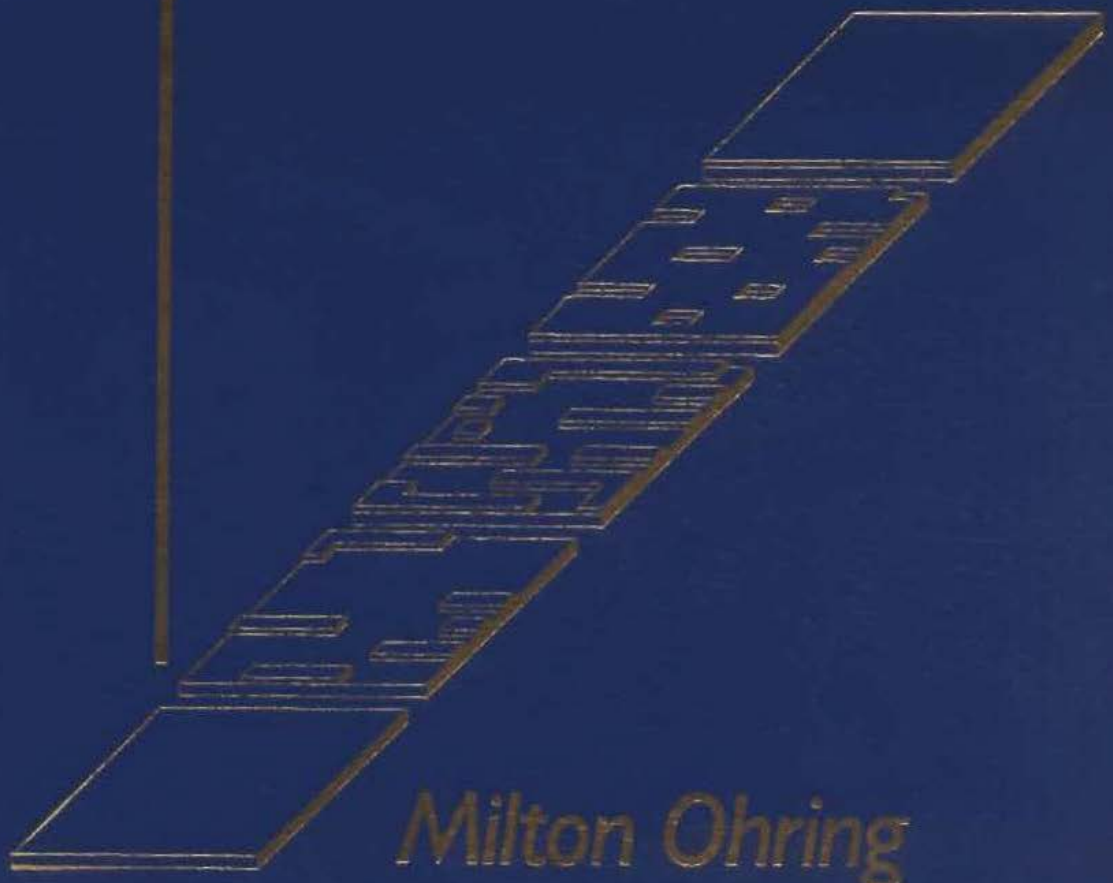
The MATERIALS SCIENCE of THIN FILMS



Milton Ohring

IP Bridge Exhibit 2048
Western Digital v. IP Bridge
IPR2024-01447

The Materials Science *of* Thin Films

The Materials Science *of* Thin Films

Milton Ohring

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



Academic Press
San Diego New York Boston
London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Copyright © 1992 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.

Designed by Elizabeth E. Tustian

ACADEMIC PRESS
A Division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495

United Kingdom Edition published by
ACADEMIC PRESS LIMITED
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Ohring, Milton, date.
The materials science of thin films / Milton Ohring.
p. cm.
Includes bibliographical references and index.
ISBN 0-12-524990-X (Alk. paper)
1. Thin films. I. Title.
TA418.9.T45047 1991
620'.44—dc20 91-9664
CIP

Printed in the United States of America
99 00 01 02 03 MV 11 10 9 8 7



Contents

Foreword xi

Preface xiii

Acknowledgments xvii

Thin Films — 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. Bonding of Materials 14
1.5. Thermodynamics of Materials 21
1.6. Kinetics 33
1.7. Nucleation 40
1.8. Conclusion 43
Exercises 43
References 46

Chapter 2

Vacuum Science and Technology 49

2.1. Kinetic Theory of Gases 49
2.2. Gas Transport and Pumping 55

2.3. Vacuum Pumps and Systems 62
Excercises 75
References 77

Chapter 3

Physical Vapor Deposition 79

3.1. Introduction 79
3.2. The Physics and Chemistry of Evaporation 81
3.3. Film Thickness Uniformity and Purity 87
3.4. Evaporation Hardware and Techniques 96
3.5. Glow Discharges and Plasmas 101
3.6. Sputtering 109
3.7. Sputtering Processes 118
3.8. Hybrid and Modified PVD Processes 132
Exercises 140
References 144

Chapter 4

Chemical Vapor Deposition 147

4.1. Introduction 147
4.2. Reaction Types 149
4.3. Thermodynamics of CVD 155
4.4. Gas Transport 162
4.5. Growth Kinetics 167
4.6. CVD Processes and Systems 177
Exercises 190
References 193

Chapter 5

Film Formation and Structure 195

5.1. Introduction 195
5.2. Capillarity Theory 198
5.3. Atomistic Nucleation Processes 206
5.4. Cluster Coalescence and Depletion 213
5.5. Experimental Studies of Nucleation and Growth 219
5.6. Grain Structure of Films and Coatings 223
5.7. Amorphous Thin Films 234
Exercises 243
References 246

Chapter 6

Characterization of Thin Films 249

6.1. Introduction 249
6.2. Film Thickness 252
6.3. Structural Characterization 265
6.4. Chemical Characterization 275
Exercises 300
References 305

Chapter 7

Epitaxy 307

7.1. Introduction 307
7.2. Structural Aspects of Epitaxial Films 310
7.3. Lattice Misfit and Imperfections in Epitaxial Films 316
7.4. Epitaxy of Compound Semiconductors 322
7.5. Methods for Depositing Epitaxial Semiconductor Films 331
7.6. Epitaxial Film Growth and Characterization 339
7.7. Conclusion 350
Exercises 351
References 353

Chapter 8

Interdiffusion and Reactions in Thin Films 355

8.1. Introduction 355
8.2. Fundamentals of Diffusion 357
8.3. Interdiffusion in Metal Alloy Films 372
8.4. Electromigration in Thin Films 379
8.5. Metal–Semiconductor Reactions 385
8.6. Silicides and Diffusion Barriers 389
8.7. Diffusion During Film Growth 395
Exercises 398
References 401

Chapter 9

Mechanical Properties of Thin Films 403

9.1. Introduction 403
9.2. Introduction to Elasticity, Plasticity, and Mechanical Behavior 405
9.3. Internal Stresses and Their Analysis 413
9.4. Stress in Thin Films 420
9.5. Relaxation Effects in Stressed Films 432

9.6. Adhesion 439
Exercises 446
References 449

Chapter 10

Electrical and Magnetic Properties of Thin Films 451

10.1. Introduction to Electrical Properties of Thin Films 451
10.2. Conduction in Metal Films 455
10.3. Electrical Transport in Insulating Films 464
10.4. Semiconductor Contacts and MOS Structures 472
10.5. Superconductivity in Thin Films 480
10.6. Introduction to Ferromagnetism 485
10.7. Magnetic Film Size Effects — M_s versus Thickness and Temperature 489
10.8. Magnetic Thin Films for Memory Applications 493
Exercises 502
References 505

Chapter 11

Optical Properties of Thin Films 507

11.1. Introduction 507
11.2. Properties of Optical Film Materials 508
11.3. Thin-Film Optics 524
11.4. Multilayer Optical Film Applications 531
Exercises 542
References 544

Chapter 12

Metallurgical and Protective Coatings 547

12.1. Introduction 547
12.2. Hard Coating Materials 551
12.3. Hardness and Fracture 561
12.4. Tribology of Films and Coatings 570
12.5. Diffusional, Protective, and Thermal Coatings 580
Exercises 585
References 587

Chapter 13

Modification of Surfaces and Films 589

13.1. Introduction 589
13.2. Lasers and Their Interactions with Surfaces 591

13.3. Laser Modification Effects and Applications 602
13.4. Ion-Implantation Effects in Solids 609
13.5. Ion-Beam Modification Phenomena and Applications 616
Exercises 624
References 626

Chapter 14

Emerging Thin-Film Materials and Applications 629

14.1. Film-Patterning Techniques 630
14.2. Diamond Films 635
14.3. High-T_c Superconductor Films 641
14.4. Films for Magnetic Recording 645
14.5. Optical Recording 650
14.6. Integrated Optics 654
14.7. Superlattices 661
14.8. Band-Gap Engineering and Quantum Devices 669
14.9. Conclusion 678
Exercises 678
References 681

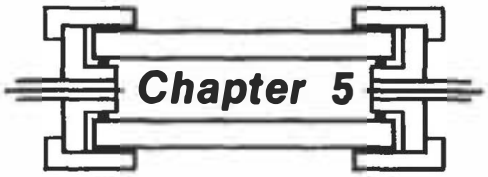
Appendix 1

Physical Constants 685

Appendix 2

Selected Conversions 687

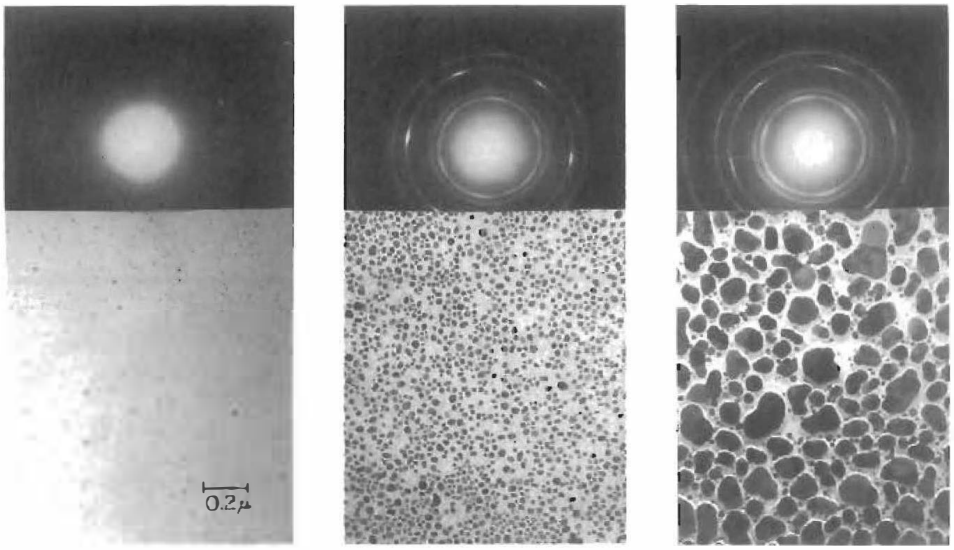
Index 689



Film Formation and Structure

5.1. INTRODUCTION

Interest in thin-film formation processes dates at least to the 1920s. During research at the Cavendish Laboratories in England on evaporated thin films, the concept of formation of nuclei that grew and coalesced to form the film was advanced (Ref. 1). All phase transformations, including thin-film formation, involve the processes of nucleation and growth. 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. Although numerous high-resolution transmission electron microscopy investigations have focused on the early stages of film formation, it is doubtful that there is a clear demarcation between the end of nucleation and the onset of nucleus growth. The sequence of nucleation and growth events can be described with reference to the micrographs of Fig. 5-1. 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



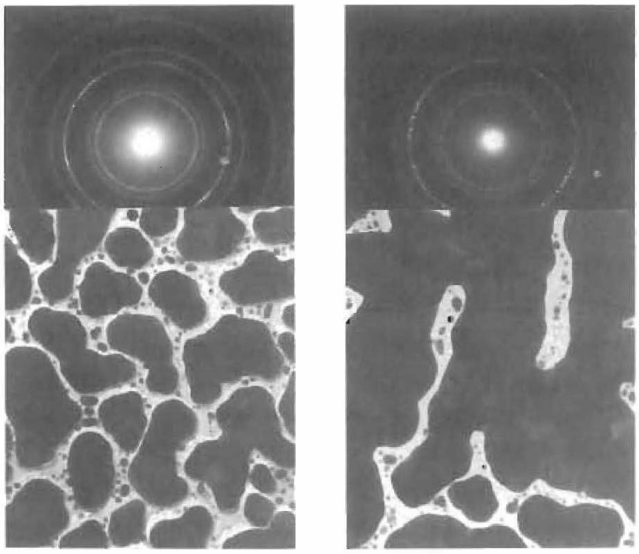


Figure 5-1. Nucleation, growth and coalescence of Ag films on (111) NaCl substrates. Corresponding diffraction patterns are shown. (Courtesy of R. W. Vook. Reprinted with permission from the Metals Society, from R. W. Vook, *Int. Metals Rev.* **27**, 209(1982).

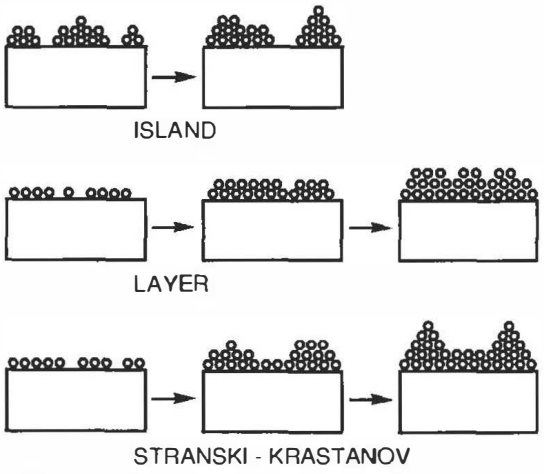


Figure 5-2. Basic modes of thin-film growth.

that is liquidlike in character 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. Coalescence continues until a connected network with unfilled channels in between develops. 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 collective set of events occurs during the early stages of deposition, typically accounting for the first few hundred angstroms of film thickness.

The many observations of 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. 5-2. 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. Many systems of metals on insulators, alkali halide crystals, graphite, and mica substrates display this mode of growth.

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. The most important example of this growth mode involves single-crystal epitaxial growth of semiconductor films, a subject treated extensively in Chapter 7.

The layer plus island or Stranski–Krastanov (S.K.) growth mechanism is an intermediate combination of the aforementioned modes. In this case, after forming one or more monolayers, subsequent layer growth becomes unfavorable and islands form. The transition from two- to three-dimensional growth is not completely understood, but any factor that disturbs the monotonic decrease in binding energy characteristic of layer growth may be the cause. For example, due to 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. This growth mode is fairly common and has been observed in metal–metal and metal–semiconductor systems.

At an extreme far removed from early film formation phenomena is a regime of structural effects related to the actual grain morphology of polycrystalline films and coatings. This external grain structure together with the internal defect, void, or porosity distributions frequently determines many of the engineering properties of films. For example, columnar structures, which interestingly develop in amorphous as well as polycrystalline films, have a profound effect on magnetic, optical, electrical, and mechanical properties. In this chapter we discuss how different grain and deposit morphologies evolve as a function of deposition variables and how some measure of structural control can be exercised. Modification of the film structure through ion bombardment or laser processing both during and after deposition has been a subject of much research interest recently and is treated in Chapters 3 and 13. Subsequent topics in this chapter are:

5.2 Capillarity Theory
5.3 Atomistic Nucleation Processes
5.4 Cluster Coalescence and Depletion
5.5 Experimental Studies of Nucleation and Growth
5.6 Grain Structure of Films and Coatings
5.7 Amorphous Thin Films

References 1–5 are recommended sources for much of the subject matter in this chapter.

5.2. Capillarity Theory

5.2.1. Thermodynamics

Capillarity theory possesses the mixed virtue of yielding a conceptually simple qualitative model of film nucleation, which is, however, quantitatively inaccu-

rate. The lack of detailed atomistic assumptions gives the theory an attractive broad generality with the power of creating useful connections between such variables as substrate temperature, deposition rate, and critical film nucleus size. An introduction to the thermodynamic aspects of *homogeneous* nucleation was given on p. 40 and is worth reviewing. In a similar spirit, we now consider the *heterogeneous* nucleation of a solid film on a planar substrate. Film-forming atoms or molecules in the vapor phase are assumed to impinge on the substrate, creating aggregates that either tend to grow in size or disintegrate into smaller entities through dissociation processes.

The free-energy change accompanying the formation of an aggregate of mean dimension r is given by

$$\Delta G = a_3 r^3 \Delta G_V + a_1 r^2 \gamma_{vf} + a_2 r^2 \gamma_{fs} - a_2 r^2 \gamma_{sv} . \tag{5-1}$$

The chemical free-energy change per unit volume, ΔG_V, drives the condensation reaction. According to Eq. 1-39, any level of gas-phase supersaturation generates a negative ΔG_V without which nucleation is impossible. 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. For the cap-shaped nucleus in Fig. 5-3, 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 $a_1 = 2\pi(1 - \cos\theta)$, $a_2 = \pi \sin^2\theta$, $a_3 = \pi(2 - 3\cos\theta + \cos^3\theta)/3$. Consideration of the mechanical equilibrium among the interfacial tensions or forces yields Young's equation

$$\gamma_{sv} = \gamma_{fs} + \gamma_{vf}\cos\theta . \tag{5-2}$$

Therefore, the contact angle θ depends only on the surface properties of the involved materials. The three modes of film growth can be distinguished on the

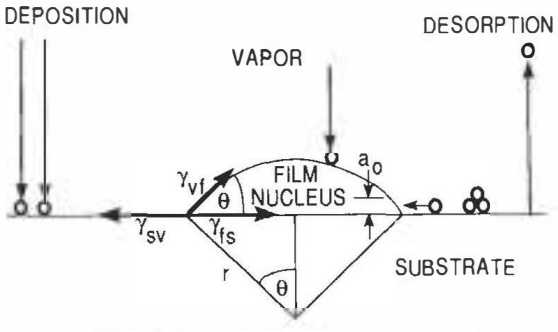


Figure 5-3. Schematic of basic atomistic processes on substrate surface during vapor deposition.

basis of Eq. 5-2. For island growth, $\theta > 0$, and therefore

$$\gamma_{sv} < \gamma_{fs} + \gamma_{vf}. \tag{5-3}$$

For layer growth the deposit "wets" the substrate and $\theta = 0$. Therefore,

$$\gamma_{sv} = \gamma_{fs} + \gamma_{vf}. \tag{5-4}$$

A special case of this condition is ideal homo- or *autoepitaxy*. Because the interface between film and substrate essentially vanishes, $\gamma_{fs} = 0$. Lastly, for S.K. growth,

$$\gamma_{sv} > \gamma_{fs} + \gamma_{vf}. \tag{5-5}$$

In this case, the strain energy per unit area of film overgrowth is large with respect to γ_{vf}, permitting nuclei to form above the layers. In contrast, a film strain energy that is small compared with γ_{vf} is characteristic of layer growth.

Returning now to Eq. 5-1, we note that any time a new interface appears there is an increase in surface free energy, 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. The critical nucleus size r^* (i.e., the value of r when $d\Delta G/dr = 0$) is given by differentiation, namely,

$$r^* = \frac{-2(a_1\gamma_{vf} + a_2\gamma_{fs} - a_2\gamma_{sv})}{3a_3\,\Delta G_V}. \tag{5-6}$$

Correspondingly, ΔG evaluated at $r = r^*$ is

$$\Delta G^* = \frac{4(a_1\gamma_{vf} + a_2\gamma_{fs} - a_2\gamma_{sv})^3}{27a_3^2\,\Delta G_V^2}. \tag{5-7}$$

Both r^* and ΔG^* scale in the manner shown in Fig. 1-19. 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 that similarly lowers ΔG. In heterogeneous nucleation the accommodating substrate catalyzes vapor condensation and the energy barrier ΔG^* depends on the contact angle. After substitution of the geometric constants, it is easily shown that ΔG^* is essentially the product of two factors; i.e.,

$$\Delta G^* = \left(\frac{16\pi\gamma_{vf}^3}{3\,\Delta G_V^2}\right)\left(\frac{2 - 3\cos\theta + \cos^3\theta}{4}\right). \tag{5-8}$$

The first is the value for ΔG^* derived for homogeneous nucleation. It is modified by the second term, a wetting factor that 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 $a_3 r^3 \Delta G_s$, where ΔG_s is the strain energy per unit volume, would be appropriate. In the calculation for ΔG^*, the denominator of Eq. 5-7 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—then stress relieval during nucleation would be manifested by a reduction of ΔG^*. Substrate charge and impurities similarly influence ΔG^* by affecting terms related to either surface and volume electrostatic, chemical, etc., energies.

5.2.2. 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 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. 5-3, energetic vapor atoms that impinge on the substrate may immediately desorb, but usually they remain on the surface for a length of time τ_s given by

$$\tau_s = \frac{1}{\nu} \exp \frac{E_{\text{des}}}{kT}. \tag{5-9}$$

The vibrational frequency of the adatom on the surface is ν (typically 10^{12} sec^{-1}), and 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, in the course of their migration, may 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. Changes in E_{des} are particularly expected 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 number of atomic bonds available at these accommodating sites leads to higher E_{des} values. For this reason, 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, depending on type and distribution of atoms or molecules involved.

We now exploit some of these microscopic notions in the capillarity theory of the nucleation rate $\dot{N}$. Reproducing Eq. 1-41, we obtain the expression for $\dot{N}$:

$$\dot{N} = N^* A^* \omega \text{ nuclei/cm}^2\text{-sec}. \tag{5-10}$$

(The Zeldovich factor, included in other treatments, is omitted here for simplicity.) Based on the thermodynamic probability of existence, the equilibrium number of nuclei of critical size per unit area of substrate is given by

$$N^* = n_s \exp - \Delta G^*/kT. \tag{5-11}$$

The quantity n_s represents the total nucleation site density. A certain number of these sites are occupied by adatoms whose surface density, n_a, is given by the product of the vapor impingement rate (Eq. 2–8) and the adatom lifetime, or

$$n_a = \tau_s P N_A / \sqrt{2\pi MRT}. \tag{5-12}$$

Surrounding the cap-shaped nucleus of Fig. 5-3 are adatoms poised to attach to the circumferential belt whose area is

$$A^* = 2\pi r^* a_0 \sin\theta. \tag{5-13}$$

Quantities r^* and θ were defined previously, and a_0 is an atomic dimension.

Lastly, the impingement rate onto area A^* requires adatom diffusive jumps on the substrate with a frequency given by $\nu \exp -E_S/kT$, 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

$$\omega = \frac{\tau_s P N_A \nu \exp - E_S/kT \,(\text{cm}^{-2}\text{sec}^{-1})}{\sqrt{2\pi MRT}} \tag{5-14}$$

There is no dearth of adatoms that 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

$$X = \sqrt{2 D_S \tau_s}. \tag{5-15}$$

The surface diffusion coefficient D_S is essentially

$$D_S = (1/2) a_0^2 \nu \exp - E_S/kT$$

and therefore

$$X = a_0 \exp \frac{E_{des} - E_S}{2kT} \tag{5-16}$$

Large values of E_{des} coupled with small values of E_S serve to extend the nucleus capture radius.

Upon substitution of Eqs. 5-11, 5-13, and 5-14 in Eq. 5-10, we obtain

$$\dot{N} = 2\pi r^* a_0 \sin\theta \frac{PN_A}{\sqrt{2\pi MRT}} n_s \exp\frac{E_{des} - E_S - \Delta G^*}{kT} \tag{5-17}$$

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 previously, a high nucleation rate encourages a fine-grained, or even amorphous, structure, whereas a coarse-grained deposit develops from a low value of $\dot{N}$.

