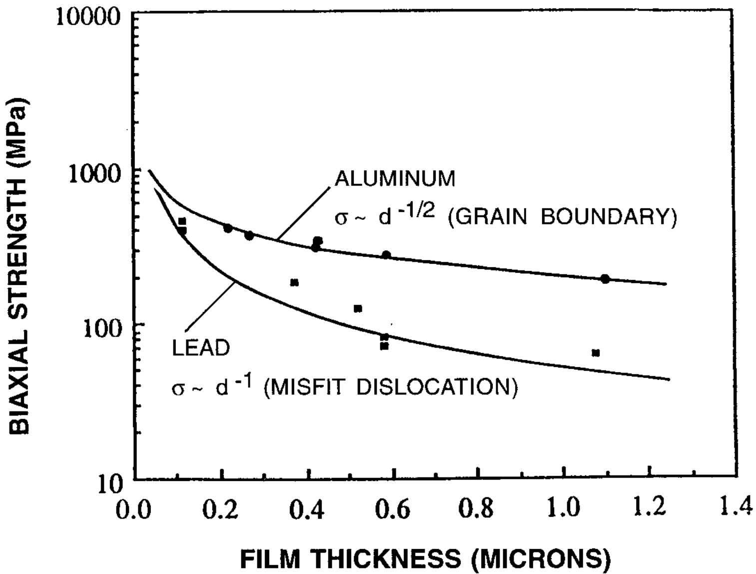


Figure 12-5 Biaxial yield strengths for lead and aluminum films on silicon as a function of film thickness. The strengths vary as d^{-1} and $d^{-1/2}$, respectively. (From Ref. 2. Reprinted with the permission of Professor W. D. Nix.)

Variants of this equation link ultimate tensile strength σ_{UTS} to grain size as well. Therefore,

$$\sigma_{\mathrm{UTS}} = \sigma_{\mathrm{i}} + K l_{\mathrm{g}}^{-1/2}. \tag{12-8}$$

In a modified version d replaces l_{g}. This is a justifiable substitution since the two quantities are directly related (see Section 9.4.2.1). In Eq. 12-8 σ_{i} is an intrinsic stress level and K is a constant. Again on the basis of this formula, thinner films are expected to be stronger.

As revealed in Fig. 12-5 the biaxial yield strengths for lead and aluminum films on silicon are film-thickness dependent varying as d^{-1} and $d^{-1/2}$, respectively. That the Al films were stronger than predicted by Eq. 12-6 was attributed to significant additional grain-size (i.e., Hall–Petch) strengthening. In a study by Kang and Ho (Ref. 18), the tensile behavior of thin Al film/polyimide composite specimens was measured. As the Al film thickness decreased from 480 to 60 nm the tensile strength increased from 196 to 408 MPa. By separating the individual contributions to film strengthening it was concluded, in accord with Eq. 12-8, that small columnar grain sizes were the most influential factor; film thickness and texture effects exerted a lesser influence on strength.

In the scaling down from micron-thick foils to nanometer-thick films, it is not certain whether there is a thinness beyond which no further strengthening occurs. Experimental data have not resolved this issue.

12.2.6 Hardness and Strength of Multilayer Films

We close this section with a brief discussion of the mechanical properties of multilayers and superlattices. A concise summary of hardness (H) data as a function of bilayer period spacing (Λ) for a number of epitaxial as well as polycrystalline layers is plotted in Fig. 12-6. Metal/metal, metal/nitride and nitride/nitride multilayers were studied (Ref. 19) and it is evident that shorter period spacings yield higher hardnesses in virtually all cases, a result suggestive of the previous section on film strengthening. Therefore, if in Eq. 12-8, H is substituted for σ_{UTS} and Λ for l_g, a Hall–Petch equation for hardness can be imagined. Replotting H vs $\Lambda^{-1/2}$, which is left to the reader as an exercise, generally reveals a linear fit to the data for large values of Λ; however, significant deviations from Hall–Petch behavior exist below 5–10 nm.

Interestingly, multilayers exhibit a hardness enhancement above that predicted by the rule of mixtures, i.e., the average hardness of the two individual films. There are several possible causes for this extra strengthening. In 1970 Koehler (Ref. 20) suggested a couple of recipes to produce strong solids. The first was to layer materials having widely differing shear moduli (μ). In such cases the shear stress required to move dislocations across layer interfaces is directly proportional to the product of $\Delta\mu$ (the modulus difference) and $\cos\theta$, where θ is the smallest angle between glide planes and the interface. Furthermore, he suggested that the layers be thin to prevent the operation of dislocation sources. The increased shear stress required to activate dislocation motion, relative to bulk material, varies inversely with layer thickness, precisely as indicated by Eq. 12-6. Dislocation pileups at layer interfaces also serve to repel preexistent dislocations from penetrating them. The net result is added strength because dislocation motion is so restricted in thin layers.

12.3 ANALYSIS OF INTERNAL STRESS

12.3.1 Introduction

Implicit in the discussion to this point is that the stresses and the effects they produce are the result of *externally* applied forces. After the load is

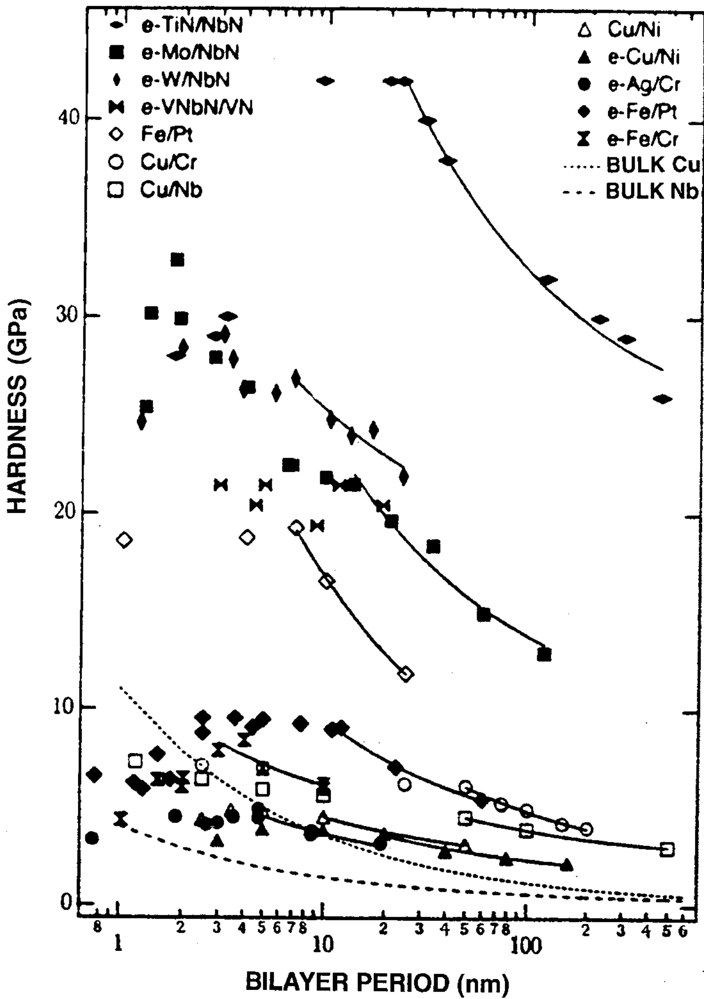


Figure 12-6 Hardness as a function of bilayer period spacing for several multilayer systems. Predicted Hall–Petch behaviors for Cu and Nb are dotted in. The designation e refers to epitaxial. (From Ref. 19. Reprinted with the permission of B. M. Clemens.)

removed, the stresses are expected to vanish. On the other hand, thin films are stressed even without the application of externally imposed forces and are said to possess internal or residual stresses. The existence of internal stresses in thin electrodeposited films has been known since 1858 when the English chemist Gore noted (Ref. 21), "In electrodeposits generally the inner

and outer surfaces are in unequal states of cohesive tension frequently in so great a degree as to rend the deposit extensively and raise it from the cathode in the form of a curved sheet with its concave side to the anode." The concave bending of the cantilevered cathode-electrode implies, as we shall later see, a tensile stress in the deposit. Despite the passage of years, the origins of stress in electrodeposited films are still not completely understood. A similar state of affairs exists in both physical and chemical vapor-deposited thin films; stresses exist, they can be measured, but their origins are not always known with certainty.

Residual stresses are, of course, not restricted to composite film-substrate structures, but occur universally in all classes of homogeneous materials under certain conditions. A state of nonuniform plastic deformation is required and this frequently occurs during mechanical or thermal processing. For example, when a metal strip is reduced slightly by rolling between cylindrical rolls, the surface fibers are extended more than the interior bulk. The latter restrains the fiber extension and places the surface in compression while the interior is stressed in tension. This residual-stress distribution is locked into the metal, but can be released like a jack-in-the-box. Machining a thin surface layer from the rolled metal will upset the mechanical equilibrium and cause the remaining material to bow. Residual stresses arise in castings, welds, machined and ground materials, and heat-treated glass. The presence of residual stresses is usually undesirable, but there are cases where it is beneficial. Tempered glass and shot-peened metal surfaces rely on residual compressive stresses to counteract harmful tensile stresses applied in service.

12.3.2 Mechanical Equilibrium

A model for the generation of internal stress during the deposition of films is illustrated in Fig. 12-7. Regardless of the internal-stress distribution which prevails, it must be self-equilibrating. Maintenance of mechanical equilibrium therefore requires that the *net* force (F) and bending moment (M) vanish on any film/substrate cross section. Thus

$$F = \int \sigma \, dA = 0 \tag{12-9a}$$

and

$$M = \int \sigma y \, dA = 0 \tag{12-9b}$$

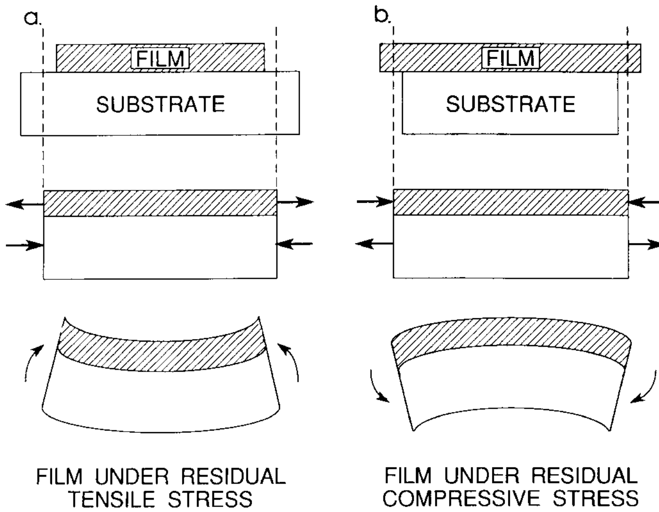


Figure 12-7 Sequence of events leading to (a) residual tensile stress in film; (b) residual compressive stress in film.

where A is the sectional area and y is the moment lever arm. Intuitive use is made of these basic equations here and they are applied analytically in deriving the Stoney formula in the next section. First consider the behavior shown in Fig. 12-7a, where the growing film initially shrinks relative to the substrate. Surface tension forces are one reason why this might happen; the misfit accompanying epitaxial growth is another. Compatibility, however, requires that both the film and substrate have the same length. Therefore, the film is constrained and stretches, while the substrate accordingly contracts. The tensile forces developed in the film are balanced by the compressive forces in the substrate. However, the combination is still not in mechanical equilibrium because of the uncompensated end moments. If the film/substrate pair is not restrained from moving, it will elastically bend as shown to counteract the unbalanced moments. Thus, films containing internal tensile stresses bend the substrate concave upward. In an entirely similar fashion compressive stresses develop in films that tend to initially expand relative to the substrate (Fig. 12-7b). Internal compressive film stresses, therefore, bend the substrate convex outward. These results are perfectly general regardless of the specific mechanisms which cause films to stretch or shrink relative to substrates. Sometimes the tensile stresses are sufficiently large to cause film fracture. Similarly, excessively high compressive stresses can cause film wrinkling and local loss of adhesion to the

substrate. Examples of both of these effects are shown in Fig. 12-8. A discussion of the causes of internal stress will be deferred until Section 12.5 as we now turn our attention to a calculation of film stress as a function of substrate bending.

12.3.3 THE STONEY FORMULA

The formulas which have been used in virtually all experimental determinations of film stress are variants of an equation first given by Stoney in 1909 (Ref. 22). This equation can be derived with reference to Fig. 12-9, which shows a composite film/substrate combination of unit width w. The film thickness and Young's modulus are d_f and Y_f, respectively, while the corresponding substrate values are d_s and Y_s. Because of lattice misfit, differential thermal expansion, film growth effects, etc., mismatch forces arise at the film/substrate interface. In the free-body diagrams of Fig. 12-9b each interfacial set of forces can be replaced by the statically equivalent combination of a force and moment; F_f and M_f in the film, F_s and M_s in the substrate, where $F_f = F_s$. Force F_f can be imagined to act uniformly over the cross-sectional area ($d_f w$) giving rise to the film stress. The moments are responsible for the bowing of the film/substrate composite. Equation 12-9b implies equality of the clockwise and counterclockwise moments, a condition expressed by

$$\frac{(d_f + d_s)F_f}{2} = M_f + M_s. \qquad (12\text{-}10)$$

Consider now an isolated beam bent by moment M as indicated in Fig. 12-9c. In this case, the deformation is assumed to consist entirely of the extension or contraction of longitudinal beam fibers by an amount proportional to their distance from the central or neutral axis, which remains unstrained in the process. The stress distribution reflects this by varying linearly across the section from maximum tension ($+\sigma_m$) to maximum compression ($-\sigma_m$) at the outer beam fibers. In terms of the beam radius of curvature R, and angle θ subtended, Hooke's law yields

$$\sigma_m = \frac{Y\{(R + d/2)\theta - R\theta\}}{R\theta} = \pm\frac{Yd}{2R}. \qquad (12\text{-}11)$$

Corresponding to this stress distribution is the bending moment across the beam section

$$M = 2\int_0^{d/2} \sigma_m w\left(\frac{y}{d/2}\right) y\,dy = \sigma_m d^2 w/6 = Yd^3 w/12R. \qquad (12\text{-}12)$$

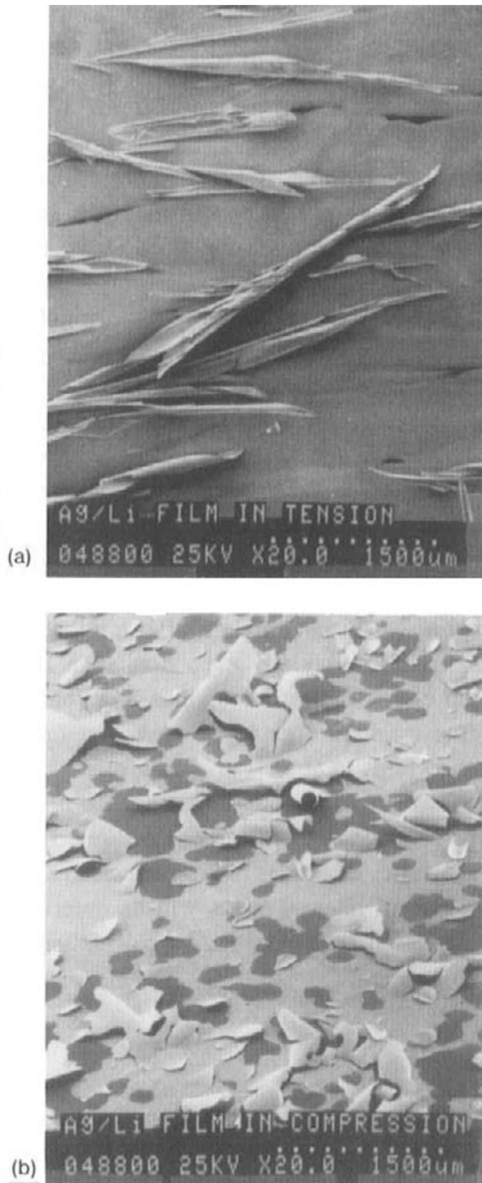


Figure 12-8 Stresses in silver-lithium thin films: (a) Tensile film failures during deposition; (b) compressive film failures during aging in Ar. (Courtesy of R. E. Cuthrell, Sandia National Laboratory.)

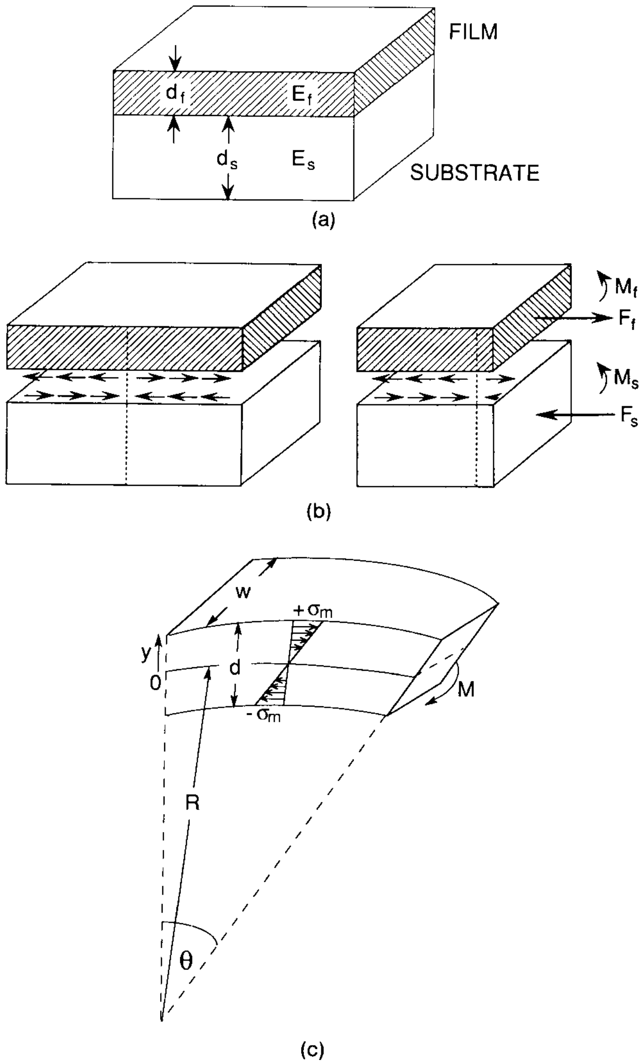


Figure 12-9 Stress analysis of film/substrate combination: (a) Composite structure; (b) free-body diagrams of film and substrate with indicated interfacial forces and end moments; (c) elastic bending of beam under applied end moment.

By extension of this result, we have

$$M_f = \frac{Y_f d_f^3 w}{12R} \qquad \text{and} \qquad M_s = \frac{Y_s d_s^3 w}{12R}.$$

Lastly, in order to account for actual biaxial-stress distributions in films, rather than the uniaxial stresses assumed for ease of derivation, it is necessary to replace Y_f by $Y_f/1 - \nu_f$, and similarly for Y_s. Substitution of these terms in Eq. 12-10 yields

$$\frac{(d_f + d_s)F_f}{2} = \frac{wY_f d_f^3}{12R(1 - \nu_f)} + \frac{wY_s d_s^3}{12R(1 - \nu_s)}. \tag{12-13}$$

Since d_s is normally much larger than d_f, the film stress σ_f is, to a good approximation, given by

$$\sigma_f = \frac{F_f}{d_f w} = \frac{Y_s d_s^2}{6R(1 - \nu_s)d_f}. \tag{12-14}$$

This is the Stoney formula. Values of σ_f are determined through measurement of R. The reader should be certain that literature values of σ_f include the $1 - \nu_s$ correction factor.

