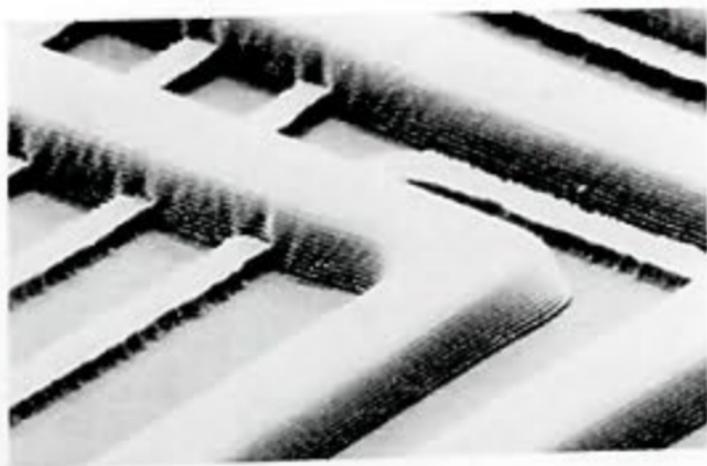


**SEMICONDUCTOR MATERIALS
AND
PROCESS TECHNOLOGY
HANDBOOK**

for
Very Large Scale Integration (VLSI)
and
Ultra Large Scale Integration (ULSI)



Edited by
Gary E. McGuire

NOYES PUBLICATIONS

**SEMICONDUCTOR MATERIALS AND
PROCESS TECHNOLOGY HANDBOOK**

MATERIALS SCIENCE AND PROCESS TECHNOLOGY SERIES

Editors

Rointan F. Bunshah, University of California, Los Angeles (*Materials Science and Technology*)

Gary E. McGuire, Microelectronics Center of North Carolina (*Electronic Materials and Processing*)

DEPOSITION TECHNOLOGIES FOR FILMS AND COATINGS; Developments and Applications: by *Rointan F. Bunshah et al*

CHEMICAL VAPOR DEPOSITION FOR MICROELECTRONICS; Principles, Technology, and Applications: by *Arthur Sherman*

SEMICONDUCTOR MATERIALS AND PROCESS TECHNOLOGY HANDBOOK; For Very Large Scale Integration (VLSI) and Ultra Large Scale Integration (ULSI): edited by *Gary E. McGuire*

SOL-GEL TECHNOLOGY FOR THIN FILMS, FIBERS, PREFORMS, ELECTRONICS, AND SPECIALTY SHAPES: edited by *Lisa C. Klein*

HYBRID MICROCIRCUIT TECHNOLOGY HANDBOOK; Materials, Processes, Design, Testing and Production: by *James J. Licari* and *Leonard R. Enlow*

HANDBOOK OF THIN FILM DEPOSITION PROCESSES AND TECHNIQUES; Principles, Methods, Equipment and Applications: edited by *Klaus K. Schuegraf*

Related Titles

ADHESIVES TECHNOLOGY HANDBOOK: by *Arthur H. Landrock*

HANDBOOK OF THERMOSET PLASTICS: edited by *Sidney H. Goodman*

HANDBOOK OF CONTAMINATION CONTROL IN MICROELECTRONICS; Principles, Applications and Technology: edited by *Donald L. Tolliver*

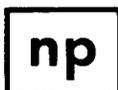
Semiconductor Materials and Process Technology Handbook

**for
Very Large Scale Integration (VLSI)
and
Ultra Large Scale Integration (ULSI)**

Edited by

Gary E. McGuire

Microelectronics Center of North Carolina
Research Triangle Park, North Carolina



**NOYES PUBLICATIONS
WILLIAM ANDREW PUBLISHING, LLC**

Norwich, New York, U.S.A.

Copyright © 1988 by Noyes Publications

**No part of this book may be reproduced in any form
without permission in writing from the Publisher.**

Library of Congress Catalog Card Number: 87-31529

ISBN: 0-8155-1150-7

Printed in the United States

**Published in the United States of America by
Noyes Publications/William Andrew Publishing, LLC
13 Eaton Avenue, Norwich, New York 13815**

10 9 8 7 6

Reprint Edition

Library of Congress Cataloging-in-Publication Data

Semiconductor materials and process technology.

Bibliography: p.

Includes index.