5.2.3. Nucleation Dependence on Substrate Temperature and Deposition Rate

Substrate temperature and deposition rate $\dot{R}$ (atoms/cm^2-sec) are among the chief variables affecting deposition processes. Calculating their effect on r^* and ΔG^* is both instructive and simple to do. First we define

$$\Delta G_V = -\frac{kT}{\Omega} \ln\frac{\dot{R}}{\dot{R}_e} \tag{5-18}$$

by analogy to Eq. 1-38, where $\dot{R}_e$ is the equilibrium evaporation rate from the film nucleus at the substrate temperature T, and Ω is the atomic volume. Assuming an inert substrate, i.e., $\gamma_{vf} = \gamma_{fs}$, direct differentiation of Eq. 5-6 yields (Ref. 3)

$$\left(\frac{dr^*}{dT}\right)_{\dot{R}} = \frac{2}{3}\left|\frac{\gamma_{vf}(\partial \Delta G_v/\partial T) - (a_1 + a_2)\,\Delta G_v(\partial \gamma_{vf}/\partial T)}{a_3\,\Delta G_V^2}\right|. \tag{5-19}$$

Assuming typical values, $\gamma_{vf} \approx 1000$ ergs/cm^2 and $\partial\gamma_{vf}/\partial T = -0.5$ erg/cm^2-K. An estimate for $\partial \Delta G_v/\partial T$ is the entropy change for vaporization, which is roughly 8×10^7 ergs/cm^3-K for many metals. If $|\Delta G_v| < 1.6 \times 10^{11}$ ergs/cm^3, then direct substitution gives

$$(\partial r^*/\partial T)_{\dot{R}} > 0. \tag{5-20}$$

Similarly, by the same assumptions and arguments,

$$(\partial \Delta G^*/\partial T)_{\dot{R}} > 0. \tag{5-21}$$

It is also simple to show that

$$\left(\partial r^*/\partial \dot{R}\right)_T < 0. \tag{5-22}$$

Direct chain-rule differentiation using Eqs. 5-6 and 5-18 yields

$$\frac{\partial r^*}{\partial \dot{R}} = \left(\frac{\partial r^*}{\partial \Delta G_V}\right)\left(\frac{\partial \Delta G_V}{\partial \dot{R}}\right) = \left(-\frac{r^*}{\Delta G_V}\right)\left(-\frac{kT}{\Omega \dot{R}}\right).$$

Since ΔG_v is negative, the overall sign is negative. Similarly, it is easily shown that

$$\left(\partial \Delta G^*/\partial \dot{R}\right)_T < 0. \tag{5-23}$$

The preceding four inequalities have interesting implications and summarize a number of common effects observed during film deposition. From Eq. 5-20 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. 5-21) 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 $\dot{N}$ 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. 5-22, 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 formation, this is equivalent to high substrate temperatures and low deposition rates. Alternatively, 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. 5-4 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, $r^* = -2\gamma/\Delta G_V$, is used, where $\Delta G_V = -kT/\Omega \ln P_v/P_s$ (Eq. 1-38). Assuming $\Omega = 20 \times 10^{-24}\text{cm}^3/\text{at}$, $\gamma = 1000$ ergs/cm^2, $P_v = 10^{-3}$ torr,

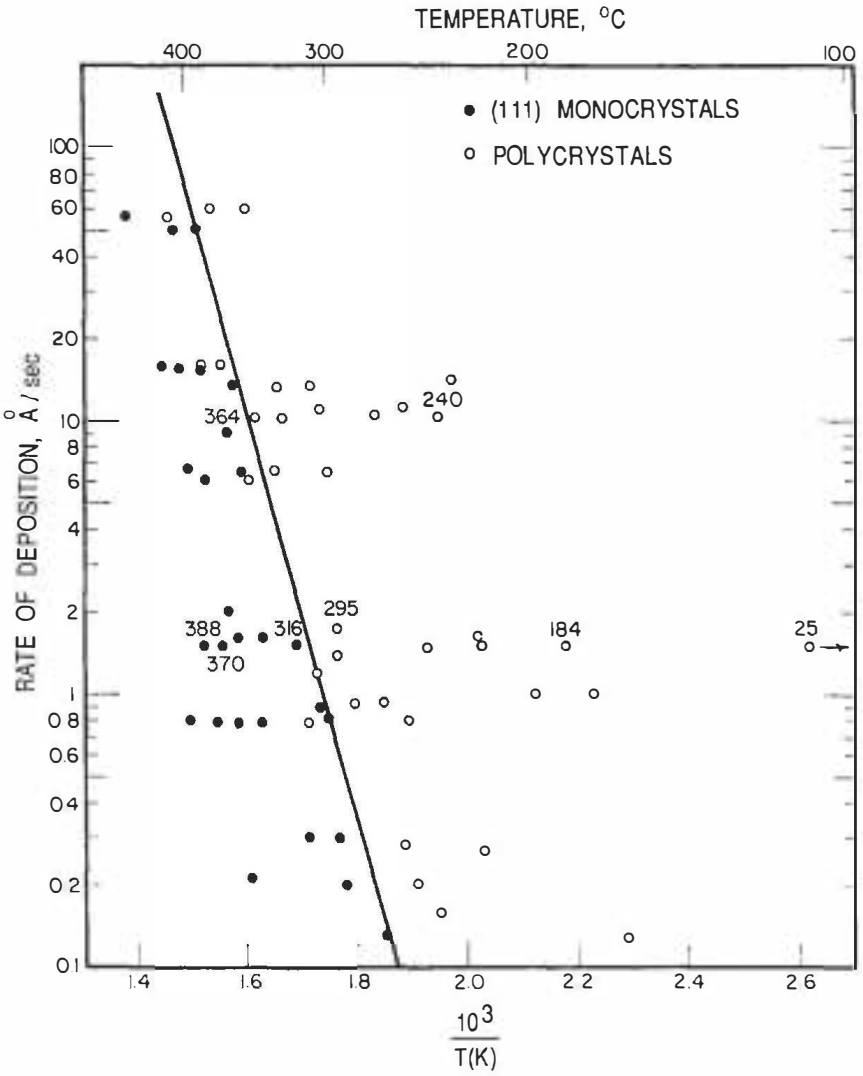


Figure 5-4. Dependence of microstructure on deposition rate and substrate temperature for Cu films on (111) NaCl. (Some substrate temperatures are noted.) (Reprinted with permission from the Metals Society, from R. W. Vook, *Int. Metals Rev.* **27**, 209, 1982).

and $P_s = 10^{-10}$ torr, then

$$r^* = \frac{-2(1000)}{-1.38 \times 10^{-16}(300/20 \times 10^{-24})\ln(10^{-3}/10^{-10})} = 6 \times 10^{-8}\ \text{cm}.$$

A nucleus of this size contains just a few atoms, and it is therefore doubtful that continuum concepts like surface tension and nucleus radius have much significance in such a case. It is more realistic to interpret vapor condensation

phenomena in terms of a theory of heterogeneous nucleation based on an atomistic model. This is considered in the next section.

5.3. Atomistic Nucleation Processes

5.3.1. The Walton – Rhodin Theory

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 milestone in the atomistic approach to nucleation was the theory advanced by Walton *et al.* (Ref. 6), 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

$$\frac{N_{i*}}{n_0} = \left|\frac{N_1}{n_0}\right|^{i^*} \exp\frac{E_{i*}}{kT}, \tag{5-24}$$

which expresses the chemical equilibrium between the clusters and monomers. In Eq. 5-24, 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. 5-12, is given by $N_1 = \dot{R}\tau_s$. Hence,

$$N_1 = \dot{R}\nu^{-1}\exp E_{\text{des}}/kT. \tag{5-25}$$

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. 5-16,

$$\dot{R}X^2 = \dot{R}a_0^2\exp\frac{E_{\text{des}} - E_S}{kT}. \tag{5-26}$$

By combining Eqs. 5-24–5-26, we obtain the critical nucleation rate (cm^{-2} sec^{-1})

$$\dot{N}_{i*} = \dot{R}a_0^2 n_0\left(\frac{\dot{R}}{\nu n_0}\right)^{i^*}\exp\left(\frac{(i^*+1)E_{\text{des}} - E_S + E_{i*}}{kT}\right). \tag{5-27}$$

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. It has the advantage of expressing the nucleation rate in terms of measurable parameters rather than macroscopic quantities such as ΔG^*, γ, or θ, which are not known with certainty, nor are they easily estimated in the capillarity theory. The uncertainties are now 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. 5-5, 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., $\dot{N}^* \propto \dot{R}^2$, 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 nucleus orientation were preserved with subsequent deposition and cluster impingement.

A thermally activated nucleation rate whose activation energy depends on the size of the critical nucleus is predicted by Eq. 5-27. This suggests the existence of critical temperatures where the nucleus size and orientation undergo change. For example, the temperture $T_{1\to 2}$ at which there is a

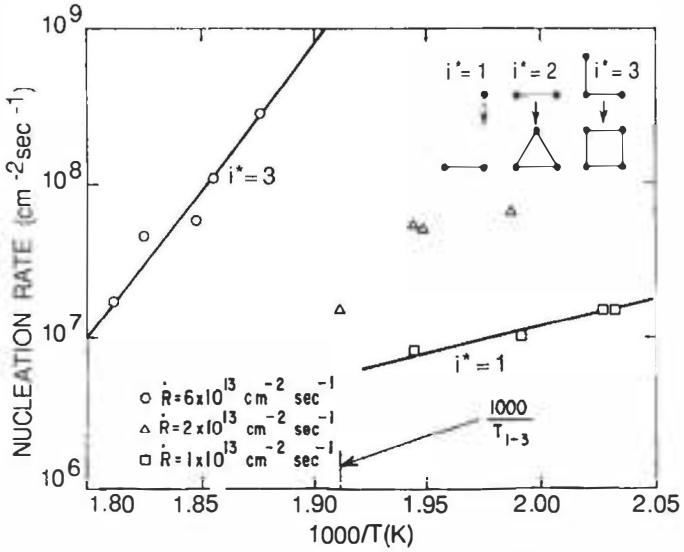


Figure 5-5. Nucleation rate of Ag on (100) NaCl as a function of temperature. Data for three different deposition rates are plotted. Also shown are smallest stable epitaxial clusters corresponding to critical nuclei. (From Ref. 6).

transition from a one- to two-atom nucleus is given by equating the respective nucleation rates:

$$\dot{N}_{i^*=1} = \dot{N}_{i^*=2}.$$

After substitution and some algebra, we have

$$T_{1\to 2} = -\frac{E_{\text{des}} + E_2}{k \ln(\dot{R}/\nu n_0)} \tag{5-28a}$$

or

$$\dot{R} = \nu n_0 \exp -\left|\frac{E_{\text{des}} + E_2}{kT_{1\to 2}}\right|. \tag{5-28b}$$

Equation 5-28b can be used to analyze the data in Fig. 5-4. 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 ($i^* = 1$). The 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-sec or 1 Å/sec, and $\nu n_0 = 6.9 \times 10^{27}$ atoms/cm^2-sec, the epitaxial transition temperature can be calculated to be 577 K. Equations similar to 5-28 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. For smooth layer-by-layer growth on high-symmetry crystal surfaces, adatoms must reach growth ledges

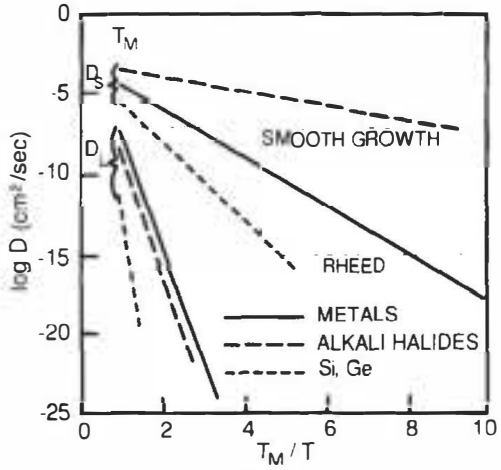


Figure 5-6. Lattice (D_L) and surface diffusivities (D_S) as a function of T_M / T for metals, semiconductors, and alkali halides. (From Ref. 7).

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 sec. By Eq. 5-15 a surface diffusivity greater than $\sim 10^{-8}$ cm^2/sec is, therefore, required irrespective of the material deposited. This means a different critical epitaxial growth temperature (T_E), depending on whether semiconductor, metal, or alkali halide films are involved. In the Arrhenius plot of Fig. 5-6, D_S values are graphed versus T_M/T, where T_M is the melting point. Similar relationships for lattice, grain-boundary, and dislocation diffusivities will be introduced in Chapter 8. At the critical value of $D_S = 10^{-8}$, Fig. 5-6 indicates that $T_E \sim 0.5T_M$, $\sim 0.3T_M$, and $\sim 0.1T_M$ for layer growth on group IV semiconductors, metals, and alkali halides, respectively. These temperatures agree qualitatively with experimental observations.

5.3.2. Kinetic Models of Nucleation

More recent 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 rate of change of cluster densities in terms of the processes that 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

$$\frac{dN_1}{dt} = \dot{R} - \frac{N_1}{\tau_s} - K_1 N_1^2 - N_1 \sum_{i=2}^{\infty} K_i N_i . \tag{5-29}$$

The equation states that the time rate of change of the monomer density is given by the incidence rate, minus the 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

$$dN_i/dt = K_{i-1} N_1 N_{i-1} - K_i N_1 N_i , \tag{5-30}$$

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 depends on direct impingement from the vapor as well as desorption through their link back to Eq. 5-25. Inclusion of diffusion terms, i.e., d^2N_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 preceding rate equations for a variety of physical situations and the results are summarized below starting with the case for which $i^* = 1$.

1. The nucleation and growth kinetics for the one-atom critical nucleus has been treated under simplifying assumptions by Robinson and Robins (Ref. 8). They divided the deposition conditions into two categories. At high temperatures and/or low deposition rates, the reevaporation rate from the surface will control the adatom density and exceed the rate of diffusive capture into growing nuclei. Here the adsorption–desorption equilibrium is rapidly established where $N_1 = \dot{R}\tau_s$, and *incomplete* condensation is said to occur. When reevaporation is not important, however, because the energy to desorb adatoms (E_{des}) is large, then we speak of *complete* condensation. This regime dominates at low tempertures and/or high deposition rates where the monomer capture rate by growing nuclei exceeds the rate at which they are lost due to reevaporation. These same condensation regimes also refer to situations where $i^* > 1$. The division between these high- and low-temperature regions is not abrupt like $T_{1\to 2}$ (Eq. 5-28a). It can be characterized by a dividing temperature T_D, calculated to be

$$T_D = \frac{2E_S - 3E_{\text{des}}}{k \ln\left[\left(C\alpha^2/\beta\right)\left(\dot{R}/\nu n_0\right)\right]}, \tag{5-31}$$

where C is the number of pair formation sites ($C = 4$ for a square lattice), and α and β are dimensionless constants with typical values of ~ 0.3 and 4, respectively. Relatively simple closed-form analytic expressions for the time (t) dependence of the transient density of stable nuclei $N(t)$ as well as the saturation value of $N(t = \infty)$ or N_S in terms of measurable quantities were obtained at temperatures above or below T_D.

For $T > T_D$,

$$N(t) = N_S \tanh\left(\frac{\dot{N}_o(0)\,t}{N_S}\right), \tag{5-32}$$

$$N_S = \left(\frac{Cn_0}{\beta\nu}\right)^{1/2} \dot{R}^{1/2} \exp\frac{E_{\text{des}}}{2kT}. \tag{5-33}$$

For $T < T_D$,

$$N(t) = N_S\left\{1 - \exp - 3\eta^2\frac{\dot{N}(0)t}{N_S^3}\right\}^{1/3}, \tag{5-34}$$

$$N_S = \left(\frac{Cn_0^2}{\alpha\beta\nu}\right)^{1/3}\dot{R}^{1/3}\exp\frac{E_S}{3kT}. \tag{5-35}$$

Clearly, as $t \to \infty$, $N(\infty) \to N_S$. For all temperatures,

$$\dot{N}(0) = \frac{C\dot{R}^2}{n_0\nu}\exp\frac{2E_{\text{des}} - E_S}{kT}. \tag{5-36}$$

Only $\dot{N}(0)$, the initial nucleation rate (i.e., $\dot{N}(0) = dN(t)/dt$ at $t = 0$) and η, which is equal to $(n_0/\alpha)\exp(E_S - E_{\text{des}}/kT)$, have not been defined before.

The equations 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}, depending on deposition rate and substrate temperature. With further deposition, N decreases due to nucleus coalescence. Furthermore, N_S is larger the lower the substrate temperature, as shown schematically in Fig. 5-7. This is an obvious consequence of Eqs. 5-33 and 5-35. When the temperature range spanned exceeds T_D, N_S data as a function of T directly yield $E_{\text{des}}/2$ on an Arrhenius plot. Similarly, $E_S/3$ can be extracted from such data when $T < T_D$. In this manner, values of E_{des} and E_S for noble metal–alkali halide systems have been obtained and the results summarized in Table 5-1. An experimental verification of Eq. 5-36 is shown in Fig. 5-8. That

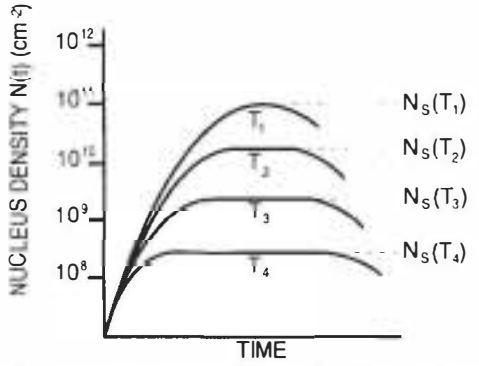


Figure 5-7. Schematic dependence of $N(t)$ with time and substrate temperature. ($T_4 > T_3 > T_2 > T_1$) From Ref. 4.

Table 5-1. Nucleation Parameter Values for Noble Metal–Alkali Halide Systems

System	E_{des} (eV)	E_S (eV)	ν (10^{12}/sec)
Au–NaCl	0.70	0.29	1.65
Au–KCl	0.68	0.25	1.26
Au–KBr	0.83	0.44	~ 24
Au–NaF	0.64	0.093	0.91
Ag–NaCl	0.63	0.093	~ 0.8
Ag–KCl	0.48	0.2	0.24
Ag–KBr	0.48	0.22	~ 0.1

From Ref. 5.

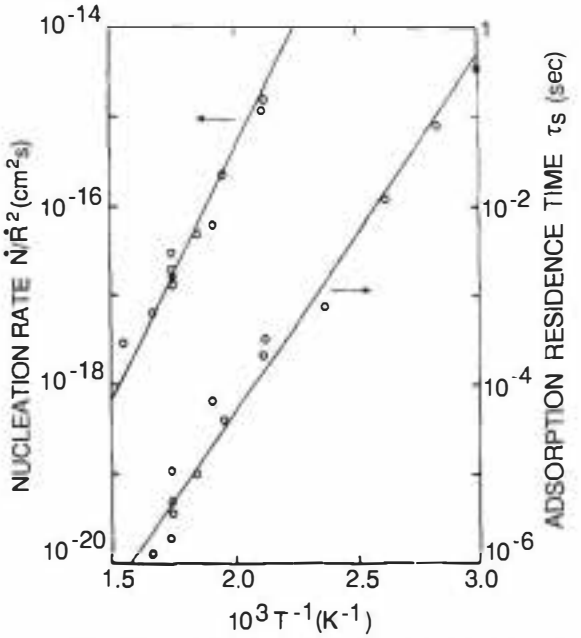


Figure 5-8. Plot of $\dot{N}/\dot{R}^2$ vs. $1/T$ (left) and τ_s vs. $1/T$ (right) for Au on KCl. (From Ref. 5, with permission from IOP Publishing Ltd.)

$\dot{N}(0)/\dot{R}^2$ is observed to vary linearly as a function of $1/T$ is confirmation that Au monomers are stable nuclei on KCl substrates.

2. Venables *et al.* (Ref. 5) have neatly summarized nucleation behavior for cases where i^* assumes any integer value. In general, the stable cluster density is given by

$$N_S = An_0\left|\frac{\dot{R}}{n_o\nu}\right|^p \exp\frac{E}{kT}, \tag{5-37}$$

Table 5-2. Nucleation Parameters p and E in Eq. 5-37

Regime	3D Islands	2D Islands
Extreme incomplete	$p = (2/3)i^*$	i^*
	$E = (2/3)[E_{i^*} + (i^* + 1)E_{des} - E_S]$	$E_{i^*} + (i^* + 1)E_{des} - E_S$
Initially incomplete	$p = 2i^*/5$	$i^*/2$
	$E = (2/5)(E_{i^*} + i^*E_{des})$	$(1/2)(E_{i^*} + i^*E_{des})$
Complete	$p = i^*/(i^* + 2.5)$	$i^*/(i^* + 2)$
	$E = \dfrac{E_{i^*} + i^*E_S}{i^* + 2.5}$	$\dfrac{E_{i^*} + i^*E_S}{i^* + 2}$

From Ref. 5.

where A is a calculable dimensionless constant dependent on the substrate coverage. Parameters p and E depend on the condensation regime and are summarized in Table 5-2. Three regimes of condensation and two types of island nuclei are considered. The *complete* and *extreme incomplete* condensation categories parallel those considered previously, but intermediate incomplete condensation regimes may also be imagined, depending on deposition conditions. 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, and values for the energies of desorption, diffusion, and cluster binding have been extracted from them. The Walton *et al.* nucleation theory is seen to be a special case of the more general rate theory. For extremely incomplete condensation of 2-D islands, the p and E parameters are seen to vary in the same way as Eq. 5-27.

Although the major application of the kinetic model has been to island growth, the theory is also capable of describing S.K. growth.

5.4. Cluster Coalescence and Depletion

The results of the kinetic theories of nucleation indicate that in the initial stages of growth the density of stable nuclei increases with time up to some maximum level before decreasing. In this section, the coalescence processes that are operative beyond the cluster saturation regime are examined. Coalescence of nuclei is generally characterized by the 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 liquidlike in nature with islands merging and undergoing shape changes after the fashion of liquid droplet motion.
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 these are discussed in turn.

5.4.1. Ostwald Ripening (Ref. 2)

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 time evolution of the distribution of island sizes has been considered both from a macroscopic surface diffusion–interface transfer viewpoint as well as from statistical models involving single atom processes. The former is simply driven by a desire to minimize the surface free energy of the island structure. To understand the process, consider two isolated islands of different size in close proximity, as shown in Fig. 5-9a. For simplicity they are assumed to be spherical with radii r_1 and r_2. The surface free energy per unit area of a given

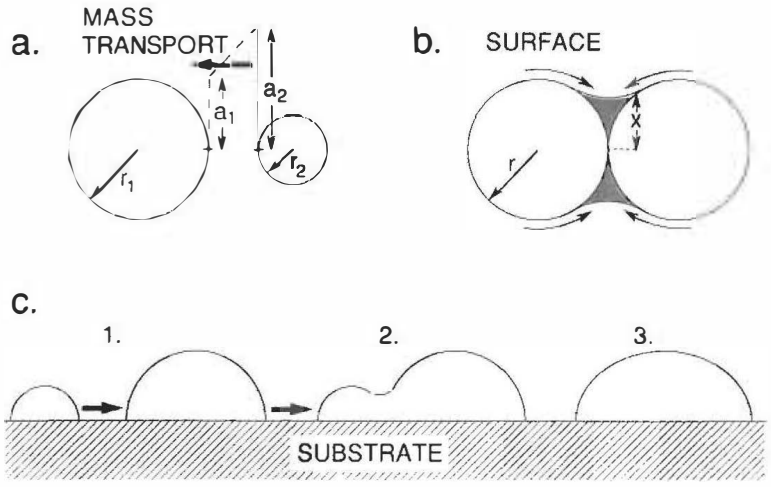


Figure 5-9. Coalescence of islands due to (a) Ostwald ripening, (b) sintering, (c) cluster migration.

island is γ, so the total energy $G_\gamma = 4\pi r_i^2\gamma$. 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 as dG_γ/dn_i in this application, we have, after substitution,

$$\mu_i = \frac{8\pi r_i\gamma\, dr_i}{4\pi r_i^2\, dr_i/\Omega} = \frac{2\Omega\gamma}{r_i}$$

Making use of Eq. 1-9 ($\mu_i = \mu_0 + kT \ln a_i$), we see that the Gibbs–Thomson relation

$$kT \ln\frac{a_i}{a_\infty} = \frac{2\Omega\gamma}{r_i} \qquad \text{or} \qquad a_i = a_\infty \exp\frac{2\Omega\gamma}{r_i kT} \tag{5-38}$$

directly follows. This equation states that atoms in an island of radius r_i can be in equilibrium only with a substrate adatom activity or effective concentration a_i. The quantity a_∞ may be interpreted as the adatom concentration in equilibrium with a planar island ($r_i = \infty$) or, alternatively, with the vapor pressure of island atoms at temperature T. When the island surface is convex (r_i is positive), atoms have a greater tendency to escape, compared with atoms situated on a planar surface, because there are relatively fewer atomic bonds to attach to. Therefore, $a_i > a_\infty$. Conversely, at a concave island surface, r_i is negative and $a_\infty > a_i$. These simple ideas have significant implications not only with respect to Ostwald ripening but to the sintering mechanisms of coalescence that are treated in the next section.

The establishment of the concentration gradient of adatoms situated between the two particles of Fig. 5-9a can now be understood. Diffusion of individual adatoms will proceed from the smaller to larger island until the former disappears entirely. A mechanism has thus been established for coalescence without the islands having to be in direct contact. In a multi-island array the kinetic details are complicated, but ripening serves to establish a quasi-steady-state island size distribution that changes with time. Ostwald ripening processes never reach equilibrium during film growth since the theoretically predicted narrow distribution of crystallite sizes is generally not observed.

5.4.2. Sintering (Ref. 9)

Sintering is a coalescence mechanism involving islands in contact. It can be understood by referring to Fig. 5-10, depicting a time sequence of coalescence events between Au particles deposited on molybdenite (MoS_2) at 400 °C and photographed within the transmission electron microscope (TEM). Within tenths of a second a neck forms between islands and then successively thickens

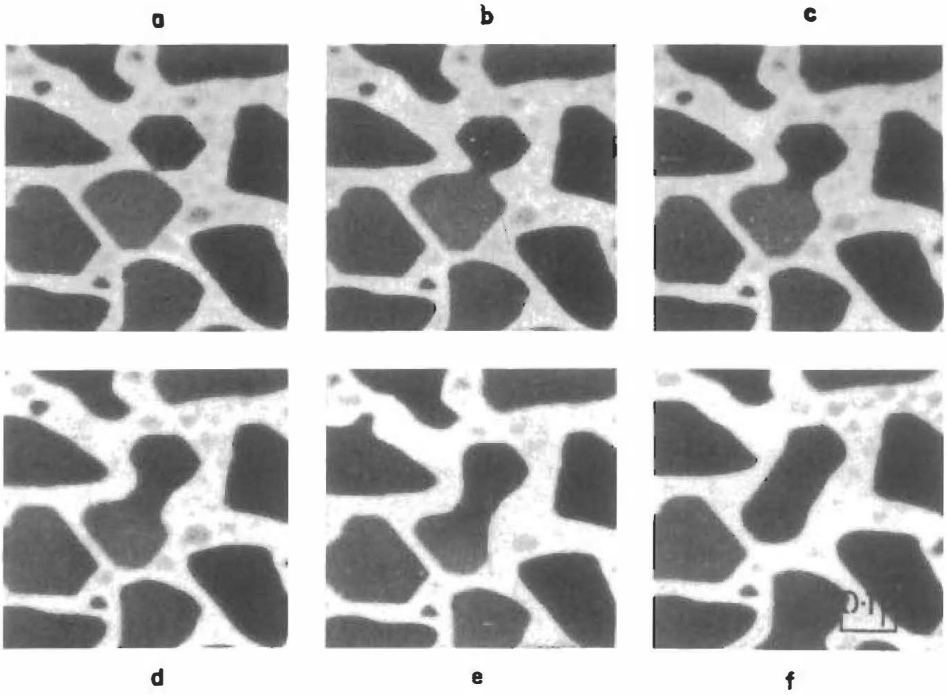


Figure 5-10. Successive electron micrographs of Au deposited on molybdenite at 400 °C illustrating island coalescence by sintering (a) arbitrary zero time, (b) .06 sec, (c) 0.18 sec, (d) 0.50 sec, (e) 1.06 sec, (f) 6.18 sec. (From Ref. 10).

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 atoms on the convex island surfaces have a greater activity than atoms situated in the concave neck, an effective concentration gradient between these regions develops. 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. Of the several mechanisms available for mass transport, the two most likely ones involve self-diffusion through the bulk or via the surface of the islands. In the case of sintering or coalescence of two equal spheres of radius r (Fig. 9b), theoretical calculations in the metallurgical literature have shown that the sintering kinetics is given by $x^n/r^m = A(T)t$, where x is the neck radius, $A(T)$ is a temperature-dependent constant, and n and m are constants. Explicit expressions for the bulk and surface diffusion mechanisms are

$$\frac{x^5}{r^2} = \frac{10\pi D_L \gamma \Omega t}{kT} \tag{5-39}$$

and

$$\frac{x^7}{r^3} = \frac{28 D_S \gamma (\Omega)^{4/3} t}{kT}, \tag{5-40}$$

respectively, and D_L and D_S are the lattice and surface diffusion coefficients, respectively. In principle, experimental determination of m and n and the activation energy for diffusion would serve to pinpoint the transport mechanism, but insufficient data have precluded such an analysis. A simple relative comparison between the two mechanisms can, however, be made to order to predict which dominates. The ratio of the times required to reach a neck radius $x = 0.1r$, for example, is given by

$$\frac{t_S}{t_L} = \frac{\pi r}{280} \frac{D_L}{D_S \Omega^{1/3}}. \tag{5-41}$$

In the case of Au at 400 °C, the ratio of D_L/D_S can be directly read off Fig. 5-6. At the value $T_M/T = 1336/673 = 1.99$, $D_L/D_S \sim 10^{-13}/10^{-6} = 10^{-7}$. Substituting $\Omega^{1/3} = 2.57 \times 10^{-8}$ cm and $r = 10^{-5}$ cm, $t_S/t_L = 4.4 \times 10^{-7}$. Therefore, surface diffusion is expected to control sintering coalescence. This is also true for any plausible combination of r and T values in films.