For a stack of n layered, continuous films, since moments are additive the total curvature is just the sum of the individual film contributions. Stoney's formula then yields

$$\frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n} = \frac{6(1 - \nu_s)(\sigma_1 d_1 + \sigma_2 d_2 + \cdots + \sigma_n d_n)}{Y_s d_s^2}. \tag{12-15}$$

12.3.4 Stress in Epitaxial Superlattices

Because of their scientific and increasingly technological importance, it is appropriate to consider the issue of stress in these layered structures. In particular, following Tu *et al.* (Ref. 23), let us consider a given A–B strained-layer superlattice where the lattice constants a_A and a_B differ, but the layer spacings are equal, i.e., $d_A = d_B = d$, and the elastic properties are identical, i.e., $Y_A = Y_B = Y$. The question is which substrate will reduce superlattice strain energy (E_S) more: one which has the same lattice constant as material A (case 1), or one which has a lattice constant a_{AB} midway between them, i.e., $a_{AB} = (a_A + a_B)/2$ (case 2)? The choice of resulting structures is shown in Fig. 12-10. Assuming pseudomorphic growth, only B

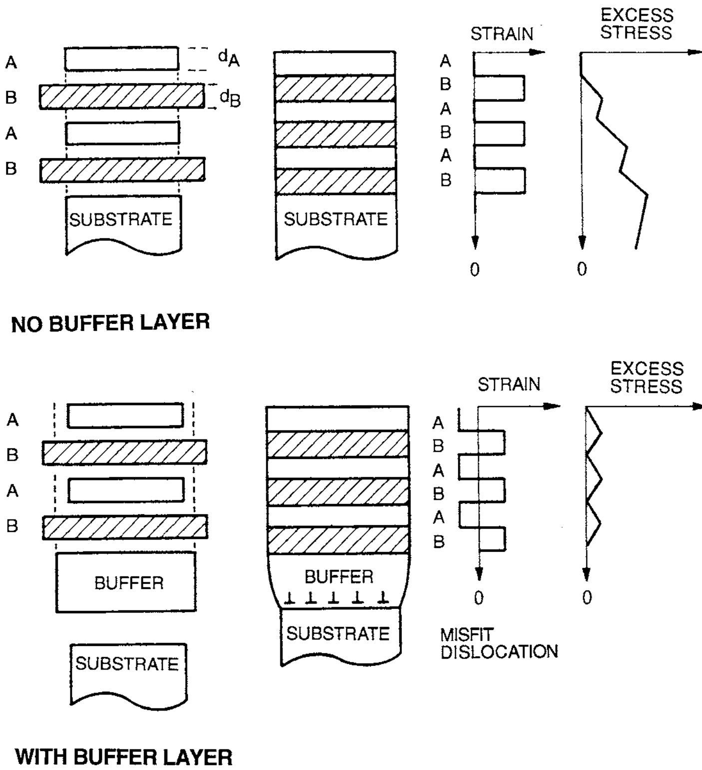


Figure 12-10 Schematic diagram of strained-layer superlattice illustrating layer stress for (a) no buffer layer and layer A lattice matched to the substrate (Case 1), and (b) with buffer layer lattice matched to average lattice constant of layers A and B (Case 2). (From S. T. Picraux, B. L. Doyle, and J. Y. Tsao, in *Strained-Layer Superlattices*, ed. T. P. Pearsall, Vol. 33, *Semiconductors and Semimetals*. Academic Press, New York, 1990. Reprinted with permission.)

layers are strained in case 1. By Eq. 1-45, $E_S = \frac{1}{2}Y\varepsilon^2$, where ε is the misfit strain, and hence

$$E_S(1) = \frac{nd}{2}\left\{\frac{Y}{2}\left(\frac{a_B - a_A}{a_A}\right)^2\right\} \tag{12-16}$$

where n is the total number of A and B layers. However, in case 2 both layers are strained, and therefore,

$$E_S(2) = \frac{nd}{2}\left\{\frac{Y}{2}\left[\left(\frac{a_A - a_{AB}}{a_{AB}}\right)^2 + \left(\frac{a_B - a_{AB}}{a_{AB}}\right)^2\right]\right\}. \tag{12-17}$$

Algebraic expansion readily shows that the ratio $E_S(1)/E_S(2) = 2a_{AB}^2/a_A^2$. Inasmuch as a_{AB} is only slightly different from a_A, the strain energy is basically halved in case 2. Because the strain energy depends quadratically on strain, it is better for all film layers to be partially strained than for half of the films to be strained to maximum levels.

This result is the basis for the use of partially relaxed buffer layers shown in Fig. 12-10 that are grown on substrates prior to the deposition of strained epitaxial layers. As a result of this step, film defect densities are normally reduced (see Fig. 8-23).

12.3.5 Thermal Stress in Homogeneous and Bilayer Structures

Thermal effects make important contributions to film stress. Films and coatings prepared at elevated temperatures and then cooled to room temperature will be thermally stressed; so will films which are thermally cycled or cooled from the ambient to cryogenic temperatures. In Section 1.8.3 we saw that the magnitude of the thermal stress was given by Eq. 1-48 or $\sigma = Y\alpha(T - T_0)$ where α is the coefficient of thermal expansion and $T - T_0$ is the temperature difference.

Consider now the bilayer film/substrate combination of Fig. 12-9 subjected to a temperature differential ΔT. The film and substrate strains are respectively given by

$$\varepsilon_f = \alpha_f \Delta T + \frac{F_f(1 - \nu_f)}{Y_f d_f w} \tag{12-18a}$$

$$\varepsilon_s = \alpha_s \Delta T - \frac{F_f(1 - \nu_s)}{Y_s d_s w}. \tag{12-18b}$$

But strain compatibility requires that $\varepsilon_f = \varepsilon_s$, and therefore the thermal-mismatch force is

$$F_f = \frac{w(\alpha_s - \alpha_f)\Delta T}{\{(1 - \nu_f)/d_f Y_f + (1 - \nu_s)/d_s Y_s\}}. \tag{12-19}$$

If $d_s Y_s/1 - \nu_s \gg d_f Y_f/1 - \nu_f$, the thermal stress in the film is

$$\sigma_f(T) = \frac{F_f}{d_f w} = \frac{(\alpha_s - \alpha_f)\Delta T\, Y_f}{1 - \nu_f}. \tag{12-20}$$

Note that the signs are consistent with dimensional changes in the film/substrate. Films prepared at elevated temperatures will be residually com-

pressed when measured at room temperature ($\Delta T < 0$) if $\alpha_s > \alpha_f$. In this case the substrate shrinks more than the film. Overall, the system must contract a fixed amount but a compromise is struck; the substrate is not allowed to contract fully and is, therefore, placed in tension, while the film, hindered from shrinking, is consequently forced into compression. An example of this occurs in TiC coatings on steel. At the CVD deposition temperature of 1000°C, it is assumed the coating and substrate are unstressed. For the values $\alpha_{\text{steel}} = 11 \times 10^{-6}\ \text{K}^{-1}$, $\alpha_{\text{TiC}} = 8 \times 10^{-6}\ \text{K}^{-1}$, $Y_{\text{TiC}} = 450$ GPa, and $\nu_f = 0.19$, the compressive stress calculated for TiC, using Eq. 12-20, is 1.67 GPa at 0°C.

Lastly, by equating Eqs. 12-14 and 12-20 we obtain

$$\frac{1}{R} = \frac{6Y_f(1-\nu_s)d_f(\alpha_s-\alpha_f)\Delta T}{Y_s(1-\nu_f)d_s^2}. \tag{12-21}$$

This modification of Stoney's equation represents the extent of bowing when differential thermal-expansion effects cause the stress. In order to convert the measured deflection into film stress, we note that the curvature is related to the second derivative of the beam displacement (y), i.e., $1/R = d^2y(x)/dx^2$. After integration, $y(x) = x^2/2R$. For a cantilever of length L, if the free-end displacement is δ, then $\delta = L^2/2R$ and Stoney's formula yields

$$\sigma_f = \frac{\delta Y_s d_s^2}{3L^2(1-\nu_s)d_f}. \tag{12-22}$$

12.3.6 Thermal Stress in Three-Layer Structures

There are a number of somewhat more complicated but quite important thermal-stress problems that arise in electronic packaging applications. These involve layers of different components that are bonded by means of an intermediate thin adhesive layer. An example is the trilayer sandwich structure consisting of a silicon chip (die) and substrate (the adherents) bonded by a thin solder or epoxy adhesive depicted in Fig. 12-11. The thermal stress distribution in such structures subjected to a temperature variation ΔT has been treated rigorously by Suhir (Ref. 24). Simplified solutions (Refs. 1, 25) for this two-dimensional, three-layer geometry are available and presented below assuming that:

(a) the adhesive is linearly elastic and the shear stress does not vary through its thickness
(b) both top and bottom adherent layers are uniformly stressed through the thickness

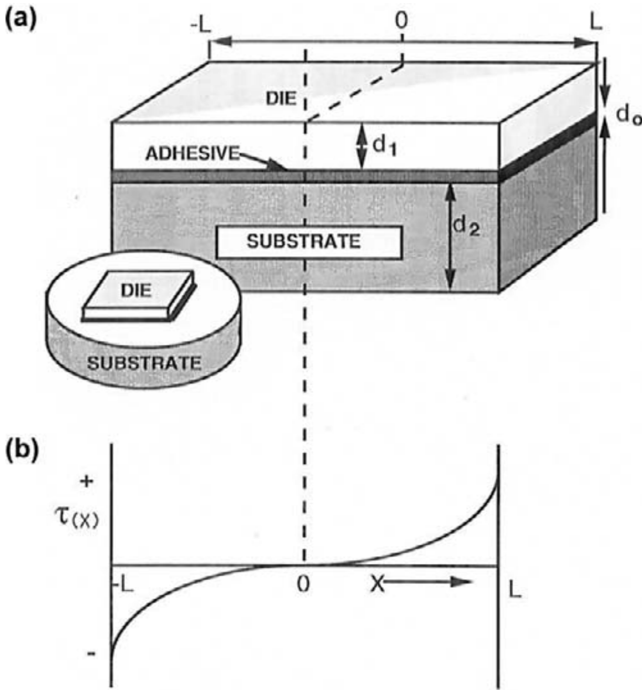


Figure 12-11 (a) Two-dimensional model of a trilayer structure subjected to thermal stresses. (b) Simplified solution for shear stress as a function of distance.

(c) bending of the three-layer structure is neglected
(d) the temperature is uniform in all layers

The shear stress (τ) in the adhesive film is given by

$$\tau(x) = \frac{\mu[\alpha(1) - \alpha(2)]\Delta T}{d_0 \beta} \frac{\sinh \beta x}{\cosh \beta L}, \tag{12-23}$$

where μ is the shear modulus and $\beta^2 = \mu/d_0\{[Y(1)d(1)]^{-1} + [Y(2)d(2)]^{-1}\}$. Its distribution is plotted in Fig. 12-11b where symbols x, d_0, d_1, d_2, and L are all defined. Normally, we are only interested in the maximum shear stress which occurs at $x = L$. Therefore,

$$\tau(\max) = \frac{\mu}{d_0 \beta}[\alpha(1) - \alpha(2)]\Delta T \tanh \beta L. \tag{12-24}$$

Aside from assorted constants this formula for thermal stress has essentially the same form as Eq. 12-20. It should be noted that the shear stress at

the interface depends directly on the thermal-mismatch strain, the modulus of the adhesive, and the length of the assembly joint. But, the shear stress is independent of the thermal expansion coefficient of the adhesive material. In addition to interfacial shear stresses in the adhesive, the adherents are subject to normal and peeling stresses. For example, the normal or tensile stress that acts along the chip plane is large in the center and vanishes at its outer edges. On the other hand transverse normal or peeling stresses arise from the constraint that forces the assembled components to bend jointly and have the same curvature at all sections. Peeling of films from substrates can be analyzed in similar ways. The magnitudes of these three thermal stresses depend on the involved material properties, geometries, and values for ΔT; not infrequently they are large enough to induce cracking in the die and adhesive. Matching the thermal-expansion coefficients of the adherents is the most widespread strategy employed to minimize thermal stresses.

12.4 TECHNIQUES FOR MEASURING INTERNAL STRESS IN FILMS

There are differences between measuring stress in an externally loaded film and detecting its sign and magnitude when present in outwardly unloaded films. In the former case, the measured stress is the direct response to an imposed strain and there is a correspondence between cause (ε) and effect (σ). Conversely, stress in internally stressed films is normally revealed by measuring the strain it induces. In common, the attachment to a massive substrate makes extracting σ or ε information in films an often difficult task. There are basically two general categories of experimental techniques for measuring internal stresses in thin films. One relies on measuring the deflection or curvature of substrates to which the films are attached; the second involves the direct determination of elastic strains by means of X-ray diffraction methods. Many variants of these methods have been developed and continue to evolve, employing different specimen geometries and novel instrumental techniques. A number of the more important of these methods will be discussed next.

12.4.1 SUBSTRATE DEFLECTION OR CURVATURE METHODS

12.4.1.1 Cantilever Beams

Two film/substrate configurations have been primarily used. In the first, the substrate is fashioned into the shape of a cantilever beam. The film is

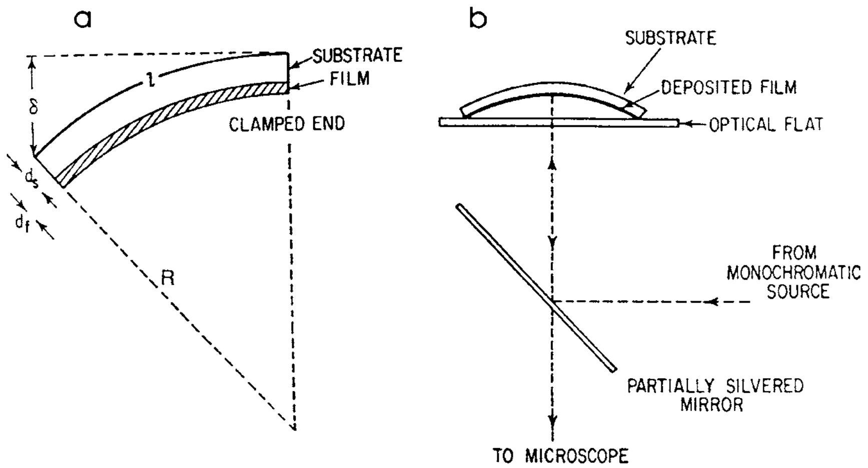


Figure 12-12 Schematic diagrams of film stress measurement techniques. (a) Bending of cantilever beam. (b) Bowing of circular plate. (From Ref. 30.)

deposited on one surface and the deflection of the free end of the bent beam is then determined (Fig. 12-12a). As a result the force per unit width S is defined by the product of the film stress and thickness. By Eq. 12-22,

$$S = \sigma_f d_f = \frac{\delta Y_s d_s^2}{3L^2(1 - \nu_s)}. \tag{12-25}$$

This definition circumvents difficulties in the magnitude of σ_f for very thin films since d_f is no longer in the denominator. S is simply proportional to the deflection; in order to obtain σ_f, d_f must be independently known.

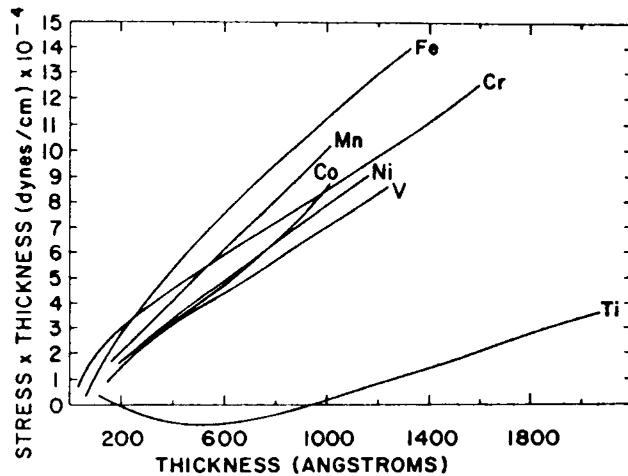
Measurements of stress are frequently made in real time during film formation and growth. Sensitive electromechanical or magnetic restoration of the null, or undeflected position of the beam, in combination with measurement of the restoring force enables continuous monitoring. Null methods have a couple of advantages in film stress determinations relative to techniques in which beam (or plate) deflections are measured. One is the lack of stress in the film because the substrate is effectively restrained from deflecting. The second is that the frequently unknown value of Young's modulus is, surprisingly, *not* required to evaluate the stress in the film. Since there is no deflection, Stoney's formula is inappropriate. Rather, the restoring force establishes a moment, and from an equation of the type $M = \sigma_m d^2 w/6$ (Eq. 12-12), the stress can be evaluated.

Measurements for a number of metals evaporated onto MgF_2 coated glass cantilever substrates at room temperature (Ref. 26) are shown in Fig. 12-13; the results will be discussed subsequently. In addition to σ_f, which can be determined at any value of d_f, the slope at any point of the S–d_f plot yields a quantity known as the incremental or instantaneous stress $\sigma_f(z)$. The latter is the stress present in a layer thickness dz at a distance z from the film/substrate interface and is useful in characterizing dynamic changes in film stress. Readers should be alert to the fact that S, σ_f, and $\sigma_f(z)$ have all been used in the literature to report stress values. Unless otherwise noted, the single term σ_f will generally be used for the remainder of the chapter.

12.4.1.2 Circular Plates

The second type of substrate configuration frequently employed is a circular plate. In this case Eq. 12-22 is assumed to describe the resultant stress where L is now the plate radius and δ represents the center deflection. Not only is the plate an important test geometry, but it has obvious applications in optical components. As a practical example, consider a circular glass plate window with a diameter-to-thickness ratio of 16:1 having the elastic properties $Y = 63.7$ GPa and $\nu = 0.25$. If one face is coated with a 1000 Å thick single-layer antireflection film of MgF_2, a stress of 333 MPa is found to develop after deposition. The depression of the center relative to the circumference is calculated to be 7.5×10^{-8} m. For a wavelength $\lambda = 5000$ Å the extent of bowing corresponds to 0.15λ. This change from a planar surface to a paraboloid of revolution is significant in the case of high-precision optical surfaces where the tolerance may be taken to be 0.05λ. In the case of thin lenses, stress distortions can also change critical spacings in multielement optical systems. A typical arrangement for stress measurement of plates is shown in Fig. 12-12b where the change in the optical fringe pattern between the film substrate and an optical flat is used to measure the deformation. Interference fringe patterns (Ref. 27) illustrating different states of biaxial stress in films can be made visible as shown in Fig. 12-14. Alternatively, as in bulge testing (Section 12.2.3) a calibrated optical microscope can be used to measure the extent of bowing.

One of the most popular techniques for determining film stress makes use of laser-scanning methods to measure substrate curvature (Ref. 2). It was primarily developed to serve the needs of the semiconductor industry where wafer bowing is an issue of considerable concern. Stressed thin films deposited or grown on much thicker circular wafers bend them only slightly. Therefore, high sensitivity is required to detect changes in wafer curvature,



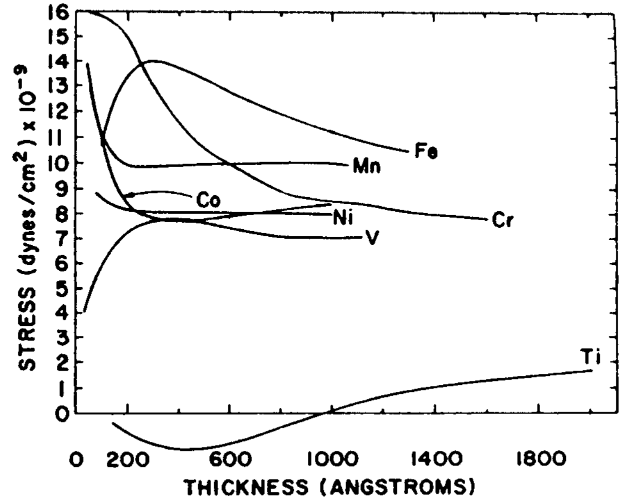


Figure 12-13 Internal stress values in a number of evaporated metal thin films: (a) $S = \sigma_f d_f$ vs film thickness, (b) σ_f vs film thickness. (From Ref. 26, reprinted by permission of the Electrochemical Society.)

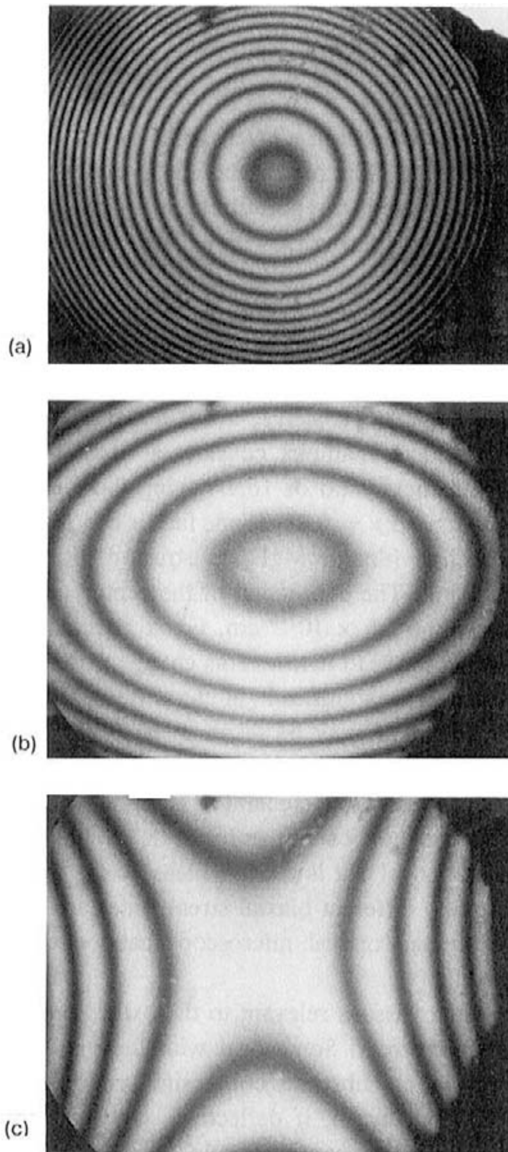


Figure 12-14 Interference fringe patterns in biaxially stressed sputtered Mo films. (From Ref. 27.) (a) Balanced biaxial tension or compression ($\sigma_x = \sigma_y$); (b) unbalanced biaxial tension or compression ($\sigma_x \neq \sigma_y$); (c) one component tensile, one component compressive.