**1. Integrated circuits--Very large scale
integration--Design and construction--Handbooks,
manuals, etc. I. McGuire, G.E.**

TK7874.S4178 1988 621.395 87-31529

ISBN 0-8155-1150-7

Contributors

Kenneth E. Bean

Texas Instruments Incorporated
Dallas, TX

Bruce E. Deal

Research Center
Fairchild Semiconductor Corp.
Palo Alto, CA

William C. Dautremont-Smith

AT&T Bell Laboratories
Murray Hill, NJ

Richard B. Fair

Microelectronics Center of
North Carolina
Research Triangle Park, NC

Richard A. Gottscho

AT&T Bell Laboratories
Murray Hill, NJ

Paul S. Ho

IBM Thomas J. Watson Research
Center
Yorktown Heights, NY

Gary E. McGuire

Microelectronics Center of North
Carolina
Research Triangle Park, NC

William C. O'Mara

Aeolus Laboratory
Palo Alto, CA

R. Fabian Pease

Stanford University
Stanford Electronics Laboratory
Department of Electrical
Engineering
Stanford, CA

Ronald J. Schutz

AT&T Bell Laboratories
Murray Hill, NJ

John A. Thornton

University of Illinois
Department of Materials Science and
Coordinated Science Laboratory
Urbana, IL

NOTICE

To the best of the Publisher's knowledge the information contained in this book is accurate; however, the Publisher assumes no responsibility nor liability for errors or any consequences arising from the use of the information contained herein. Final determination of the suitability of any information, procedure, or product for use contemplated by any user, and the manner of that use, is the sole responsibility of the user. The book is intended for informational purposes only. Semiconductor raw materials and processes are potentially hazardous and due caution should always be exercised in the handling of materials and equipment. Expert advice should be obtained at all times when implementation is being considered.

Contents

INTRODUCTION	1
<i>Gary E. McGuire</i>	
References	7
1. SILICON MATERIALS TECHNOLOGY	8
<i>William C. O'Mara</i>	
Introduction	8
Silicon Crystal Growth	8
Crystal Growth Equipment and Process	9
Dopant and Impurity Incorporation	15
Incorporation of Oxygen	19
Incorporation of Carbon	21
Wafer Preparation	22
Mechanical Shaping Procedures	22
Wafer Etching	23
Polishing	25
Cleaning	25
Material Properties	26
Crystal Structure	26
Electrical Properties	27
Optical Properties	30
Mechanical Properties	31
Process-Induced Defects	32
Oxidation-Induced Stacking Faults	32
Saucer Pits	33
Carbon-Related Defects	34
Oxygen in Silicon	35
Oxygen in As-Grown Silicon	35
Quantitative Analysis of Oxygen in Silicon	35

Interpretation of Infrared Absorption Spectra	36
Solid Solubility	38
Diffusion Coefficient	39
Donor Formation	40
Precipitation from Solid Solution	41
Denuded Zone Formation	42
Device Application	43
References	43
2. THE THERMAL OXIDATION OF SILICON AND OTHER SEMI- CONDUCTOR MATERIALS	46
<i>Bruce E. Deal</i>	
Introduction and Background	46
Silicon Thermal Oxidation Kinetics	47
General Relationship	48
Thin Oxide Formation	54
Properties of Thermal Oxides	54
Process Variable/Oxidation Reaction Dependencies	55
Effects of the Oxidation Reaction on Surface Properties	55
Oxide Charges	55
Dopant Redistribution	61
Effects of Surface Properties on the Oxidation Reaction	62
Silicon Orientation	62
Dopant Concentration	63
Surface Preparation	64
Effects of Ambients on the Oxidation Reaction	65
Ambient Type	65
Chlorine Additions	65
Nitridation	66
Oxidant Pressure	67
Oxidation Mechanism	68
Atomic Reactions	68
Structure of the Si-SiO ₂ Interface	70
Other Oxidation Processes	72
Assisted Oxidation	72
Silicon-Containing Materials	73
Other Semiconductors	74
Future Trends	75
References	77
3. CHEMICAL VAPOR DEPOSITION OF SILICON AND ITS COMPOUNDS	80
<i>Kenneth E. Bean</i>	
Introduction	80
Epitaxial Deposition	82
HCl in Situ Etching	86
Gettering	91
Selective Deposition	94

CVD of Dielectric Films	105
X-Ray Lithography Mask Fabrication	115
References	125
4. CHEMICAL ETCHING AND SLICE CLEANUP OF SILICON	126
<i>Kenneth E. Bean</i>	
Introduction.	126
Orientation Dependent Cleaving of Silicon	129
Orientation Dependent Etching and Orientation Dependent Deposition.	149
(110) Orientation Dependent Effects	157
Defect Delineation Etching	172
Slice Cleanup	183
Precleanup Solvent Rinse	187
Choline Cleanup Process	189
References	190
5. PLASMA PROCESSING: MECHANISMS AND APPLICATIONS	191
<i>W.C. Dautremont-Smith, Richard A. Gottscho and R.J. Schutz</i>	
Introduction.	191
Fundamental Aspects.	192
Plasmas and Sheaths.	193
Response Time and Screening Distance	193
Equivalent Circuits	195
Feedstock Composition	210
Pressure, Flow-Rate, and Residence Time.	213
Power Density	217
Plasma-Surface Chemistry	217
Chemical Vapor Transport	218
Plasma Modified Chemical Vapor Transport	224
Mechanisms	230
Modeling	241
Plasma Etching	242
Introduction.	242
Outline	242
Pattern Definition and Transfer	242
An Illustration of Plasma Etch Patterning.	244
Equipment.	244
Parallel Plate Etchers	244
The Hexagonal Cathode Etcher	247
Single Wafer Etcher	248
Endpoint Detection	249
Voltage/Power	249
Optical Emission Spectra	249
Laser Interferometry	249
Defining Process Parameters and Goals	249
Material to Be Etched.	251
Feature Edge Profiles.	252