Surface energy and diffusion-controlled mass-transport mechanisms undoubtedly influence liquidlike coalescence phenomena involving islands in contact, yet other driving forces are probably also operative. For example, sintering mechanisms are unable to explain

1. Observed liquidlike coalescence of metals on substrates maintained at 77 K where atomic diffusion is expected to be negligible
2. Widely varying stabilities of irregularly shaped necks, channels, and islands possessing high curvatures at some points
3. A large range of times required to fill visually similar necks and channels
4. An observed enhanced coalescence in the presence of an applied electric field in the substrate plane

5.4.3. Cluster Migration (Ref. 2)

The last mechanism for coalescence considered deals with migration of clusters on the substrate surface (Fig. 9-c). Coalescence occurs as a result of collisions between separate islandlike 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 up to 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 upon 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, 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. Observations of coalescence in a conservative system include a decreased density of particles, increased mean volume of particles, a particle size distribution that increases in breadth, and a decreased coverage of the substrate.

The surface migration of a cap-shaped cluster with projected radius r is characterized by an effective diffusion coefficient $D(r)$ with units of cm^2/sec. 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. 11)

$$D(r) = \frac{B(T)}{r^s} \exp - \frac{E_C}{kT}, \tag{5-42}$$

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 that can distinguish among the mechanisms. In fact, it is difficult to distinguish cluster mobility coalescence from Ostwald ripening based on observed particle size distributions.

The interesting effect an applied electric field has in enhancing coalescence is worthy of brief comment. Chopra (Ref. 12) has explained the effect on the basis of the interaction of the field with electrically charged islands. The assumed island charge is derived from ionized vapor atoms and/or the potential at the substrate interface. For a spherical particle of radius r, which already possesses a surface free energy, the presence of a charge q contributes additional electrostatic energy (i.e., q^2/r). The increase in total energy is accommodated by an increase in surface area. Therefore, the sphere distorts into a flattened oblate spheroid, the exact shape being determined by the balance of various free energies. The net effect of charging is then to promote

further coalescence by ripening, sintering, or cluster mobility processes. However, with greater charge or higher fields, the cluster may break up in much the same way that a charged droplet of mercury does.

5.5. Experimental Studies of Nucleation and Growth

A full complement of microscopic and surface analytical techniques has been employed to reveal the physical processes of nucleation and test the theories used to describe them. In this section we focus on just a few of the more important experimental techniques and the results of some studies.

5.5.1. Structural Characterization

The most widely used tool, particularly for island growth studies, is the conventional TEM. The technique consists of depositing metals such as Ag and Au onto cleaved alkali halide single crystals (e.g., LiF, NaCl, and KCl) in an ultrahigh-vacuum system for a given time at fixed $\dot{R}$ and T. Then the deposit is covered with carbon, and the substrate is dissolved away outside the system, leaving a rigid carbon replica that 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, that does not condense in the absence of prior noble metal deposition. This technique renders visible otherwise invisible clusters that may contain as few as two atoms, thus making possible more precise comparisons with theory. A disadvantage of these postmortem step-by-step observations is that much useful information on the structure and dynamical behavior of the intermediate stages of 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 at the expense of decreased resolution. The contamination of substrates from hydrocarbons present in the specimen chamber is a general problem in such studies necessitating the use of high-vacuum, oil-less pumping methods. In addition to direct imaging of film nucleation and growth, the electron-diffraction capability of the TEM provides additional information on the crystallography and orientation of deposits. Diffraction halos and continuous, as well as spotty, diffraction rings characterize stages of growth where clusters, which are initially randomly oriented, begin to acquire some preferred orientation as shown in Fig. 5-1. When a

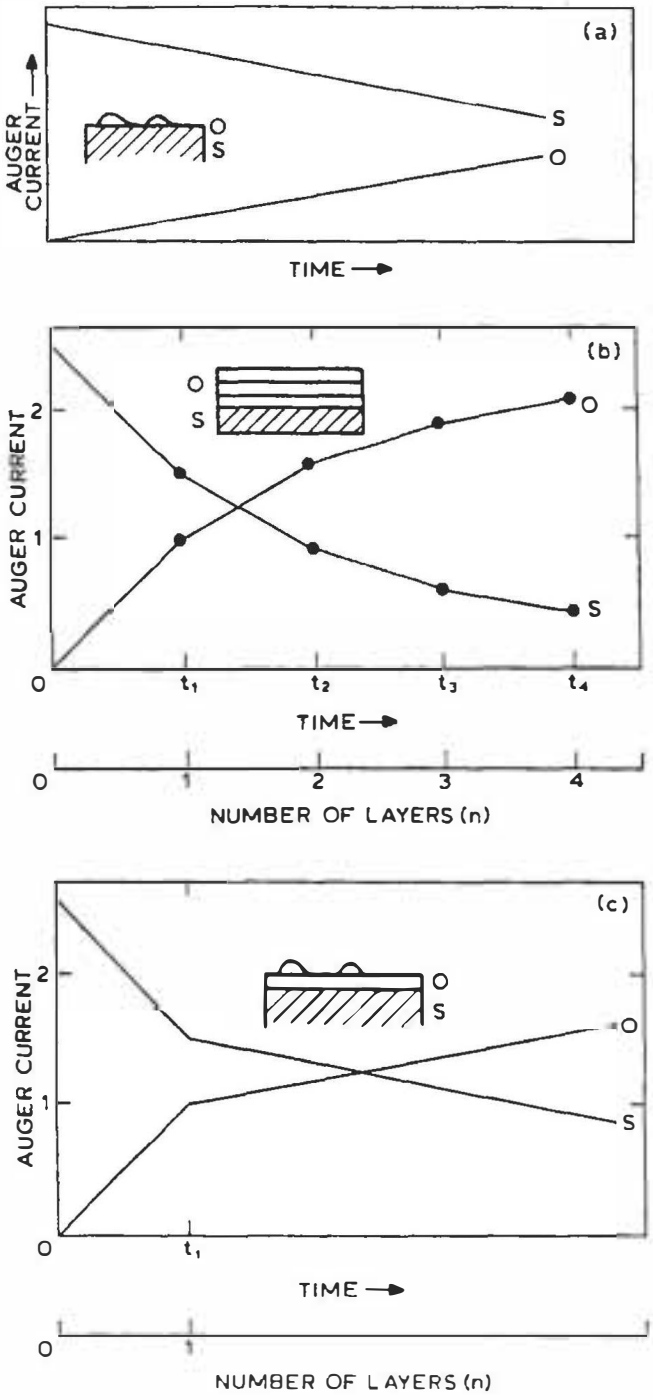


Figure 5-11. Schematic Auger signal currents as a function of time for the three growth modes: (a) island, (b) planar, (c) S.K. O = overlayer, S = substrate. (From Ref. 2).

continuous epitaxial single-crystal film eventually develops, then individual diffraction spots appear.

5.5.2. Auger Electron Spectroscopy (AES)

The AES technique is based on the measurement of the energy and intensity of the Auger electron signal emitted from atoms located within some 5–15 Å of a surface excited by a beam of incident electrons. The subject of AES will be treated in more detail in Chapter 6, but 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 them. The magnitude of the AES signal is related directly to the abundance of the atoms in question. Consider now the deposit substrate combinations corresponding to the three growth modes. 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. 5-11, assuming a sticking coefficient of unity (Ref. 13).

In the case of island growth, the signal from the deposit atoms builds slowly while that from the substrate atoms correspondingly falls. For S.K. growth 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, are formed. The interpretation of the AES signal in the case of layer growth 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 sensitivity in detecting specific elements. 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 the overall response is therefore segmented as indicated in Fig. 5-11b. It would be misleading to suggest that all AES data fit one of the categories in Fig. 5-11. Complications in interpretation arise from atomic contamination, diffusion, and alloying between deposit and substrate, and transitions between two- and three-dimensional growth processes.

5.5.3. Some Results for Metal Films

Studies of the nucleation and growth of metals, especially the noble ones, on assorted substrates have long provided a base for understanding epitaxy and film formation processes. An appreciation of the scope of past as well as present research activity on metal–substrate systems can be gained by referring to Table 5-3. Only epitaxial Au film–substrate combinations are entered in this

Table 5-3. Substrates on Which Epitaxial Gold Deposits Have Been Observed

1. Au–Metal Halides

CaF_2	(2)	KI	(10)	NaI	(1)
CdI_2	(1)	LiF	(7)	$PbCl_2$	(1)
KBr	(10)	NaBr	(2)	RbBr	(1)
KCl	(35)	NaCl	(131)	RbI	(1)
KF	(2)	NaF	(2)		

2. Au–Metals

Ag	(42)	Fe	(3)	Pt	(5)
Al	(1)	Mo	(6)	W	(10)
Cr	(2)	Ni	(2)	Zn	(1)
Cu	(17)	Pb	(2)		
Cu_3Au	(1)	Pd	(17)		

3. Au–Selected Semiconductors and Chalcogenides

C (graphite)	(5)	MoS_2	(20)	ZnS	(1)
Ge	(2)	PbS	(6)		
Si	(16)	PbSe	(3)		
GaSb	(3)	PbTe	(2)		
GaAs	(5)	SnTe	(1)		

4. Au–Carbonates, Oxides, Mica

$CaCO_3$	(3)	MgO	(20)	Mica	(23)
Al_2O_3 (sapphire)	(2)	SiO_2 (quartz)	(4)		
$BaTiO_3$	(2)	ZnO	(2)		

Numbers of references dealing with particular Au–substrate system are in parentheses. From Ref. 14.

recent compilation (Ref. 14), and yet they correspond to over 300 references to the research literature. Other epitaxial metal film–substrate systems have been comprehensively tabulated (Ref. 15) together with deposition methods and variables. The sheer numbers and varieties of metals and substrates involved point to the fact that epitaxy is a common phenomenon. In the overwhelming number of studies, island growth is involved. Perusal of Table 5-3 reveals that epitaxial Au films can be deposited on metallic, covalent, and ionic substrates. Although the majority of substrate materials listed have cubic crystal structures, this is not an essential requirement for epitaxy that occurs, for example, on hexagonal close-packed Zn as well as on monoclinic mica. That epitaxy is possible between materials of different chemical bonding and crystal structure means that its origins are not simple. The long-held belief that a small difference in lattice constant between film and substrate is essential for epitaxy is mistaken; small lattice mismatch is neither a necessary nor 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 discussed in Chapter 7. Very close lattice matching is maintained in the planar epitaxial systems.

The following specific findings briefly characterize the numerous studies of epitaxy of metal films on ionic substrates (Ref. 12).

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 of atoms 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. Temperature T_E depends on the nature of the substrate as well as the deposition rate. For example, T_E for Ag on LiF, NaCl, KCl, and Kl was determined to be 340 °C, 150 °C, 130 °C and 80 °C, respectively. The progressive decrease in T_E correlates with increases in lattice parameter and enhanced ionic (both positive and negative) polarizabilities. The latter 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 $\dot{R} \leq \text{const}\, e^{-E/kT_E}$. 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 $E_{\text{des}} + E_2$ in Eq. 5-28b. The reader should compare this criterion for T_E with those proposed earlier.

4. Contamination. The effect of contamination is a source of controversy. It has been reported that epitaxy of FCC metals is more difficult on ultrahigh-vacuum-cleaved alkali halide substrates than on air-cleaved crystals. Apparently air contamination increases the density of initial nuclei inducing earlier coalescence.

5.6. Grain Structure of Films and Coatings

5.6.1. Zone Models for Evaporated and Sputtered Coatings

Until now the chapter has largely focused on the early stages of the formation of both polycrystalline and single-crystal films. In this section the leap is made

to the regime of the fully developed grain structure of thick polycrystalline films and coatings. 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 positions in the lattice by bulk diffusive motion. This atomic odyssey involves four basic processes: shadowing, surface diffusion, bulk diffusion, and desorption. The last three are quantified by the characteristic diffusion and sublimation activa-

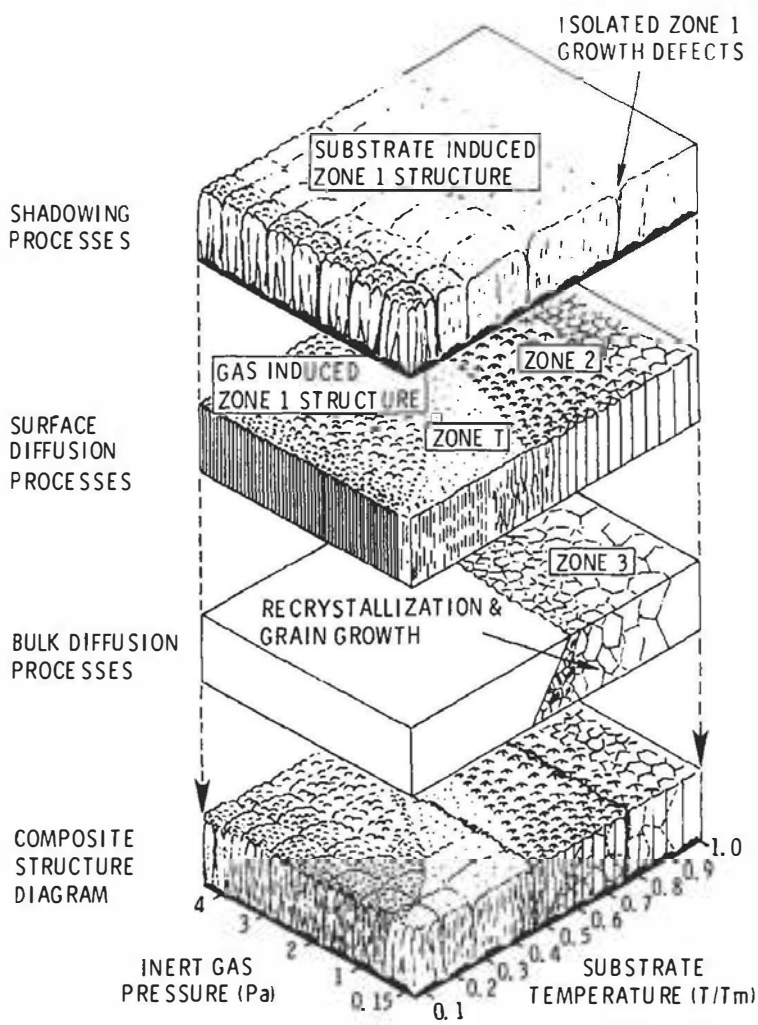


Figure 5-12. Schematic representation showing the superposition of physical processes which establish structural zones. (Reprinted with permission from Ref. 17, © 1977 Annual Reviews Inc.).

tion 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 the zone structure models that have been developed to characterize film and coating grain structures.

The earliest of the zone models was proposed by Movchan and Demchishin (Ref. 16), based on observations of very thick evaporated coatings (0.3 to 2 mm) of metals (Ti, Ni, W, Fe) and oxides (ZrO_2 and Al_2O_3) 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). A similar zone scheme was introduced by Thornton (Ref. 17) for sputtered metal deposits, but with four zones (1, T, 2, 3). His model is based on structures developed in 20- to 250-μm-thick magnetron sputtered coatings deposited at rates ranging from 50 to 20,000 Å/min. The exploded view of Fig. 5-12 illustrates the effect of the individual physical processes on structure and how they depend on substrate temperature

Table 5-4. 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 to 0.4 at 0.15 Pa, ~ 0.4 to 0.5 at 4 Pa	Fibrous grains, dense grain boundary arrays.	High dislocation density, hard, high strength, low ductility.
2 (E)	0.3 to 0.5	Columnar grains, dense grain boundaries.	Hard, low ductility.
2 (S)	0.4 to 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.

and inert sputtering gas (Ar) pressure. A comparison between zone structures and properties for evaporated and sputtered coatings is made in Table 5-4. 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 that overcome limited adatom surface diffusion. In the zone 2 regime, structures are the result of surface diffusion-controlled growth. Lattice diffusion dominates at the highest substrate temperatures, giving rise to the equiaxed recrystallized grains of zone 3.

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.

5.6.2. Zone Model for Evaporated Metal Films

In a recent study (Ref. 18), a zone model for thin evaporated metal films 1000 Å thick has been developed. The results for 10 elemental films are shown in Fig. 5-13a, where the maximum and minimum grain size variation with T_S is shown. 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 as shown schematically in Fig. 5-13b. Although the same zone classification scheme has been used for both sputtered and evaporated films, the grain morphology in zones 1 and T differ. Zones 1 and T (a transition zone) possess structures produced by continued renucleation of grains during deposition and subsequent grain growth. The result is the bimodal grain structure of zone T. Zone 2 structures are the result of granular epitaxy and grain growth. The variation in grain structure in zones 1, T, and 2 presumably arises because different grain boundaries become mobile at different temperatures. In zone 1, virtually all grain boundaries are immobile, whereas in zone 2 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 3. Bulk grain growth and surface recrystallization occur at the highest temperatures with the largest activation energies. This is evident in Fig. 5-13a, which shows the steep dependence of grain size with T_S for $T_M / T_S < 3$.

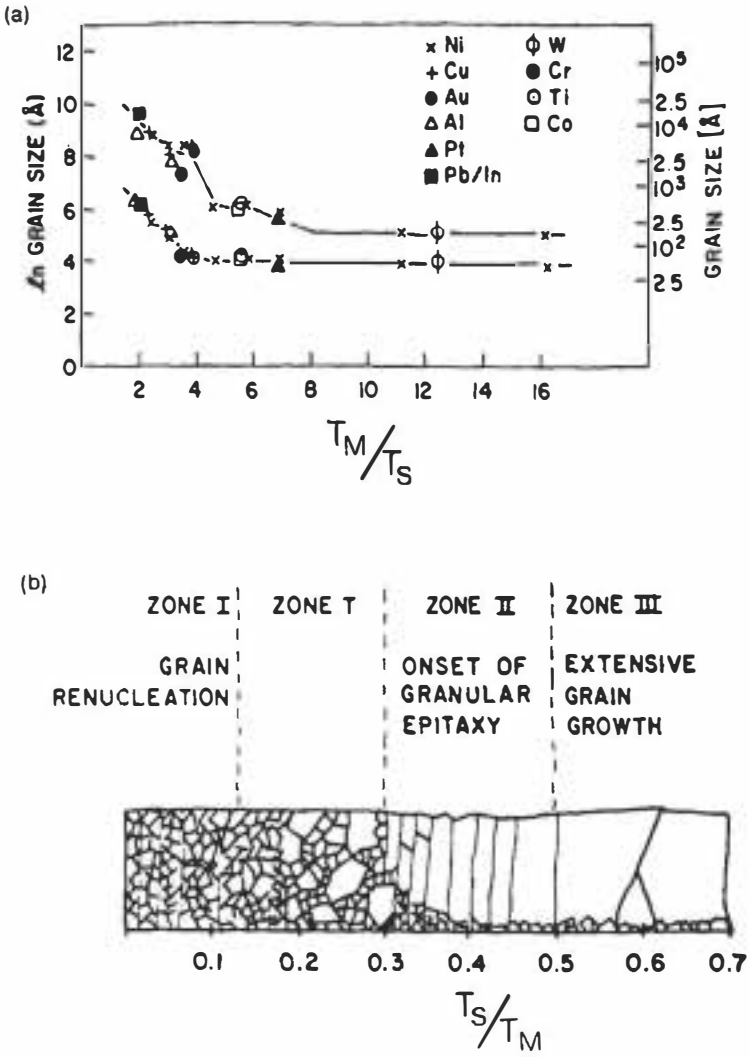


Figure 5-13. (a) Plot of maximum and minimum grain size variation with homologous substrate temperature for 10 different evaporated metals. (b) Zone model for evaporated metal films. (From Ref. 18).

5.6.3. Columnar Grain Structure

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 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 non-noble 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 temperature. Inasmuch as grain boundaries are axiomatically absent in amorphous films, it is 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. 5-14. Particularly noteworthy are the structural similarities among varied materials deposited by different processes, suggesting common nucleation and growth mechanisms.

An interesting observation (Ref. 20) on the geometry of columnar grains has been formulated into the so-called tangent rule expressed by Eq. 5-43. Careful measurements on obliquely evaporated Al films reveal that the columns are oriented toward the vapor source, as shown in the microfractograph of Fig. 5-15. The angle β between the columns and substrate normal is universally observed to be somewhat less than the angle α, formed by the source direction and substrate normal. An experimental relation connecting values of α and β, obtained by varying the incident vapor angle over a broad range ($0 < \alpha < 90°$), was found to closely approximate

$$\tan \alpha = 2 \tan \beta. \tag{5-43}$$

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.

Recently, a closer look has been taken of the detailed microstructure of columnar growth in sputtered amorphous Ge and Si, as well as TiB_2, WO_3, BN, and SiC thin films (Ref. 21). Interestingly, an evolutionary development of columnar grains ranging in size from ~ 20 to 4000 Å occurs. 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

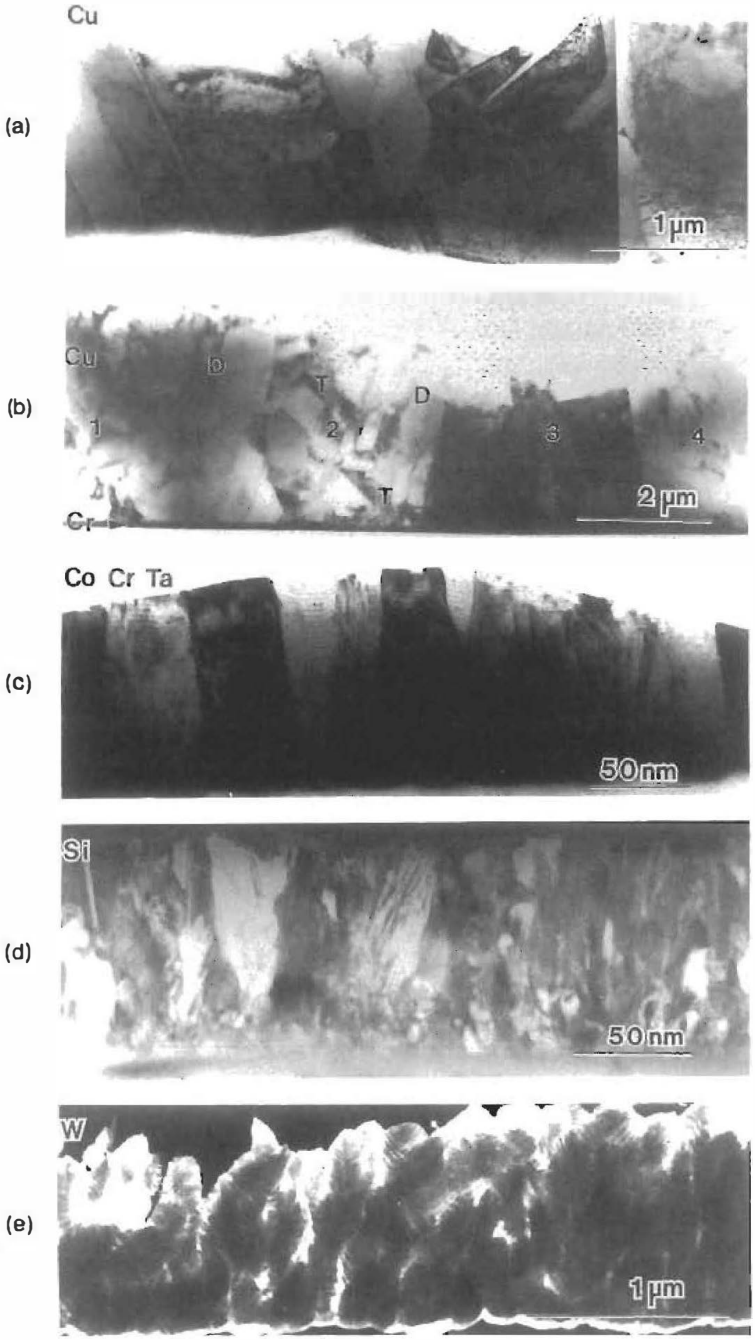


Figure 5-14. Representative set of cross-sectional transmission electron micrographs of thin films illustrating variants of columnar microstructures. (a) acid-plated Cu, (b) sputtered Cu, (c) sputtered Co–Cr–Ta alloy, (d) CVD silicon (also Fig. 4-12), (e) sputtered W, D = dislocation, T = twin. (Courtesy of D. A. Smith, IBM T. J. Watson Research Lab. Reprinted with permission from Trans-Tech Publication, from Ref. 19).

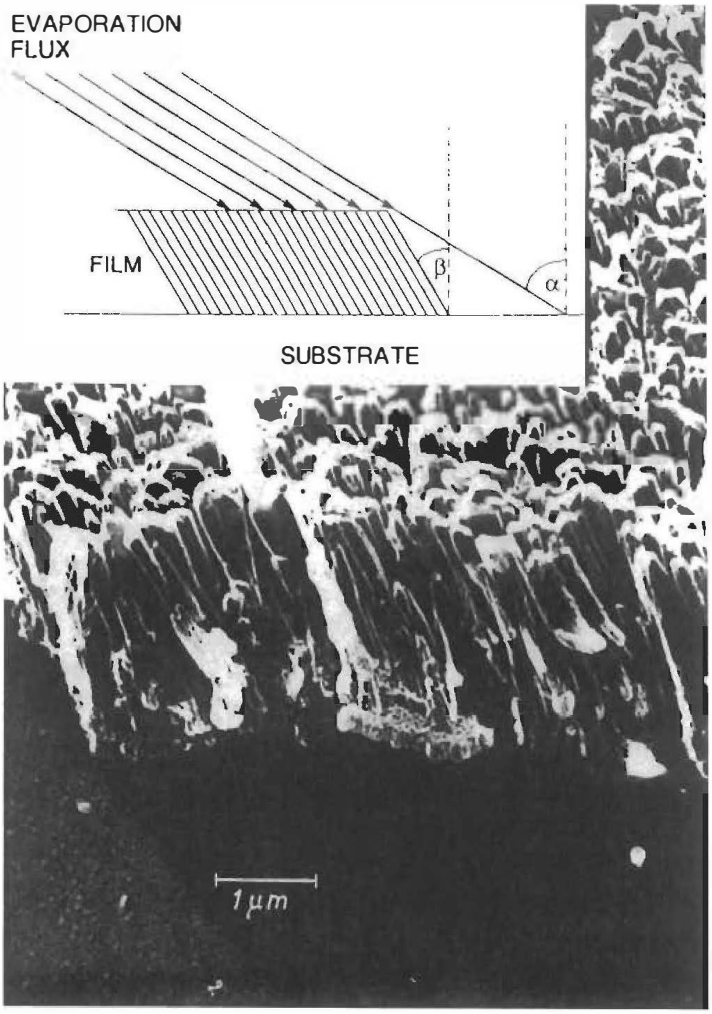


Figure 5-15. Electron micrograph of a replica of a ~2 μm-thick Al film. Inset shows deposition geometry. (From Ref. 20).

these interrelated, nested columns is shown in Fig. 5-16. It is very likely that the columnar grains of zones 1 and T in the Thornton scheme are composed of nano- and microcolumns.