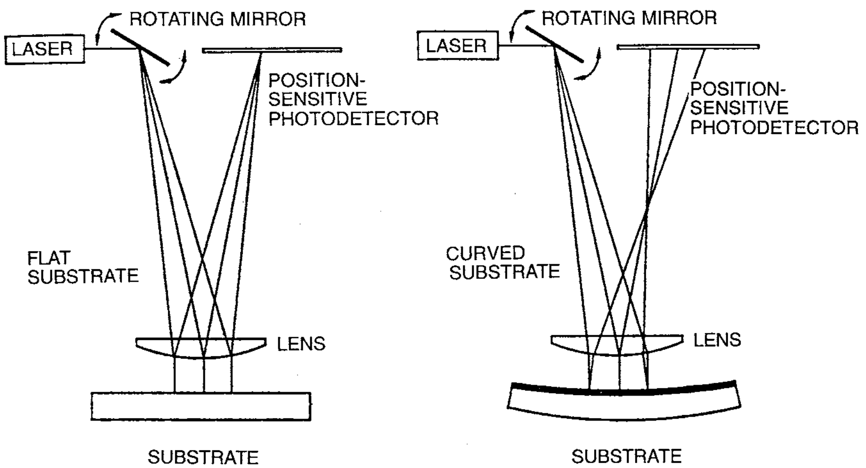


Figure 12-15 Schematic diagram of a laser scanning instrument used to measure substrate curvature and associated film stress. (From Ref. 2. Reprinted with the permission of Professor W. D. Nix.)

a capability provided by the system schematically illustrated in Fig. 12-15. By means of a rotating mirror, an incident laser beam is scanned across the film surface. A lens is interposed to ensure that the beam impinges perpendicular to the substrate while a position-sensitive photodetector records the reflected beam. The angle of reflection simply depends on the surface curvature at the film point being probed. For a stress-free film/substrate the surface is perfectly flat, and the reflected light converges to the same point irrespective of beam position. But if the wafer is bowed, light is reflected to different detector positions in concert with the film/surface sites sequentially scanned by the beam. The photodetector information is converted to angle, then to curvature, and ultimately to stress. Commercially available laser-scanning instruments are capable of measuring silicon wafer radii-of-curvature equal to 10 km, corresponding to typical stress sensitivity of $\sim$0.2 MPa. Uncoated wafers often have a nonconstant curvature and are sometimes deformed into a potato-chip shape. Nevertheless, the average stress in subsequently deposited films usually can be determined with good precision by subtracting curvature data of the initially bare substrate from that obtained after coating it, both scanned over identical paths. Laser scanning is generally limited to postdeposition stress determinations, not real-time monitoring of stress evolution. An example of laser-scan monitoring of stress in a thermally cycled aluminum film will be discussed in Section 12.6.4.

12.4.2 X-RAY DIFFRACTION

A number of useful methods for determining stress and related properties such as texture (Section 9.4.4.2) and epitaxial strains (Section 10.3.4) are based on X-ray diffraction methods. These techniques measure stress by detecting the slight changes in interplanar spacing of crystal planes (Refs. 9, 28). As an example consider a polycrystalline film containing an isotropic biaxial tensile stress distribution in the xy plane ($\sigma_z = 0$). The film contracts in the z direction (Section 1.8.1) by an amount

$$\varepsilon_z = (-\nu/Y)(\sigma_x + \sigma_y) = -\nu(\varepsilon_x + \varepsilon_y). \quad (12\text{-}26)$$

By measuring the lattice spacing (a) in stressed films relative to the unstressed bulk lattice (a_0) in the z direction with X-rays, Bragg's law ($\lambda = 2a \sin\theta$) immediately yields ε_z, i.e., $\varepsilon_z = (a - a_0)/a_0$. Since $\sigma_x = \sigma_y = \sigma_f$,

$$\sigma_f = -\frac{Y(a - a_0)}{2\nu a_0}. \quad (12\text{-}27)$$

The accuracy of the X-ray technique is considerably extended with high-precision lattice-parameter determinations. Precise determination of a and a_0 is complicated by line broadening due to small grain size, dislocations, twins, stacking faults, and nonuniform microstrains.

A noteworthy advance in technique has the important objective of obtaining X-ray diffraction data in structures having linear dimensions ranging from tenths to hundreds of microns. These microbeam X-ray diffraction methods (Ref. 29) collimate the X-ray beam with tapered glass capillaries. The latter act as waveguides along which X-rays reflect at glancing angles to yield sufficiently intense laboratory beams. Applications include microdiffraction as well as microfluorescence of thin film (NiFe) magnetic recording heads, and patterned surface as well as buried interconnect grain structures in integrated circuits. In particular, stress distributions have been mapped over film regions extending $\sim 100\,\mu$m employing a $\sim 10\,\mu$m diameter probe beam.

12.4.3 ACCURACY AND SENSITIVITY OF STRESS DETERMINATIONS

There are a number of issues relevant to the determination of accurate film stress values irrespective of the experimental technique employed. Some deal with the validity of elastic theory itself in treating the film/substrate

composite properties, geometries, and deflections. In general, if deflections are small compared to the substrate beam or plate thickness, the simple theory suffices. Substrate elastic-constant values should be chosen with care and checked to determine whether they are isotropic. In this regard, special care must be exercised when determining the stress in epitaxial films because elastic constants of both films and substrates are anisotropic.

The smallest detectable stress is a common measure of the sensitivity of measurement. In cantilevers typical values for the different experimental techniques range from approximately ~0.05 MPa (electromechanical, interferometric) to 25 MPa (magnetic restoration, optical). Except for the laser scanning technique (~0.2 MPa), the sensitivity for disks (~100 MPa) using pressure (bulge) and optical methods is relatively low. For X-ray methods the smallest detectable stress is ~50 MPa (Ref. 30). In all cases 100 nm thick films are assumed.

12.5 INTERNAL STRESSES IN THIN FILMS AND THEIR CAUSES

12.5.1 Intrinsic Stress in Thin Films

To begin with, a distinction sometimes made between internal or residual and *intrinsic* stress (σ_i) should be appreciated. Internal stress is the more inclusive term and includes the unavoidable thermal or *extrinsic* stress, $\sigma(T)$, in addition to σ_i. Thus the magnitude of the internal film stress, σ_f, is expressed by the algebraic sum of both contributions or

$$\sigma_f = \sigma(T) + \sigma_i. \tag{12-28}$$

If σ_i is desired then the thermal stress given by Eq. 12-20 must be subtracted from the measured value of σ_f. Nix (Ref. 2) has suggested that intrinsic stress arises primarily from the processes of film growth and epitaxy. Strains induced by the latter are described by Eq. 8-1. If the film density changes during growth as it bonds to the substrate, the resulting elastic accommodation strain (ε) is given by $\varepsilon = -\varepsilon_T/3$, where ε_T is the dilatational transformation strain associated with densification. Thus, a film that densifies is pulled in biaxial tension in order to accommodate to substrate dimensions.

Results of residual stress measurements for films, largely employed in microelectronics and deposited by evaporation, sputtering, and CVD, are entered in Table 12-1 and reviewed in subsequent sections. Unless noted otherwise, room temperature values of σ_f are meant. Despite the apparent simplicity of both the experimental techniques and corresponding defining stress equations, the measured values of intrinsic stress display a bewildering

Table 12-1

Residual Stress Values in Films Encountered in Si Devices[a]

Film	Process	Conditions	Stress (GPa)
SiO_2	Thermal	900–1200°C	−0.2 to −0.3
SiO_2	CVD 400°C	40 nm/min	+0.13
SiO_2	$SiH_4 + O_2$	400 nm/min	+0.38
SiO_2	CVD	450°C	+0.15
	TEOS	725°C	+0.02
SiO_2	TEOS	685°C	+0.38
	TEOS + 25% B, P		−0.02
SiO_2	Sputtered		−0.15
Si_3N_4	CVD	450–900°C	+0.7 to +1.2
Si_3N_4	Plasma	400°C	−0.7
		700°C	+0.6
Si_3N_4	Plasma 13.56 MHz	150°C	−0.3
		300°C	+0.02
Si_3N_4	Plasma 50 kHz	350°C	−1.1
Poly Si	LPCVD	560–670°C	−0.1 to −0.3
$TiSi_2$	PECVD	As-deposited	+0.4
		Annealed	+1.2
$TiSi_2$	Sputtered		+2.3
$CoSi_2$	Sputtered		+1.3
$TaSi_2$	Sputtered	800°C anneal	+3.0
$TaSi_2$	Sputtered		+1.2
W	Sputtered	200 to 400 W power	+2 to −2
W	Sputtered	5 to 15 mtorr Ar pressure	−3 to +3
Al			+0.5 to ~+1

[a] Data are mostly from Ref. 31.

variation as a function of deposition variables, nature of film/substrate combination, and film thickness. From this body of work (Refs. 8, 31) the following trends may be gleaned:

(a) In evaporated metal films the stress is invariably tensile with a typical magnitude of ~1 GPa.
(b) The stress in sputtered metal films appears to be 2–3 times higher than for evaporated metals. Such stress values considerably exceed those for the yield stress in bulk metals.

(c) There is no apparent strong dependence of stress on the nature of the substrate.
(d) The magnitude of the stress in nonmetallic films is typically 100–300 MPa.
(e) Both compressive and tensile stresses arise in dielectric films.

A simple way to rationalize the difference in behavior between metals and nonmetals is to note that metals are strong in tension, but offer little resistance to compression, whereas insulators and semiconductors are strong in compression and weak in tension. Therefore, metals prefer to be in tension, whereas nonmetals are best deposited in a state of compression. Internal stress issues peculiar to specific classes of films and deposition methods will be aired in the next few sections.

12.5.2 Stress in Evaporated Metal Films

The early stress data of Fig. 12-13 provide a valuable means for evaluating distinctions and similarities among different metals because of the common *in situ* measurement technique, comparable film-deposition rates, and the absence of a thermal-stress contribution. With few exceptions, the film stress is always tensile. Hard refractory metals and those with high melting points generally tend to exhibit higher residual stresses than softer, more easily melted metals. Normally, for high melting point materials (e.g., refractory metals) the intrinsic stress depends strongly on deposition conditions and often exceeds the thermal mismatch stress. For low melting point materials (e.g., Al, Cu) the intrinsic stress is often less than the thermal mismatch stress (Ref. 32). Appreciable film stress arises after only 100 Å or so of deposition after which large stresses continue to develop up to a thickness of roughly 600 Å. This thickness increment range is, not surprisingly, coincident with typical coalescence and channel stages of growth leading to continuous-film formation. There is thus good reason for the use of the term growth stress to denote intrinsic stress.

With respect to the last item, the extensive research by Abermann and co-workers (Refs. 33, 34), and review by Koch (Ref. 3), has extended the connections between intrinsic stress and the mechanisms of film growth. In Fig. 12-16 a set of force data, continuously recorded from cantilever-beam deflections, is plotted as a function of film thickness (or deposition time). Such *in situ* stress measurements were made on a variety of metal films deposited onto amorphous and crystalline substrates under ultrahigh-vacuum conditions. Subsequent transmission electron microscope observations tracked simultaneous film-formation processes.

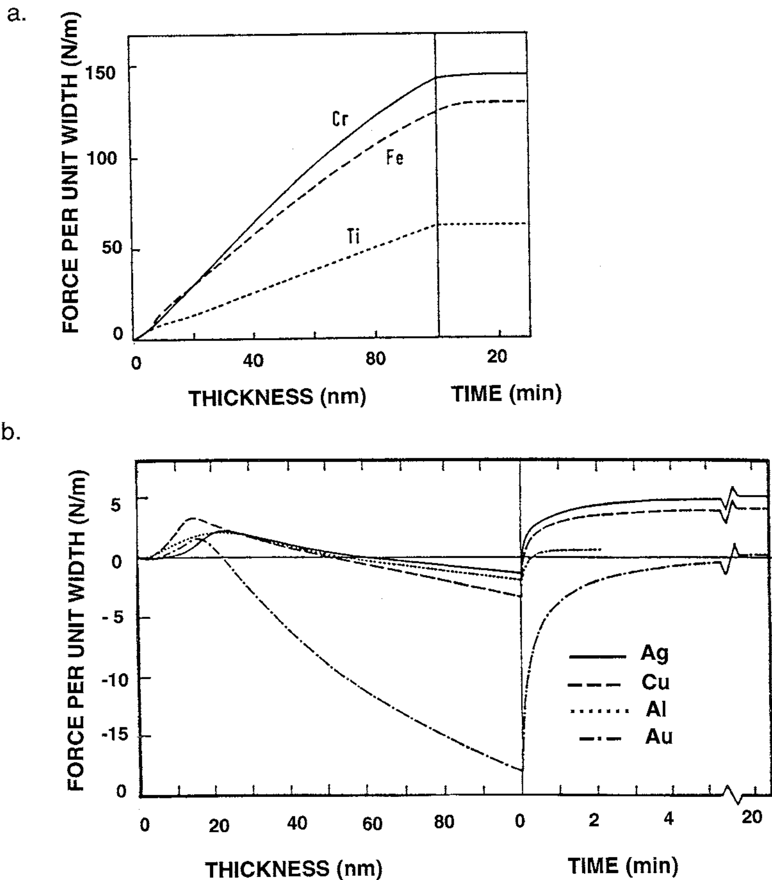


Figure 12-16 Film force per unit width vs film thickness (left-hand side), and time after completion of deposition (right-hand side) for (a) low-mobility growth metals: Ti, Fe and Cr; (b) high-mobility growth metals: Ag, Cu, Al, and Au. (From Ref. 3. Reprinted with the permission of the Institute of Physics Publishing and Dr. R. Koch.)

In the more refractory metals Ti, Fe, and Cr (Fig. 12-16a), the force rose linearly indicating development of intrinsic-stress levels of 0.6, 1.3, and 1.4 GPa, respectively, which remained constant after deposition ceased. These metals display what was termed *low-mobility* Volmer–Weber growth (see Fig. 7-12) and exhibited columnar growth with tensile stresses presumably developed at grain boundaries. On the other hand, *high-mobility* Volmer–Weber growth was observed in the lower melting point metals Ag, Cu, Al, and Au (Fig. 12-16b); here the force response rose initially, peaked, and then reversed sign as a function of film thickness. Intrinsic stresses are typically

more than an order of magnitude smaller in these metals. The initial tensile stress was apparently due to the incomplete coalescence of islands and the formation of low-angle grain boundaries. With further film growth, recrystallization processes served to enlarge the grain size leading to the maximum in the force curve. It is known that in isolated metal particles the lattice parameter (a) is smaller than the bulk value (a_0), but with further film growth a_0 is gradually reached. If this happens when particles are anchored to the substrate by adhesion, capillarity forces arise to induce compressive stresses in them. Therefore we observe film compression beyond the point where the films are continuous. Interestingly, large changes in stress occurred at room temperature after film deposition was complete.

Furthermore, the low-mobility Volmer–Weber force-film thickness response in Cr was essentially converted to high-mobility metal behavior when the substrate temperature was raised. Conversely, high-mobility Volmer–Weber behavior in Ag resembled the typical refractory-metal, low-mobility response when the temperature was reduced. These observations convincingly demonstrate that intrinsic stress in polycrystalline films is strongly dependent on the mobility of atoms during film growth. Heated substrates relax the intrinsic growth stresses by promoting defect annealing and the processes of recrystallization and grain growth if the temperature is high enough. Diffusion of impurities into and out of the film is also accelerated and this can give rise to substantial stress change. Changing the oxygen partial pressure in the deposition chamber has altered film-stress levels and even caused reversals in stress sign (Ref. 3). One implication of stress reversal is the possibility of depositing films with low or even near-zero stress levels. This is frequently a desirable feature, e.g., for magnetic thin films. To achieve the critical properties required, film composition, deposition rates, and substrate temperature must be optimally adjusted.

12.5.3 Sputtered Films

Unlike evaporated films, generalizations with respect to stress are difficult to make for sputtered metal-films because of the complexity of the plasma environment and deposition process. A credible theory for residual stress in sputtered films must explain the dependence on sputtering conditions and film properties, the role of gas pressure, and the observed stress reversals during deposition. Each is addressed in turn.

12.5.3.1 Dependence on Sputtering Variables and Film Properties

Windischmann (Ref. 35) has comprehensively reviewed the subject of intrinsic stress in sputtered films and discussed attempts to explain the

origins of stress. Based on the idea that the volumetric distortion, i.e., strain, is related to the fraction of atoms displaced from equilibrium sites during sputtering, the following equation for film stress was derived:

$$\sigma = \frac{k\Phi\sqrt{E}MY}{(1-\nu)\rho}, \tag{12-29}$$

where Φ and E are the ion flux and energy, respectively, and k is a constant. At low substrate temperatures compressive stresses are often observed. Moreover, at fixed sputtering conditions a linear correlation between the magnitude of compressive stress and product of the film biaxial-elastic modulus and molar volume, i.e., $\sigma \sim [Y(\mathrm{f})/1-\nu(\mathrm{f})](M/\rho)$, has been verified for a number of ion beam-, magnetron-, triode-, and diode-sputtered films (Ref. 36). In this process-independent correlation, quantities M and ρ are the film molecular weight and density, respectively. Furthermore, the effect of substrate temperature on film stress has been investigated for a variety of materials and sputtering variables. Not surprisingly, high temperatures are observed to reduce the magnitude of compressive stress. That the temperature dependence of intrinsic stress is independent of material, stress level, or deposition technique attests to the generality of stress-relief mechanisms such as grain growth, grain-boundary relaxation, and recrystallization.

12.5.3.2 The Effect of Sputtering Gas Pressure

Since the extent of film compression varies directly with the amount of trapped sputtering gas the latter is viewed as an important source of stress. In this regard the more than decade-long collaboration between Hoffman and Thornton on studies of stress in sputtered films laid the foundation of our understanding of this subject. A perspective and review of this research, which was initially motivated by cracking of decorative chromium films sputtered onto plastic automotive parts, has been provided by Hoffman (Ref. 37). The results on magnetron-sputtered metal films are particularly instructive since the internal stress was shown to correlate directly with microstructural features and physical properties. It was found that two distinct regimes, separated by a relatively sharp boundary, exist where the change in film properties is almost discontinuous. The transition boundary can be thought of as a multidimensional space of the materials and processing variables involved. On one side of the boundary the films contain compressive intrinsic stresses and entrapped gases, but exhibit near-bulklike values of electrical resistivity and optical reflectance. This side of the boundary occurs at low sputtering pressures, with light sputtering gases, high-mass targets, and low deposition rates. On the other hand, elevated

sputtering pressures, more massive sputtering gases, light target metals, and oblique incidence of the depositing flux favor the generation of films possessing tensile stresses containing lesser amounts of entrapped gases.

Viewed from the plasma, low gas-pressures mean few collisions so that neutralized or reflected working gas molecules can directly peen the film surface, compressing it. High gas pressures, however, mean many scattering collisions so that energetic-atom bombardment of the growing film is hindered; tensile stresses then arise in the film. From the standpoint of the zone structure of sputtered films (Section 9.2.1.3), elevated working gas pressures are conducive to development of columnar grains with intercrystalline voids (zone 1). Such a structure exhibits high resistivity, low optical reflectivity, and tensile stresses. At lower pressures the development of the zone 1 structure is suppressed. For zone T-type dense films, energetic-particle bombardment, mainly by sputtered atoms, apparently induces compressive film stress by an atomic-peening mechanism.

12.5.3.3 Stress Reversals

Internal stress as a function of the Ar pressure is shown in Fig. 12-17a for planar-magnetron-deposited Cr, Mo, Ta, and Pt, and in all cases there is a reversal in internal stress (Ref. 38). The pressure at which the reversal occurs is plotted in Fig. 12-17b vs the atomic mass of the metal. In the case of Mo the pressure where stress changes sign was observed to decrease as the mass of the inert sputtering gas increased, i.e., from Ne, Ar, Kr to Xe. Although the bulk of the research was carried out on metals, a film-stress transition was also observed in sputtered amorphous silicon. One of the significant findings of this body of work is that almost every sputter-deposition variable spans a range where a stress reversal can be induced. A collection of these variables and their general effects are entered in Table 12-2. For example, increasing negative substrate-bias levels in RF diode sputtered molybdenum at first induced tensile stresses which peaked at -80 V and then became compressive beyond -100 V (Ref. 39). Similarly, in RF-diode sputtered tungsten films a tension to compression stress-reversal was achieved in no less than three ways (Ref. 40), i.e.:

(a) Raising the power level above 30 W at zero substrate-bias

(b) Reversing the DC bias from positive to negative

(c) Reducing the argon pressure

In conclusion, self-bombardment by depositing particles appears to play the dominant role producing dense, compressive films; backscattered neutral bombardment, on the other hand, plays a secondary role in this regard.

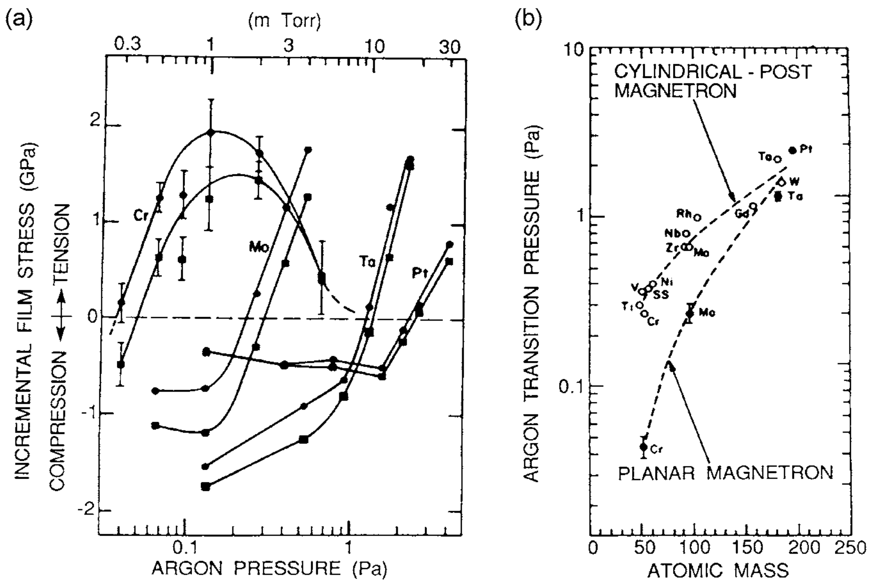


Figure 12-17 (a) Biaxial internal stresses as a function of Ar pressure for Cr, Mo, Ta, and Pt films sputtered onto glass substrates: ●, parallel and ■ perpendicular to long axis of planar cathode. (From Refs. 37, 38.) (b) Ar transition pressure vs atomic mass of sputtered metals for tensile-to-compressive stress reversal. (From Refs. 37, 38.)