Computer simulations (Ref. 22) have contributed greatly to our understanding of the origin of columnar grain formation and the role played by shadowing. By serially "evaporating" individual hard spheres (atoms) randomly onto a growing film at angle α, the structural simulations in Fig. 5-17 were obtained. The spheres were allowed to relax following impingement into the nearest triangular pocket formed by three previously deposited atoms, thus maximizing close atomic packing. The simulation shows that limited atomic

mobility during low-temperature deposition reproduces features observed experimentally. As examples, film density decreases with increasing α, high-density columnlike regions appear at angles for which $\beta < \alpha$, and film densification is enhanced at elevated temperatures. 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 succeed 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

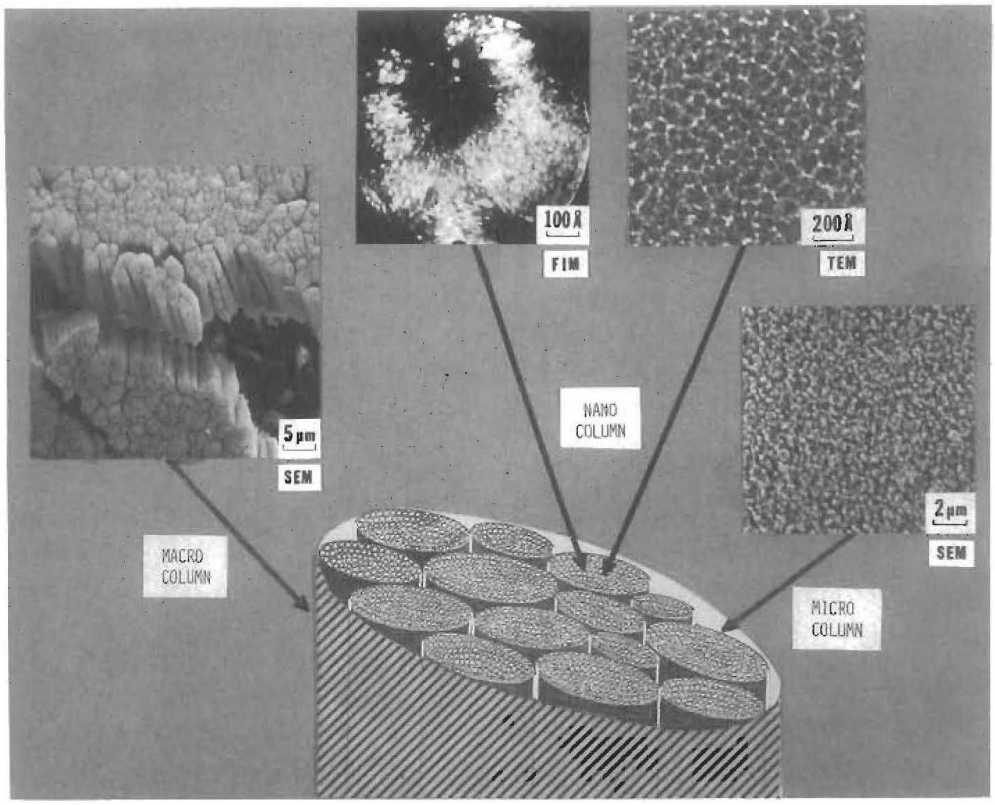


Figure 5-16. Schematic representation of macro, micro and nano columns for sputtered amorphous Ge films. (Courtesy of R. Messier, from Ref. 21).

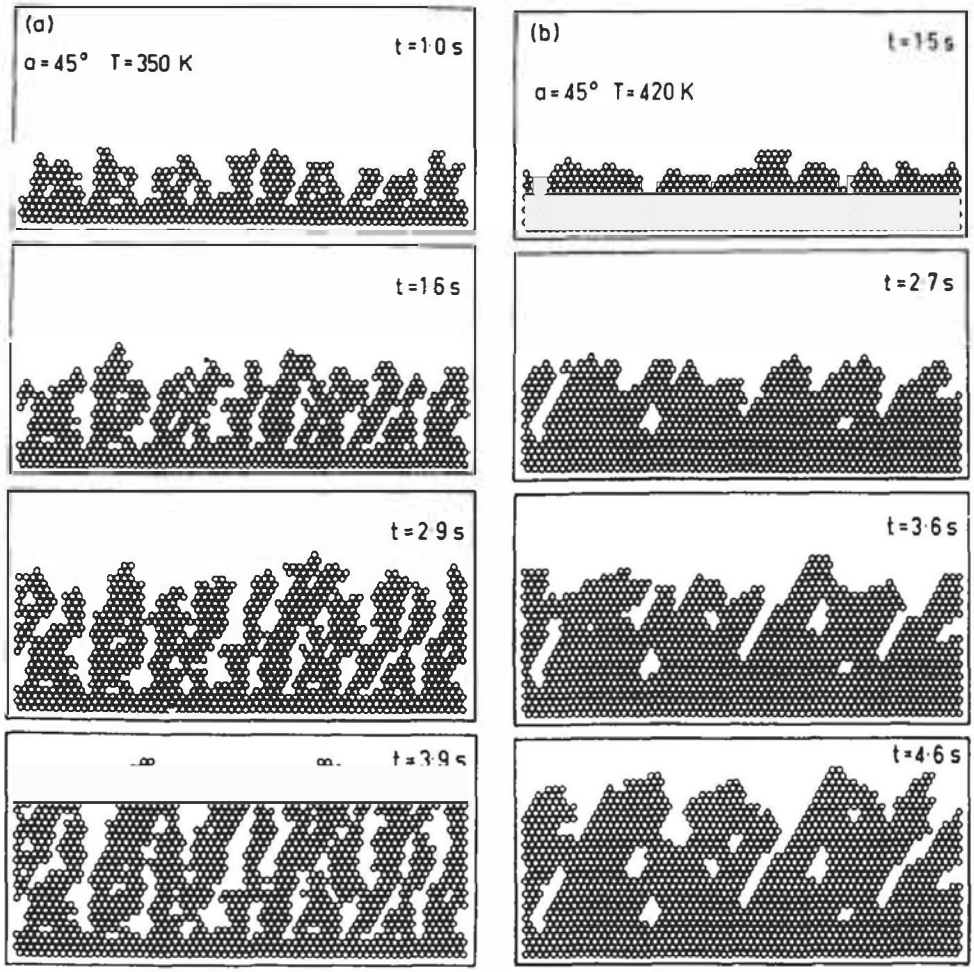


Figure 5-17. Computer-simulated microstructure of Ni film during deposition at different times for substrate temperatures of (a) 350 K and (b) 420 K. The angle of vapor deposition α is 45°. (From Ref. 22).

index of refraction and absorption coefficient. Moisture-induced degradation has plagued optical film development for many years. A promising remedy for this problem is ion bombardment, which serves to compact the film structure. This approach is discussed further in Chapters 3 and 11.

5.6.4. 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 film mass per unit area and thickness. Among the experimental findings related to film density are the following (Ref. 23):

1. The density of both metal and dielectric films increases with thickness and reaches a plateau value that asymptotically approaches that of the bulk density. The plateau occurs at different thicknesses, depending on material deposition method and conditions. In Al, for example, a density of 2.1 g/cm^3 at 250 Å rises to 2.58 g/cm^3 above 525 °C and then remains fairly constant thereafter. As a reference, bulk Al has a density of 2.70 g/cm^3. The gradient in film density is thought to be due to several causes, such as higher crystalline disorder, formation of oxides, greater trapping of vacancies and holes, pores produced by gas incorporation, and special growth modes that 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, defined as

$$P = \frac{\text{volume of solid}}{\text{total volume of film (solid + voids)}} . \tag{5-44}$$

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, we can increase P 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 TEM phase (defocus) contrast techniques (Ref. 24). 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. 5-18. 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

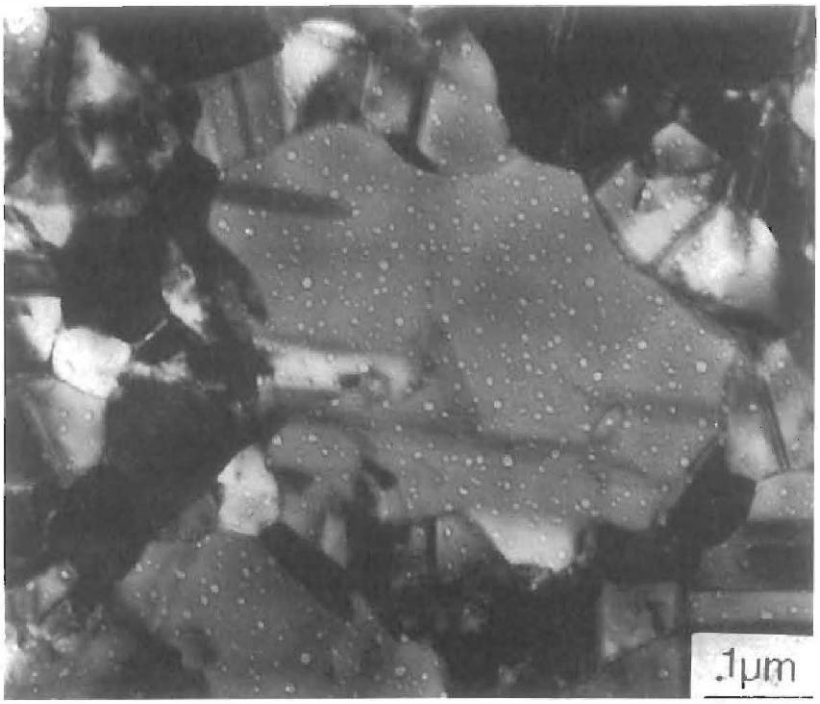


Figure 5-18. Transmission electron micrograph showing microvoid distribution in evaporated Au films. (Courtesy of S. Nakahara, AT&T Bell Laboratories.)

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 in fostering interdiffusion in thin-film couples where the Kirkendall effect has been observed (see Chapter 8). The natural tendency to decrease the vacancy concentration through annihilation is manifested by such film changes as stress relaxation, surface faceting, adhesion failure, recrystallization and grain growth, formation of dislocation loops and stacking faults, and decrease in hardness.

5.7. Amorphous Thin Films

5.7.1. Systems, Structures, and Transformations

Amorphous or glassy materials have a structure that exhibits only short-range order or regions where a predictable placement of atoms occurs. However,

within a very few atom spacings, this order breaks down, and no long-range correlation in the geometric positioning of atoms is preserved. Although bulk amorphous materials such as silica glasses, slags, and polymers are well known, amorphous metals were originally not thought to exist. An interesting aspect of thin-film deposition techniques is that they facilitate the formation of amorphous metal and semiconductor structures relative to bulk preparation methods.

As noted, production of amorphous films requires very high deposition rates and low substrate temperatures. The latter immobilizes or freezes adatoms on the substrate where they impinge and prevents them from diffusing and seeking out equilibrium lattice sites. By the mid-1950s Buckel (Ref. 25) produced amorphous films of pure metals such as Ga and Bi by thermal evaporation onto substrates maintained at liquid helium temperatures. Alloy metal films proved easier to deposit in amorphous form because each component effectively inhibits the atomic 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. Although they are virtually impossible to measure, vapor quench rates in excess of 10^{10} °C/sec have been estimated. From laboratory curiosities, amorphous Si, Se, GdCo, and GeSe thin films have been exploited for such applications as solar cells, xerography, magnetic bubble memories, and high-resolution optical lithography, respectively.

Important fruits of the early thin-film work were realized in the later research and development activities surrounding the synthesis of *bulk* amorphous metals by quenching melts. Today continuously cast ribbon and strip 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/sec are required to prevent appreciable rates of nucleation and growth of crystals. Heat transfer limitations restrict the thickness of these metal glasses to less than 0.1 mm. In addition to achieving the required quench rates, the alloy compositions are critical. Most of the presently known glass-forming binary alloys fall into one of four categories (Ref. 26):

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" (low-temperature) eutectics are found on the phase diagram.

Amorphous thin films of some of these alloys as well as other metal alloys

and virtually all elemental and compound semiconductors, semimetals, oxides, and chalcogenide (i.e., S-, Se-, Te-containing) glasses have been prepared by a variety of techniques. Amorphous Si films, for example, have been deposited by evaporation, sputtering, and chemical vapor deposition techniques. In addition, large doses of ion-implanted Ar or Si ions will amorphize surface layers of crystalline Si. Even during ion implantation of conventional dopants, local amorphous regions are created where the Si matrix is sufficiently damaged, much to the detriment of device behavior. Lastly, pulsed laser surface melting followed by rapid freezing has produced amorphous films in Si as well as other materials (see Chapter 13).

5.7.2. Au – Co and Ni – Zr Amorphous Films

It is instructive to consider amorphous Co–30Au films since they have been well characterized structurally and through resistivity measurements (Ref. 27). The films were prepared by evaporation 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. 5-19. The as-deposited film is rather featureless with a smooth topography, and the broad halos in the diffraction pattern cannot be easily and uniquely assigned to the known lattice spacings of the crystalline alloy phases in this system. Both pieces of evidence point to the existence of an amorphous phase whose structural order does not extend beyond the next-nearest-neighbor distance. 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–38Au (an alloy similar to Co–30Au) revealed a two-step transformation as shown in Fig. 5-20. 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. 5-19. The high resistivity of the amorphous films is due to the 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 to less thermody-

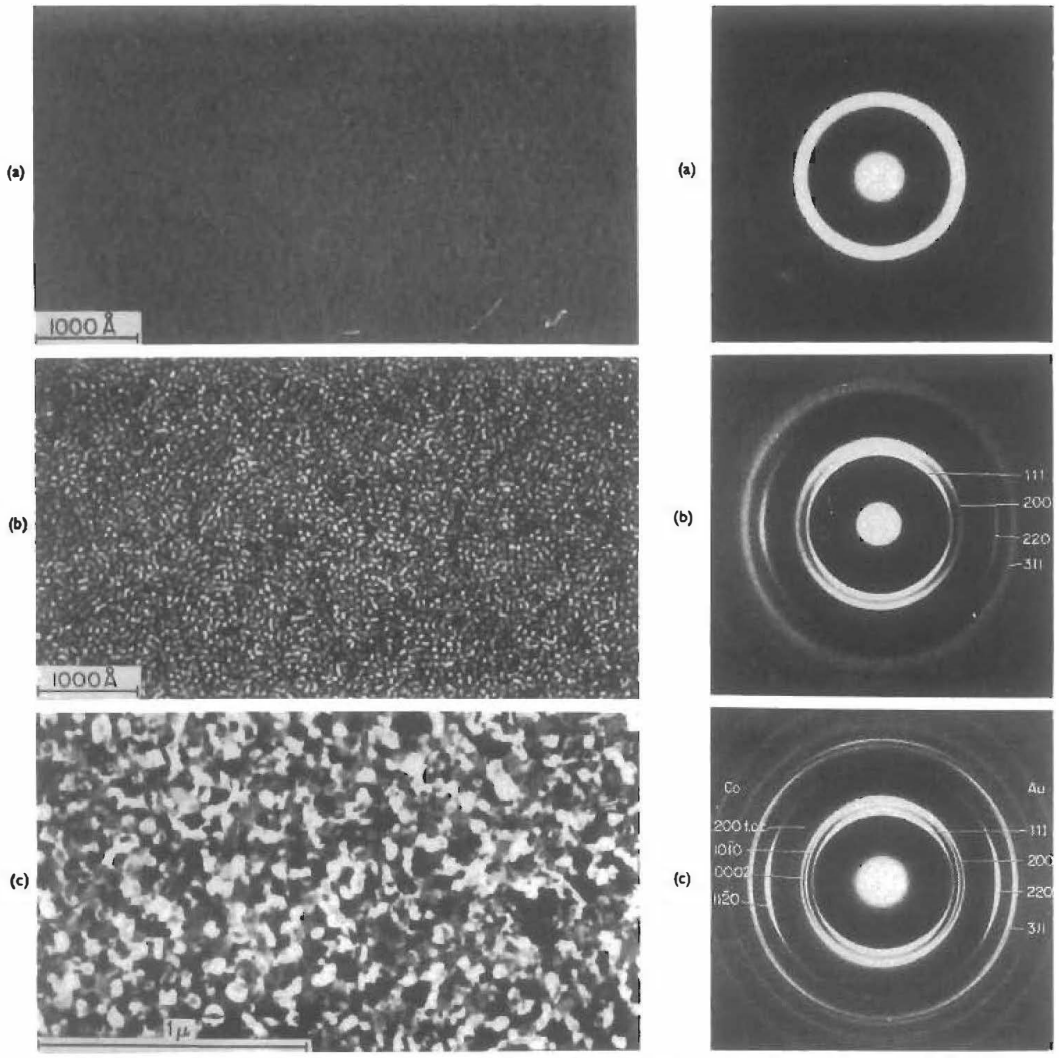


Figure 5-19. Electron micrographs and diffraction patterns of Co–30at%Au: (top) as deposited at 80 K, warmed to 300 K (amorphous); (middle) film warmed to 470 K (single-phase FCC structure); (bottom) film heated to 650 K (two-phase equilibrium). (From Ref. 27).

namically 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 exhibit other manifestations of thermodynamic instability. One is increased atomic solubility in amorphous

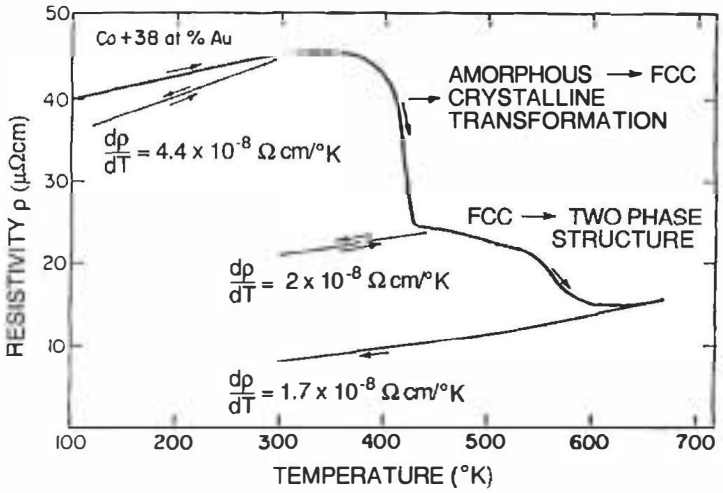


Figure 5-20. Resistivity of a Co–38at%Au film as a function of annealing temperature. Reversible values of $d\rho/dT$ in various structural states of the film are shown together with changes in ρ during phase transformation. (From Ref. 27).

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 that 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.

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. Consider Fig. 5-21, which shows the result of annealing a bilayer couple consisting of pure polycrystalline Ni and Zr films at 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 NiZr without triggering crystallization.

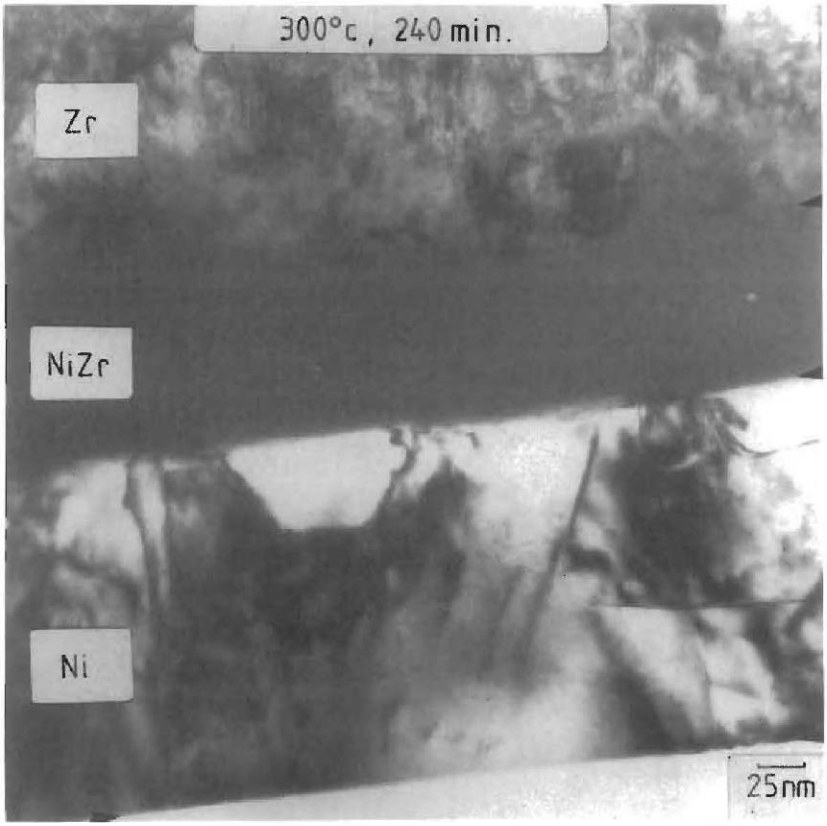


Figure 5-21. Cross-sectional electron micrograph of an amorphous Ni–Zr alloy film formed by annealing a crystalline bilayer film of Ni and Zr at 300 °C for 4 hours. (Courtesy of K. N. Tu, IBM Corp., T. J. Watson Research Lab., from Ref. 28).

5.7.3. A Model To Simulate Structural Effects in Thin Films

One of the outcomes of their research on quenched alloy films was an engaging mechanical model Mader and Nowick (Ref. 29) 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 pedogogic tool and worth presenting here. The "atoms" composing 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 collision 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, and the latter contained a perfect two-dimensional periodic array of interstices 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, they prepared reduced negatives (with an array size of about 4 mm square). 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 a 135-mm lens of a 35-mm camera. The resulting photographs are reproduced here.

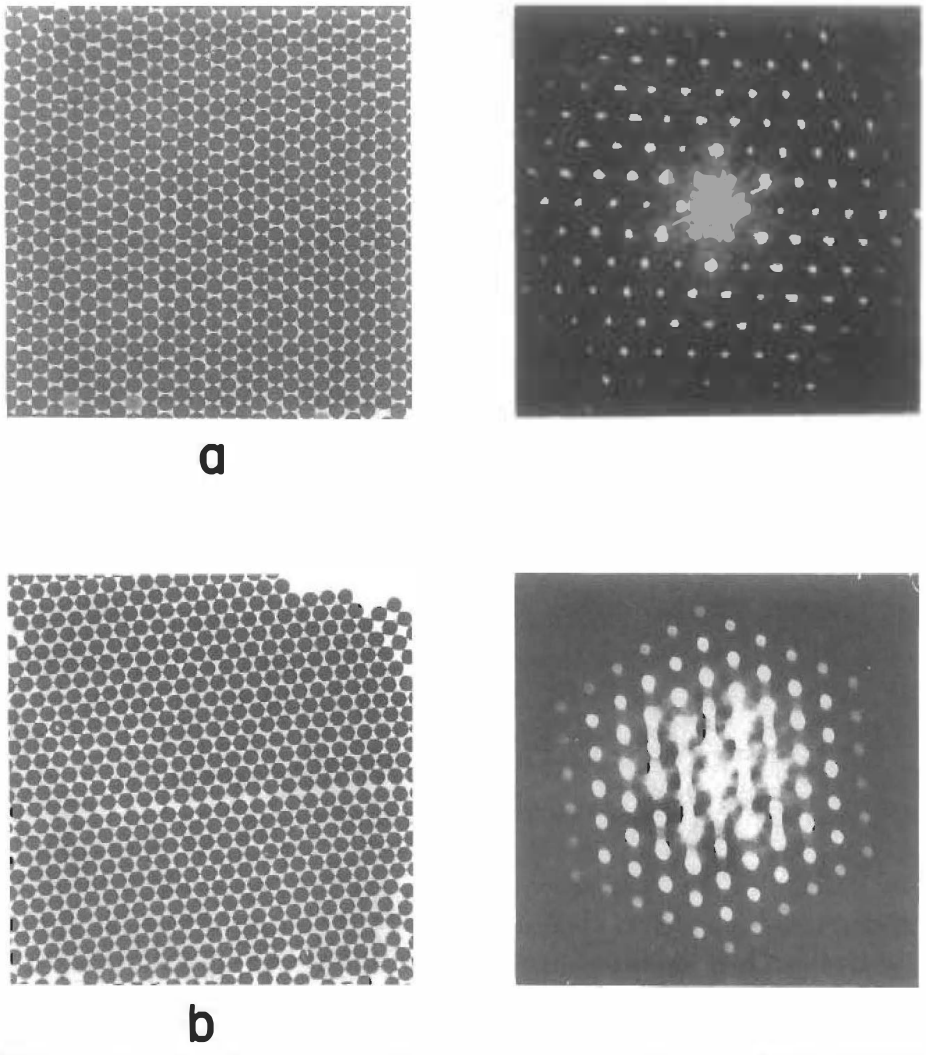


Figure 5-22. Atomic sphere film structures and corresponding Fraunhofer diffraction patterns for (a) perfect array, (b) stacking fault, (c) pure film; low deposition rate, (d) pure film; high deposition rate. (Reprinted with permission from the IBM Corp., from A. S. Nowick and S. R. Mader, *IBM J. Res. Dev.* **9**, 358, 1965).

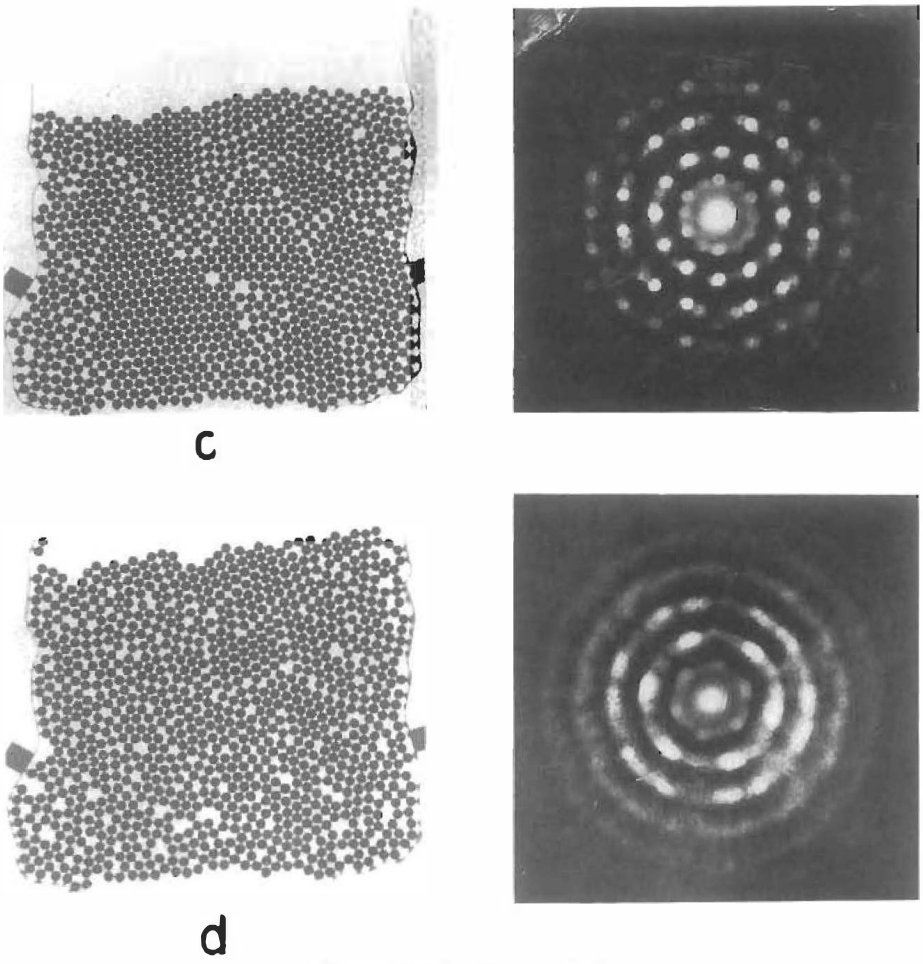


Figure 5-22. *Continued.*

The perfect array of spheres of one size is shown together with the corresponding diffraction pattern in Fig. 5-22a. A hexagonal pattern of sharp spots, very reminiscent of electron diffraction patterns of single-crystal films, is obtained, reflecting the symmetry of the close-packed array. After creation of a stacking fault in the structure, the diffraction pattern shows streaks (Fig. 5-22b). These run perpendicular to the direction of the fault in the structure.

The effect of deposition rate is shown in Figs. 5-22c and 5-22d. When the film is deposited "slowly," there are grains, vacancies, and stacking faults present in the array. Relative to Fig. 5-22a, the diffraction spots are broadened, a precurser to ring formation. In Fig. 5-22d, 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, which 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.

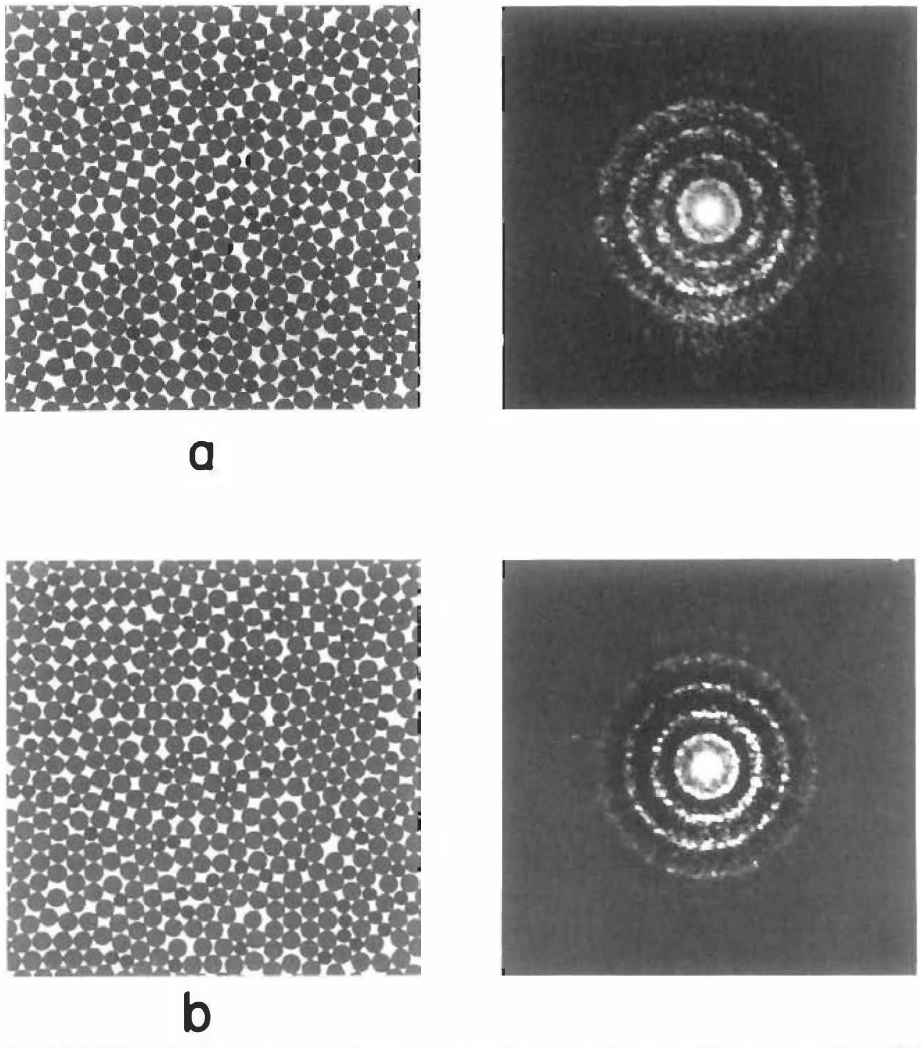


Figure 5-23. Atomic sphere film structure for concentrated alloy (50A–50B, 27% size difference: (a) as-deposited (amorphous); (b) vibration annealed. (Reprinted with permission from the IBM Corp., from A. S. Nowick and S. R. Mader, *IBM J. Res. Dev.* **9**, 358, 1965).

We now turn our attention to alloy films. For "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. 5-23.. The diffraction pattern contains broad halos. Upon vibration annealing, the film densifies slightly, but the atomic logjam cannot be broken 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 phenomena. 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 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.

The foregoing represents a sampling of the simulations of the dependence of film structure on deposition variables. Readers interested in this as well as other mechanical models of planar arrays of atoms, such as the celebrated Bragg bubble raft model (Ref. 30), should consult the literature on the subject. Much insight can be gained from them.

EXERCISES

1. Under the same gas-phase supersaturation, cube-shaped nuclei are observed to form homogeneously in the gas and heterogeneously both on a flat substrate surface and at right-angle steps on this surface. For each of these three sites calculate the critical nucleus size and energy barrier for nucleation.

2. 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 $\Delta G_V B/\pi\gamma^2 > 1/2$, ΔG monotonically decreases with r, but when $\Delta G_V B/\pi\gamma^2 < 1/2$ there is a turnaround in the curve. (The latter case corresponds to a metastable state and associated energy barrier.)

3. Cap-shaped nuclei on substrates grow both by direct impingement of atoms from the vapor phase as well as 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.

4. Two spherical nuclei with surface energy γ having radii r_1 and r_2 coalesce in the gas phase to form one spherical nucleus. If mass is conserved, calculate the energy reduction in the process. Suppose two spherical *caps* of different radii coalesce on a planar substrate to form one cap-shaped nucleus. Calculate the energy reduction.

5. 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 the diffusivity of atoms on the surface is D_S. Make simplifying assumptions as you see fit.

6. Assume that the two nuclei in Fig. 5-10 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 a value for the approximate diffusivity. Assume $\gamma = 1000$ ergs/cm^2, $T = 400$ °C, and $\Omega = 17 \times 10^{-24}$ cm^3/atom.

7. A film is deposited on a substrate by means of evaporation. In the expression for the rate of heterogeneous nucleation (Eq. 5-17), 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.

8. From data shown in Fig. 5-5 calculate values for E_{des}, E_S, and E_{i*}. (For answers consult Ref. 3, page 8-23.)

9. Three different methods for estimating the temperature for epitaxial growth of films have been discussed in this chapter.

a. Comment on the similarities and differences in the respective approaches.
b. How well do they predict the experimental findings of Fig. 5-4?

10. Derive expressions for the epitaxial transition temperatures $T_{1\to 3}$ and $T_{2\to 3}$.

11. 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.

Explain the physical reasons for these observations.

12. The formation of three-dimensional crystallites from an amorphous matrix undergoing transformation by *nucleation* and *growth* processes follows the time (t) dependent kinetics given by

$$f(t) = 1 - \exp - \frac{\pi \dot{N} v^3 t^4}{3}$$

$\dot{N}$ is the nucleation rate of crystallites (per unit volume), v is their growth velocity, and f is the fractional extent of transformation.

a. $\dot{N}$ is small near the critical transformation temperature and at low temperature, but larger in between. Why?
b. v is usually larger for higher temperatures. Why?
c. Schematically sketch $f(t)$ vs. t (or ln t) at a series of temperatures. Note that an incubation time dependent on temperature is suggested.

13. a. 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 show that the grain size will tend to grow with parabolic kinetics.

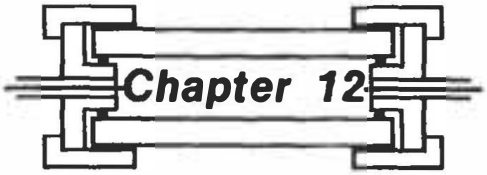
b. Part (a) is valid when the film grain size is smaller than the film thickness. Why? If the reverse is true, suggest why parabolic growth kinetics may not be observed.

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. A. Neugebauer, in *Handbook of Thin-Film Technology*, eds. L. I. Maissel and R. Glang, McGraw Hill, New York (1970).
4.* K. Reichelt, *Vacuum* **38**, 1083 (1988).
5.* J. A. Venables, G. D. T. Spiller, and M. Hanbücken, *Rep. Prog. Phys.* **47**, 399 (1984).
6. D. Walton, T. N. Rhodin, and R. W. Rollins, *J. Chem. Phys.* **38**, 2698 (1963).
7. H. M. Yang and C. P. Flynn, *Phys. Rev. Lett.* **62**, 2476 (1989).
8. V. N. E. Robinson and J. L. Robins, *Thin Solid Films* **20**, 155 (1974).
9. R. M. German, *Powder Metallurgy Science*, Metal Powder Industries Federation, Princeton, NJ (1984).
10. D. W. Pashley and M. J. Stowell, *J. Vac. Sci. Tech.* **3**, 156 (1966).
11. D. Kashchiev, *Surface Science* **86**, 14 (1979).
12. K. L. Chopra, *Thin-Film Phenomena*, McGraw-Hill, New York (1969).
13. G. E. Rhead, *J. Vac. Sci. Tech.* **13**, 603 (1976).
14. R. W. Vook and B. Oral, *Gold Bull.* **20**, (1/2), 13 (1987).
15. E. Grunbaum, in *Epitaxial Growth B*, ed. J. W. Matthews, *Academic Press*, New York (1976).
16. B. A. Movchan and A. V. Demchishin, *Phys. Met. Metallogr.* **28**, 83 (1969).
17. J. A. Thornton, *Ann. Rev. Mater. Sci.* **7**, 239 (1977).
18. H. T. G. Hentzell, C. R. M. Grovenor, and D. A. Smith, *J. Vac. Sci. Tech.* **A2**, 218 (1984).
19. 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. Publ. Switzerland (1988).

*Recommended texts or reviews.

20. J. M. Nieuwenhuizen and H. B. Haanstra, *Philips Tech. Rev.* **27**, 87 (1966).
21. R. Messier, A. P. Giri, and R. Roy, *J. Vac. Sci. Tech.* **A2**, 500 (1984).
22. K. H. Müller, *J. Appl. Phys.* **58**, 2573 (1985).
23. H. Pulker, *Coatings on Glass*, Elsevier, Amsterdam (1984).
24. S. Nakahara, *Thin Sold Films* **64**, 149 (1979).
25. W. Buckel, *Z. Phys.* **138**, 136 (1954).
26. H. S. Chen, H. J. Leamy, and C. E. Miller, *Ann. Rev. Mater. Sci.* **10**, 363 (1980).
27. S. Mader, in *The Use of Thin Films in Physical Investigations*, ed. J. C. Anderson, Academic Press, New York (1966).
28. S. B. Newcomb and K. N. Tu, *Appl. Phys. Lett.* **48**, 1436 (1986).
29. A. S. Nowick and S. R. Mader, *IBM J. Res. Dev.* **9**, 358 (1965).
30. W. L. Bragg and J. F. Nye, *Proc. Roy. Soc.* **A190**, 474 (1947).



Metallurgical and Protective Coatings

12.1. INTRODUCTION

Paralleling the dramatic development of thin-film technology in microelectronics have been the no less than remarkable advances in what may be conveniently called metallurgical and protective coatings. The unusual materials which comprise these coatings are drawn from several classes of solids and include ionic ceramic oxides (e.g., Al_2O_3, ZrO_2, TiO_2), covalent materials (e.g., SiC, BC, diamond), transition metal compounds (e.g., TiC, TiN, WC) and metal alloys (e.g., CoCrAlY, NiAl, NiCrBSi). As a whole they are characterized by extremely high hardness, very high melting points, and resistance to chemical attack, attributes that have earmarked their use in critical applications where one or more of these properties is required; correspondingly the respective categories of hard, thermal, and protective coatings denote the functions to which they are put. Hard coatings of TiN and TiC, for example, are used to extend the life of cutting tools, dies, punches, and in applications such as ball bearings to minimize wear. The collection of coated cutting tools and dies shown in Fig. 12-1 is representative of the widespread commercial use of this technology in machining and forming operations. Thermal coatings find extensive use in gas turbine engines where they help to

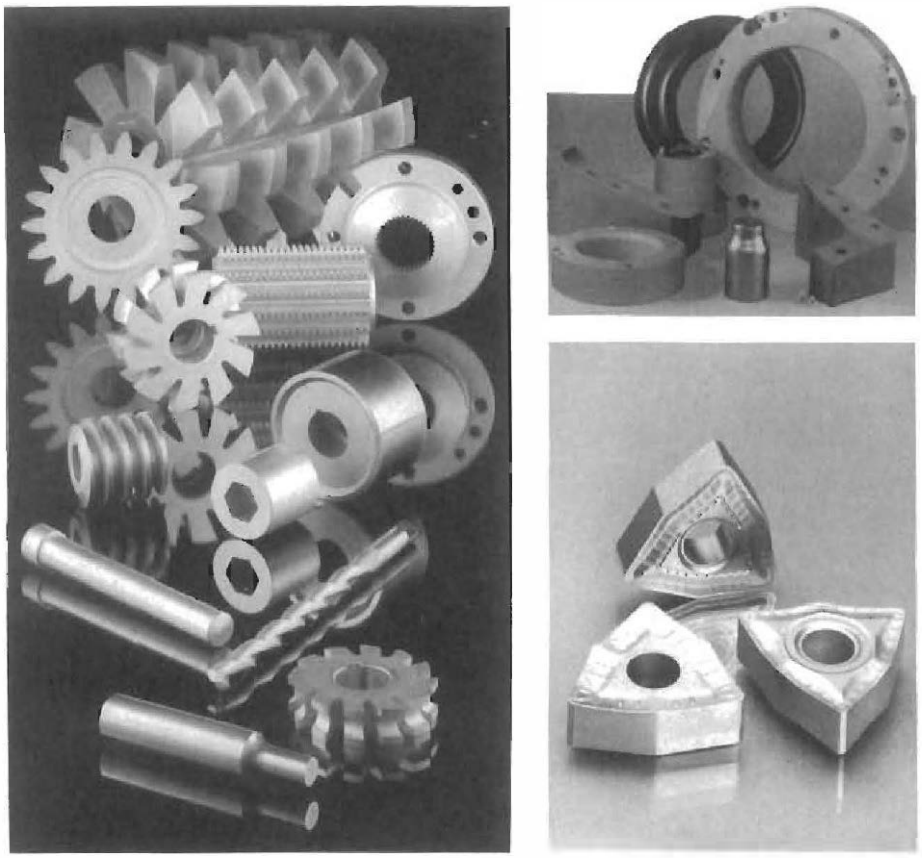


Figure 12-1. (Left) Assorted cutting and forming tools coated with TiN and multilayer coatings. (Courtesy Multi-Arc Scientific Coatings). (Top right) HSS forming and sheet metal dies coated with TiN and TiC. (Courtesy Ti Coating Inc.) (Lower right) multilayer coated cutting tool inserts. (Courtesy of S. Wertheimer, ISCAR Ltd.)

improve the performance and extend the life of compressor and turbine components. As the name implies, protective coatings are intended to defend the underlying materials, usually metals, from harsh gaseous or aqueous environments that cause corrosive attack. Such coatings have found applications in chemical and petroleum industries, coal gasification plants, as well as in nuclear reactors.

Employing coatings represents a significant departure from traditional engineering design and manufacturing practices. Processing components beyond the primary manufacturing steps of casting, forging, extrusion, machining and grinding, pressing and sintering, etc., has generally been resisted. This is due in part to a reluctance to tamper with the product, and to leave well enough alone. A compelling case was not made for the cost effectiveness of additional

treatments. However, more recently several important factors have combined to firmly establish the practice of modifying the surface properties of engineering materials and components.

1. In many critical applications the design specifications call for properties that are simply beyond the capabilities of the commonly available and routinely processed materials. The new limits of behavior demanded can be met by the use of the unusually hard, temperature- and degradation resistant materials noted earlier. However, these materials are extremely difficult to fabricate in bulk form.
2. Concerns of limited availability of strategic materials, the thrust toward energy efficiency and independence, and an increasingly competitive world economy have exerted a strong impetus to considerably tighten engineering design, improve performance, and economize on materials utilization.
3. High-quality coatings possessing fewer surface imperfections than comparable pressed and sintered bulk parts made from powder, can now be reproducibly deposited. This is due to the advances made in our basic understanding of the deposition processes and the development of improved coating and deposition techniques.
4. The commercial availability of the necessary deposition chambers or reactors, hardware, computer-controlled processing equipment, and high-purity sources of precursor gases, powders and sputtering targets has facilitated the option of employing coatings.

Various combinations of the above factors have then resulted in the marriage of coatings to the underlying base materials, each with their particular set of desirable and complementary properties. For example, many structural materials with adequate high-temperature mechanical properties simply do not have the ability to withstand high-temperature oxidation, corrosion, particle erosion, and wear. On the other hand, the materials that do possess the environmental resistance either do not qualify as structural materials because of low toughness or, if they do, are prohibitively expensive to fashion in bulk form.

Before we turn to the main subjects of the chapter, it is worth noting some of the similarities and differences between the present mechanically and environmentally functional *coatings*, and the *thin films* of prior book chapters. In common, many coatings are deposited by the same type of physical (PVD) and chemical (CVD) vapor deposition techniques. Adhesion to the substrate, development of desirable structure and properties, and meeting performance standards are universal concerns. Among the differences are the following:

1. The coatings we will be considering are far thicker than thin films. Whereas a couple of microns, at most, is the arbitrary upper limit to what we have

called films, coatings typically range from several to tens and even hundreds of microns in thickness.

2. The maintenance of precise coating thickness and uniformity is not usually a major concern. There is generally a broad range of acceptable coating thicknesses. This is in contrast to the critical thickness tolerances and uniform coverage that must be achieved in optical and microelectronic films.
3. The substrate is frequently an integral part of the coating system. In diffusion coatings, for example, metalloid as well as metal elements are diffused into the substrate, creating thick, solute-enriched layers beneath the surface.
4. For the most part, the substrates employed for hard and protective coatings are rather special metals and alloys such as tool, high-speed, and stainless steels; iron-, cobalt-, and nickel-base superalloys; sintered tungsten carbide; titanium, etc. The use of the term *metallurgical coating* is based in part on this fact.
5. There are many methods for producing coatings. In addition to the vapor phase atomistic deposition processes (PVD, CVD) for films, coatings are also formed by

   a. Deposition of particulates (e.g., by thermal spraying of metal or oxide powders either through a hot flame or an even hotter plasma)
   b. Immersion of substrates in molten baths or heated solid packs
   c. Electrolytic processes such as electroplating, fused salt electrolysis, and electroless plating
   d. Miscellaneous processes, e.g., welding and enameling

6. Except for epitaxial semiconductor films, most thin-film depositions are carried out at relatively low-substrate temperatures. Metallurgical and protective coatings, however, are frequently deposited at elevated temperatures. Certainly this is true of the CVD coatings, and, therefore, atomic interdiffusion and reactions generally occur at the interface between coating and substrate. Compositional change can either be beneficial or detrimental to adhesion and coating properties depending on the materials involved.

The bulk of the chapter will be concerned with hard coatings and issues related to them. Somewhat lesser emphasis is placed on thermal and environmental coatings. To limit the treatment to manageable proportions, we deal with properties and the phenomena they influence in a fundamental way. The more widely used vapor deposition processes will be primarily discussed to

maintain a consistency with prior chapters. Electrodeposition, for example, will not be mentioned again, since there is already a huge and accessible literature on the subject. Case histories and examples are always interesting and will be interspersed where appropriate. The specific topical outline of the rest of the chapter is

12.2. Hard Coating Materials
12.3. Hardness and Fracture
12.4. Tribology of Films and Coatings
12.5. Diffusional, Protective, and Thermal Coatings

12.2. Hard Coating Materials

12.2.1. Compounds and Properties

Hard coating materials can be divided into three categories, depending on the nature of the bonding. The first includes the *ionic* hard oxides of Al, Zr, Ti, etc. Next are the *covalent* hard materials exemplified by the borides, carbides, and nitrides of Al, Si, and B, as well as diamond. (See Section 14.2.) Finally, there are the *metallic* hard compounds consisting of the transition metal borides, carbides, and nitrides. Typical mechanical and thermal property values for important representatives of these three groups of hard materials are listed in Table 12-1. The reader should be aware that these data were gathered from many sources (Refs. 1–6) and that there is wide scatter in virtually all reported property values. Differences in processing (e.g., CVD, PVD, and sintering of powders), variations in structure (e.g., grain size, porosity, density, defects) and composition (e.g., metal–nonmetal ratio, purity), together with statistical error in measurement, contribute to the uncertainties. Perusal of this tabulated information leads to the following broad conclusions:

1. All of these compounds have extremely high hardnesses. This can be appreciated by noting that heat-treated tool steel has a hardness of about $H_v = 850$. Hardness is the most often quoted material property of hard coatings. Therefore, Section 12.3 has been specially reserved for an extensive discussion of the concept of hardness, the technique of measurement, and the significance of its magnitude in coatings.
2. These compounds have very high melting points and decomposition temperatures. For example, the decomposition temperatures of TaC, HfC, and diamond exceed the melting point for tungsten (MP = 3410 °C).
3. The modulus of elasticity is lowest for the ionic solids. In comparison, only

Table 12-1. Mechanical and Thermal Properties of Coating Materials

Material	Melting or Decomposition Temperature (°C)	Hardness (kg-mm^{-2})	$H = H_0 e^{-aT}$ (Eq. 12-4) H_0 (kg-mm^{-2})	a (10^{-4} C^{-1})	Density (g-cm^{-3})	Young's Modulus (kN-mm^{-2})	Thermal Expansion Coefficient (10^{-6} K^{-1})	Thermal Conductivity (Wm^{-1} K^{-1})	Fracture Toughness (MPa-m$^{1/2}$)
Ionic									
Al_2O_3	2047	2100	2300	7.85	3.98	400	6.5	~ 25	3.5
TiO_2	1867	1100	1250	5.99	4.25	200	9.0	9	
ZrO_2	2710	1200			5.76	200	8.0	1.5	4–12
SiO_2	1700	1100			2.27	151	0.55	2	< 1
Covalent									
C (Diamond)	3800	~ 8000			3.52	1050	1	1100	
B_4N	2450	~ 4000			2.52	660	5		
BN	2730	~ 5000			3.48	440			
SiC	2760	2600	2800	0.90	3.22	480	5.3	84	3
Si_3N_4	1900	1700	1900	2.79	3.19	310	2.5	17	4
AlN	2250	1200			3.26	350	5.7		
Metal Compounds									
TiB_2	3225	3000	3500	18.9	4.5	560	7.8	30	
TiC	3067	2800	3300	18.3	4.9	460	8.3	34	0.46
TiN	2950	2100	2100	23.5	5.4	590	9.3	30	
HfN			2000	8.57			6.9	13	
HfC	3928	2700	3000	14.7	12.3	460	6.6		
TaC	3985	1600	1800	6.75	14.5	560	7.1	23	
WC	2776	2300	2350	3.62	15.7	720	4.0	35	
Substrate Materials									
High-Speed Steel	1400	900			7.8	250	14	30	50–170
WC-6%Co		1500				640	5.4	80	11.4
Ti	1667	250			4.5	120	11	13	80
Ni Superalloys	1280				7.9	214	12	62	> 100

the stiffest metals have modulus values overlapping those of the oxides listed.

4. The linear thermal expansion coefficient generally increases in going from the covalent to metallic to ionic hard compounds. Metals tend to have thermal expansion coefficients that are higher by approximately a factor of two or more than these hard compounds.
5. The thermal conductivity of the hard metallic and covalent compounds is comparable to that of the transition metals and their alloys. Good metallic electrical conductors have proportionally higher thermal conductivities. Ceramic oxides are the poorest thermal conductors.

The last two properties have important implications for the properties and performance of coatings. An important source of coating residual stress is the thermal contribution generated by the difference in expansion between coating and substrate. The illustrative problem dealing with TiC on steel (p. 419) is worth reviewing and indicates the magnitude of possible effects.

The susceptability to cracking of coatings subjected to varying temperature histories is an important limitation to the performance of thermal coatings. To see how thermal cracking can occur, consider the rapid cooling of a high-temperature component. The surface coating contracts more than the interior, which is still relatively hot. As a result, the surface forces the interior into compression and is itself stretched in tension. The coatings we have been discussing are weak in tension and prone to fracture at incipient crack flaws. A measure of this fracture susceptibility is given by the thermal shock parameter S_T. This composite quantity is derivable from heat transfer and thermal stress considerations that are defined respectively by the equations

$$Q = -\kappa(\Delta T/\Delta x) \tag{12-1}$$

and

$$\sigma = (E\Delta\alpha\,\Delta T)/(1-\nu). \tag{12-2}$$

In the first equation the heat flux Q is given by the product of the thermal conductivity κ and the temperature gradient. The second equation was considered previously (Eq. 9-17), where $\Delta\alpha$ is the difference in thermal expansion coefficient between coating and substrate. Elimination of ΔT yields

$$S_T^{-1} = \frac{\sigma}{Q\,\Delta x} = \frac{E\Delta\alpha}{(1-\nu)\kappa}. \tag{12-3}$$

Therefore, for a given thermal flux through a coating of thickness Δx, the stress that develops is inversely proportional to the thermal shock parameter. Ceramic oxides with the lowest S_T values are most prone to fracture by

thermal shock, and metals with the largest S_T values are least susceptible. The hard metal and covalent compounds are intermediate in this regard.

12.2.2. Deposition Processes (Ref. 7)

Hard coatings have now been commercially deposited on perhaps hundreds of millions of cemented tungsten carbide cutting tools by CVD methods since the mid-1960s. However, advances in physical vapor deposition methods in the past decade have led to the fruitful competition in many applications between these two broad deposition processes. It is instructive to briefly review and compare several aspects of these processes as they relate to hard coatings. The more extensive discussion of CVD and PVD in Chapters 3 and 4 provide the appropriate background.