Table 12-2

Effects of Sputtering Process Variables on Film Stress[a]

Variable	Compressive	Tensile
Gas pressure	Low	High
Substrate bias	Negative	Positive
Gas atomic mass	Low	High
Target atomic mass	High	Low
Angle of deposition	Normal	Oblique
Angle of emission	Oblique	Normal
Target shape	Cylindrical	Planar
Cathode power	High	Low

[a] After Hoffman (Ref. 37).

Understanding the interplay of these and other deposition variables is critical if stress and other film properties are to be controlled.

12.5.4 CVD Films

Since these films are generally deposited at high temperatures one might expect the extrinsic contribution to the residual stress at 300 K to be high (large ΔT), and the intrinsic stresses to be small (greater structural equilibration). This is a gross simplification, however, because the sign and magnitude of film stress depend on many materials and process variables (Refs. 31, 41). Among the former we cite the nature of the substrate, precursor gas compositions and their stoichiometric ratio; for the latter, the type of CVD, i.e., thermal or plasma, deposition temperature, plasma conditions, and reactor pressure are influential variables. To focus the discussion, the films considered will be restricted to those employed in silicon technology and will include semiconductor, insulator, and metallic materials deposited on a Si wafer. We consider polysilicon first.

1. *Poly Si.* Since the film and substrate have virtually the same thermal expansion coefficient ($\alpha_{Si} \simeq 4 \times 10^{-6}\,K^{-1}$), the compressive stress of 0.1 to 0.3 GPa typically observed is essentially all intrinsic in nature. Suggested sources of the stress include grain growth during annealing, oxygen contamination, and dopant incorporation (Ref. 31). Subsequent ion implantation of dopants induces compressive stresses.

2. *CVD SiO_2.* Table 12-1 indicates that these films deposit in a state of tension ranging up to +0.4 GPa. Assuming $\alpha_{SiO_2} = 0.5 \times 10^{-6}\ K^{-1}$, $Y =$ 83 GPa, $\nu = 0.17$, and a deposition temperature of 700°C, the calculated extrinsic stress contribution, using Eq. 12-20, is −0.14 GPa. The origin of the tensile stress in CVD oxides is thought to be random arrival of molecules at the surface in a loose packing followed by gradual crosslinking (Ref. 31). Stress in phosphosilicate glass films is tensile and is greatly reduced with increasing P content (Ref. 41). Plasma CVD oxides are compressively stressed, on the other hand. Energetic bombardment of the growing film by plasma species is cited for the cause of the stress.

3. *Silicon nitride.* In these much studied films it is found that thermal CVD films prepared at high temperatures (~900°C) have intrinsic tensile stress levels of 1.2–1.8 GPa, while lower compressive stresses ranging from 0.1 to 0.8 GPa are measured in plasma-deposited films (Ref. 42). For the latter, incorporated hydrogen is believed to be the source of the compressive stress. In one study (Ref. 43), the residual stress of LPCVD SiN_x films deposited in SiH_4 and NH_3 at ~750°C revealed a stress reversal from

−0.6 GPa to +0.8 GPa with increasing NH_3/SiH_4 ratio. Apparently, Si-rich (compressive) silicon nitride gives way to n-doped (tensile) silicon films in the competition between silicon crystallization and nitriding effects. Resistance to microcracking in the compressive, plasma-deposited — relative to thermal CVD — films is a significant advantage of the former.

4. *Metals*. Silicides, e.g., $TiSi_2$, deposited by CVD methods tend to exhibit large tensile stresses, a consequence of the large differential thermal expansion relative to silicon. The intrinsic stress contribution tends to be small in comparison. Stress effects are varied in tungsten films. Whereas sputtered films are generally compressive, but can be tensile as we have seen, CVD films are usually tensile as a result of vacancies left behind by process by-products.

With the exception of epitaxial films, perhaps, stress measurements in CVD films appear to be less systematic and extensive than in evaporated and sputtered films. Nevertheless, the brief sampling given here suggests a very rich array of stress effects due to the interplay of chemical and process related variables. Capitalizing on stress reversals offers possible avenues for depositing stress-free films.

12.5.5 Some Theories for Intrinsic Stress

Over the years many investigators have sought universal explanations for the origin of the constrained shrinkage that is responsible for the intrinsic stress. Buckel (Ref. 44) classified the conditions and processes conducive to internal stress generation into the following categories, most of which have already been discussed or alluded to.

1. Differences in thermal expansion coefficients of film and substrate (f-s)
2. Incorporation of atoms, e.g., residual gases, or chemical reactions (f)
3. Lattice mismatch between films and substrates during epitaxial growth (f-s)
4. Variation of the interatomic spacing with crystal size (f)
5. Recrystallization processes (f)
6. Microscopic voids and special arrangements of dislocations (f, f-s)
7. Phase transformations (f, f-s)

More recently, Koch (Ref. 3) has divided the mechanisms into those that are operative within films (f) and at the film–substrate interface (f-s). Examples of these are identified above. Let us consider the case of a film–substrate combination where a phase transformation yields precipitates solely within the film matrix. Volume differences between the involved phases may mean

localized compression of the precipitates and corresponding tension in the matrix, or vice versa; no bending of the film–substrate occurs if the combination is unconstrained. However, if film and substrate react to produce an interfacial phase, bending would probably occur.

12.5.5.1 Recrystallization Effects

One of the mechanisms, not previously discussed, which explains the large intrinsic tensile stresses observed in metal films is related to item 5 in the preceding classification. In the model by Klokholm and Berry (Ref. 26) the stress arises from annealing and shrinkage of disordered material buried behind the advancing surface of the growing film. The magnitude of the stress reflects the amount of disorder present on the surface layer before it is covered by successive condensing layers. If the film is assumed to grow at a steady-state rate of $\dot{G}$ monolayers/s, the atoms will on average remain on the surface for a time $\dot{G}^{-1}$. In this time interval, thermally activated atom movements occur to crystallographically order (recrystallize) the film surface. These processes occur at a rate r described by an Arrhenius behavior

$$r = v_0 \exp - (E_R/RT_S) \tag{12-30}$$

where v_0 is a vibrational frequency factor, E_R is an appropriate activation energy, and T_S is the substrate temperature. On this basis it is apparent that high growth stresses correspond to the condition $\dot{G} > r$ and low growth stresses to the reverse case.

At the transition between these two stress regimes, $\dot{G} = r$, and $E_R/RT_S = 32$, if $\dot{G}$ is 1 s^{-1} and v_0 is taken to be 10^{14} s^{-1}. Experimental data in metal films generally show a steep decline in stress when $T_M/T_S = 4.5$ where T_M is the melting point. Therefore, $E_R = 32RT_M/4.5 = 14.1T_M$. In Section 11.2.2 it was shown that for FCC metals the self-diffusion activation energies are proportional to T_M as $34T_M$, $25T_M$, $17.8T_M$, and $13T_M$ for lattice, dislocation, grain boundary, and surface diffusion mechanisms, respectively. The conclusion is that surface diffusion of vacancies most likely governs the temperature dependence of film growth stresses by removing the structural disorder at the surface of film crystallites.

12.5.5.2 Stress Due to Crystallite Coalescence at Grain Boundaries

Hoffman was the first (Ref. 45) to address stress development due to coalescence of isolated crystallites when forming a grain boundary (GB). Through deposition neighboring crystallites enlarge until a small gap exists between them. The interatomic forces acting across this gap cause a constrained relaxation of the top layer of each surface as the grain boundary

forms. As a result a stress given by

$$\sigma = \frac{Y\,\Delta p}{(1-\nu)l_g} \tag{12-31}$$

was predicted where l_g is the average grain size, Δ is the constrained relaxation length or gap, and p is the packing density of the film.

This coalescence mechanism has been reconsidered by Nix and Clemens (Ref. 46), who visualized it in thermodynamic terms. They assumed a hexagonal array of isolated crystallites shown in Fig. 12-18 (top) whose free energy per unit film/substrate interface area may be expressed by

$$G_1 = G_0 + \frac{2d\gamma_{sv}}{(l_g/2)}, \tag{12-32}$$

where G_0 is the contribution to G_1 from the film and film/substrate interface. The second term is the free energy per unit area from the crystallite side faces, where γ_{sv} is the surface free energy. After coalescence (Fig. 12-18 (bottom)), the free energy (G_2) becomes

$$G_2 = G_0 + \frac{d\gamma_{GB}}{(l_g/2)} + \frac{Yd(\Delta/l_g)^2}{1-\nu} \tag{12-33}$$

where the second term is the GB free energy, with γ_{GB} the interfacial grain-

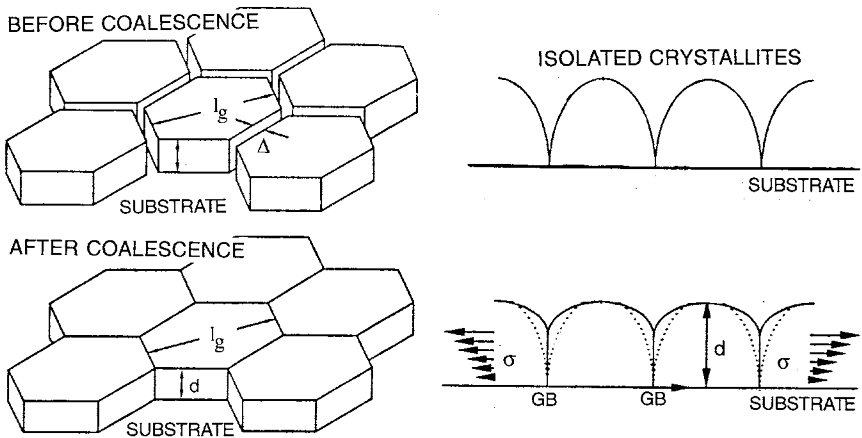


Figure 12-18 Polycrystalline film grains (top) before and (bottom) after crystallite coalescence. On the left the coalescing crystallites are seen. The elastic displacements and stresses associated with forming a continuous film are shown on the right. (After Ref. 46. Reprinted with the permission of the authors.)

boundary energy, and the third term is the strain energy, all per unit film area. As in the Hoffman theory the quantity Δ/l_g may be viewed as the biaxial strain produced on closing the GB gaps by elastic displacements. By equating G_1 and G_2, a *maximum* value of Δ is estimated to be

$$\Delta_{max} = [2l_g(2\gamma_{sv} - \gamma_{GB})(1 - \nu)/Y]^{1/2}.$$

Substitution into Eq. 12-31 yields a corresponding maximum film stress of

$$\sigma_{max} = \frac{[2(2\gamma_{sv} - \gamma_{GB})Y]^{1/2}}{l_g(1 - \nu)}, \qquad (12\text{-}34)$$

assuming $p = 1$. For $Y/(1 - \nu) = 100$ GPa and $l_g = 10$ nm, $\gamma_{sv} = 1$ J/m^2, and $\gamma_{GB} = 0.5$ J/m^2, calculation yields $\sigma_{max} = 5$ GPa and $\Delta_{max} = 0.55$ nm.

These same authors have additionally proposed a novel model for the crystallite coalescence process which is also depicted in Fig. 12-18. We can imagine a surface-energy reduction when isolated elliptical crystallites contact and form a grain boundary. But this necessarily establishes an internal tensile-stress distribution in the film plane that serves to physically open the boundary starting at the cusp-like crack. The equilibrium pitting GB formation energy against elastic strain-energy reduction determines how much grain boundary unzipping occurs. Using a fracture mechanics approach, an equation similar to Eq. 12-31 was derived.

A truly quantitative theory for film stress has yet to be developed, and it is doubtful that one will emerge soon that is valid for different film materials and methods of deposition. Uncertain atomic compositions, structural arrangements, and interactions in crystallites and at the film/substrate interface are not easily amenable to a description in terms of macroscopic elastic concepts.

12.6 MECHANICAL RELAXATION EFFECTS IN STRESSED FILMS

12.6.1 INTRODUCTION

Until now we have primarily considered stresses arising during film-formation processes. During subsequent use, the grown-in elastic-plastic state of stress in the film may or may not remain stable. When films are exposed to elevated temperatures or undergo relatively large temperature excursions, internal stresses frequently relax by time-dependent deformation processes induced by thermally activated motion of atoms and defects. As a result, defect concentrations are altered, film surface topography can change,

and stress as well as strain magnitudes may be reduced. In a hierarchy of levels we shall discuss mechanical relaxation in thin films from a dislocation viewpoint, consider phenomena such as creep that embody the collective response of many dislocations, and finally develop macroscopic models for hillock formation and film–substrate fracture. These topics will be explored in materials as diverse as lead-alloy films in superconducting junction devices and thermally grown SiO_2 films in integrated circuits. Relaxation in SiO_2 films is a good place to start because it is readily understood in terms of classical viscoelastic models used in describing silica glasses.

12.6.2 STRESS AND STRAIN RELAXATION IN THERMALLY GROWN SiO_2 FILMS

It was previously noted in Section 11.2.4.2 that a volume change of some 220% occurs when Si is converted into SiO_2. This expansion is constrained by the adhesion in the plane of the Si wafer surface. Large intrinsic compressive stresses are, therefore, expected to develop in SiO_2 in the absence of stress relaxation. A value of 30 GPa has, in fact, been estimated (Ref. 47), but such a stress level would cause fracture of both the Si and SiO_2. Not only does oxidation of Si occur without catastrophic failure, but virtually no intrinsic stress is measured in SiO_2 grown above 1000°C.

1. *Stress relaxation.* To explain the paradoxical lack of stress, let us consider the Maxwell viscous-flow model to analyze the uniaxial compressive stress-relaxation in a slab of SiO_2. The mechanical response reflects that of a series combination of an elastic spring with modulus Y and a dashpot with viscosity η, which maintains an overall constant strain. Assuming an initial stress (σ_i) in such a constrained film, the well-known solution (Ref. 48) to this model yields a time (t)-dependent exponential decay of stress given by

$$\sigma(t) = \sigma_i \exp - (Yt/\eta). \tag{12-35}$$

Substituting $Y = 66$ GPa and $\eta = 280$ GPa-s at 1100°C, the time it takes for the initial stress to decay to σ_i/e is a mere 4.2 s. Oxides grown at this temperature are, therefore, expected to be unstressed. A more advanced treatment of stress relaxation in SiO_2 that incorporates the effects of a stress-dependent viscosity has been published by Rafferty *et al.* (Ref. 49).

2. *Strain relaxation.* In this case the solid can be modeled by a spring and dashpot connected in parallel combination (Voigt model). Under the application of an applied tensile stress, σ, the spring wishes to instantaneously extend, but is restrained from doing so by the viscous response of

the dashpot. As a result, strain relaxation is manifested by creep, an irreversible time-dependent stretching (or contraction) of the solid. It is left for the reader to show that the strain, ε, has a time dependence given by

$$\varepsilon = \sigma/Y[1 - \exp - (Yt/\eta)]. \qquad (12\text{-}36)$$

12.6.3 DISLOCATION MECHANISMS OF RELAXATION IN FILMS

Interest in this subject primarily stems from the generation of misfit dislocations in epitaxial films that have grown thicker than the critical value d_c. As noted in Section 8.3.1, it is the strain energy release that drives dislocation formation, motion, and multiplication parallel to and normal to the film-substrate interface. What is essential for these relaxation processes is that dislocations move. In this regard, the dislocation velocity v has been measured in epitaxial GaAs, InAs, GaInAs, and Ge–Si films (Ref. 50). In general weaker interatomic bonds in compound semiconductors than in covalent elemental semiconductors favor faster relaxation processes in the former. The dependence of v on the effective stress σ_e acting on dislocations is usually given by

$$v = v_0(\sigma_e)^m \exp - (E_v/k_B T) \qquad (12\text{-}37)$$

where v_0 is a constant and m is a number, typically 1 to 2 in Ge and Si but possibly higher in other films. Reflecting the thermally activated nature of dislocation motion is the Boltzmann factor, where E_v is the activation energy. Because of the threat of relaxation in strained layers of Ge–Si, these films have attracted a lot of attention insofar as dislocation dynamics studies is concerned. For (100) Ge–Si films of assorted stoichiometries and thicknesses, E_v was measured to be ~ 2.2 eV for isolated threading dislocations (Ref. 51). Similarly, values of E_v equal to ~ 1 eV were measured in compound semiconductors (Ref. 50).

The kinetic sequence of dislocation events unleashed after the critical epitaxial film thickness is exceeded has been modeled (Ref. 52) by Dodson and Tsao. Qualitatively, relaxation is initially very sluggish because the film stress is relatively low and dislocation multiplication is inhibited by the few sources present. As a result there is a relaxation incubation period. But as the film thickens, the internal stress rises, increasing dislocation velocities. More defects are then produced and through multiplication, exponentially increasing densities of dislocations arise. Strain relaxation rapidly occurs. Eventually, a large fraction of the lattice mismatch is relieved and the film stress drops. This causes a slowing of dislocation motion and defect

multiplication until the film strain gradually approaches some equilibrium strain state. An S-shaped curve resembling that for phase transformations (Fig. 11-20a) tracks the relaxation process.

Most generally, film strains are relaxed by mechanisms that integrate the role of assorted defects, e.g., grain boundaries, dislocations, and vacancies, with processes such as motion, multiplication, interaction, and diffusion that are associated with them. The primary dislocation mechanisms that are involved include the following:

1. *Deflectless flow*. When the stresses are very high, slip planes can rigidly move over neighboring planes. A shear stress equal to $\sim\mu/20$ has been theorized for such flow. Stresses in excess of this value essentially cause very large strain rates. Below the theoretical shear-stress limit the plastic strain rate is zero. Deflectless flow is dominant when the normalized tensile stress (σ/μ) is greater than $\sim 9 \times 10^{-2}$ but is not normally accessed in films.

2. *Dislocation glide*. Under stresses sufficiently high to cause plastic deformation, dislocation glide is the dominant mechanism in ductile crystalline materials. Dislocation motion is impeded by the presence of obstacles such as impurity atoms, precipitates, and other dislocations. In thin films additional obstacles present are the native oxide the substrate and grain boundaries. Thus, the film thickness, d, and grain size, l_g, may be thought of as obstacle spacings in Eq. 12-7. An empirical law for the dislocation glide strain rate $\dot{\varepsilon}_2$ as a function of stress and temperature is

$$\dot{\varepsilon}_2 = \dot{\varepsilon}_0 \left(\frac{\sigma}{\sigma_0}\right) \exp - \frac{E}{k_B T} \tag{12-38}$$

where σ_0 is the flow stress at absolute zero temperature, E is the energy required to overcome obstacles, $\dot{\varepsilon}_0$ is a preexponential factor, and $k_B T$ has the usual meaning.

3. *Dislocation climb*. When the temperature is raised sufficiently, dislocations can acquire a new degree of motional freedom. Rather than be impeded by obstacles in the slip plane, dislocations can circumvent them by climbing vertically and then gliding. This sequence can be repeated at new obstacles. The resulting strain rate of this so-called climb-controlled creep is strongly dependent on stress and is given by

$$\text{at } T > 0.3T_M; \quad \dot{\varepsilon}_3 = A_3 \frac{\mu b}{k_B T} D_{GB} \left(\frac{\sigma}{\mu}\right)^5 \tag{12-39}$$

$$\text{at } T > 0.6T_M; \quad \dot{\varepsilon}_4 = A_4 \frac{\mu b}{k_B T} D_L \left(\frac{\sigma}{\mu}\right)^7. \tag{12-40}$$

Here, D_{GB} and D_L are the thermally activated grain-boundary and lattice diffusion coefficients, respectively, and A_3 and A_4 are constants.

4. *Diffusional creep.* Viscous creep in polycrystalline films can occur by diffusion of atoms within grains (Nabarro–Herring creep) or by atomic transport through grain boundaries (Coble creep). The respective strain rates are given by

$$\dot{\varepsilon}_5 = \frac{A_5 \mu \Omega D_L (\sigma/\mu)}{k_B T d l_g} \tag{12-41}$$

and

$$\dot{\varepsilon}_6 = \frac{A_6 \mu \Omega \delta D_{GB} (\sigma/\mu)}{k_B T d^2 l_g}, \tag{12-42}$$

where in addition to constants A_5 and A_6, Ω is the atomic volume and δ is the grain-boundary width. It is instructive to think of the last two equations as variations on the theme of the Nernst–Einstein equation (Eq. 1-35). The difference is that in the present context the applied stress (force) is coupled to the resultant rate of straining (velocity). Unlike the nonlinear dependence on stress observed for dislocation climb processes, diffusional creep exhibits a linear coupling between $\dot{\varepsilon}$ and σ.

In any relaxation process the dominant mechanism is the one which relaxes strain the fastest. A useful way to represent the operative regime for a given deformation mechanism is through the use of deformation-mechanism maps, first developed for bulk materials (Ref. 53) and then extended to thin films by Murakami *et al.* (Ref. 54). Such a map for a Pb–In–Au film is shown in Fig. 12-19 where the foregoing strain-relaxation mechanisms are taken into account. In constructing it the process exhibiting the largest strain-relaxation rate is calculated at each point in the field of the normalized stress–temperature space. The field boundaries are determined by equating pairs of rate equations for the dominant mechanisms and solving for the resulting stress dependence on temperature. In the next section we shall use this map to track strain relaxation in superconducting films.

12.6.4 Relaxation Effects in Metal Films during Thermal Cycling

Here we consider stress and strain relaxation effects but in different applications where metal films undergo large cyclic temperature excursions.

1. *Stress relaxation.* Integrated-circuit metallizations are often subjected to thermal cycling, e.g., deposition at elevated temperature, cooling to ambient temperature, reheating during subsequent CVD of a dielectric film, and recooling. As an example of stress relaxation effects, consider a pure polycrystalline-aluminum film on a silicon-wafer substrate that is thermally cycled between room temperature and 450°C (Ref. 2). Shown in Fig. 12-20

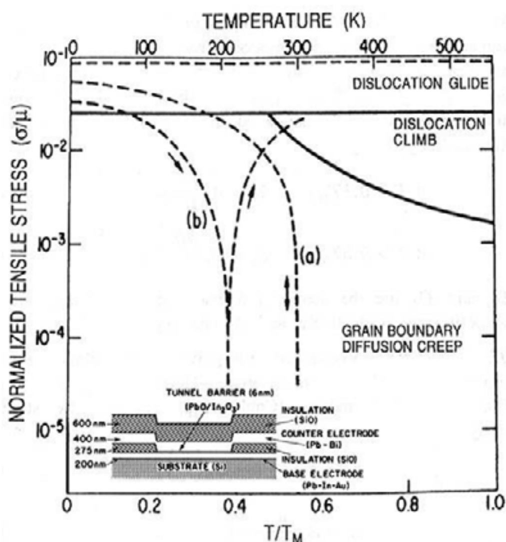


Figure 12-19 Deformation mechanism map for Pb–In–Au thin films. (From Ref. 55.) (Inset) Schematic cross section of Pb alloy Josephson junction device. (From Ref. 54.)

are the results of wafer curvature (Section 12.4.1.2) measurements in a film initially possessing a residual tensile stress of 280 MPa. After subjecting it to thermal cycling, the film is effectively mechanically loaded in a nonisothermal manner. Because $\alpha(\text{Al}) > \alpha(\text{Si})$ and $\Delta T > 0$, the thermal stress becomes increasingly less tensile with temperature and actually turns compressive above $\sim 175°\text{C}$. From Eq. 12-20 the slope of the initial linear response is $Y(\text{Al})\Delta\alpha/1 - \nu(\text{Al})$. Of note is the low magnitude of the yield stress at elevated temperatures where the strength of Al declines. At still higher temperatures the flow stress peaks and then falls. During cooling, $\Delta T < 0$, and the Al film again acquires a tensile stress that reaches its original value at room temperature. Interestingly, repeated thermal stress cycles trace out a reversible stress hysteresis response. However, in alloyed films some combination of structural change, precipitation, densification, compound formation, and stress relaxation may occur, leading to more complex, irreversible stress–temperature–time histories.

2. *Strain relaxation.* An interesting application of *strain relaxation* effects is found in Josephson superconducting tunnel-junction devices (Refs. 54, 55). A schematic cross section of such a device is shown in the inset of Fig. 12-19. The mechanism of operation need not concern us. However, their very fast switching speeds, e.g., $\sim 10^{-11}$ s, combined with low-power dissipation levels, e.g., $\sim 10^{-6}$ W/device, offered promise for ultrahigh-speed

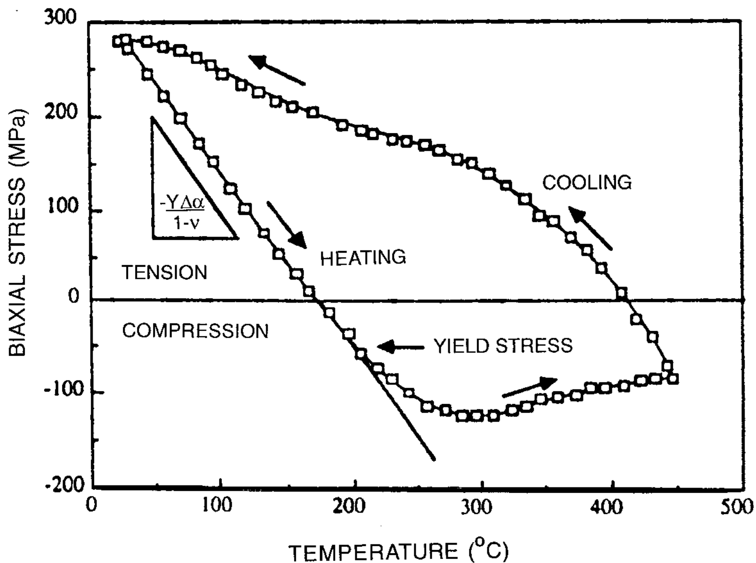


Figure 12-20 Stress–temperature plot for a 590 nm thick Al film on Si thermally cycled between 20 and 450°C. Both elastic and plastic deformation occur during film heating and cooling. (From Ref. 2. Reprinted with the permission of the authors.)

computing at low temperatures; this was prior to the discovery of high-temperature superconductors. The junction basically consists of two superconducting electrodes separated by a 60 Å thick tunnel barrier. Lead-alloy films serve as the electrode materials primarily because they have a relatively high superconducting transition temperature and are easy to deposit and pattern. A serious materials-related concern with this junction structure surfaced during thermal cycling between room temperature and liquid helium temperature (4.2 K) where the device is operated. Failure of some devices occurred by the rupture of the ultrathin tunnel barrier due to thermal-expansion mismatch between Pb alloys and the Si substrate on which the device was fabricated. During temperature cycling the thermal strains, relaxed by the plastic-deformation processes considered earlier, resulted in harmful dimensional changes.

Let us now consider the mechanical history of an initially unstressed Pb film as it is cooled to 4.2 K. Assuming no strain relaxation, path a in Fig. 12-19 indicates that the grain-boundary creep field is traversed from 300 K to 200 K followed by dislocation glide at lower temperatures. Because cooling rates are high at 300 K there is insufficient thermal energy to cause

diffusional creep. Therefore, dislocation glide within film grains is expected to be the dominant deformation mechanism on cooling. If, however, no strain relaxation occurs, the film could then be rewarmed and the $\sigma - T$ path would be reversibly traversed if, again, there is no diffusional creep. Under these conditions the film could be thermally cycled without apparent alteration of the state of stress and strain. If, however, a relaxation of the thermal strain by dislocation glide did occur upon cooling, then the path followed during rewarming would be along b. Because the coefficient of thermal expansion for Pb exceeds that of Si, a large tensile stress initially develops in the film at 4.2 K. As the temperature is raised, dislocation glide rapidly relaxes the stress so that at 200 K the tensile stress effectively vanishes. Further warming from 200 K to 300 K induces compressive film stresses. These provide the driving force to extrude micron-sized protrusions, so-called hillock or stunted whisker growths, from the film surface. This manifestation of strain relaxation is encouraged because grain-boundary diffusional creep is operative in Pb over the sub-room-temperature range.

One way to prevent troublesome hillocks from forming is to strengthen the electrode film. This will minimize the dislocation glide that originally set in motion the train of events leading to hillock formation. Practical methods for strengthening bulk metals include alloying and reducing the grain size in order to create impediments to dislocation motion. Indeed, by alloying Pb with In and Au fine intermetallic compounds form, which both harden the films and refine the grain size. The result was a suppression of strain-relaxation effects and the elimination of hillock formation. Overall, a dramatic reduction in device failure due to thermal cycling was realized.

12.6.5 Stress Relief by Film Cracking

The most extreme form of stress relaxation in thin film structures is cracking or fracture of either the film, the substrate, or the interfacial region between the two. Most contemporary models of fracture mechanics are based on an elegantly simple analysis originally formulated by Griffith in 1920 (Ref. 56) to account for cracking of brittle materials, and glass filaments in particular. In this theory preexistent flaws or cracks were assumed to play a leading role in influencing strength. With suitable modification, fracture mechanics is applicable to all classes of materials, film–substrate combinations, and crack geometries.

The classical analysis assumes the presence of a flat elliptical crack of length $2l$ within a layer of thickness d. An applied or internal tensile stress

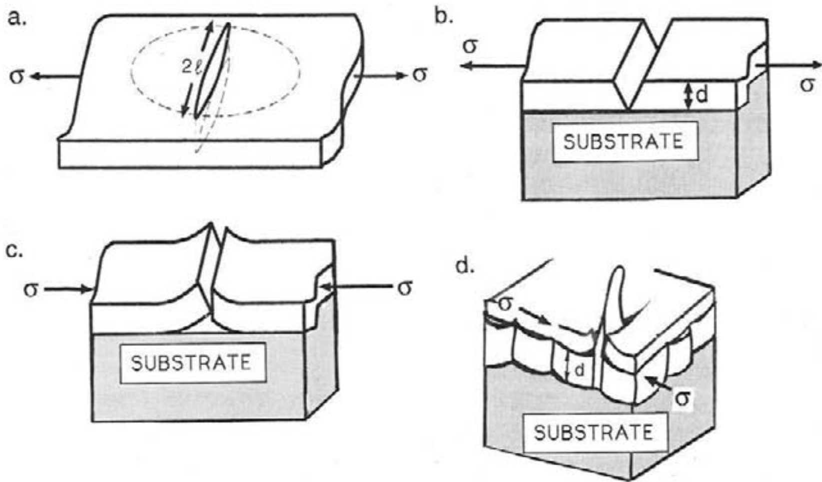


Figure 12-21 Cracking of thin films: (a) Flat elliptical through-crack lying in the film plane loaded uniaxially in tension; (b) crack leading to coating fracture; (c) crack leading to coating delamination; (d) mechanism of metal whisker formation and cracking of overlying brittle film layer.

σ acts in a direction perpendicular to the crack axis as shown in Fig. 12-21a. The process of crack growth releases strain energy in a circular disk volume roughly assumed to be $\pi l^2 d$. Since the elastic strain energy per unit volume is $\sigma^2/2Y$ (Eq. 1-45), the reduction of strain energy (E_S) is equal to $-\pi l^2 d\sigma^2/2Y$. Simultaneously, energy is absorbed at the propagating crack front in breaking bonds between atoms and overcoming the cohesive energy of the material. Because the total crack area is $(2l + 2l)d$ or $4ld$, and it possesses an effective surface energy per unit area γ, the energy (E_γ) contribution from this source is $4ld\gamma$. When the crack is on the verge of growth the rate of decrease of elastic strain energy precariously balances the rate at which surface energy increases. This equilibrium, determined by setting $d(E_S + E_\gamma)/dl = 0$, establishes a relation between the critical stress ($\sigma = \sigma^*$) and l expressed by

$$\sigma^* = (4Y\gamma/\pi l)^{1/2}. \tag{12-43}$$

Mechanical instability and fast fracture occurs when $\sigma > (4Y\gamma/\pi l)^{1/2}$, whereas stability prevails if the inequality is reversed. It is common to express Eq. 12-43 as $\sigma^* = K_{1C}/(\pi l)^{1/2}$, where $K_{1C} = (4Y\gamma)^{1/2}$ is a material property pronounced "kay-one-see" and defined as the critical stress-intensity factor, or simply the fracture toughness. Reflecting their high toughness,

metals have typical K_{1C} value in range of ~ 30 to $200\,\text{MPa-m}^{1/2}$, whereas in brittle ceramics and oxides K_{1C} is usually less than $3\,\text{MPa-m}^{1/2}$.

Instead of an isolated layer let us now consider the more complex problem of a film of thickness d attached to a substrate. For a through crack, either fracture (Fig. 12-21b) or delamination of the film (Fig. 12-21c) could occur under stress relaxation. A simple condition for the former possibility is $\sigma d^{1/2} > K_{1C}(\text{c})/\pi^{1/2}$ where $K_{1C}(\text{c})$ is the fracture toughness of the film. Similarly, film delamination will occur when the stress normal to the surface reaches a value such that $\sigma d^{1/2} > K_{1C}(i)/\pi^{1/2}$. In this expression $K_{1C}(i)$ is equal to $(4Y\gamma_{\text{i}})^{1/2}$ with γ_{i} interpreted as an interfacial debonding energy.

Another interesting case of film cracking involves multilayer films. Oxide films that surround interconnections serve to conformally constrain the powered metal conductors. During electromigration, hillocks and whiskers have been observed to extrude from grain boundaries through cracked insulating overlayers as suggested in Fig. 12-21d. Compressive stress relief in this case is detrimental to reliability because penetration of whiskers can electrically short remote conductors.

12.6.6 Crack Propagation and Spacing

A more rigorous analysis of stress relief due to thin-film cracking has been advanced by Thouless (Refs. 57, 58), who has pointed out interesting parallels between crack advance and dislocation propagation in epitaxial films (see Fig. 8-11). In both mechanisms film strain energy is relaxed as defects (i.e., crack/misfit dislocation) are introduced. Furthermore, both involve creation of an elastic singularity (i.e., crack tip/dislocation core) near the film–substrate interface. Finally, by analogy to the dislocation case, there is a critical film thickness (d_{c}) below which cracking can not occur for a given residual strain ε_0, namely,

$$d_{\text{c}} = \frac{\gamma_{\text{f}}(1 - v)}{(1 + v)Y\varepsilon_0^2}. \qquad (12\text{-}44)$$

In this equation it is assumed that film and substrate have the same elastic constants, and v and γ_{f} are Poisson's ratio and surface energy of the film, respectively.

The spacing of cracks is an issue of interest in brittle films and coatings. Stress relaxation of a crack prevents subsequent cracks from developing unless they are sufficiently far away for the strain energy to build to a critical level. Analysis (Ref. 57) shows that if film and substrate have identical elastic properties, cracks will negligibly interact if they are spaced apart by more

than eight times the film thickness. If they are closer, cracks are not independent and contribute to the strain energy of the film. It was found that the equilibrium spacing (s) of cracks varies as

$$s \sim 5.6d \left\{ \frac{\gamma_f(1-\nu)}{[(1+\nu)Y\varepsilon_0^2 d]} \right\}^{1/2}. \tag{12-45}$$

12.7 ADHESION

12.7.1 INTRODUCTION

The term adhesion refers to the interaction between the closely contiguous surfaces of adjacent bodies, i.e., a film and substrate. According to the American Society for Testing and Materials (ASTM), adhesion is defined as the condition in which two surfaces are held together by valence forces, by mechanical anchoring, or by both together. Adhesion to the substrate is certainly the first attribute a film must possess before any of its other properties can be manifested or exploited. Even though it is of critical importance, adhesion is one of the least understood properties. The lack of a broadly applicable method for quantitatively measuring adhesion makes it virtually impossible to test any of the proposed theories for it. This state of affairs has persisted for years and has essentially spawned two attitudes with respect to the subject (Ref. 59). The "academic" approach is concerned with the nature of bonding and the microscopic details of the electronic and chemical interactions at the film–substrate interface. Clearly, a detailed understanding of this interface is essential to better predict the behavior of the macrosystem, but atomistic models of the former have thus far been unsuccessful when extrapolated to describe the continuum behavior of the latter. For this reason the "pragmatic" approach to adhesion by the thin-film technologist has naturally evolved. The primary focus here is to view the effect of adhesion on film quality, durability, and environmental stability. Whereas the atomic binding energy may be taken as a significant measure of adhesion for the academic, the pragmatist favors the use of large-area mechanical tests to measure the force or energy required to separate the film from the substrate. Both approaches are, of course, valuable in dealing with this difficult subject. We shall adopt aspects of these contrasting viewpoints in the ensuing discussion of adhesion mechanisms, measurement methods, and strategies for enhancing adhesion. All of these topics have been reviewed by Baglin (Ref. 60).

12.7.2 ENERGETICS OF ADHESION

12.7.2.1 A Macroscopic View

From a thermodynamic standpoint, the work (W_A) required to separate a unit area of two phases forming an interface is expressed by

$$W_A = \gamma_f + \gamma_s - \gamma_{fs}. \tag{12-46}$$

The quantities γ_f and γ_s are the specific surface energies of film and substrate, and γ_{fs} is the interfacial energy. A positive W_A denotes attraction (adhesion) while a negative W_A implies repulsion (de-adhesion). W_A is largest and adhesion strongest when materials of high surface energy come into contact, such as metals with high melting points. Conversely, W_A is smallest and adhesion weakest when low surface energy materials such as polymers are brought into contact. When f and s are identical, then an interfacial grain boundary forms where $\gamma_f + \gamma_s > \gamma_{fs}$. Under these circumstances $\gamma_f = \gamma_s$ and γ_{fs} is relatively small, e.g., $\gamma_{fs} = \frac{1}{3}\gamma_s$ in metals. If, however, a homoepitaxial film is involved, then $\gamma_{fs} = 0$ by definition, and $W_A = 2\gamma_s$.

Attempts to separate an epitaxial film from its substrate will likely cause a *cohesion* failure through the bulk rather than an *adhesion* failure at the interface. When the film/substrate combination is composed of different materials, γ_{fs} may be appreciable, thus reducing the magnitude of W_A. Interfacial adhesion failures tend to be more common under such circumstances. In general, the magnitude of W_A increases in the order: (a) immiscible materials with different types of bonding, e.g., metal–polymer, (b) solid-solution formers, and (c) same materials.

12.7.2.2 A Microscopic View

The adsorption theory of adhesion is generally accepted and suggests that when sufficiently intimate contact is achieved at the interface between film and substrate, the surfaces will adhere because of the pairwise interaction of the involved atoms or molecules. It is believed that a large contribution to the adhesion energy is provided by van der Waals (physiosorption) forces. These are classified into London, Debye, and Keesom types depending, respectively, on whether neither, one, or both of the paired atoms possess electric dipoles. Interaction energies between film and substrate atoms typically fall off as the sixth power of the separation distance. The resulting forces are weak and secondary bonding is said to exist with energies of 0.1 eV per atomic pair. In addition to van der Waals forces, chemical interactions (chemisorption) also contribute to adhesion. Stronger primary covalent, ionic, and metallic binding

forces are involved now and bond energies ranging from 1 to 10 eV can be expected.

For a typical interface containing some 10^{15} primary bonds per cm^2 at 1 eV per bond, the total energy is 10^{15} eV/cm^2 or 1.6 J/m^2. Bonding forces can be obtained from the bond-energy variation with separation distance. Suppose, for example, the adhesion energy drops to zero when the surfaces are parted by some 5 Å. The specific adhesion force (stress) is then $F_A = 1.6\ J/m^2/5 \times 10^{-10}$ m or 3.2×10^9 Pa. In contrast, van der Waals adhesion forces are expected to be more than an order of magnitude less or roughly 0.1 GPa. Although secondary bonding forces alone may result in adequate adhesion, the presence of primary bonds can considerably strengthen the joint. Surface-specific analytical techniques such as X-ray photoelectron spectroscopy and SIMS have yielded strong evidence that primary interfacial bonding significantly promotes intrinsic adhesion.

Exchange of charge across film/substrate interfaces also contributes to adhesion. In a capacitor-like configuration, electrical double layers consisting of oppositely charged sheets develop and exert adhesive forces. The attractive force per unit area is $Q^2/2\varepsilon_0$, where Q is the charge density and ε_0 is the permittivity of free space. A value of $Q = 10^{11}-10^{13}$ charges/cm^2, yields attractive forces of 10^3-10^7 Pa. These are small compared to other adhesion force contributions.

12.7.2.3 Adhesion at Various Materials Interfaces

12.7.2.3.1 Metal–Metal

Numerous theories and analyses have appeared over the years (Ref. 61) that either predict values related to W_A or correlate them with other physical properties, e.g., melting point. Invariably the energy of adhesion is calculated as a function of surface separation employing either classical or, more exactly, quantum-mechanical methods. One such equation for the variation of W_A with atomic distance a takes the explicit form (Ref. 61)

$$W_A(a) = -\Delta E[1 + (a - a_0)/S] \exp - [(a - a_0)/S)], \tag{12-47}$$

where ΔE and a_0 are the equilibrium binding energy and atomic separation, respectively, and S is a scaling factor. A universal curve that strongly resembles the one for pairwise interaction energy between atoms vs separation distance (e.g., Fig. 1-8b), emerges. Proceeding in this way adhesion energies between Al–Al, Zn–Zn, Mg–Mg, Al–Zn, Al–Mg, and Mg–Zn pairs have been obtained (Ref. 62). In a similar vein, adhesive energy vs layer separation curves have been reported for Ir–Ir, Ir–Au, and Ir–Al pairs

(Ref. 63). Attempts were made to correlate the calculation with atomic-scale metal adhesion by scanning tunneling microscopy measurements employing atomically sharp Ir tips. It was claimed that for the Ir–metal pairs tested the results were well described by bulk adhesion theory.

In general, measured values of adhesion differ from those of W_A because of contributions from chemical interactions, interdiffusional effects, internal film stresses, interfacial impurities, imperfect contact, etc.

12.7.2.3.2 *Metal–Oxide*

Since metal films are often deposited on nonmetallic substrates there has been much interest in adhesion between these dissimilar materials (Ref. 60). Not unexpectedly, values of W_A vary widely. For example, on Al_2O_3 substrates W_A ranges from close to 3 J/m^2 for Fe–15Cr and Co–40Ni, but is less than 0.5 J/m^2 for Cu, Ni, and Fe–40Ni; adhesion is excellent in the former alloys but poor in the latter metals. Theoretical modeling of adhesion energies employing universal energy relations has also been applied to Ag–MgO and Al–MgO pairs. Both theoretically as well as practically, interfacial impurities can either have dramatic beneficial or deleterious adhesion implications.

12.7.3 FILM–SUBSTRATE INTERFACES

The type of interfacial region formed during deposition will depend not only on W_A, but also on the substrate morphology, chemical interactions, diffusion rates, and nucleation processes. At least four types of interfaces can be distinguished, and these are depicted in Fig. 12-22.