Since hard coatings are deposited on varied substrates (e.g., carbides, high-speed tool steels, ball-bearing steels, watch cases, jewelry, etc.), a range of substrate temperatures is involved, 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

$$TiCl_4 + CH_3CN + (5/2)H_2 \rightarrow Ti(C, N) + CH_4 + 4HCl.$$

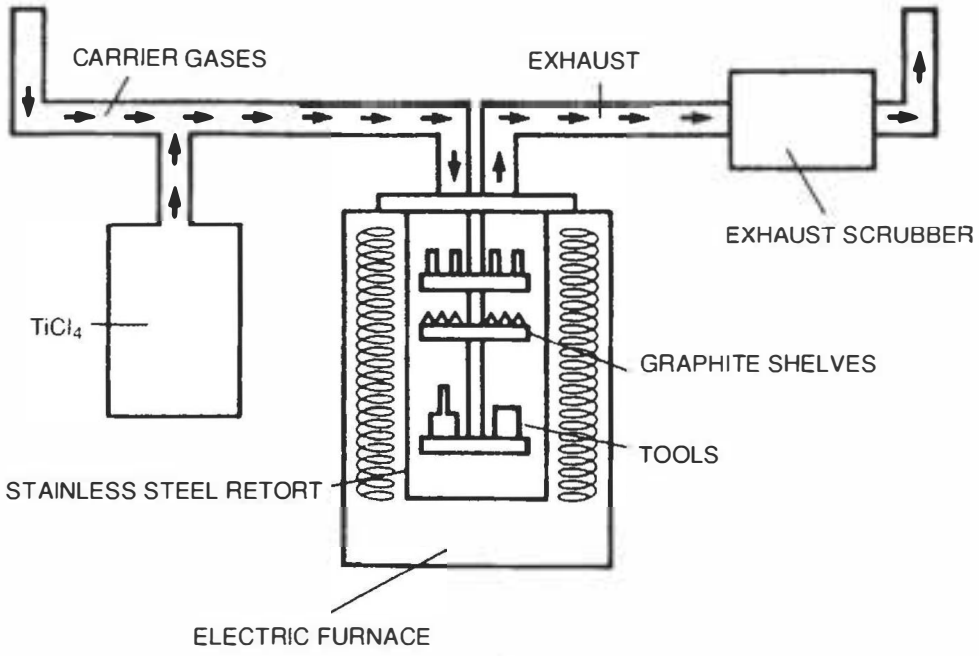
3. Low temperature: 700 °C > T > 300 °C

$$2TiCl_4 + 2N_2 + 7H_2 + \text{Ar plasma} \rightarrow 2TiN + 8HCl + 2NH_3.$$

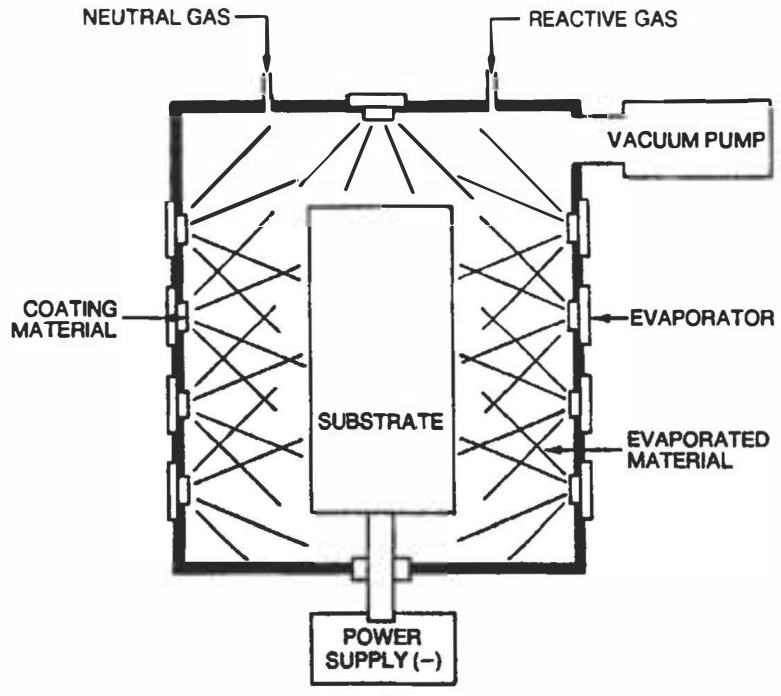
Analagous reactions involving gaseous carbon compounds, such as CH_4 result in TiC deposits. Precursor gases and operating temperatures for these as well as other coatings have been given previously in Chapter 4.

One of the important advantages of CVD methods is the ability to batch-coat large numbers of small tools at one time. The schematic of such a commercial system capable of individual or sequential TiC, TiN, and Al_2O_3 depositions is shown in Fig. 12-2a. Typical gas feed proportions for these materials are

TiC–$TiCl_4$ (1–2%);	CH_4 (3–6%);	H_2 (96–92%)
TiN–$TiCl_4$ (1–2%);	N_2 (22–26%);	H_2 (77–72%)
Al_2O_3–$AlCl_3$ (0.18–0.3%);	CO (0.2–0.5%);	H_2 (99%)



(a)



(b)

Figure 12-2. (a) Schematic view of commercial CVD reactor for deposition of TiC, TiN and Al_2O_3 on carbide cutting tools. (b) Cathodic-arc process for reactive evaporation of TiN coatings. (Courtesy of A. Gates, Multi-Arc Scientific Coatings Inc.)

Temperatures within the range 700–1100 °C are maintained to ±2 °C. As many as 10–20,000 WC tool inserts (1/2 in square) can be simultaneously given a ~ 6-μm-thick coating during a several-hour treatment cycle. The fracture surface morphologies of single-layer CVD TiC, TiN and TiCN as well as HfN coatings deposited on cemented carbide substrates are shown in Fig. 12-3. Rough transverse fracture surfaces containing columnar-like grains rising from a sharply defined substrate interface, typically characterize these coating structures. In the coating plane a mountainous relief of jagged crystals is evident.

The three basic PVD coating processes are evaporation, sputtering, and ion plating together with reactive variants incorporating chemical reactions (Refs.

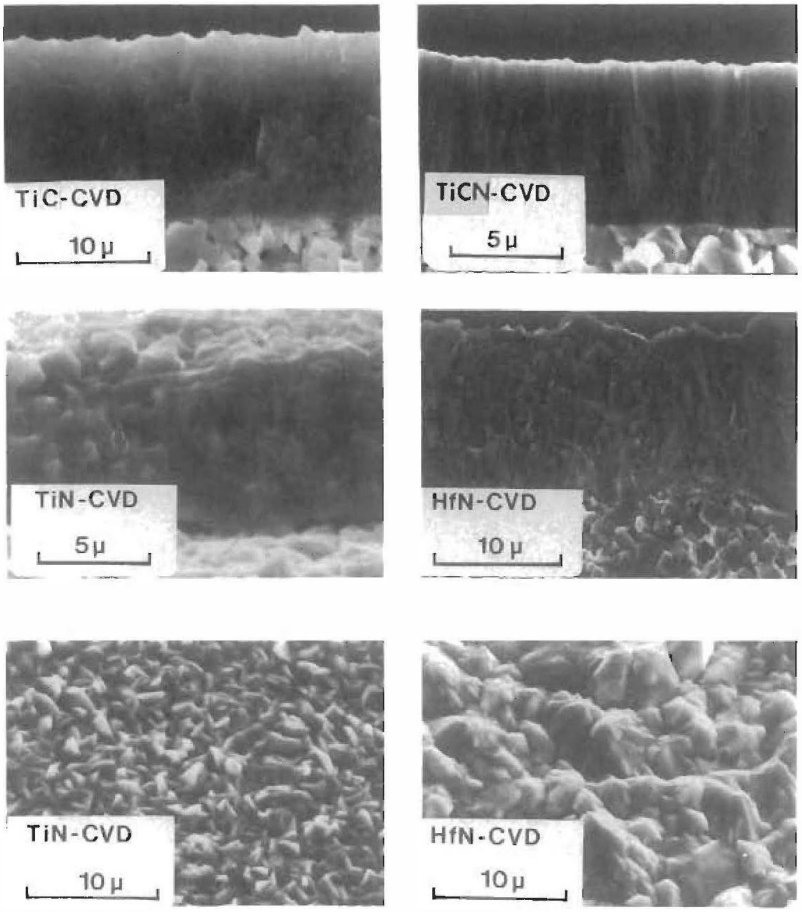


Figure 12-3. Morphology of representative CVD coatings viewed in the SEM. Top–Transverse fracture surfaces of TiC and TiCN. Middle–Transverse fracture surfaces of TiN and HfN. Bottom–Lateral surface structures of TiN and HfN. Substrate is cemented carbide. (Courtesy of D. T. Quinto, Kennametal, Inc., from Ref. 8).

8, 9). An important advantage of PVD is that substrate temperatures in the range of 200–600 °C, rather than 800–1200 °C for CVD, are utilized. This leads to somewhat finer coating grain sizes possessing slightly higher hardnesses. Because of the high deposition temperatures in CVD, substrate steels are sometimes softened, requiring subsequent hardening and tempering treatments. Undesirable part distortion occasionally results from this sequence of thermal treatments. In addition, substrate elements such as iron, chromium, and carbon diffuse into the coating altering adhesion as well as substrate integrity. These difficulties do not arise in the low-temperature PVD coatings. An added advantage of commercial PVD processes is the generally higher deposition rates of 4–6 μm/hour relative to typical CVD rates of 1–2 μm/hour. Nevertheless, throughput of small parts is substantially greater for CVD because substrate coverage is not limited by line-of-sight gas impingement as in PVD. On the other hand, the hazard of handling toxic and corrosive gases is considered a disadvantage of CVD processes. For larger tools such as drills and milling cutters, the PVD techniques of reactive evaporation and sputtering are widely employed. Commercially, uniform coating of tools and dies is accomplished through multiple evaporators arranged to surround the work as shown in Fig. 12-2b. While single (e.g., TiC), double (e.g., TiC–TiN)

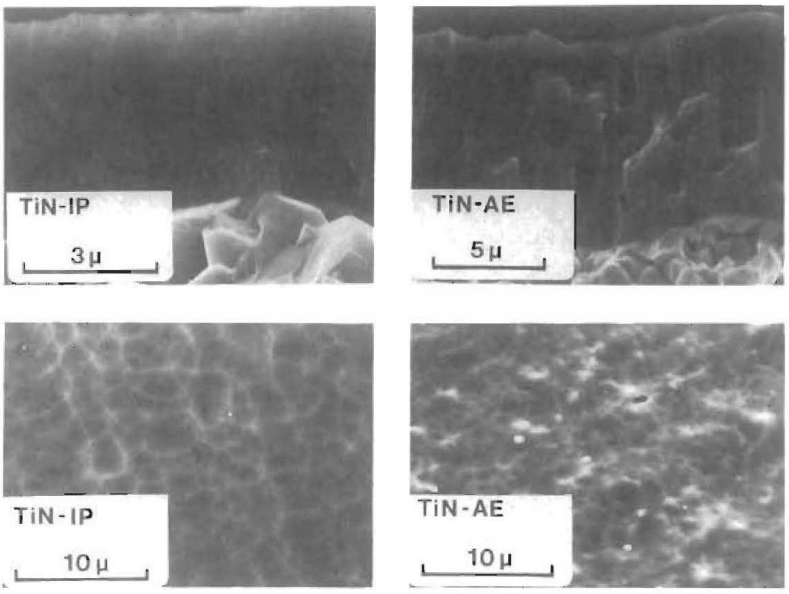


Figure 12-4. Morphology of PVD coatings of TiN prepared by ion plating (IP) and arc evaporation (AE) viewed in the SEM. Top–Transverse fracture surfaces. Bottom–Lateral surface structures. Substrate is cemented carbide. (Courtesy of D. T. Quinto, Kennametal, Inc., from Ref. 8).

and even triple (e.g., $TiC-Al_2O_3-TiN$) hard layers are commonly produced by CVD, commercial PVD coatings are largely limited to single TiN layers. Transverse fracture and lateral surface morphologies of TiN prepared by two different PVD processes are reproduced in Fig. 12-4. The smooth fracture surfaces composed of densely packed, fine columnar grains should be compared with the microstructures of CVD TiN (Fig. 12-3). Differences in grain morphology can be attributed to varying conditions of deposition temperature and energetic bombardment by vapor phase atoms.

12.2.3. Alloy Compounds (Refs. 3, 10)

Just as metallic alloys display properties superior to each of the component elements, so there has been considerable interest in improving the properties of the transition metal compounds through alloying. Fortunately, this is possible because the binary carbides and nitrides tend to be miscible in the solid state. For example, at elevated temperature, TiC reacts with VC, ZrC, NbC, HfC,

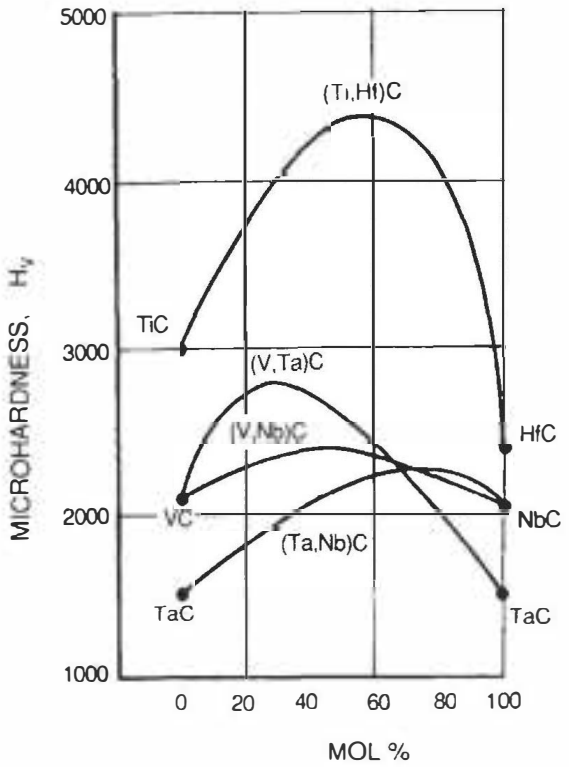


Figure 12-5. Microhardness of mixed carbides due to solid solution and precipitation hardening (From Ref. 3).

and TaC to produce mixed carbide solid solutions, such as (Ti, V)C, (Ti, Zr)C, (Ti, Nb)C. Since these carbides have the same crystal structures, TiC–MC (M = V, Zr, etc.) phase diagrams resemble the one for Ge–Si shown in Fig. 1-12, at least at high temperatures. Similarly, TiN can be alloyed with VN, ZrN, NbN, and HfN to produce solid solutions of mixed nitrides. The increase in hardness for the mixed carbides can be quite extensive, as shown in Fig. 12-5. Substitution on the metal sublattice results in maximum hardening at specific alloy concentrations. Similar increases in hardness have been observed in the mixed metal nitride solid solutions as well as in metal carbonitrides, e.g., Ti(C, N), Hf(C, N). Property benefits also extend to oxide alloys. For example, aluminum oxynitride has greater chemical and thermal stability than Al_2O_3 alone and is used to coat tools.

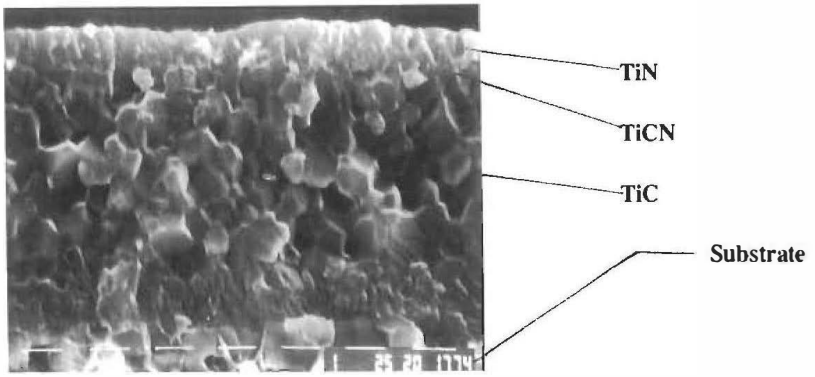
Quaternary alloys arising from mixtures of carbides and nitrides of different metals (e.g., Ti–V–C–N) have also been investigated and appear to provide a promising basis for the selection of yet other new composite coating materials. Subsequent heat treatment of the alloyed carbides, nitrides, and carbonitrides can lead to complex low-temperature phase transformations and precipitation reactions. Dispersions of hard phases such as TiB_2 or B_4C in TiC or TiN matrices have produced coatings that show better wear resistance and adhesion as well as higher toughness than single-phase layers.

12.2.4. Multilayer Coatings (Refs. 3, 11)

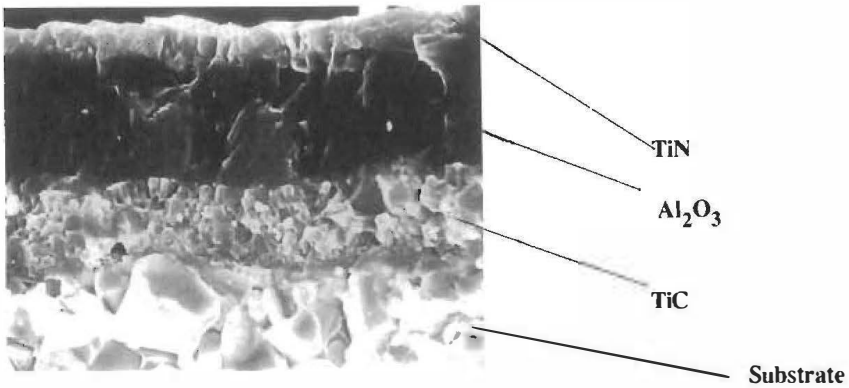
Distinct from *multicomponent single* layers is the *multilayer* concept. Several strategies have been adopted in creating composite multilayers:

1. Mutually soluble layers, e.g., TiC–TiN or Al_2O_3–AlN, are sequentially deposited. Alloys with enhanced properties are produced this way.
2. Compounds possessing coherent interfaces such as TiC–TiB_2 or TiN–TiB_2 are sequentially deposited. Adhesion between layers is promoted by low energy bonding between common elements.
3. Separate layers of largely noninteracting compounds are sequentially deposited. Each layer has a specific purpose, such as interfacial adhesion, diffusion barrier, abrasion resistance, etc.

The last approach is one that has attained considerable commercial success. Cross sections of such coatings on WC inserts for assorted cutting tools are shown in Fig. 12-6. They contain some combination of TiC, TiN, and Al_2O_3 layers deposited by CVD and are typically 6–9 μm thick. The multilayer of Fig. 12-6a is tough and suitable for rough, semifinishing, and interrupted



S.E.M. x 5500



S.E.M. x 3500

Figure 12-6. SEM images of CVD multilayer coatings for cutting tool inserts. (a) Carbide substrate/TiC/TiCN/TiN (5500 ×). (b) Carbide substrate–TiC–Al_2O_3–TiN (3500 ×). (Courtesy of S. Wertheimer, ISCAR Ltd.)

machining of plain carbon, alloy, tool, and stainless steels. The multilayer of Fig. 12-6b containing Al_2O_3 has many of the same characteristics but is capable of cutting at higher speeds without tool failure. Because of its low thermal conductivity, Al_2O_3 shields the cemented carbide substrate and enables the generated heat to be diverted to the chip instead. Among the reasons for the use of TiN as the outer coating layer are its high lubricity, chemical stability, and the lack of adherence of abrasive oxide particles to it. As indicated in Section 12.4.3 tools coated with multilayers suffer considerably less wear than tools with only single-layer TiC or TiN coatings.

12.3. Hardness and Fracture

12.3.1. Hardness

Hardness is an important material property of concern in films utilized for electronic and optical as well as mechanically functional applications. It affects wear resistance and plays an important role in the friction and lubrication of surface films in contact. Hardness, a complex property related to the strength of interatomic forces, apparently depends on more than one variable. Hard materials are generally modeled by a deep potential energy well (Fig. 1-8b) with steep walls. These characteristics imply a combination of high cohesive energy and short bond length. Measures of the cohesive energy are the heat of sublimation and enthalpy of compound formation (ΔH_{298}). If some account is

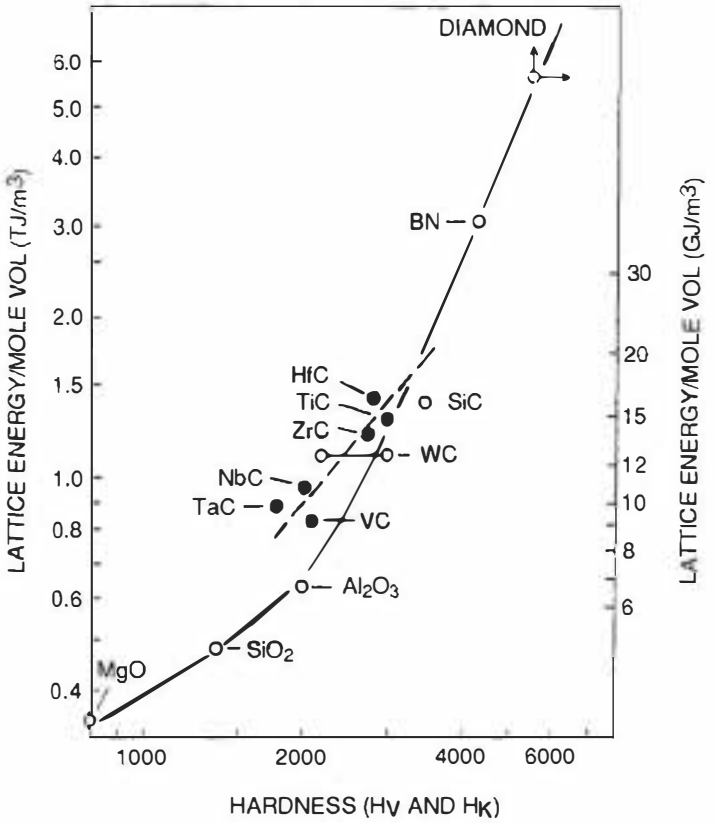


Figure 12-7. Relation between hardness and cohesive energy per molar volume expressed in terms of heat of sublimation (○-Left axis) and heat of formation (●-Right axis).

taken of the bond length by dividing the cohesive energy by the molar volume of the material, then the correlations with hardness shown in Fig. 12-7 result. It is apparent that the hardest materials are covalently bonded, and that increasing the ionic character of the bond leads to reduced hardness. At a microscopic level, directional bonds more readily resist distortion and rupture by concentrated loads than do ionic bonds.

The hardness 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 compression test. A correlation between hardness H and yield strength σ_y is therefore expected. Typically $H \approx 3\sigma_y$ in bulk metals, but the correlation has not been directly verified in films or hard coatings. It is well known that σ_y for bulk materials decreases with temperature, a fact that facilitates hot mechanical forming operations. Therefore, it can be anticipated that the hot hardness of coatings will be lower than that at ambient temperatures for which values are usually quoted. It is these hot hardness values that are important in thermal coatings and in applications such as machining. A functional relationship between H and T given by

$$H = H_0 e^{-aT} \tag{12-4}$$

has been proposed (Ref. 6). Values for the constants H_0 and a, determined from experimental data, are given in Table 12-1.

12.3.2. Hardness Testing

Hardness testing of films and coatings is a relatively simple (though deceptively so) measurement to perform. The most frequently employed methods are modifications of techniques having long standing in the metallurgical community. The Vickers hardness 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°. H_v, the Vickers hardness number, is obtained as the ratio of the applied load L to the surface area of the resulting indentation. It is given by

$$H_v = 2\cos 22° \frac{L}{l_v^2} = \frac{1.854L}{l_v^2} \text{ kg/mm}^2, \tag{12-5}$$

where l_v is the indentation diagonal. A related test is the Knoop microhardness test, which employs a rhombic-based diamond pyramid indenter where the length ratio of the major to minor indentation diagonal is 7.11; furthermore,

the minor diagonal is four times the penetration depth. The Knoop hardness is given by

$$H_k = 14.22L/l_k^2 \text{ kg/mm}^2 \tag{12-6}$$

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

Testing loads applied to the indenter are selected to produce an indent whose lateral dimensions are large enough to be measured with an optical microscope or SEM, and whose depth is small compared with the film thickness. Caution must be exercised so that the plastically deformed zone under the indenter does not extend to the substrate where stress interactions may be reflected in erroneous hardness readings. Film-thickness-to-indentation-depth ratios ranging from 5 to 10 appear to be safe in most situations. For example, suppose that in TiN coatings, a value of 7 has been established as safe. In the Vickers test this means that the indent geometry requires that the film thickness must be at least larger than l_v. For a TiN coating where $H_v = 2000$ kg/mm^2, a 10-g load generates an l_v value of ~ 3 μm. This would be the minimum acceptable film thickness for hardness testing in this case. Unusually small indents of less than a few microns in size should be avoided because of the difficulty in resolving dimensions. In TiN films, for example, the same indent measured by the same investigator optically and with an SEM gave values of 3238 and 2359 kg/mm^2, respectively (Ref. 12). Hardness comparisons should, therefore, only be made when testing methods and film thicknesses are maintained approximately constant.

Further complications arise in interpreting hardness values in hard materials. First, indentation-induced cracking phenomena make it difficult to measure indent dimensions precisely. Second, and perhaps more importantly, the quoted hardness values are load-dependent. For small loads, it is observed that the indentation dimension l varies with load as $L = K_1 l^m$, where K_1 is a constant and $1 < m \leq 2$. Elimination of l_v (or l_k) in Eq. 12-5 (or 12-6) yields

$$\frac{dH}{dL} = \frac{m-2}{m} K_2 \left(\frac{K_1}{L}\right)^{2/m} \tag{12-7}$$

where K_2 is either 1.854 or 14.22. A drop in hardness with increasing indenter load is thus predicted because $m - 2$ is negative. In polycrystalline SiC, WC–Co cemented carbide, and bearing steels, m equals 1.70, 1.88, and 1.92, respectively (Ref. 13). It is this load dependence of hardness that may be partially responsible for claims of unusually outstanding hardness values in films and coatings.

It is appropriate to end this section with brief mention of *nano*, rather than *micro*, hardness indentation measurements. The Nanoindenter, which was introduced previously (Section 9.2.4) is capable of reproducibly making indentations as shallow as 500 Å and as small as 5000 Å across. Unlike the Vickers and Knoop tests, the Nanoindenter produces a triangular pyramidal indentation. Small load ranges of 0–20 mN and 0–0.120 mN, make it possible to space indents ~ 1 μm apart when determining the hardness profile across a coating surface.

12.3.3. Effect of Microstructure on Hardness

A fundamental cornerstone of materials science is the relationship between microstructure and properties. Following Sundgren and Hentzell (Ref. 4) connections between the hardness of films and coatings and various microstructural characteristics will be discussed.

12.3.3.1. Grain-Size and Grain-Boundary Structure. In metal and alloy films the hardness correlates well with the Hall–Petch relation (Eq. 9-4), which may be written in the variant form

$$H = H_i + K_H l_g^{-1/2}, \tag{12-8}$$

where H is the hardness, H_i is the intrinsic hardness of a single crystal, l_g is the grain size, and K_H is a material constant. In refractory compound coatings, however, Eq. 12-8 appears to be invalid because the hardness of fine-grained films is frequently close to the bulk material value.

What does apparently affect the hardness and strength of refractory compounds is the perfection of the grain boundaries. Porosity and fine microcracks are very deleterious to such coatings and lower their strength and hardness significantly. Metal films are, however, somewhat more tolerant of such defects, whose influence can be blunted by plastic deformation effects. Increasing the substrate temperature is the simplest and most common way to reduce the grain-boundary defect structure and enhance the hardness of these compounds. The effect can be rather dramatic. For example, an increase in substrate temperature from 100 to 600 °C raised the hardness of magnetron-sputtered TiN from 1300 to 3500 kg/mm^2 (Ref. 14). Elevated temperatures eliminate void networks and evidently promote the strengthening of grain boundaries. The grain size increases as does hardness, contrary to the Hall–Petch prediction. With further grain growth at still higher temperatures, the hardness drops and reaches bulk levels. Metal films display a rather

different character with respect to the influence of substrate temperature. Raising the latter increases thermal stresses and grain size, but reduces the intrinsic stress and hardness.