1. The *abrupt* interface is characterized by a sudden change from the film to the substrate material within an atomic spacing distance ($\sim$1–3 Å). Such interfaces arise because of the lack of interaction between film and substrate atoms, and low interdiffusion rates. In this type of interface, stresses and defects are confined to a narrow planar region where stress gradients are high. Film adhesion in this case will be low because of easy interfacial-fracture modes. Roughening of the substrate surface will tend to promote better adhesion.

2. The *compound* interface is characterized by a layer or multilayer structure many atomic dimensions thick which is created by chemical reaction and diffusion between film and substrate atoms. Because of the high stresses generated by volumetric changes accompanying reaction, the compounds formed are frequently brittle. Such interfaces arise in oxygen-active metal films on oxide substrates or between intermetallic compounds and

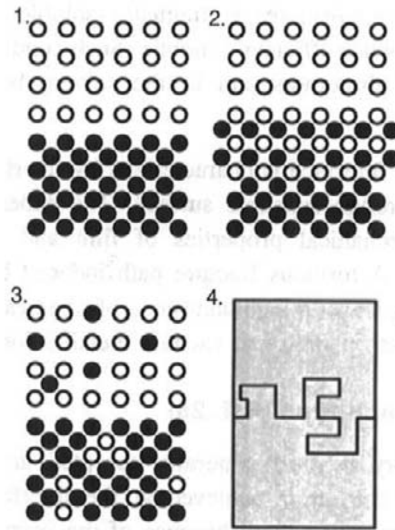


Figure 12-22 Different interfacial layers formed between film and substrate: (1) abrupt interface; (2) compound interface; (3) diffusion interface; (4) mechanical anchoring at interface.

metals. Adhesion is generally good if the interfacial layer is thin, but poorer if thicker layers form.

3. The *diffusion* interface is characterized by a gradual change in composition between film and substrate. Mutual solubility of film and substrate precludes the formation of interfacial compounds. Differing atomic mobilities may cause void formation due to the Kirkendall effect (Section 11.5.4.2). Although this effect tends to weaken the interface, interdiffusion usually promotes good adhesion. Important examples of interdiffusion adhesion are to be found in polymer systems which are widely used as adhesives. In view of the above, interdiffusion of polymer chains across an interface requires that the adhesive and substrate be mutually soluble and that the macromolecules or segments be sufficiently mobile. Such conditions are easily met in the autoadhesion of elastomers and in the solvent bonding of compatible amorphous plastics.

4. The *mechanical* interface is characterized by interlocking of the depositing material with a rough substrate surface. In this case the adhesion strength depends primarily on the mechanical properties of film and substrate as well as interfacial geometry. A tortuous fracture path induced by rough surfaces and mechanical anchoring leads to high adhesion. Mechanical interlocking is relied upon during both electroplating and vacuum metallization of polymers.

12.7.4 STRATEGIES FOR IMPROVING FILM ADHESION

12.7.4.1 Glue Layers

Attaining good adhesion is often an art. Adhesion theories generally do not always provide guidelines on how to achieve good film adhesion in practice. Conventional wisdom, for example, suggests using very clean substrates. This may not necessarily work for metal films on glass substrates because optimum adhesion appears to occur only when the metal contacts the substrate through an oxide bond. Thus Al adheres better when there is some Al_2O_3 present between it and the glass substrate. It is not surprising that strong oxide formers adhere well to glass. Intermediate oxide layers can be produced by depositing metals with large heats of oxide formation such as Cr, Ti, Mo, and Ta. Reactions of the type given by Eq. 1-16a proceed at the interface resulting in good adhesion. Conversely, noble metals such as Au, Ag, and Cu do not form oxides readily and accordingly, adhere poorly to glass, a fact reflected in low film stresses. To promote adhesion it is common practice, therefore, to first deposit a few hundred angstroms of an intermediate oxygen-active metal to serve as the "glue" between the film and substrate. This is the basis of several multilayer-metallization contact systems, including Ti–Au, Ti–Pd–Au and Ti–Pt–Au, Cr–Au, Cr–Pd–Au, Cr–Ag, and Mo–Au. After deposition of the intermediate glue layer, the second film should be deposited without delay, for otherwise the glue metal may oxidize and impede adhesion of the covering metal film. Added layers to improve adhesion are unwelcome because additional deposition steps are costly and often introduce new reliability problems.

12.7.4.2 Use of Ion Beams

That ion bombardment of substrates improves film adhesion has already been mentioned, e.g., Section 4.6.4.3. Baglin (Refs. 64, 65) has reviewed the various mechanisms of ion-beam enhanced-film adhesion and neatly summarized them in Fig. 12-23. Each is briefly considered in turn.

1. *Low-energy presputtering*. Cleaning substrates by a presputtering treatment with ions having energies between 100 eV and a few keV is a good way to enhance subsequent film adhesion. This cleaning procedure removes most contaminants including surface-bonded polymer chains, oil films, and terminal oxide layers. Ion bombardment also roughens the surface and can thus strengthen film–substrate interfaces. At ion energies less than 500 eV, preferential sputtering of compound substrates can enable optimal compositional tailoring as a precursor to film adhesion.

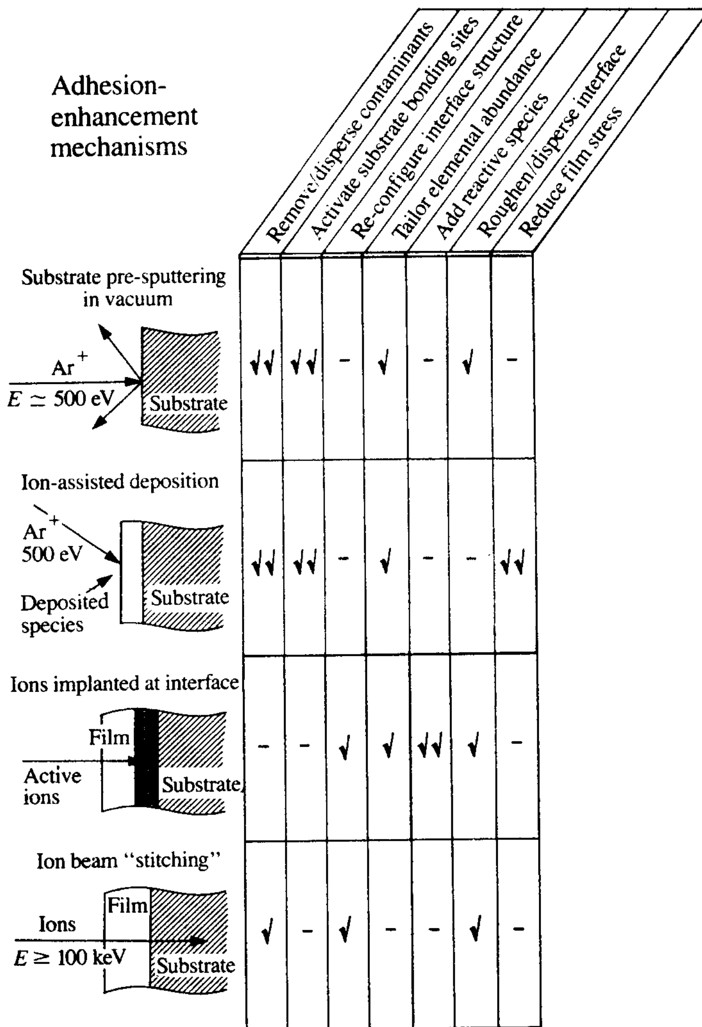


Figure 12-23 Mechanisms for ion-beam enhancement of thin-film adhesion. (After Ref. 65. Reprinted with the permission of the author.)

2. *Ion-beam-assisted deposition.* In this case ion beams with several keV energy at fluxes of $\sim 5 \times 10^{15}$ ions/cm^2 accompany film deposition. Therefore, in addition to the benefits of presputtering, the film is densified. The increased kinetic activity at the surface means near-equilibrium atomic packing and films that are generally freer of stress and excess strain-energy

than otherwise. Large-area Kaufman ion-beam sources are available for such purposes. Under this category, we may include film deposition under simultaneous ion bombardment present during sputtering or ion plating. The interfacial "pseudodiffusion" layer produced is a type of structurally disordered transition zone consisting of backscattered atoms sputtered from the substrate, incoming vapor atoms, and high concentrations of point defects. What condenses is a metastable phase in which the solubility of the involved components exceeds equilibrium limits. All in all, diffusion occurs between materials that do not naturally mix or adhere, strongly promoting adhesion.

3. *Reactive ion implantation.* Whereas the first two processes are carried out at low energies, ion implantation requires an expensive accelerator. However, there are some advantages to reactive ion implantation, including the creation of compositionally graded layers and ballistic mixing at the interface. These effects tend to produce interface layers exhibiting high fracture toughness and integrity.

4. *Ion-beam stitching.* As in the previous method, high-energy ion beams are required. Deep penetration of ions results in sheathlike "collision cascades" that contain much electronic excitation and structural disorder. At the film–substrate interface locally mixed clusters eventually link with other similarly intermixed zones to create a continuous ion-beam mixed or "stitched" layer. Experimentally, ion stitching has improved adhesion in a great many material pairs including metal films and SiO_2, Al_2O_3, semiconductor, and Teflon substrates; the reason it works in such diverse system pairs may be due to the randomness of its action.

A list of publications dealing with ion beam enhanced adhesion in metal–ceramic (oxide, carbide) and metal–polymer systems is given in Ref. 60.

12.7.4.3 Promoting Adhesion to Polymers

In view of the broad use of plastics in diverse applications there is often a need to deposit thin metal films on polymeric substrates. This often proves a challenge because of great differences in atomic bonding, melting points, and thermal expansion coefficients between the involved materials. Nevertheless, metal adhesion has been successfully achieved in applications ranging from web coatings (Section 3.5.3) to polyimide and polyethylene dielectrics employed in electronic packaging. Strategies to improve adhesion broadly attempt to modify polymer surfaces by either physical or chemical means.

Alteration of surface topography, mechanical abrasion, and plasma treatment of surfaces are examples of the first approach. We shall only touch

on plasma modification of surfaces, a subject that has been extensively studied over the years (Refs. 66–68). Plasmas with their collective assortment of ions, electrons, and photons enhance adhesion by removing contaminants, roughening surfaces, promoting crosslinking, and generally introducing or removing reactive chemical groups. In particular, the greater chemical reactivity of the surface often enhances nucleation of deposited films. Both oxygen and nitrogen plasmas are often employed to activate polymer surfaces. Oxygen plasmas tend to make the surface more acidic by creating C=O groups. On the other hand, nitrogen or ammonia treatments result in basic surfaces due to the formation of amine and imine groups. Extensive chain scission must be avoided, however, because it weakens surface and interfacial regions. The trick is thus to incorporate adhesion-promoting entities without degrading the structural integrity of near-surface layers.

Alternatively, treatment with solvents, caustic, and acidic solutions (Ref. 69) and graft polymerization of polar monomers are examples of chemical methods to foster adhesion. Thus, using certain reducing-agent solutions, polyimide surfaces are activated to mediate electron transfer for depositing metals seeds, e.g., Pd, Pt, Ni, and Cu from metalorganic compounds. Thick layers are then electrolytically deposited.

12.7.5 Adhesion Tests

Although there are no ways to directly measure interfacial atomic-bond strengths, there are a number of tests which characterize adhesion practically (Refs. 70, 71). Two types of tests can be distinguished depending on whether tensile or shear stresses are generated at the interface during testing.

12.7.5.1 Tensile-Type Adhesion Tests

The simplest of these include direct pull-off as well as so-called topple tests, and both are used primarily for coatings. As Fig. 12-24a indicates, force is applied to a member glued or soldered to the coating, and the load to cause interfacial separation is measured. Misalignment problems associated with normal pulling are partially overcome by applying a torque in the topple test. The value of F_A is equal to F, the applied force at separation, divided by the contact area A and has stress units.

Acceleration tests also generate tensile stresses in the coating but without the disadvantage of glues and mechanical linkages. In the ultracentrifugal method a coated cylinder is levitated electromagnetically and spun at ever-increasing speed until the coating debonds from the substrate.

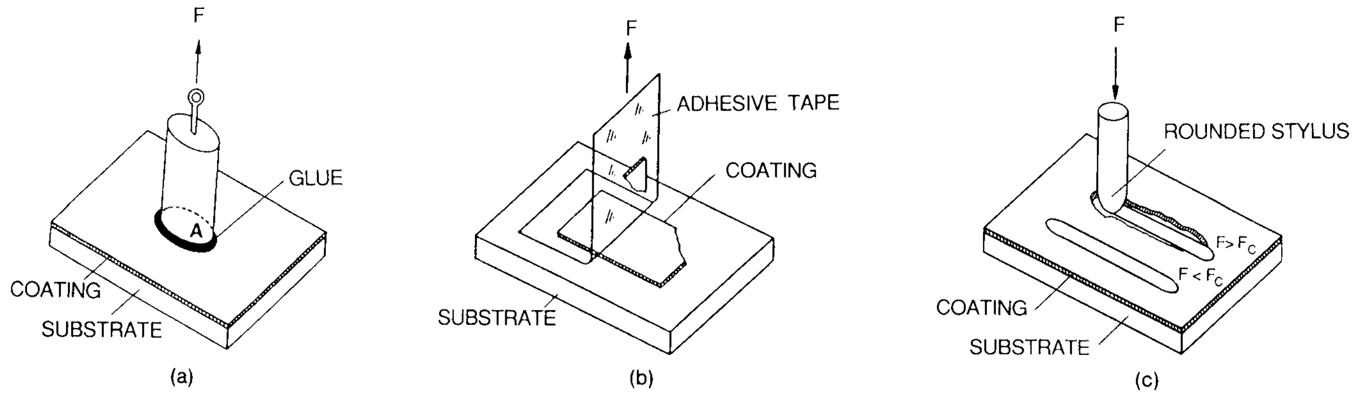


Figure 12-24 Adhesion test methods. (a) pull-off test (Ref. 70); (b) adhesive tape test (Ref. 70); (c) scratch test. (After H. K. Pulker, *Coatings on Glass.* Elsevier, Amsterdam, 1984.)

Pulsed lasers have also been used to measure adhesion effects. When the back of the substrate is exposed to the laser pulse, successive compressive and tensile shock waves rapidly flex the substrate backward and then forward, detaching the coating in the process. Adhesion is characterized by the energy density absorbed per unit area.

12.7.5.2 Shear-Type Tests

The adhesive-tape test developed well over a half-century ago provides the simplest and quickest qualitative measure of film or coating adhesion. Schematically indicated in Fig. 12-24b, the test can, with a little bit of discrimination, distinguish between complete lifting, partial lifting, or complete adhesion. The test can also be made semiquantitative by controlling the angle of pull and the rate of pulling. Using improved adhesives the force required to peel the tape is measured as a function of angle, and the force extrapolated to zero angle is a measure of the adhesion. In such tests it is necessary that the tape/film bond be stronger than the film/substrate bond.

12.7.5.3 Scratch Tests

Scratch testing, shown schematically in Fig. 12-24c, is widely used to evaluate film adhesion. The test consists of drawing a stylus or indenter with known radius of curvature over a film or coating under increasing vertical loads. Resultant scratches are then observed under an optical or scanning electron microscope in order to estimate the minimum or critical load required to scribe away the film and leave behind a clear channel or visible substrate. The elastoplastic deformation is complicated, however, and films can be thinned and appear translucent while still adhering to the substrate. Alternatively, films can remain opaque when detached. Commercial equipment is available to enable measurement of the critical load on the basis of a single scratch. By ramping the indenting load between set limits, followed by visual examination of the scratch, the critical load F_c which just causes adhesion failure can be determined. The scratching process is also accompanied by the emission of acoustic signals which are small when the film adheres at low loads. The onset of large acoustic emission caused by shearing or fracture at the film/substrate interface has been taken as a measure of the critical de-adhesion load. Theoretical analyses relating the critical load, stylus geometry, and scratch dimensions to the specific adhesion force have been made. One such relation is

$$F_A = KH_vF_c/\pi R^2 \tag{12-48}$$

where the magnitude of coefficient K depends on the model details (K can range between 0.2 and 1), H_v is the Vickers hardness, and R is the stylus radius.

At present there is little quantitative agreement in F_A values obtained from different adhesion test methods. Rather, individual tests are well suited to internal comparisons of the same film/substrate combination prepared in different ways.

12.8 CONCLUSION

As in other solids the mechanical behavior of thin films depends on both applied and internal stresses, but unlike them the influence of the latter is paramount in most thin-film applications. That is why a major focus of the chapter is internal stress and its implications. While bulk materials usually acquire residual stresses in response to thermomechanical processing operations, such stresses in films originate during the earliest stages of deposition. That the magnitude of the stresses can exceed typical flow stresses in bulk is surprising. Crystallite coalescence and recrystallization effects are two of the mechanisms that help to explain the origins of *intrinsic* stress in polycrystalline films. And if we recall strained-layer epitaxial films, lattice-misfit strains give rise to (intrinsic) stresses that initiate at the film–substrate interface and grow as the film thickens. In this as well as other deposition and postdeposition treatments, thermal-expansion mismatch between films and substrates provide an important additional source of stress, in this case *extrinsic*. Even though so few atoms are involved, their stresses are sufficient to bow substrates and cause film peeling and fracture. Sensitive measuring techniques involving electromechanical sensors, X-rays, and lasers have made such stress determinations routine. Less routine, perhaps, are the mechanical tests performed to determine film strength or hardness employing nanoindentation and nanoloading methods.

Subjecting films to large temperature excursions is the simplest way to dramatically alter film stress. In metal films where dislocation motion is triggered at relatively low stresses and temperatures, processes such as dislocation glide, climb, and creep allow films to mechanically relax. During thermal cycling of films, alternate strengthening (at low temperatures) and softening (at elevated temperatures) results in scientifically interesting and technically important mechanical hysteresis effects.

Mechanical behavior and film stress cannot generally be discussed apart from the influence of the substrate; this is true of adhesion where forces and stresses basically act normal to the film/substrate interface. It is appropriate

to end the chapter and book on the subject of adhesion and ways to understand and improve it. Without film adhesion to substrates, topics treated in previous chapters would have little practical significance.

EXERCISES

1. Identical metal films of equal thickness, deposited on both sides of a thin substrate strip, are found to possess a residual tensile stress.
   (a) Is the overall strip deformed? Explain.
   (b) One of the films is completely removed by sputter-etching. Qualitatively describe how the remaining film–substrate combination deforms or bows.
2. It is desired to grow epitaxial films of GaSb on AlSb substrates by deposition at 500°C. Refer to Table 8-1.
   (a) What is the expected lattice mismatch at 500°C?
   (b) What thermal stress can be expected in the film at 20°C if Y(GaSb) = 91.6 GPa and ν(GaSb) = 0.3?
3. (a) Based on tensile data for bulk Al the value of the Hall–Petch slope K is 5×10^4 N/m$^{3/2}$. What is the corresponding value of K for Al films, and what accounts for the difference?
   (b) Compare the experimental tensile strengths of bulk and thin-film Al and rationalize the difference in values.
   (c) Calculate the predicted value of the yield stress of Al films by direct substitution in Eq. 12-6. Is it consistent with the stress magnitudes given in part (b)?
4. A 0.4 μm thick $YBa_2Cu_3O_7$ film was deposited on a 0.5 mm thick Si wafer at 700°C by CVD methods. To develop superconducting properties in the film it must be cooled to 90 K. What is the sign and magnitude of wafer bowing under these conditions? Assume for $YBa_2Cu_3O_7$: $Y = 96$ GPa, $\alpha = 11.8 \times 10^{-6}\,°C^{-1}$, and $\nu = 0.25$. For Si assume $Y = 160$ GPa, $\alpha = 4 \times 10^{-6}\,°C^{-1}$, and $\nu = 0.3$.
5. An engineer wishes to determine whether there will be more bow at 20°C in a Si wafer with a 1 μm thick SiO_2 film, or with a 1 μm thick Si_3N_4 film. Both films are deposited at 500°C on a 0.75 mm thick Si wafer. At the deposition temperature the intrinsic stresses are −300 MPa for SiO_2 and −600 MPa for Si_3N_4. If the moduli are $Y(SiO_2) =$ 73 GPa, and $Y(Si_3N_4) = 150$ GPa, and the thermal expansion coefficients are $\alpha(SiO_2) = 0.55 \times 10^{-6}\,°C^{-1}$ and $\alpha(Si_3N_4) = 3.0 \times 10^{-6}\,°C^{-1}$, calculate the radius of curvature for each wafer. Assume Y(Si) = 160 GPa, α(Si) = $4 \times 10^{-6}\,°C^{-1}$, and Poisson's ratio for film and substrate is

0.3. What is the expected radius of curvature in a 30 cm diameter wafer? Calculate the difference in height between the edge and center of the wafer.

6. A 5000 Å thick Al film is deposited stress free on a 20 cm diameter Si wafer (0.5 mm thick) at 250°C such that there is no stress relaxation on cooling to 20°C. Next, the Al–Si combination is heated to 500°C where Al completely relaxes. A 2 μm thick Si_3N_4 film is then deposited with an intrinsic compressive stress of 700 MPa. What is the final radius of curvature after cooling to 20°C? Note for Al, $Y = 66$ MPa and $\alpha = 23 \times 10^{-6}\,°C^{-1}$. Other material constants are given in the previous exercise. Assume $\nu = 0.3$ for all materials.