12.3.3.2. Metastable Structures. Metastable phases are frequently observed in refractory compound films. As in the case with metastable metal alloy films, high deposition rates and low substrate temperatures are conducive to the formation of nonequilibrium structures and a fine grain size. Manifestations of the metastability are the incorporation of C and N in interstitial lattice sites and the generation of supersaturated solid solutions. This generally occurs during PVD rather than CVD, which is usually carried out under conditions closer to thermodynamic equilibrium. The incorporated interstitials tend to distort the lattice and the subsequent difficulty in initiating dislocation motion is reflected in increased hardness. The effect can be quite large. Hardnesses of 2500 to 3500 kg/mm^2 are generally found in reactively sputtered HfN films compared to 1600 in CVD-grown films. Apparently the bombardment of the growing HfN film by the sputtering gas forces N into tetrahedral interstitial positions. This creates a high compressive stress in the plane of the film and thus a higher hardness.

In general, metastable phases and structures can be frozen in up to temperatures of $0.3T_M$ (T_M is the melting point). As a consequence, metastable hard coating systems can be used at temperatures up to 550–800 °C (for T_M = 2500–3300 °C) (Ref. 10).

12.3.3.3. Impurities. Since hard PVD coatings are generally grown in medium vacuum or under even higher pressure ambients, the incorporation of noble gases, C, N, and O from residual gases, and impurities from chamber hardware and walls is not uncommon. The deposit impurities are located in both substitutional, interstitial as well as grain-boundary sites at total levels up to a few atomic percent. Even at such low concentrations, the effect on hardening can be pronounced. The mechanism of hardening due to impurities apparently involves the electrostatic attachment of the latter to charged dislocations in ionic materials and to dangling dislocation bonds in covalent compounds. Such interactions limit dislocation mobility by pinning effects.

12.3.3.4. Film Texture. By texture we mean the preponderance of one (or more) crystallographic planes oriented parallel to the film surface compared with the case of randomly oriented planes. In the latter, isotropic behavior may

be expected. Films grown by PVD and CVD techniques usually display a preferred orientation, however, with low index planes lying parallel to the substrate surface, creating a texture that is strongly dependent on virtually all deposition and process variables. Factors of as much as 2 in hardness have been observed in cubic coatings (e.g., TiC, TiN, ZrC) as a function of the preferred orientation plane [e.g., (111), (100) and (110)]; similar hardness anisotropy on basal (0001) and prismatic (1100) planes of hexagonal materials (e.g., WC, SiC) has been reported.

12.3.4. Fracture

It is a fact of nature that materials that are extremely hard are simultaneously brittle and prone to fracture. The phenomenon of fracture is of cardinal concern in a great many materials engineering applications; once fracture of a component or structure occurs, other issues quickly assume secondary importance. Films and coatings are no exception. For example, with the exception of oxidation wear, all wear mechanisms are based on some sort of crack development that creates new surfaces, from which particles can be detached by fracture processes. In a similar vein, high-temperature fracture or spalling of coatings leaves the underlying substrate unprotected and at the mercy of harsh corrosive atmospheres.

At the outset it is important to distinguish between brittle and ductile fracture. The latter occurs after the material has undergone some plastic deformation. Metals tend to undergo ductile fracture upon overloading. Brittle fracture, on the other hand, occurs rapidly, without warning and in such a way that the broken pieces can usually be neatly fitted together. Generally, brittle fracture occurs more readily in materials having small tensile strengths compared with their compressive strengths. Thus, glasses, ceramic oxides, and covalent, as well as hard metal compounds, to a lesser extent, are particularly prone to brittle fracture.

Two approaches to fracture of coatings will be presented next. The paramount assumption is that the materials involved possess an intrinsic collection of structural flaws distributed laterally as well as through the coating thickness. Voids, porosity, interconnected voids or porosity, voided or grooved grain boundaries, local regions of de-adhesion, etc. may be viewed, in a broader context, as flaws and even incipient cracks. Under either external or residual internal stressing, each flaw will locally concentrate stress and the surrounding material will tend to deform. If the stresses are sufficiently large, they can ultimately destroy the coating by crack propagation if tensile, and by wrinkling or buckling if compressive.

12.3.5. Statistical Strength of Brittle Materials

A useful concept for interpreting the strength of brittle solids is that of the "weakest link." Under load the worst flaw is assumed to govern the strength and useful life of the object. Since the flaw size geometry and distribution varies from sample to sample, there will be a distribution in fracture strength. The Swedish engineer Weibull suggested a way of handling the statistics of strength. He defined the survival probability $P(V_0)$ as the fraction of identical samples each of volume V_0 that survive loading to a tensile strength σ,

$$P(V_0) = \exp\left[-(\sigma/\sigma_f)^m\right], \tag{12-9}$$

where σ_f and m are constants. When $\sigma = 0$, then $P(V_0) = 1$ and all samples survive. At large stress levels all samples fail and $P(V_0) \to 0$. When $\sigma = \sigma_f$, then $P(V_0) = 1/e = 0.37$ and 37% of the samples survive. The constant m, known as the Weibull modulus, reflects how rapidly the strength falls as σ_f is approached. The greater the m value, the greater is the probability that the material will fail at very nearly the same stress level. In a material like steel, $m = 100$ and the reader can show that a plot of P vs. σ resembles a step function. Thus, for $\sigma < \sigma_f$, $P = 1$ and for $\sigma > \sigma_f$, $P = 0$. This expresses the fact that steel has a rather well-defined fracture stress, σ_f. For brittle ceramics, however, m may be 5–10, and there is a much greater variability in strength, as evidenced by a statistically large spread in hardness values. These facts are displayed in Fig. 12-8 on the Weibull probability axes. If the sample volume

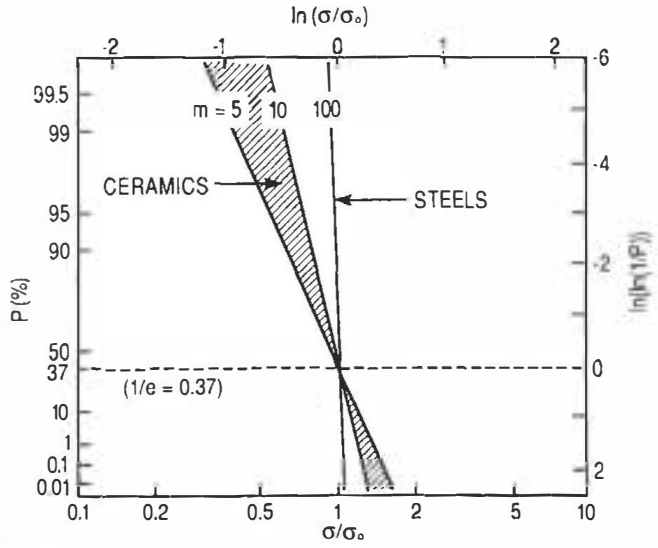


Figure 12-8. Survival probability of various materials plotted on Weibull probability axes. (Reprinted with permission. From Ref. 1; © 1986; Pergamon Press).

(V) is n times larger than V_0, i.e., $V = nV_0$, the probability of incorporating a fatal defect is enhanced leading to a smaller value of P. Under these circumstances, $P(V) = [P(V_0)]^n$ and

$$P(V) = \exp\left[-\frac{V}{V_0}\left(\frac{\sigma}{\sigma_f}\right)^m\right]. \tag{12-10}$$

Equation 12-10 is more strictly valid for bulk solids, but it has been applied to the study of macrocrack distributions in TiC coatings 2–10 μm thick (Ref. 15). As expected, more cracks are observed in thicker deposits, a reason for restricting coating thickness in practice. To analyze such situations, we can substitute film thickness for volume in Eq. 12-10.

12.3.6. Griffith Crack Theory and Fracture Toughness

We now turn to the issue of the stability of a *single* crack under load. Fracture mechanics provides a basis for addressing this situation. The Griffith theory considers a thin flat crack of length $2l$ and depth d in a plate (coating) that is assumed to be elastically isotropic. Tensile stresses act to open the crack, as shown in Fig. 12-9a. As the applied stress increases, more and more elastic strain energy E_ε is released from the coating. At any stress level this energy has a magnitude of $(1/2)\sigma\varepsilon$ (ε is the strain) per unit volume and represents

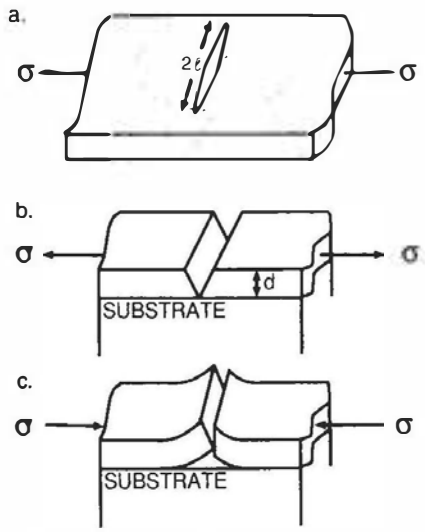


Figure 12-9. (a) Flat elliptical crack of length $2l$ in a coating loaded uniaxially in tension. (b) Crack leading to fracture of coating. (c) Crack leading to delamination of coating.

the area under the elastic stress–strain curve. By Hooke's law, $\sigma = E\varepsilon$ and

$$E_\varepsilon = \sigma^2/2E. \tag{12-11}$$

The elastic forces are resisted by interatomic bonds that must be broken for further crack extension. At any instant the total energy E_T associated with these opposing tendencies is given by

$$E_T = -\frac{\sigma^2}{2E}\pi l^2 d + 4\gamma l d. \tag{12-12}$$

The elastic strain energy is assumed to be roughly concentrated in a disk of volume $\pi l^2 d$ surrounding the crack. The second term represents the effective energy associated with broken bonds, where γ is the surface energy per unit area of crack. Rapid extension of the crack occurs when the rate of strain energy release exceeds the rate at which energy is absorbed by bonds at the crack tip. The instability occurs when $dE_T/dl = 0$, and simple differentiation then yields a critical stress for crack propagation given by

$$\sigma_c = \sqrt{4\gamma E/\pi l}\,. \tag{12-13}$$

This important formula connects *Design*, *Processing and Materials*; σ_c can be thought of as a *design* stress, l as the flaw size in the coating, an unavoidable consequence of *processing* (deposition) and γE as the intrinsic *material* property. It is common to refer to $\sqrt{4\gamma E}$ as K_c, a material property known as the fracture toughness. Fast fracture will then occur when

$$\sigma\sqrt{\pi l} \geq K_c\,. \tag{12-14}$$

As a simple application of Eq. 12-14, consider a coating of thickness d with a through crack to the substrate. Either fracture (Fig. 12-9b) or delamination of the coating (Fig. 12-9c) could occur under stress application. The condition for the former possibility is $\sigma_c d^{1/2} > K_c(c)/\sqrt{\pi}$, where $K_c(c)$ is the fracture toughness of the coating. Similarly, coating delamination will occur when the stress normal to the surface reaches a value such that $\sigma_c d^{1/2} > K_c(i)/\sqrt{\pi}$. In this expression $K_c(i)$ is equal to $\sqrt{4\gamma E}$, with γ interpreted as an interfacial debonding energy.

In microhardness tests on brittle materials and coatings, cracks usually spread out from the indentation crater along low-index cleavage planes. Through measurement of the indentation crack dimensions and hardness, values of $K_c(c)$ and $K_c(i)$ have been estimated through complex, semiquantitative relationships (Ref. 16). K_c has the peculiar units of psi-in$^{1/2}$ or MPa-$\sqrt{m}$, and values for several important coating materials and substrates are entered in Table 12-1. Metals are typically more than an order of magnitude tougher than either ceramic oxides or covalent materials in resisting fracture.

As an example, let us estimate the critical stress required to propagate microdefect cracks 10^4 Å in size within an Al_2O_3 coating. For Al_2O_3, $K_c(c) \approx 3.5$ MPa-$\sqrt{m}$ and therefore by Eq. 12-14, $\sigma_c = 3.5/\sqrt{\pi 10^4 \times 10^{-10}} = 1.97 \times 10^3$ MPa. This value is equivalent to 200 kg/mm^2 or 2×10^{10} dynes/cm^2, which is an order of magnitude larger than expected internal stresses. Therefore, it can be concluded that flaws of this size are subcritical. The benevolence of nature in generally providing for internal compressive (growth) stresses in brittle coatings should be appreciated. However, through-thickness cracks in thicker coatings are much more susceptible to propagation leading to eventual coating fracture, particularly in the presence of tensile stresses. These are commonly generated in situations where coatings are stretched during contact with other material surfaces. We now turn our attention to such effects.

12.4. Tribology of Films and Coatings

Tribology is the study of phenomena associated with interacting surfaces in relative motion. Adhesion, hardness, friction, wear, erosion, and lubrication are among some of the scientific and technological concerns of this important subject. From an engineering standpoint three separate categories of behavior can be distinguished based on the relative magnitudes of friction and wear (Ref. 17):

1. Friction and wear are both low. This is the case in bearings, gears, cams, and slideways.
2. Friction is high but wear is low. This combination of properties is desired in devices that use friction to transmit power, e.g., clutches, belt drives, tires.
3. Friction is low and wear of one contacting body is high. This is the situation that prevails during material removal processes such as machining, cutting, and grinding. In these operations plastic deformation is involved, whereas in the first two categories the contacting bodies are generally elastically stressed.

We now explore the concepts of friction and wear in more detail.

12.4.1. Friction

All engineering surfaces are rough and characterized by a density of projections or asperities with some distribution of heights. Surfaces may be thought

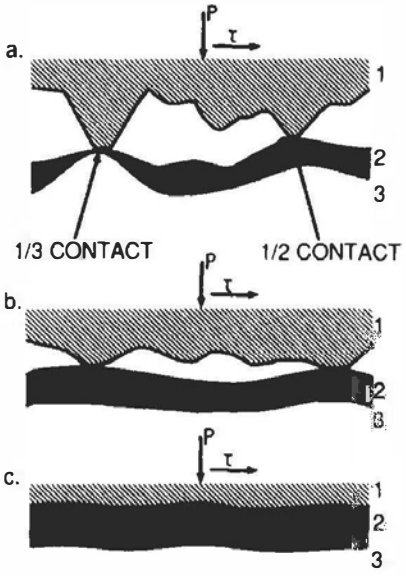


Figure 12-10. Types of contact between rough surfaces (1 and 3) through film layer (2). (Reprinted with permission from Elsevier Sequoia, S.A., from J. Halling, *Thin Solid Films* **108**, 103, 1983). (a) Film thickness is less than roughness. (b) Film thickness is of the same order as roughness. (c) Film thickness is much greater than roughness.

of as producing contact with each other at the summits of the asperities. The sum of these local contact areas represents the *real* area of contact, which may only be but a fraction of the apparent geometric contact area. In the majority of contacts surface films are present with properties different from the underlying bulk material. Surface films may be gases, fluids (e.g., oil), or deposited solid layers such as graphite, metals, or ceramic coatings. In Fig. 12-10 we consider three cases of contact between rough surfaces with an intervening film layer (Ref. 17). In the first, the film thickness is less than the roughness so that some asperities pierce the film enabling contact between bulk bodies (i.e., 1/3 contact). In Fig. 12-10b the film (2) is still thin but of the order of roughness dimensions. All contacts are now between the upper body and the film (i.e., 1/2 contact). In Fig. 12-10c the film is thicker, and all contacts are also of 1/2 type. Surface roughness is not expected to play a marked role here.

If the subscripts a and b are associated with 1/3 and 1/2 contacts, respectively, then standard notions of friction yield

$$f_a = \frac{\tau_a}{P_a}; \qquad f_b = \frac{\tau_b}{P_b} \tag{12-15}$$

In these definitions f is a nondimensional term representing the ratio of the interfacial shear stress τ to the applied normal pressure P. The overall coefficient of friction f is given by the ratio of the total real shear-to-normal forces or

$$f = \frac{\tau_a A_a + \tau_b A_b}{P_a A_a + P_b A_b}, \tag{12-16}$$

where note is taken of the total real area A for each type of contact. Furthermore, if the respective forces borne by each contact are $F_a = P_a A_a$ and $F_b = P_b A_b$, then

$$f = f_b \left(\frac{\bar{f}\bar{F} + 1}{\bar{F} + 1} \right). \tag{12-17}$$

where $\bar{f} = f_a / f_b$ and $\bar{F} = F_a / F_b$. No assumption has been made as to the nature of the film and therefore Eq. 12-17 is generally applicable to all types of contacts. For example, in the absence of a film, $F_b = 0$ and $f = f_a$ as in the simple theory of friction. In the limit of large $\bar{F}$, the value of f also tends to f_a. At the other extreme where $F_a \to 0$, $f \to f_b$ and frictional effects are governed by the film characteristics.

The dependence of frictional effects on the film thickness is of considerable technological importance. Consider a soft solid film such as lead on steel. As the film thickness increases, f decreases from f_a (steel–steel) to f_b (lead–steel), reflecting the changing nature of contact. Beyond this thickness, τ essentially stays the same but P, which is related to the effective film hardness changes. Thin lead films are harder than thicker ones because of the support provided by the steel. Therefore, P decreases as the film thickness increases, leading to an increase in f_b or f. This causes a friction variation in film thickness that passes through a minimum. The same effect occurs with thin fluid films used to lubricate metal surfaces in contact. The well-known Stribeck curve then displays a minimum in the coefficient of friction as a function of thickness or equivalent quantity (viscosity · speed/pressure). As the film thickness increases, boundary, quasi-hydrodynamic and hydrodynamic lubrication regimes are successively operative.

An important set of applications involves the dry lubrication of moving parts where fluid lubrication is not possible, e.g., at high and low temperatures or pressures. Thin, low-friction solid films composed of chalcogenides, oxides, fluorides, or carbon are used instead. In particular, MoS_2 is a favored lubricant and has been deposited by reactively sputtering MoS_2 targets in an H_2S ambient.

12.4.2. Wear Mechanisms

Wear may be defined as the progressive removal of material from surfaces that are under load in relative motion. Several different mechanisms have been identified to characterize contact wear, and these are briefly described below (Ref. 18).

12.4.2.1. Adhesive Wear. Adhesive wear occurs when applied tangential forces cause fracture between surfaces bonded at asperities. One option is for the fracture path to follow the original microwelded interface. Other paths lie above or below the interface when the strength of the bonds between asperities exceeds the cohesive strength of the bodies in contact. The result is material transfer, usually from the softer to harder body. During subsequent surface motion cycles, these particles may eventually be removed by fatigue fracture. In more severe cases of adhesive wear, smearing, galling, and seizure of surfaces may occur.

12.4.2.2. Abrasive Wear. Abrasive wear is a form of cutting wear where the material is removed by hard wear particles, by hard asperities, or by hard particles entering the interface from the environment.

12.4.2.3. Fatigue Wear. Fatigue wear occurs in situations where there is repeated loading and unloading of surfaces in contact. Failure may initiate at both surface flaws or cracks or at subsurface inhomogeneities. Crack growth eventually results in detached wear particles.

12.4.2.4. Fretting Wear. Fretting wear may be viewed as a type of fatigue wear that occurs under conditions of oscillatory movement of small amplitude (in the range of 1–200 μm), but relatively high frequency. Many sequential damage processes occur during fretting, including breakup of protective films, adhesion and transfer of material, oxidation of metal wear particles, and nucleation of surface cracks.

12.4.2.5. Delamination Wear. Delamination wear takes the form of regular detachments of thin platelike particles from wearing surfaces due to the influence of high tangential (friction) forces in the surface contact zone. During cyclic loading the cracks that develop propagate parallel to the surface at a depth governed by the material properties and coefficient of friction.

12.4.2.6. Oxidation Wear. Oxidation wear arises from the continuous rubbing and removal of surface films produced by reaction with the environment. Wear damage is modest in this case because oxide regrows soon after it is lost. However, at high temperatures where chemical reactions are accelerated oxidation wear is aggravated.

Quantifying the various wear mechanisms is a challenge that continues to engage the efforts of workers in the field. However, expressions for adhesive wear have been developed based on the probability that a junction between the two sliding surfaces produces a wear particle. A commonly employed phenomenological formula is

$$V = K_W Fl/H, \tag{12-18}$$

where V is the volume of material removed, F is the normal force, l is the length of travel, H is the indentation hardness of the softer body, and K_W is a dimensionless wear coefficient. A similar formula holds for wear of films attached to substrates. The value of K_W depends on a number of factors. Most important is the nature of the materials in contact, but the extent and type of lubrication, surface temperature, and nature of the chemical environment also play a significant role. In order of increasing wear are the material pairs: nonmetal–metal, nonmetal–nonmetal, dissimilar metals, and similar metals.

12.4.3. Tool Wear

In the important application of hard coatings for tools, wear under cutting conditions appears to be governed by two processes: abrasive wear and chemical dissolution. The latter process is important at tool temperatures above 1000 °C, which are typical for the high-speed machining of steel with coated WC tools. Dissolution of the coating atoms in the hot chips softens the cutting edge and leads to tool degradation. For the case of abrasive wear, a simple semiquantitative model to predict the relative performance of tool coatings in machining metals has been proposed and is worth presenting here. The results should be applicable to machining at low speeds or at low temperatures. The basic assumption is that three-body abrasion describes abrasive wear. In this model the abrasive particles are free to roll between two solid bodies translating relative to each other. One can imagine that abrasive inclusions at the (two-body) chip-tool interface approximate this condition.

In three-body abrasion Rabinowicz (Ref. 19) has found that the wear behavior can be divided into three regions. These depend on the ratio of the hardness of the abrasive inclusions (e.g., oxides, compounds, precipitates, hard phases, etc.) in the workpiece to that of the surface being abraded (i.e.,

the tool coating). Defining the former by H_a and the latter by H_t, the wear volume (V_t) removed from the tool is essentially expressed in each hardness regime by

$$V_t = \frac{K_W F l}{3H_t}; \qquad \frac{H_t}{H_a} < 0.8, \tag{12.19a}$$

$$V_t = \frac{K_W F l}{5.3H_t}\left(\frac{H_t}{H_a}\right)^{-2.5}; \qquad 1.25 > \frac{H_t}{H_a} > 0.8, \tag{12.19b}$$

$$V_t = \frac{K_W F l}{2.43H_t}\left(\frac{H_t}{H_a}\right)^{-6}; \qquad \frac{H_t}{H_a} > 1.25. \tag{12.19c}$$

When the coating is soft relative to the workpiece inclusions, then Eqs. 12-18 and 12-19a are basically the same and tool wear is rapid. With harder, more wear-resistant coatings, V_t decreases dramatically as an inverse power of H_t/H_a.

As an example, let us try to predict the relative abrasive wear suffered by TiN and high-speed (HSS) tools when machining steel containing hard abrasive particles of Fe_3C. Under cutting conditions it is assumed that the tool–work interface reaches 200 °C. To apply Eqs. 12-19, we need hot hardness values, which can be calculated from Eq. 12-4 and the data in Table 12-1. The estimated hardnesses are

TiN = 1310 kg/mm^2

HSS = 900 kg/mm^2

Fe_3C = 865 kg/mm^2

For TiN, $H_{TiN}/H_{Fe_3C} = 1.51$ (region c), while for HSS, $H_{HSS}/H_{Fe_3C} = 1.04$ (region b). Therefore, substitution in Eq. 12-19 yields

$$\frac{V_{HSS}}{V_{TiN}} = \frac{2.43}{5.3}\,\frac{1310}{900}\,\frac{(1.04)^{-2.5}}{(1.51)^{-6}} = 7.2$$

Therefore, TiN-coated tools are predicted to outlast HSS tools by a factor of over 7 in this case.

There have been numerous comparisons between the performance of coated vs. uncoated tools. In one study (Ref. 20) the ratio of the number of acceptable holes drilled by TiN-coated twist drills vs. uncoated HSS drills, prior to regrinding, was determined. The ratios ranged from 3 when drilling plain carbon steel to over 5 for both spring steel and gray cast iron. Reactive ion plating was used to deposit the TiN coatings in order to minimize drill

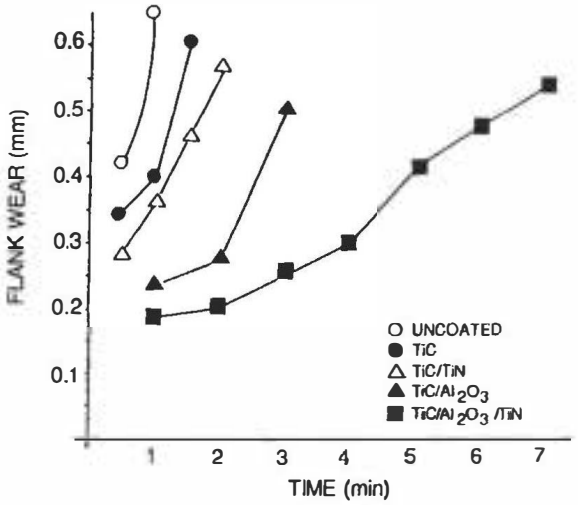


Figure 12-11. Flank wear of uncoated and coated carbide lathe tools as a function of time in machining cast iron. (Courtesy of A. Gates, Multi-Arc Scientific Coatings.)

distortion. The results of similar benefits derived from single-layer TiC coatings relative to uncoated tools are shown in Fig. 12-11. Comparisons are made for lathe tools turning cast iron. Significantly, bilayer TiC–TiN and trilayer TiC–Al_2O_3–TiN coatings are considerably longer-lived for some of the reasons noted earlier. Enhanced thermal shock and mechanical impact resistance enable these coated tools to maintain high cutting edge strength and more effectively cope with tool face, edge, and flank wear.

12.4.4. Erosion (Ref. 21)

Erosion is a type of abrasive wear in which material is lost from a surface that is in relative motion with a fluid (gas or liquid) containing solid particles. Unlike three-body wear, which was just considered, erosion is a two-body wear phenomenon. It occurs in earth-moving and agricultural equipment as well as in machinery handling slurries. We will, however, only briefly consider erosion and erosion-resistant coatings as they relate to gas-turbine engines (Ref. 22). Ingestion of particulate matter in the front end of the engine results in erosion problems downstream. Large particles at the front end typically impinge normally and wear down the leading edge of the airfoil. Further downstream in the compressor, the particles tend to be smaller and are centrifuged toward the rotor periphery. The lower impingement angle affects the midchord and even trailing edge of the airfoil. Factors such as particle

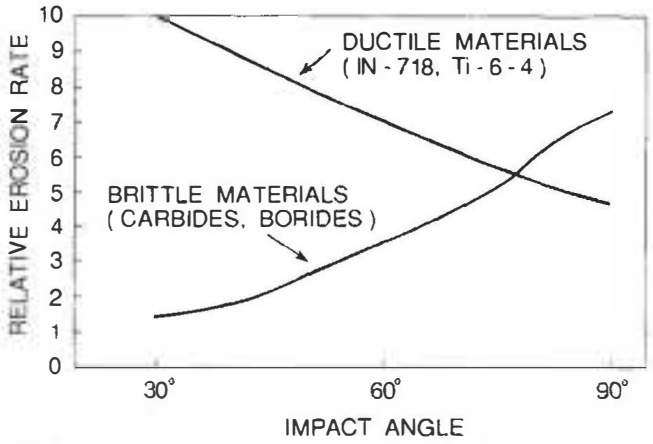


Figure 12-12. Erosion behavior of ductile and brittle materials. (From Ref. 22).

velocity, size, shape, hardness, and concentration all directly affect the magnitude of erosion. However, it is the angle of impingement of the erodent that appears to have the greatest influence on erosion. Ductile materials exhibit maximum erosion at impingement angles of 15° to 30° and erode less at higher angles. Apparently, low-angle impingement results in a tensile component that stretches the material to fracture, whereas compression and hardening of the surface by peening are the result of normal impact. Conversely, brittle materials exhibit a continuously increasing erosion rate with impingement angle and erode the maximum rate at 90°. These opposing trends are illustrated in Fig. 12-12.