7. Consider thin-film multilayers where for adjacent films 1 and 2 the corresponding film thicknesses, moduli, and unstrained lattice parameters are $d(1)$, $Y(1)$, and $a_0(1)$ and $d(2)$, $Y(2)$, and $a_0(2)$. There is a common lattice parameter, $\langle a_0 \rangle$, at the interface between films.
   (a) What is the strain in each film?
   (b) What are the corresponding stresses?
   (c) If the forces are equilibrated, show that

$$\langle a_0 \rangle = a_0(1)a_0(2)\left(\frac{1 + \dfrac{Y(2)d(2)}{Y(1)d(1)}}{a_0(2) + \dfrac{Y(2)d(2)a_0(1)}{Y(1)d(1)}}\right).$$

8. Consider a 250 Å Si/75 Å $Ge_{0.4}Si_{0.6}$ superlattice. The [100] moduli for Ge and Si are $Y(\text{Ge}) = 141$ GPa and $Y(\text{Si}) = 181$ GPa, while $a_0(\text{Ge}) = 5.66$ Å and $a_0(\text{Si}) = 5.43$ Å are the corresponding lattice parameters. If the properties of $Ge_{0.4}Si_{0.6}$ are assumed to be derived from weighted composition averages of pure component properties, find
   (a) the common interfacial (in-plane) lattice parameter using the results of the previous problem
   (b) the strains and stresses in the Si and $Ge_{0.4}Si_{0.6}$ layers
   (c) the strained lattice parameters normal to the film layers
   Assume Poisson's ratio is 0.37.

9. During the evaporation of thin films suppose the competition between atom deposition (at rate $\dot{G}$) and recrystallization processes changes the instantaneous concentration (C_d) of disordered film regions by

$$dC_d/dt = K_1\dot{G} - K_2C_d,$$

   where K_1 is a constant. The second term on the right reflects a first-order annealing reaction where constant K_2 varies as $\exp - (E_R/RT_S)$.
   (a) Solve for C_d as a function of time.

(b) Assume C_d is proportional to the film stress. Physically explain the development of stress over the time interval from 0 to $\dot{G}^{-1}$ and as a function of substrate temperature.

10. (a) The force per unit width (S) in a metal film was monitored during evaporation as a function of film thickness (d) deposited. It was observed that the behavior parallelled that of Fig. 12-16a at low substrate temperatures, whereas at elevated substrate temperatures it resembled that of Fig. 12-16b. Furthermore, the slope of the S–d response was generally steeper with higher background O_2 partial pressures, and films relaxed less at room temperature with time after deposition halted. Explain these findings.
(b) Rationalize the effects of gas pressure, substrate bias, gas atomic mass, and target atomic mass on the sign of the stress in sputtered films.

11. (a) Derive Eqs. 12-35 and 12-36 for stress and strain relaxation, respectively. Sketch the respective stress–time and strain–time responses for each.
(b) Real solids often behave as combinations of Maxwell and Voigt models. Consider a parallel spring–dashpot pair in series with a spring. A tensile load is applied to this structure. Sketch the resultant strain–time response.
(c) If film stress relaxation occurs by dislocation motion (i.e., Eq. 12-37), how is this fact specifically incorporated into Eq. 12-35?

12. Consider a substrate of thickness d_s containing deposited films of thickness d_f on either side that are uniformly stressed in tension to a level of σ_f. The substrate is assumed to be uniformly compressed. Assume film and substrate have the same elastic constants.
(a) Determine the substrate stress, assuming force equilibrium prevails.
(b) Show that under the foregoing conditions the net moment with respect to an axis at the center of the substrate vanishes.
(c) One film is rapidly annealed so that its stress vanishes. The substrate and other film are unaffected in the process. What is the net force imbalance or resultant force? What is the net moment imbalance or resultant moment?
(d) In the absence of external constraints the film–substrate will elastically deform to find a new equilibrium stress distribution with zero resultant force and moment. A uniform force as well as a moment (arising from a linear force distribution through the film–substrate cross section) is required to counter the mechanical imbalance of part (c). What is the stress contribution to the remaining film from

the uniform force? What is the maximum stress contribution to the remaining film from the moment?

(e) What is the new maximum stress in the remaining film and what sign is it?

13. Stress relaxation due to creep is analogous to the outdiffusion or "leakage" of a supersaturated species. The process is governed by the diffusion equation where stress σ replaces concentration, i.e.,

$$\frac{d\sigma(x,t)}{dt} = \frac{D\,d^2\sigma(x,t)}{dx^2}.$$

Consider the outdiffusion of stress from a film of thickness d, initially compressed to stress level σ_i. If stress-free surfaces at $x = 0$ and $x = d$ are maintained, show that the stress relaxes according to

$$\sigma(x,t) = \frac{4\sigma_i}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin[(2n+1)\pi x/d \exp - [(2n+1)^2\pi^2 Dt]/d^2.$$

14. (a) Using Eq. 12-47, plot the energy of adhesion vs $(a - a_0)/S$ over the range -1 to 10 for the case of Ir–Ir surfaces where $\Delta E = 0.594$ eV and $a_0 = 0.268$ nm. Assume $S = 1$.

(b) Estimate a value for Young's modulus if the interface layer undergoes tensile straining. Assume $Y = (a_0)^{-1}(d^2W_A/da^2)_{a_0}$.

15. If the yield stress (σ) normalized to its maximum value (σ_0) in sputtered metal films is plotted as σ/σ_0 vs normalized temperature (T/T_M), where T_M is the melting point, the data roughly fall between the curves $\sigma/\sigma_0 = 1 - 11(T/T_M)^2$ and $\sigma/\sigma_0 = 1 - 4(T/T_M)^2$. Ordinate values range between $\sigma/\sigma_0 \simeq 0.05$ near T_M, and 1 when $T/T_M \simeq 0.1$.

(a) If stress levels of $\sigma/\sigma_0 = 0.3$ can be tolerated, what is the maximum substrate temperatures you would recommend when sputtering Mo and Cu?

(b) Suppose maximum substrate temperatures of $0.4T_M$ are practical when sputtering Au and Al. What film stress levels would be reached for these metals?

REFERENCES

1. M. Ohring, *Reliability and Failure of Electronic Materials and Devices*. Academic Press, Boston, 1998.
2. W. D. Nix, *Metallurg. Trans.* **20A**, 2217 (1989).
3. R. Koch, *J. Phys.: Condens. Matter* **6**, 9519 (1994).
4. D. A. Hardwick, *Thin Solid Films* **154**, 109 (1987).
5. R. P. Vinci and J. J. Vlassak, *Ann. Rev. Mater. Sci.* **26**, 431 (1996).

6. The theme of the July 1992 issue of the *MRS Bulletin* (Vol. XVII) is the mechanical behavior of thin films.
7. G. M. Pharr and W. C. Oliver, *MRS Bull.* **XVII**(7), 28 (1992).
8. R. W. Hoffman, in *Physics of Thin Films*, Vol. 3, eds. G. Hass and R. E. Thun. Academic Press, New York, 1966.
9. I. C. Noyan, T. C. Huang, and B. R. York, *Crit. Rev. Solid State Mater. Sci.* **20**(2), 125 (1995).
10. C. A. Neugebauer, *J. Appl. Phys.* **32**, 1096 (1960).
11. L. E. Trimble and G. K. Celler, *J. Vac. Sci. Tech.* **B**7, 1675 (1989).
12. R. P. Keatch and B. Lawrenson, *Microelectron. J.* **25**, 393 (1994).
13. T. P. Weihs, S. Hong, H. C. Bravman, and W. D. Nix, *J. Mater. Res.* **3**, 931 (1988).
14. K. E. Petersen and C. R. Guarnieri, *J. Appl. Phys.* **50**, 6761 (1979).
15. A. J. Griffin, F. R. Brotzen, and C. F. Dunn, *Thin Solid Films* **220**, 265 (1992).
16. J.-A. Schweitz, *MRS Bull.* **XVII**(7), 34 (1992).
17. L. B. Freund, *J. Appl. Mech.* **54**, 553 (1987).
18. Y.-S. Kang and P. S. Ho, *J. Electron. Mater.* **26**(7), 805 (1997).
19. B. M. Clemens, H. Kung, and S. A. Barnett, *MRS Bull.* **24**(2), 20 (1999).
20. J. S. Koehler, *Phys. Rev. B* **2**, 547 (1970).
21. G. Gore, *Trans. Roy. Soc. (London) Part 1*, 185 (1858).
22. G. C. Stoney, *Proc. Roy. Soc. London* **A32**, 172 (1909).
23. K.-N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science for Electrical Engineers and Materials Scientists*. Macmillan, New York, 1992.
24. E. Suhir, in *Advances in Thermal Modeling of Electronic Components and Systems*, Vol 1, eds. A. Bar-Cohen and A. D. Krauss. Hemisphere, New York, 1988.
25. E. A. Johnson, W. T. Chen, and C. K. Lim, in *Principles of Electronic Packaging*, eds. D. P. Seraphim, R. Lasky, and C.-Y. Li. McGraw-Hill, New York, 1989.
26. E. Klokholm and B. S. Berry, *J. Electrochem. Soc.* **115**, 823 (1968).
27. R. E. Cuthrell, D. M. Mattox, C. R. Peeples, P. L. Dreike, and K. P. Lamppa, *J. Vac. Sci. Technol.* **A6**(5), 2014 (1988).
28. B. M. Clemens and J. A. Bain, *MRS Bull.* **XVII**(7), 46 (1992).
29. J. L. Jordan-Sweet, I. C. Noyan, E. Liniger, S. K. Kaldor, and P.-C. Wang, *Adv. X-Ray Anal.* **42**, 569 (2000).
30. D. S. Campbell, in *Handbook of Thin Film Technology*, eds. L. I. Maissel and R. Glang. McGraw-Hill, New York, 1970.
31. S. M. Hu, *J. Appl. Phys.* **70**, R60 (1991).
32. J. M. E. Harper and K. P. Rodbell, *J. Vac. Sci. Technol.* **B15**, 763 (1997).
33. G. Thurner and R. Abermann, *Thin Solid Films* **192**, 277 (1990).
34. R. Abermann and R. Koch, *Thin Solid Films* **129**, 71 (1985).
35. H. Windischmann, *Crit. Rev. Solid State Mater. Sci.* **17**(6), 547 (1992).
36. H. Windischmann, *J. Appl. Phys.* **62**, 1800 (1987).
37. D. W. Hoffman, *J. Vac. Sci. Technol.* **A12**(4), 943 (1994).
38. D. W. Hoffman and J. A. Thornton, *J. Vac. Sci. Technol.* **20**, 355 (1982).
39. J. A. Thornton, in *Semiconductor Materials and Process Technology Handbook*, ed. G. E. McGuire. Noyes, Park Ridge, NJ, 1988.
40. R. W. Wagner, A. K. Sinha, T. T. Sheng, H. J. Levinstein, and F. B. Alexander, *J. Vac. Sci. Technol.* **11**, 582 (1974).
41. W. Kern, G. L. Schnable, and A. W. Fisher, *RCA Rev.* **37**, 3 (1976).
42. A. Sherman, *Chemical Vapor Deposition for Microelectronics*, p. 121. Noyes, Park Ridge, NJ, 1987.
43. P. Temple-Boyer, C. Rossi, E. Saint-Etienne, and E. Scheid, *J. Vac. Sci. Technol.* **A16**(4),

2003 (1998).
44. W. Buckel, *J. Vac. Sci. Technol.* **6**, 606 (1969).
45. R. W. Hoffman, *Thin Solid Films* **34**, 185 (1976).
46. W. D. Nix and B. M. Clemens, *J. Mater. Res.* **14**, 3467 (1999).
47. E. A. Irene, E. Tierney and J. Angilello, *J. Electrochem. Soc.* **129**, 2594 (1982).
48. M. Ohring, *Engineering Materials Science*. Academic Press, San Diego, 1995.
49. C. S. Rafferty, L. M. Landsberger, R. W. Dutton, and W. A. Tiller, *Appl. Phys. Lett.* **54**, 151 (1989).
50. R. Hull and J. C. Bean, *Crit. Rev. Solid State Mater. Sci.* **17**(6), 507 (1992).
51. W. D. Nix, D. B. Noble, and J. E. Turlo, *Mater. Res. Soc. Symp. Proc.* **188**, 315 (1990).
52. B. W. Dodson and J. Y. Tsao, *Appl. Phys. Lett.* **51**, 1325 (1987).
53. M. F. Ashby, *Acta Met.* **20**, 887 (1972).
54. M. Murakami, T. S. Kuan, and I. A. Blech, in *Treatise on Materials Science and Technology*, Vol. 24, eds. R. Rosenberg and K. N. Tu. Academic Press, New York, 1982.
55. C. J. Kircher and M. Murakami, *Science* **208**, 944 (1980).
56. A. A. Griffith, *Phil. Trans. Roy. Soc.* **A221**, 163 (1920).
57. M. D. Thouless, *Ann. Rev. Mater. Sci.* **25**, 69 (1995).
58. M. D. Thouless, *IBM J. Res. Develop.* **38**(4), 367 (1994).
59. B. N. Chapman, *J. Vac. Sci. Technol.* **11**, 106 (1974).
60. J. E. E. Baglin, in *Materials and Processes for Surface and Interface Engineering*, ed. F. Pauleau. Kluwer, Dordrecht, 1995.
61. V. Gupta, *MRS Bull.* **XVI**(4), 39 (1991).
62. D. H. Buckley, J. Ferrante, M. D. Pashley, and J. R. Smith, *Mat. Sci. Eng.* **83**, 177 (1986).
63. U. Dürig, *IBM J. Res. Develop.* **38**(4), 347 (1994).
64. J. E. E. Baglin, in *Handbook of Ion Beam Processing Technology*, eds. J. J. Cuomo, S. M. Rossnagel, and H. R. Kaufman. Noyes Publications, Park Ridge, NJ, 1989.
65. J. E. E. Baglin, *IBM J. Res. Develop.* **38**, 413 (1994).
66. R. d'Agostino, ed., *Plasma Deposition, Treatment and Etching of Polymers*. Academic Press, Boston (1990).
67. Special issues of *J. Adhesion Sci. Technol.* **7**, 8 (1993, 1994).
68. F. D. Egitto and L. J. Matienzo, *IBM J. Res. Develop.* **38**(4), 423 (1994).
69. K.-W. Lee and A. Viehbeck, *IBM J. Res. Develop.* **38**(4), 457 (1994).
70. P. A. Steinman and H. E. Hintermann, *J. Vac. Sci. Technol.* **A7**, 2267 (1989).
71. J. Valli, *J. Vac. Sci. Technol.* **A4**, 3007 (1986).

Copyrighted Materials
Copyright © 2002 Elsevier Retrieved from www.knovel.com

INDEX

IndexTerms | **Links**

A

Abrupt interface 767
Absorption
  coefficient 569
  of a photon 447
Accelerated ion-assisted etching 238
AC effects in plasmas 160
AC sputtering. *See* Sputtering, AC (RF)
Active reactive evaporation (ARC) 256
Adatom arrays, crystallography of 367
Adhesion
  defined 764
  film-substrate interfaces 767
  metal-metal 766
  metal oxide 767
  methods for improving 769
  tests 772
  views of 765
Adsorption
  Gibbs 666
  reactions on solid surfaces 373
AES. *See* Auger electron spectroscopy
Ag, nucleation and 406

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Alloys		
evaporation of	103	
interdiffusion in miscible	662	
sputtering of	180	
Aluminum		
graphoepitaxy and	422	
tensile strength	722	
Al-Au, compound formation in	669	
Al-Si reactions	682	
Amorphization of films	195	
Amorphous		
carbon	341	
silicon	336	680
silicon nitride	336	544
solids	7	
to-crystalline, in CoSi2	677	
Amorphous films	540	
at atomic scale	542	
Au-Co	545	
metal alloy systems	544	
Ni-Zr and Co-Zr	547	
silicon nitride and SiO2	544	
simulation model	548	
structure of semiconductor	543	
Analog simulations of film growth	518	
Anisotropy, etching	243	
Anode	152	
disappearing	230	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Anodic arcs	259	
Arcs	148	
defined	150	
plasma deposition	258	
Area defects	14	
Arrhenius plot	41	
Aston dark space	151	
Atmospheric-pressure chemical vapor deposition (APCVD)	278	300
high temperature	313	
low temperature	316	
Atomic force microscopy (AFM)		
applications in film and surface characterization	600	605
nucleation and use of	408	
Atomic layer epitaxy (ALE)	463	
Atomic movement	39	
Atomistic models of nucleation rate	388	
Au		
-Co amorphous films	545	
nucleation and	406	
stress-strain curve	714	
Auger electron spectroscopy (AES)		
film chemical characterization and use of	607	616
film thickness and use of	578	
nucleation and use of	401	
Avrami equation	399	677

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	**Links**		
B			
Backscattered electrons	587		
Backscatter Kikuchi diffraction (BKD)	587		
Barriers, diffusion	684		
Beer's law	131		
Berkovich diamond indentor	718		
β-$FeSi_2$-silicon	424		
Blanket films, patterning	532		
Bonding			
at atomic level	14		
in solids	15		
Boron nitride films	343		
Bragg's law	7	598	714
Bravais lattices	2		
Bridging	693		
Brownian motion	153		
Bulge testing	716		
Burgers vector	11	721	
C			
Cantilever beams	735		
Capillary (droplet) theory	377	378	
Carbon			
nitride films	343		
-Zr amorphous films	547		
CARIS (constant-angle reflection interference spectroscopy)	564	567	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Cathode		
arcs	259	
dark space	152	
glow	151	
hollow	210	
sheath	152	
Cauchy approximation	569	
Ceramic films, deposited by pulsed		
laser deposition	133	
Channeling	193	628
Chemical beam epitaxy (CBE)	470	
Chemical etching	238	
Chemical-mechanical polishing (CMP)	238	
Chemical reaction rate theory	43	169
Chemical reactions		
plasmas and	169	
thermodynamics and	25	
Chemical vapor deposition (CVD)	9	96
See also under type of		
advantages of	278	
applications	277	
boron nitride films	343	
carbon nitride films	343	
compound formation reactions	283	
diamond and diamond-like		
carbon films	337	
disproportionation reactions	284	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links

Chemical vapor deposition (CVD) (*Cont.*)
- equilibrium conditions 289
- film growth kinetics 303
- gases, list of hazardous 348
- gas transport 293
- oxidation reactions 282
- pyrolysis 281
- rapid-thermal 466
- reaction feasibility 287
- reduction reactions 281
- reversible transfer reactions 284
- safety issues 347
- selective deposition 345
- silicon films 334
- silicon nitride films 336
- steps 278
- stress in films 750
- thermal processes 312
- thermodynamics of 287
- types of 278
- ultrahigh-vacuum 472
- versus transport 287

Chemical vapor deposition (CVD)
- -based epitaxy 457
- additional processes 464
- atomic layer 463
- metalorganic processes 458

Chemisorption 373

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Child–Langmuir equation	163	
Circular plates	737	740
Clausius–Clapeyron equation	98	
Closed-spaced vapor transport (CSVT)	297	
Coalescence and depletion, cluster	393	
features of	394	
grain size	398	
migration	397	
Ostwald ripening	395	
sintering	396	
Coating uniformity	255	
Cobalt disilicide, amorphous-to -crystalline transformation in	677	
Coble creep	758	
Cold plasmas	154	
Collimation	233	
Collisions, elastic versus inelastic	164	
Colloidal epitaxy	476	
Columnar grain structure	504	
Compound formation	654	
aluminum-gold system	669	
first	674	
intermetallic	669	
reactions	283	
room temperature	671	
silicide	690	
thermodynamics of	672	
Compound interface	767	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links

Compound semiconductors
- applications 447
- bandgap energy 441
- designing epitaxial film-substrate combinations 443
- energy bandgaps, direct and indirect 440
- epitaxy and 439
- lattice parameter 443
- light-emitting devices 451
- optical communications 450
- properties 439
- selective epitaxy 481
- thermal expansion coefficient 443

Computer simulations. *See* Film structure, simulations of

Conductance
- of a circular aperture 62
- gas 65

Conformal coverage 113
Constrained film structure 533
Continuity equation 299
Convection 298
Covalent bonding 19
Cracking
- propagation and spacing 763
- stress relief by film 761

Crookes space 152
Crucible, evaporation source 120

IndexTerms	**Links**	
Cryopumps	76	78
Crystalline solids	2	
Crystallite coalescence, stress due to	752	
Crystallography		
film texture and	528	
notation of heteroepitaxy systems	423	
recrystallization effects	752	
of silicon surfaces	371	
of substrate meshes and adatom arrays	367	
Curvature methods	735	
CVD. *See* Chemical vapor deposition		
Cyclotron frequency	158	

D

Damascene	238	
DC sputtering. *See* Sputtering, DC		
Debye forces	765	
Debye length	158	
Defects. *See* Lattice misfit and defects		
Deflectless flow	757	
Dense random packing (DRP)	543	
Density, film	508	
Deposition		
arc plasma	258	
geometry	106	
ion-beam-assisted	138	261

This page has been reformatted by Knovel to provide easier navigation.

Deposition (*Cont.*)
ionized cluster beam 264
long-throw 232
techniques 126
Deposition rate, nucleation and substrate 383
Diamond and diamond-like carbon films 337
Diamond pyramid hardness (DPH) test 50
Differential scanning calorimetry (DSC) 672
Diffusion
See also Interdiffusion
ambipolar 155
barriers 684
in bulk versus thin-film solids 642
couples 653
in gases 296
grain-boundary 646 663
interface 768
mechanisms 643
Diffusional creep 758
Diffusional transport 39
Diffusion pumps 74
Discharges, types and structures of 149
Dislocations 11
climb 757
formation of misfit 437

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	

Dislocations (*Cont.*)

glide 757

loops 436

misfit 429

models for strength 720

relaxation in films and 756

from substrate 435

threading 437

Disproportionation reactions 284

Double-crystal diffractometer 598

E

Edge dislocation 11

Elasticity 49

Electrically heated evaporation sources 118

Electromigration

failure 703

manifestations of 697

modeling 698

stress, effects of 702

structural and thermal effects 701

Electron backscatter diffraction (EBSD) 533 587

Electron-beam evaporation 121

Electron energy loss spectroscopy (EELS) 592

Electronic nature of surfaces 361

Electron microprobe (EMP) 615

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**		
Electron motion, in electric and magnetic fields	155		
Ellingham diagrams	26		
Ellipsometry	563	564	571
Energy band diagrams	21		
Energy dispersive x-ray (EDX) analysis	607	615	
Energy equation	299		
Energy gap	22		
Entrapment pumps	71		
Epitaxial semiconductor films, high-temperature depositing methods			
CVD-based	457		
from melts	453		
Epitaxial semiconductor films, low-temperature depositing methods			
less common processes	474		
molecular beam	467		
silicon heteroepitaxy	471		
Epitaxial superlattices, stress in	730		
Epitaxy	7		
applications	417		
atomic layer	463		
colloidal	476		
compound semiconductors	439		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links		

Epitaxy (*Cont.*)			
CVD-based	457		
defined	420		
film growth	476		
graphoepitaxy	421		
heteroepitaxy	418	423	471
homoepitaxy	418	476	
island	358	381	420
lattice misfit and defects	429		
liquid-phase	453		
molecular beam	467	474	
rapid-thermal	466		
research on	419		
selective	479		
strained-layer	419		
tilted-layer	427		
use of the term	417		
van der Waals	475		
vapor-levitation	464		
Equilibrium conditions, chemical vapor deposition	289		
Etching			
See also Plasma etching			
description of	236		
Evaporated films			
stress in	744		
structure-zone model for	498	501	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links		
Evaporation			
of alloys	103		
characteristics of materials	127		
of compounds	100		
difference between sputtering and	95	145	197
electron-beam	121		
hardware	118		
point source	106		
processes and applications	128		
rate	97		
reactive	256		
surface source	108		
vapor pressure of elements	98		

F

Faraday dark space	152		
Fe, nucleation and	406		
Fermi energy	22		
Fick's law	155	696	
Film growth			
analog simulations of	518		
axial growth-rate uniformity	303		
epitaxial	476		
historical development	357		
influence of thermodynamics	309		
island	358	381	420
lateral epitaxial	455		

This page has been reformatted by Knovel to provide easier navigation.

Film growth (*Cont.*)
layer 358 381
modes 358 381
pseudomorphic 419
radial growth-rate uniformity 306
rate for MOCVD 462
Stranski–Krastanov 358 381
temperature dependence 308
Films, thin
applications 559
conformal coverage 113
density 508
deposition geometry 106
purity 115
stress in 51 742
techniques for the
characterization of 559
Film structure
amorphous 540
columnar grain structure 504
comparison of grain and bulk 495
constrained 533
film density 508
grain growth 521
microstructure control 524
microtexture 533
structure-zone diagrams 497
texture 528

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**		
Film structure, simulations of	510		
analog	518		
molecular dynamics	515		
Monte Carlo	513		
Tangent Rule	511		
Film surface composition			
applications	619	632	634
auger electron spectroscopy (AES)	607	616	
energy dispersive x-ray (EDX) analysis	607	615	
fingerprinting atoms through electron transitions	610		
methods	606		
Rutherford backscattering law (RBS)	560	607	621
secondary ion mass spectrometry (SIMS)	578	607	628
summary of	634		
x-ray photoelectron spectroscopy (XPS)	607	618	
Film surface morphology and structure			
hierarchies	583		
scanning electron microscopy	584		
scanning tunneling microscopy	599		
transmission electron microscopy	589		
x-ray diffraction	594		
Film thickness, mechanical techniques for measuring			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**		
Film thickness, mechanical techniques for measuring (*Cont.*)			
profilometry	576		
quartz crystal microbalances	579		
ultrasonic multilayer-film metrology	580		
Film thickness, optical methods for measuring			
ellipsometry	563	564	571
interferometry	563	564	
list of	564		
Film thickness, uniformity	109		
Fizeau fringes	565		
Focused ion mean (FIB) microscopy	595		
Fringes of equal chromatic order (FECO)	564	566	
Fringes of equal thickness (FET)	564		

G

GaAs		
auger electron spectroscopy and	619	
deposition of	100	102
film growth	477	
heteroepitaxy	425	
x-ray photoelectron spectroscopy and	620	
GaInAsP, film growth	479	
GaN light-emitting semiconductor devices	452	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	

Gases
- conductance 65
- diffusion in 296
- flow regimes 63
- hazardous, employed in chemical vapor deposition 348
- impingement on surfaces 61
- kinetic theory of 58
- pumping speed 69
- transfer pumps 70
- transport in chemical vapor deposition systems 293

Gas-source molecular beam epitaxy (GSMBE) 470

Ge_xSi_{1-x}-Si films, defects in 431

Gibbs adsorption 666

Gibbs free-energy function 24 375

Gibbs phase rule 29

Glow discharges 148
- abnormal 150
- dc 151
- negative 152
- normal 150
- types and structures of 149

Gold films, grain-boundary self-diffusion in 650

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Grain boundaries	14	521
stress due to crystallite coalescence at	752	
Grain-boundary diffusion		
in alloy films	663	
estimating	650	
in gold films	650	
modeling	647	
reducing	666	
Grain growth	521	
Graphoepitaxy	421	
Growth. *See* Film growth		
H		
Hall–Petch equation	721	723
Hardness testing	50	
of multilayer films	723	
nanohardness	718	
Heteroepitaxy		
compared to homoepitaxy	418	
crystallographic notation	423	
geometrical features of metal-semiconductor	424	
silicon	471	
Homoepitaxy		
compared to heteroepitaxy	418	
on silicon	476	
Hooke's law	49	727

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	

I

In situ film characterization	482	
Interdiffusion		
effects of ambients	667	
features	659	
in miscible alloy systems	662	
Interfaces, substrate	767	
Interferometry	563	564
Ion-beam-assisted deposition (IBAD)	138	261
Ion-beam mixing	194	
Ion beams, adhesion and		
assisted deposition	770	
implantation	771	
presputtering	769	
stitching	771	
Ion bombardment		
modification of growing films	184	
molecular dynamics and, of films	516	
structure-zone diagram and the effect of	503	
-surface interactions	170	
Ionic bonding	18	
Ion implantation	191	
plasma-immersion	267	
Ionized cluster beam (ICB) deposition	264	516
Ionized physical vapor deposition (I-PVD)	233	
Ion plating	253	

This page has been reformatted by Knovel to provide easier navigation.

Island mode of film growth 358 381 420

J

Josephson superconducting
tunnel-junction devices 759

K

Kaufman ion-beam sources 771
Keesom forces 765
Kinetics
atom movements 39
gases and 58
macroscopic transport 36
nonlinear effects 696
nucleation and 386
Kirkendall effect 509 654 689
768
Knoop hardness test 50
Knudsen cells 107
Knudsen number 64
Kramers–Kronig relations 569

L

Langmuir, Irving 147
Langmuir isotherm 375
Laser-enhanced chemical vapor
deposition (LECVD) 278 321
Laser-scanning methods 737 740

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links		
Lateral epitaxial film growth			
over insulators	455		
Lateral epitaxial film growth over			
oxide (LEGO)	455		
Lattice misfit and defects			
dislocation loops	436		
equilibrium theory of	429		
formation of dislocations	437		
in Ge_xSi_{1-x}-Si films	431		
low-angle grain boundaries			
and twins	437		
stacking faults	436		
from substrate	435		
types and sources of	433		
Lattice parameter property	443		
Layer mode of growth	358	381	
Lennard–Jones potential	515		
Lever rule	32		
Light-emitting semiconductor devices	451		
Line defects	11		
Liquid-phase epitaxy (LPE)	453		
Lithographic techniques	234		
London forces	765		
Lorenz force	88	156	158
Low-energy electron diffraction (LEED)	370	372	482
	483		

This page has been reformatted by Knovel to provide easier navigation.

Low-pressure chemical vapor deposition (LPCVD) 73 278 297 300 317

M

Magnetron sputtering. *See* Sputtering, magnetron
Mask, generation of 235
Maxwell–Boltzmann relation 37 58 163
Mechanical equilibrium 725
Mechanical interface 768
Mechanical properties 47
  adhesion 764
  issues in 711
  relaxation effects 754
  strength, dislocation models for 720
  strength, testing methods 713
  stress (internal), analysis of 723
  stress (internal), causes of 742
  stress (internal), measuring methods 735
Mesotaxy 475
Metal alloy systems, amorphous 544
Metal films, nucleation and 401 403
Metallic bonding 17
Metal-metal adhesion 766
Metalorganic chemical vapor deposition (MOCVD) 283 285 318

IndexTerms	**Links**	

Metalorganic chemical vapor deposition (MOCVD) (*Cont.*)		
film growth rate	462	
gas reactions	460	
precursors grown by	459	
reactors	460	
selective	482	
for semiconductor epitaxy	458	
Metalorganic molecular beam epitaxy (MOMBE)	470	476
selective	482	
Metal-oxide-semiconductors (MOS)	24	
Metal-oxides, adhesion	767	
Metals, electronic nature of	361	
Metal-semiconductor heteroepitaxy	424	
Microbeam testing	719	
Microstructure control	524	
Microtexture	533	
Migration, cluster	397	
Miller indices	5	
Misfit dislocations	429	
MOCVD. *See* Metalorganic chemical vapor deposition		
Molecular beam allotaxy (MBA)	474	
Molecular beam epitaxy (MBE)	467	
in situ film characterization	482	
selective	481	
Molecular dynamics (MD) simulations	515	

This page has been reformatted by Knovel to provide easier navigation.

Molecular flow 64
Molecular velocities 58
Monte Carlo simulations 513
Multilayer films, hardness and
strength of 723
Multiple-angle-of-incidence (MAI)
ellipsometry 574
Multiple beam interferometry 564

N

Nabarro–Herring creep 758
Nakamura, S. 452
Nanoindentation testing 716
Nanoindentor 718
Navier–Stokes equation 299
Nernst–Einstein equation 43 522 696
Newton's law 154
Ni
nucleation and 406
-Si 690
-Zr amorphous films 547
Nichia Chemical Industries 452
Nonlinear mass transport effects 696
Nucleation 44
capillary (droplet) theory 377 378
cluster coalescence and depletion 393
dependence on temperature and

IndexTerms | **Links**

Nucleation (*Cont.*)
deposition rate 383
experimental studies of 400
film growth modes 381
kinetic processes in 386
stages 357
surface energies 377
thermodynamics of 376
Nucleation rate
atomistic models of 388
defined 386

O

OMVPE 471
One-component system 29
Opaque films 564
Optical communications 450
Organometal chemical vapor deposition (OMCVD). *See* Metalorganic chemical vapor deposition (MOCVD)
Orientation imaging microscopy (OIM) 533
Ostwald ripening 395 474
Ostwald rule 545
Oxidation reactions 282

P

Particle bombardment 532

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Paschen's law	149	
Passive compound barriers	687	
Patterning		
See also Plasma etching		
example of	236	
lithographic techniques	234	
Pauli exclusion principle	15	174
Phase diagrams	29	
Phase transformations	676	
amorphous-to-crystalline, in $CoSi_2$	677	
epitaxial regrowth of amorphous silicon	680	
Photonic crystal	476	
Photoresist layers	236	
Physical vapor deposition (PVD)	96	
hybrid and modified	252	
ionized	233	
lining and filling trenches by	535	
Physisorption	373	
Plasma-enhanced chemical vapor deposition (PECVD)	278	
applications	323	327
chemical reactions	331	
materials deposited by	323	
plasma modification of metal surfaces	329	
reactors	324	

This page has been reformatted by Knovel to provide easier navigation.

Plasma etching
  anisotropy 243
  categories of 238
  loading 244
  mechanisms 238
  rate 242
  reactors 245
  selectivity 243
  temperature increase during 249
Plasma-immersion ion
    implantation (PIII) 267
Plasmas
  ac effects in 160
  chemistry 168
  collective charge effects 158
  currents and diffusion in 154
  defined 147
  early applications of 147
  electrode sheaths 161
  electron motion in electric
    and magnetic fields 155
  particle energies and temperatures 153
  reactions in 164
  species 152
Plastic behavior 49
Point defects 11
Poisson's ratio 48
Polarimetry 571

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	**Links**	

Polarization spectroscopy	571	
Polycides	692	
Polymerization	236	
Polymers, adhesion in	771	
Positive column	152	
Pressure		
deflection response	716	
filling	538	
gas	60	
gauges	86	
Printing	235	
Profilometry	576	
Pseudomorphic growth	419	
Pulsed laser deposition (PLD)	128	145
Pump-down time	70	
Pumping speed	69	
Pumps, vacuum	70	
Pyrolysis	281	321

Q

Quantum corral	604
Quartz crystal oscillators and microbalance method	579

R

Radial distribution function (RDF)	543
Raoult's law	104
Rapid-thermal CVD (RTCVD)	466

IndexTerms	Links		
Reactions			
Al-Si	682		
diffusion barriers	684		
polycides	692		
salicides	692		
silicides	688		
Reactive evaporation processes	256		
Reactive-ion etching	238		
Reactive sputtering. *See* Sputtering, reactive			
Reactors, MOCVD	460		
Reciprocal lattice	370		
Reconstruction	367		
of silicon surfaces	371		
Recrystallization effects	752		
Reduction reactions	281		
Reflection ellipsometry	573		
Reflection high-energy electron diffraction (RHEED)	370	482	485
Refractory metal sheet, evaporation source	118		
Relaxation effects	754		
crack propagation and spacing	763		
dislocation mechanisms	756		
film cracking	761		
in metal films during thermal cycling	758		
in SiO_2 films	755		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links		
Residual gas analysis (RGA)	87		
Resist removal	236		
Resputtering	186		
Reversible transfer reactions	284		
Roots pumps	73		
Rotary mechanical pumps	71		
Royer, L.	417		
Rutherford backscattering (RBS)	560	607	621

S

IndexTerms	Links	
Sacrificial barriers	687	
Salicides	692	
Scanning electron microscopy (SEM)		
applications in film and surface characterization	584	
microtexture and	533	
nucleation and use of	400	
Scanning tunneling microscopy (STM)	370	
applications in film and surface characterization	599	
nucleation and use of	404	
reconstructed silicon surfaces	371	
Schrödinger equation	361	362
Scission	236	
Scratch tests	774	
Screw dislocation	11	
Sculptured films	538	

IndexTerms	**Links**		
Secondary ion mass spectrometry (SIMS)	578	607	628
Seeman–Bohlin diffraction geometry	596		
Selective area epitaxy (SAE)	480		
Selective epitaxial growth (SEG)	479		
SEM. *See* Scanning electron microscopy			
Semiconductors, electronic nature of	363		
Separation by implanted oxygen (SIMOX)	475		
Shadowing	498		
Shear-type tests	774		
Sidewall-protected ion-enhanced etching	238		
Silicides	688	751	
Silicon			
β-$FeSi_2$-	424		
epitaxial regrowth of amorphous	680		
films	334		
heteroepitaxy	424	471	
homoepitaxy	476		
metal	424		
nitride films	336	544	750
poly	750		
residual stress values	743		
selective epitaxy	480		
stress	750		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
SiO_2	544	
CVD	750	
growth of	656	
stress and strain relaxation in	755	
Silicon surfaces		
nucleation and	404	
reconstructed	371	
W_xN on SiO_2	407	
SIMS. *See* Secondary ion mass spectrometry		
Simulated emission of a photon	448	
Sintering	396	
Solid phase epitaxial (SPE)	680	
Solids		
banding in	21	
bonding in	15	
defects in	10	
structure of	2	
Solute segregation	667	
Spectral reflectometry	564	568
Spectroscopic ellipsometry (SE)	564	574
Spontaneous emission of a photon	447	
Sputtered films		
stress in	746	
structure-zone model for	499	503
Sputter etching	238	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**		

Sputtering
- of alloys 180
- difference between evaporation and 95 145 197
- effects of 181
- ion-surface interactions 170
- knock-ons 175
- linear collision cascades 176
- targets 206
- yields 174

Sputtering, AC (RF) 211
- discharges 212
- RF issues 214
- target self-bias 213

Sputtering, DC 207
- disadvantages of 208
- hollow cathode 210
- substrate bias 210
- triode 209

Sputtering, magnetron
- advantages of 222
- cylindrical-post 224
- dual 229
- film uniformity 232
- particulate contamination 231
- planar 223
- pulsed-power 226
- sputter gun (s-gun) 226

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**		
Sputtering, magnetron (*Cont.*)			
target integrity	230		
unbalanced	227		
Sputtering, reactive			
compounds for	216		
film deposition rates, stoichiometry and properties	221		
hysteresis effects	217	218	
modeling	217	219	
Sputter ion pumps	78	80	
Stacking faults	436		
Stefen–Boltzmann law	120		
Step coverage	113	232	534
Step gauges	568		
STM. *See* Scanning tunneling microscopy			
Stoichiometry, maintaining melt	105		
Stoney formula	727		
Strain	48		
relaxation	759		
in SiO_2 films	755		
Strained-layer epitaxy	419		
Stranski–Krastanov mode of growth	358	381	
Strength, dislocation models for	720		
Strength, testing methods	713		
bulge	716		
of multilayer films	723		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	

Strength, testing methods (*Cont.*)		
nanoindentation	716	
tensile	714	
Stress	47	
affects of	711	
in chemical vapor deposition films	750	
electromigration and	702	
in evaporated metal films	744	
relief by film cracking	761	
in SiO_2 films	755	
in sputtered films	746	
theories for intrinsic	751	
in thin films	51	742
Stress (internal), analysis of	723	
in epitaxial superlattices	730	
mechanical equilibrium	725	
Stoney formula	727	
thermal, in homogeneous and bilayer structures	732	
thermal, in three-layer structures	733	
Stress (internal), measuring methods		
accuracy and sensitivity of methods	741	
substrate deflection or curvature methods	735	
x-ray diffraction	741	
Stress (intrinsic), in thin films	742	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**	
Structure-zone diagrams (SZDs) and models (SZMs)		
condensation processes	497	
for evaporated films	498	501
ion bombardment	503	
revised and new	499	
for sputtered films	499	503
Stuffed barriers	687	
Sublimation furnaces, evaporation source	119	
Substrate deflection	735	
Substrate interfaces	767	
Substrate surfaces		
adsorption reactions on solid surfaces	373	
defect propagation from	435	
electronic nature of	361	
energies (tensions)	377	
film texture affected by	530	
reasons for studying	360	
reconstructed silicon	371	
structures of	366	
temperature and deposition rate	383	
Superlattices, stress in epitaxial	730	

T

Tamm states	362	
Tangent Rule	506	511

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	**Links**

TEM. *See* Transmission electron microscopy

Temperature

compound formation and room 671

increase of films during plasma etching 249

nucleation and substrate 383

of plasmas 153

of resistance heaters, estimating 120

Tensile testing 714

for adhesion 772

Tests

See also Strength, testing methods

adhesion 772

Texture 528

micro 533

Thermal chemical vapor deposition processes 312

Thermal cycling 758

Thermal expansion coefficient 443

Thermal stress

in homogeneous and bilayer structures 732

in three-layer structures 733

Thermodynamics 24

chemical reactions 25

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links	
Thermodynamics (*Cont.*)		
of chemical vapor deposition	287	309
compound formation and	672	
nonlinear effects	697	
of nucleation	376	
phase diagrams	29	
Threading dislocations	437	
Threshold energy	175	
Tilted-layer epitaxy	427	
Townsend equation	148	
Transmission electron microscopy (TEM)		
applications in film and surface characterization	589	
nucleation and use of	400	
Transparent films	566	
Trenches, lining and filling	535	
Triode sputtering	209	
Tungsten wire, evaporation source	118	
Turbomolecular pumps	76	
Two-component systems	30	
Two-step trench-filling process	536	

U

IndexTerms	Links	
Ultrahigh-vacuum chemical vapor deposition (UHV/CVD)	472	476
Ultrasonic multilayer-film metrology	580	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links	
V		
Vacancies	11	
Vacuum		
deposition system	81	
leaks	85	
monitoring	86	
pumps	70	
as a science	57	
systems	81	
VAMFO (variable-angle monochromatic fringe observation)	564	567
van der Waals		
adhesion	765	
bonding	20	472
epitaxy	475	
Vapor-levitation epitaxy (VLE)	464	
Vapor pressure of elements	98	
Vickers hardness test	50	
Viscous flow regime	64	294
Volmer–Weber growth	745	
W		
Waffer bonding	472	
Wavelength dispersive analysis (WDX)	615	
Web coating	135	
W_xN on SiO2, nucleation and	407	

This page has been reformatted by Knovel to provide easier navigation.

X

X-ray diffraction 6
  applications in film and surface
    characterization 594
  measuring stress with 741
  tensile data and use of 714
X-ray photoelectron spectroscopy (XPS) 607 618

Y

Young's equation 379 381
Young's modulus 49 713 716
  720 727 736

Z

Zr
  amorphous films 547
  Ni- and Co- 547

This page has been reformatted by Knovel to provide easier navigation.