Coatings have been used to counter erosion problems, but the selection of materials that can provide protection over the entire angular range has proven elusive. For example, compressor blade materials such as Ti–6V–4Al or Inconel 718 are ductile and can withstand high-angle impingement. On the other hand, carbide or boride coatings, which may be perhaps, 10 times more resistant to erosion at low angles, provide little or no protection at high angles. Composite coatings have been proposed as a solution.

12.4.5. Ball-Bearing Coatings

One of the most critical applications of hard coatings has been in ball bearings. Used in virtually all precision rotating machinery, the four basic-bearing components are the inner race, the outer race, the balls, and the cage as shown in Fig. 12-13a. The surface quality of the metallic components in rolling and sliding contact is of great importance. In particular, the roughness of the

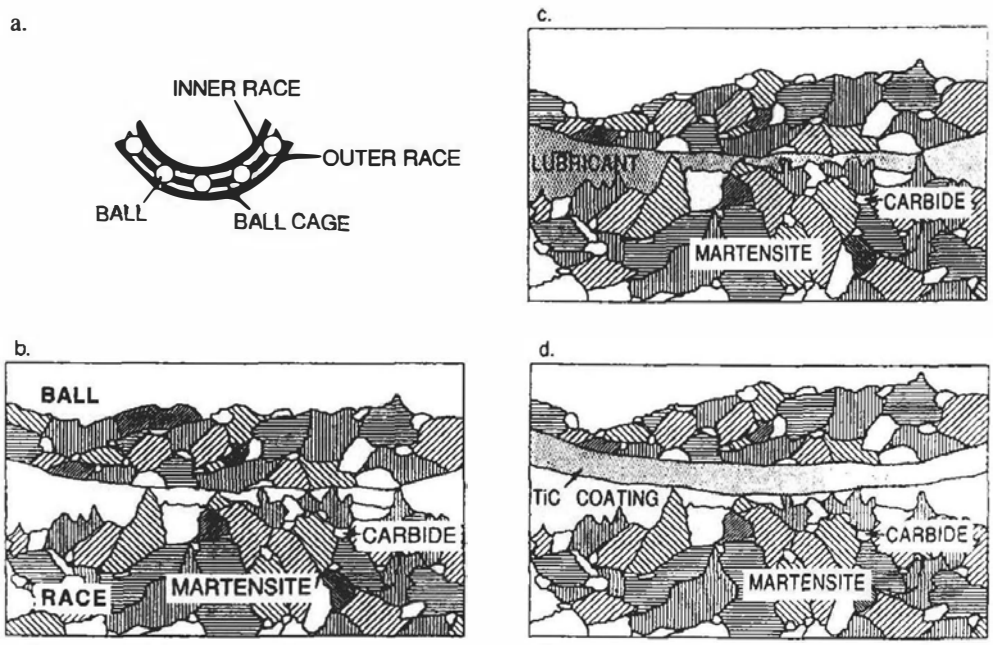


Figure 12-13. (a) Ball-bearing components: inner race or ring; outer race or ring; ball; cage or ball retainer. (b) Schematic representation of unlubricated steel ball-to-steel race contact. (c) Schematic representation of lubricated ball-to-race contact. (d) Schematic representation of unlubricated TiC coated ball-to-race contact. (Reprinted with permission from Elsevier Sequoia, S.A., H. J. Boving and H. E. Hintermann, *Thin Solid Films* **153**, 253, 1987).

raceways and balls strongly influences the effective life of the bearing. When the roughness is very low, then the oil film carries the load as desired, and minimal metal-to-metal contact occurs. However, when the roughness is high, the surface asperities occasionally impinge to cause local contact and micro-welding. This is shown in Figs. 12-13b and c, representations of the steel ball–steel race contact in the absence and presence of oil–grease lubrication. Even though the microwelds rupture almost instantaneously, the bearing interface roughens. The more rapid deterioration of the bearing when no lubricant is present is evident. Many more microwelds form, and their fracture releases hard abrasive metal particles. Such a situation arises in bearing applications where no lubricant is permitted because of environmental restrictions. For such demands, bearings coated with several microns of hard compounds such as TiC and TiN exhibit dramatically improved behavior as illustrated in Figure 12-13d. In this case a TiC-coated ball contacts an uncoated race. Longer bearing life, accompanied by lower noise and vibration, occurs

because of several beneficial effects. When the steel impinges on TiC almost no microwelding or adhesion occurs between these dissimilar materials. Upon contact, the harder TiC tends to flatten the raceway asperities by plastic deformation. This process, accompanied by a smaller temperature increase than during microweld fracture, leads to lower wear and slower lubricant degradation.

There are several issues related to bearing coatings that deserve further comment. First is the question of what to coat—the races, the balls, or both. The 1000 °C CVD TiC coating process represents a severe thermal treatment for high-precision bearing components. Furthermore, final hardening and quenching treatments are required. The geometric and size distortions accompanying these thermal cycles are decided disadvantages, especially for large bearing rings and raceways. Tribology considerations require that only one partner of the contact couple be coated. The logical alternative to coated races is to use coated balls, an increasingly accepted option. Second, there is the challenge to high-temperature CVD by low-temperature PVD processes. Even so, CVD has significant advantages. Since the coating treatment lasts for several hours, a significant amount of diffusion between the substrate and coating occurs. This results in better adhesion, progressively graded mechanical properties across the interface and improved fatigue resistance. Finally, the as-deposited TiC coating surface is too rough for use in bearings; however, with high-precision lapping and polishing the coated balls become extremely smooth with a surface roughness considerably lower than attainable with uncoated steel.

The benefits of coated bearings in specific unusual applications have been noted in the literature (Ref. 24). One example involves the orbiting European Meteosat telescope. The positioning mechanism of this telescope contains ball bearings with TiC-coated races and steel balls operating under a vacuum of 7×10^{-9} torr at temperatures between -80 to 120 °C. The bearings have functioned perfectly for several years. In other coated-bearing applications involving nuclear reactors and navigation gyroscope motors, performance at elevated temperatures (i.e., 300 °C) and high rotational speeds (i.e., 24,000 rpm) were respectively evaluated. In both cases TiC-coated bearings considerably outperformed uncoated bearings.

From these and other testimonials on bearing behavior, it is clear that the 5 millenia evolution of ways to support and move loads from sliding to rolling friction, from single contact wheels to multiple contact rollers has entered a modern phase of coating utilization. One might say that it is a whole new "ball" game.

12.5. Diffusional, Protective, and Thermal Coatings

12.5.1. Diffusion Coatings

Diffusion coatings are not coatings in the sense normally meant in this chapter. They are produced by a type of CVD reaction in which the element of interest (e.g., C, N, B, Si, Al, or Cr) is deposited on and diffused into a metal substrate (usually steel), in which it is soluble. The corresponding carburizing, nitriding, boronizing, siliciding, aluminizing, and chromizing processes yield surfaces that are considerably harder or more resistant to environmental attack than the base metal. Doping of semiconductors in which infinitesimal levels of solute are involved should be distinguished from diffusional coating processes. Through diffusion, the surface layers are frequently enriched beyond the matrix solubility limit, and when this happens, compounds (e.g., Fe_3C, Fe_4N, Fe_2B) or intermediate phases (e.g., iron and nickel aluminides) precipitate, usually in a finely dispersed form. Sometimes, however, a continuous subsurface compound layer forms. Since these compounds and phases are frequently harder than the matrix, they strengthen the surface to a depth determined by the diffusional penetration. The lack of a readily identifiable planar interface between different materials means that there is no need to be concerned about adhesion in such diffused layers.

Carburization of steel is easily the most well-known and widely used diffusional surface treatment. Carbon-rich gases such as methane are made to flow over low-to-medium carbon steels (0.1 to 0.4 wt% C) maintained at temperatures of ~ 900 °C. Pyrolysis at the metal surface releases elemental carbon that diffuses into austenite or γ-Fe, a high-temperature, face-centered cubic phase of Fe capable of dissolving about 1.25 wt% C at 920 °C. After sufficient carbon enrichment, γ-Fe can be subsequently transformed to the hard tetragonal martensite phase simply by rapidly quenching the hot steel to ambient temperature. A hard, wear-resistant case or layer of martensite containing roughly 1 wt% C then surrounds the softer mild steel core. Many automotive parts, machine components and tools such as gears, shafts, and chisels are carburized. The hard-wearing surface is backed by the softer, but tougher matrix that is required to absorb impact loading.

In order to design practical diffusional coating treatments, we must have phase compositions and solubilities, available from phase diagrams, together with diffusivity data. For example, the subsurface carbon concentration $C(x, t)$ during carburization of mild steel of composition C_0 is given by

$$C(x, t) - C_0 = (C_S - C_0)\text{erfc}\frac{x}{2\sqrt{Dt}}, \qquad (12\text{-}20)$$

where C_S, the surface carbon concentration, depends on the solubility of carbon in the steel at the particular temperature. Other terms in Eq. 12-20 have been previously defined (*cf.* p. 35); the value for the diffusivity of C in Fe is given by $D = 0.02 \exp[-(20.1 \text{ kcal/mole})/RT]$ cm^2/sec. For typical temperatures ($\sim$ 920 °C) and times ($\sim$ 1 h) case depths of the order of 1000 μm are produced. Even harder steel surfaces on steel can be produced by nitriding. Ammonia pyrolysis at 525 °C provides the N, which then penetrates the steel with a diffusivity given by $D = 0.003 \exp[-(18.2 \text{ kcal/mole})/RT]$ cm^2/sec. After two days case layers possessing a hardness of H_v 900–1200 extending about 300 μm deep can be expected. Conventional nitriding should be compared with ion-implantation methods for introducing nitrogen into steels. This technology, discussed in Chapter 13, only modifies layers several thousand angstroms deep.

As a final, but nevertheless important, example of a diffusion-coating process we consider aluminizing. Coatings based on Al have been used for several decades to enhance the environmental resistance of materials to high-temperature oxidation, hot corrosion, particle erosion, and wear. Aluminized components find use in diverse applications—nuclear reactors, aircraft, and chemical processing and coal gasification equipment.

Metals subjected to aluminizing treatments include Ni-base as well as Fe-base superalloys, heat-resistant alloys, and a variety of stainless steels. In common these alloys all contain substantial amounts of Ni, which is required for reaction with Al. Parts to be coated are packed in a retort containing Al salts, activators, and gases capable of reacting and transporting the Al to the surface being treated, in a CVD-like process. Upon solid-state diffusion, the intermetallic compound NiAl forms on the surface. This layer is hard and lacks ductility, but exhibits low wear and friction as well as impressive high-temperature corrosion resistance to both sulfur-containing gases and liquid sodium. Beyond the outer NiAl layer is a region containing a fine dispersion of Ni_3Al precipitates that serve to strengthen and toughen the matrix. Typically both regions combined do not extend deeper than $\sim$ 100 μm from the surface.

12.5.2. Oxidation and Oxide Films

The universal response of metal surfaces exposed to oxygen-bearing atmospheres is to oxidize. The oxidation product may be a thin adherent film that protects the underlying metal from further attack, or a thicker porous layer that may flake off and offer no protection. In this section, discussion is limited to oxidation via high-temperature exposure; aqueous corrosion oxidation phenomena are already the subject of a broad and accessible literature. From the

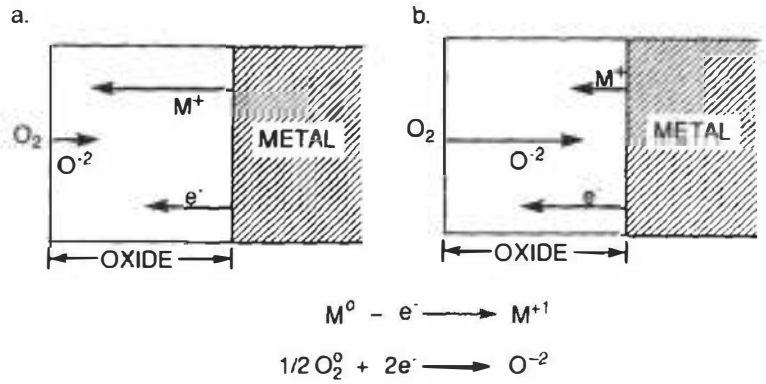


Figure 12-14. Mechanisms of oxidation: (a) oxide growth at oxide–ambient interface. (b) oxide growth at oxide–metal interface.

standpoint of thermodynamics all of the structural metals exhibit a tendency to oxidize. As noted in Chapter 1, the driving force for oxidation of a given metal depends on the free-energy change for oxide formation. What thickness of oxide will form and at what rate are questions dependent on complex kinetics and microstructural considerations, and not on thermodynamics. As shown in Fig. 12-14, two simultaneous processes occur during oxidation. At the metal–oxide interface neutral metal atoms lose electrons and become ions that migrate through the oxide to the oxide–ambient interface. The released electrons also travel through the oxide and serve to reduce oxygen molecules to oxygen ions at the surface. If metal cations migrate more rapidly than oxygen anions (e.g., Fe, Cu, Cr, Co), oxide grows at the oxide–ambient interface. On the other hand, oxide forms at the metal–oxide interface when metal ions diffuse more slowly than oxygen ions (e.g., Ti, Zr, Si). An important implication is that highly insulating oxides, such as Al_2O_3, SiO_2, do not grow readily because electron mobility, so central to the process, is low. This is what limits their growth and results in ultrathin protective native oxide films.

The model of growth kinetics developed for oxidation of Si, and Eq. 8-34 in particular, is applicable to other systems. Both parabolic oxide growth under diffusion-controlled conditions, as well as linear oxide growth when interfacial reactions limit oxidation, are frequently observed. However, not all oxidation processes fit the aforementioned categories, and other growth rate laws have been experimentally observed in various temperature and oxygen pressure regimes (Ref. 25). Specific formulas for the oxide thickness d_0, with constants $C_1, C_2, \ldots, C_7$, include

Cubic rate law

$$d_0^3 = C_1 t + C_2 \qquad \text{(e.g., Ti-400 °C)}. \tag{12-21}$$

Logarithmic

$$d_0 = C_3 \ln(C_4 t + C_5) \qquad \text{(e.g., Mg-100 °C)}. \tag{12-22}$$

Inverse Logarithmic

$$1/d_0 = C_6 - C_7 \ln t \qquad \text{(e.g., Al-100 °C)}. \tag{12-23}$$

In fact, careful plotting of data reveals that many metals and alloys apparently exhibit a number of different rates, depending on temperature. Most metals gain weight during oxidation, but, interestingly, metals like Mo and W lose weight during oxidation. The reason is that the oxide films that form (MoO_3 and WO_3) are volatile and evaporate as soon as they form.

The physical integrity of the oxide coating is the key issue that determines its ability to protect the underlying metal. If the oxide that forms is dense and thin, then it can generally be tolerated. If it is porous and continues to grow and spall off, the exposed underlying metal will undergo further deterioration. Whether the oxide formed is dense or porous can frequently be related to the ratio of oxide volume produced to the metal consumed. The quotient, known as the Pilling–Bedworth ratio, is given by

$$\frac{\text{Volume of oxide}}{\text{Volume of metal}} = \frac{M_o \rho_m}{x M_m \rho_o}, \tag{12-24}$$

where M and ρ are the molecular weight and density, respectively, of the metal (m) and oxide (o), and x is the number of metal atoms per molecule of oxide M_xO. If the ratio is less than unity, then compatability with the metal will create residual tensile stresses in the oxide. This will generally split it, much like dried wood, and make it porous, affording little protection to the underlying metal. If the ratio is close to or greater than unity, there is a good chance the oxide will not be porous; it may even be protective. On the other hand, if the ratio is much larger than unity, the oxide will acquire a residual compressive stress. Wrinkling and buckling of the oxide may cause pieces of it to spall off. For example, in the case of Al_2O_3 the Pilling–Bedworth ratio is calculated to be 1.36, whereas for MgO the ratio is 0.82. The lack of a protective oxide in the case of Mg has limited its use in structural applications.

What we have said of oxidation applies as well to the sulfidation of metals in SO_2 or H_2S ambients. Metal sulfides are particularly deleterious because of their low melting temperatures. Liquid sulfide films tend to wet grain boundaries and penetrate deeply, causing extension of intergranular cracks. Whether the elevated temperature atmosphere is oxidizing or sulfidizing, structural metals must be generally shielded by protective or thermal coatings.

12.5.3. Thermal Coatings (Refs. 26, 27)

Ever-increasing demands for improved fuel efficiency in both civilian and military jet aircraft has continually raised operating temperatures of turbine engine components. Among those requiring protection are turbine blades, stators, and gas seals. The metals employed for these critical applications are Co-, Ni-, and Fe-base superalloys, which possess excellent bulk strength and ductility properties at elevated temperatures. A widely used cost-effective way to achieve yet higher temperature resistance to degradation in the hot gas environment is to employ an additional thermal barrier coating (TBC) system. This consists of a metallic bond coat and a top layer composed primarily of ZrO_2. The bond coating, as the name implies, is the glue layer between the base metal and the outer protective oxide. Its function is not unlike that of a bond or primer coating used to prepare surfaces for painting. Typical bond coatings consist of MCrAlY or MCrAlYb, where M = Ni, Co, Fe. Original bond coating compositions such as Ni-26Cr-6Al-0.15Y (in wt%) have been continually modified in an effort to squeeze more performance from them. The role of Y or other rare earth substitutes is critical. These elements apparently protect the bond coat from oxidation and shift the site of failure from the base metal and coat interface to within the outer thermal barrier oxide. Just why is not known with certainty; it appears that these reactive metals easily diffuse along the boundaries of the plasma-sprayed particles of the bond coating, oxidize there, and limit further oxygen penetration.

The use of ZrO_2 is based on a desirable combination of properties: melting point = 2710 °C, thermal conductivity = 1.7 W/m-K, and thermal expansion coefficient = 9×10^{-6} K^{-1} (Ref. 28). However, the crystal structure undergoes transformation—from monoclinic to tetragonal to cubic—as the temperature increases, and vice versa, as the temperature decreases. A rapid, diffusionless martensitic transformation of the structure occurs in the temperature range of 950–1400 °C accompanied by a volume contraction of 3–12%. The thermal stresses so generated lead to fatigue cracking, which signifies that ZrO_2 alone is unsuitable as a TBC. The ZrO_2 overlayers are generally stabilized with 2–15 wt% CaO, MgO, and Y_2O_3. Through alloying with these oxides, a partially stable cubic structure is maintained from 25 °C to 2000 °C. Actually the tetragonal and monoclinic phases coexist together with the cubic phase, whose stabilization depends on the amount of added oxide. Cubic phase stabilization results in stress-induced transformation toughening, which can be understood as follows. If a crack front meets a tetragonal particle, the latter will transform to the monoclinic phase a process that results in a volume increase. The resultant compressive stresses blunt the advance of cracks, toughening the matrix.

Both bond and thermal barrier coatings are usually deposited by means of plasma spraying. This process is carried out in air and utilizes a plasma torch, commonly fashioned in the form of a handheld gun. An arc emanates from the gun electrodes and is directed toward the workpiece. Powders of the coating material are introduced into the plasma by carrier gases that drive them into the arc flame. There they melt and are propelled to the workpiece surface where they splat and help to build up the coating thickness. Typical bond and thermal barrier coating thicknesses are 200 and 400 μm, respectively. Exposures to temperatures of 1100 to 1200 °C, to thermal cycling, and to stresses are common in the use of TBC systems.

EXERCISES

1. a. If the potential energy of interaction between neighboring atoms in hard compounds is

$$V(r) = -A/r^m + B/r^n$$

(see problem 1-5), show that the modulus of elasticity is given by

$$E = m(n - m)A/a_0^{m+3}$$

(a_0 is the equilibrium lattice spacing).

b. Show that E is proportional to the *binding energy density* or

$$E = -mnV(r = a_0)/a_0^3.$$

c. How well does this correlation fit the data of Fig. 12-7?

2. Why are epitaxial hard coatings of TiC or TiN not practical or of particular interest?

3. A hardness indenter makes indentations in the shape of a tetrahedron whose base is an equilateral triangle that lies in the film plane.

a. If the side of this triangle has length l_t, what is the depth of penetration of the indenter?

b. What is the hardness value in terms of applied load and l_t?

4. Compare the relative abrasive wear of TiC-coated tools vs. HSS tools when machining steel containing Fe_3C particles. Assume the same wear model and machining characteristics as in the illustrative problem on p. 575.

5. Molten steel at 1500 °C can be poured into a quartz crucible resting on a block of ice without cracking it. Why? Calculate the stress generated.

6. A coating has small cracks of size l that grow by fretting fatigue. Assume the crack extension rate is given by

$$\frac{dl}{dN} = A\,\Delta K^m \qquad \text{with } \Delta K = \sqrt{\pi l}\,\Delta\sigma,$$

where $\Delta\sigma$ the range of cyclic stress, N is the number of stress cycles, and A and m are constants. By integrating this equation, derive an expression for the number of cycles required to extend a crack from l_i to l_f.

7. Speculate on some of the implications of the Weibull distribution for hard coating materials if

a. the volume can be replaced by the coating thickness.
b. the tensile strength is proportional to coating hardness (this is true for some metals).
c. the Weibull modulus is 10.

8. Oxidation rates are observed to vary as

a. $\dfrac{d(d_0)}{dt} = \dfrac{A}{d_0}$

b. $\dfrac{d(d_0)}{dt} = B\exp - Cd_0$

c. $\dfrac{d(d_0)}{Dt} = D\exp\dfrac{E}{d_0}$

A, B, C, D, E are constants. Derive explicit equations for the oxide thickness (d_0) vs. time (t) for each case.

9. Contrast the materials and processes used to coat sintered tungsten carbide lathe tool inserts and high-speed steel end-mill cutters.

10. By accident a very thin *discontinuous* rather than continuous film of TiC was deposited on the steel races of ball bearings. How do you expect this to affect bearing life?

11. The Taylor formula $Vt_l^n = c$, widely used in machining, relates the lifetime (t_l) of a cutting tool to the cutting velocity (V). Constants n and c depend on the nature of the tool, work, and cutting conditions. A TiN-coated cutting tool failed in 40 min when turning a 10-cm-diameter steel shaft at 300 rpm. At 250 rpm the tool failed after 2 h. What will the tool life be at 400 rpm?

12. Assume that K_w for adhesive wear is 10^{-10} for 52100 steel balls on 52100 steel races and an order of magnitude lower for TiC-coated bearings.

a. Approximately how many revolutions are required to generate a wear volume of 10^{-5} cm^3 in an all-steel, 2-cm-diameter bearing if $H = 900$ kg/mm^2 and $F = 200$ kg?
b. At 1000 rpm how long will it take to produce this amount of wear in the coated bearing?

13. A problem arising during the CVD deposition of TiC on cemented carbides is the loss of C from the substrate due to reaction with $TiCl_4$. This leads to a brittle *decarburized* layer (the η phase) between substrate and coating. Assuming that interstitial diffusion of C in W is responsible for the effect, sketch the expected C profile in the substrate after a 2-h exposure to 1000 °C, where the diffusivity is 10^{-9} cm^2/sec.

14. The surface of a HSS drill is exposed to a flux of depositing Ti atoms and a N_2 plasma during reactive ion plating at 450 °C. It is assumed that an effective surface concentration of 50 at% N is maintained that can diffuse into the substrate. Under typical deposition conditions roughly estimate the ratio of layer thicknesses of TiN to Fe_4N formed as a function of time.

REFERENCES

1. M. F. Ashby and D. R. H. Jones, *Engineering Materials 1 and 2*, Pergamon Press, Oxford (1980 and 1986).
2.* E. A. Almond, *Vacuum* **34**, 835 (1984).
3. H. Holleck, *J. Vac. Sci. Tech.* **A4**, 2661 (1986).
4.* J. E. Sundgren and H. T. G. Hentzell, *J. Vac. Sci. Tech.* **A4**, 2259 (1986).
5. M. Rühle, *J. Vac. Sci. Tech.* **A3**, 749 (1985).
6. B. M. Kramer and P. K. Judd, *J. Vac. Sci. Tech.* **A3**, 2439 (1985).
7.* R. F. Bunshah, ed., *Deposition Technologies For Films and Coatings*, Noyes, Park Ridge, NJ (1982).
8. D. T. Quinto, G. J. Wolfe and P. C. Jindal, *Thin Solid Films* **153**, 19 (1987).

*Recommended texts or reviews.

9.* R. F. Bunshah and C. Deshpandey, in *Physics of Thin Films*, Vol. 13, eds. M. H. Francombe and J. L. Vossen, Academic Press, New York (1987).
10. D. Holleck and H. Schulz, *Thin Solid Films* **153**, 11 (1987).
11. A. Layyous and R. Wertheim, *J. de Phys. Colloque* C5, 423 (1989).
12. W. D. Sproul, *Thin Solid Films* **107**, 141 (1983).
13. C. A. Brookes, *Science of Hard Materials*, Plenum Press, New York (1983).
14. W. D. Münz and G. Hessberger, *Vak. Tech.* **30**, 78 (1981).
15. G. Gille, *Thin Solid Films* **11**, 201 (1984).
16. P. K. Mehrotra and D. T. Quinto, *J. Vac. Sci. Tech.* **A3**, 2401 (1985).
17. J. Halling, *Thin Solid Films*, **108**, 103 (1983).
18. M. Antler, *Thin Solid Films*, **84**, 245 (1981).
19. E. Rabinowicz, *Lubr. Eng.*, **33**, 378 (1977).
20. R. Buhl, H. K. Pulker, and E. Moll, *Thin Solid Films*, **80**, 265 (1981).
21. R. C. Tucker, in *Metals Handbook*, Vol. 11, 9th ed., American Society for Metals (1986).
22. R. V. Hillery, *J. Vac. Sci. Tech.* **A4**, 2624 (1986).
23. H. J. Boving and H. E. Hintermann, *Thin Solid Films* **153**, 253 (1987).
24. H. E. Hintermann, *Thin Solid Films* **84**, 215 (1981).
25. S. Mrowec and T. Werber, *Gas Corrosion of Metals* U.S. Dept. of Commerce, NTIS TT 76-54038, Springfield, VA (1978).
26. S. Stecura, *Thin Solid Films* **136**, 241 (1986).
27. R. A. Miller and C. C. Berndt, *Thin Solid Films* **119**, 195 (1984).
28. G. Johner and K. K. Schweitzer, *Thin Solid Films* **119**, 301 (1984).