- Uniformity 253
- Selectivity 254
- Throughput 256
- Defect Introduction 257
- Radiation Damage Effects 257
- Specific Etching Processes 258
 - Silicon and Silicides 258
 - Etching of Thermal and LPCVD Oxide 260
 - Etching of Aluminum Metallization 263
 - Silicon Nitride 264
 - Etching of Group III, V Compound Semiconductors 264
- Plasma Deposition 265**
 - Introduction 265
 - Applications of PECVD Materials 268
 - General Aspects of PECVD and PECVD Reactors 268
 - Reactor Designs 273
 - Source Gases 280
 - Uniformity Considerations and Interaction of PECVD
 - Variable Parameters 281
 - Film Properties and Their Control 284
 - Materials Deposited and Their Applications 289
 - Silicon Nitride 289
 - Silicon Oxide 301
 - Amorphous Silicon 305
 - Other Semiconductors, Including Epitaxial Growth 308
 - Metals 311
 - Silicides 312
 - Other Materials 312
 - Interface Properties 315
- Summary and Conclusions 318**
- References 320**

6. PHYSICAL VAPOR DEPOSITION 329

John A. Thornton

- Introduction 329**
- The Vacuum Environment 330**
- Evaporation 336**
 - Introduction 336
 - Evaporation Rate 337
 - Evaporation Sources 340
 - Wire and Metal Foil Sources 340
 - Crucible Sources 341
 - Sublimation Sources 343
 - Baffle Type Sources 343
 - Knudsen Cell Sources 343
 - Electron Beam Sources 344
 - Other Types of Evaporation Sources 345
 - Deposit Thickness Uniformity 345

Evaporation of Alloys, Compounds and Mixtures	346
Introduction.	346
Evaporation of Alloys	347
Evaporation of Compounds.	347
Special Evaporation Methods.	348
Flash Evaporation	348
Hot-Wall Evaporation.	349
Close-Spaced Sublimation	350
Multi-Source Evaporation	351
Reactive Evaporation	352
Deposition Rate and Flux Monitors	354
Evaporation Source Material	355
Molecular Beam Epitaxy	355
Introduction.	355
Apparatus Configuration.	356
Deposition Procedure.	358
Coating Growth	360
Applications.	362
Sputtering	364
Introduction.	364
Basic Sputtering Mechanisms	365
Sputtered Species	369
The Sputtering Yield	371
Sputtering Alloys and Compounds	376
Glow Discharge Sputtering Apparatuses.	378
Planar Diodes	378
Assisted-Discharge Devices, Triodes.	381
Magnetrons	382
Ion Beam Sputtering	392
RF Sputtering.	394
Reactive Sputtering	400
Target Fabrication.	409
Thin Film Growth and Properties	410
Coating Nucleation and Growth	410
Condensation	410
Nucleation	412
Evolution of Microstructure	415
Growth of Compound Semiconductors from Multicomponent Vapors	419
The Use of Ion Bombardment for Substrate Cleaning and to Influence Coating Growth.	422
Internal Stresses	429
Metallization of Semiconductor Devices	438
Introduction.	438
Metallization Materials Considerations.	438
Step Coverage.	440
Radiation Damage.	443
References	445

7. DIFFUSION AND ION IMPLANTATION IN SILICON	455
<i>Richard B. Fair</i>	
Introduction	455
Continuum Theory	456
Special Cases	456
Predeposition	456
Redistribution or Drive-In	461
Diffusion Coefficients	463
Atomic Theory of Diffusion	465
Diffusion Mechanisms	465
The Flux Equation in Diffusivity	466
Multiple Charge State Vacancy Model	470
The Role of Point Defects in Silicon Processing	472
The Silicon Processing Balancing Act	472
Point Defects	473
The Monovacancy	473
The Silicon Self-Interstitial Atom	475
Point Defect Models of Diffusion in Silicon	476
Experimental Observations	477
Diffusion in the Presence of Excess Point Defects	479
Oxidation-Enhanced Diffusion	479
Doping Dependence of Oxidation-Enhanced Diffusion	482
Effect of Chlorine on Oxidation-Enhanced Diffusion	483
Characteristics of Silicon Self-Diffusion	485
Dopant Diffusion in Silicon	488
Arsenic Diffusion Models	491
Phosphorus Diffusion Models	496
Boron Diffusion Models	501
Design Considerations for Implanted-Diffused Layers	505
Arsenic Diffusion	506
Phosphorus Diffusion	510
Ion Implantation	512
Ion Implant System	515
Simple Range Theory	515
Nuclear Stopping	519
Electronic Stopping	519
Critical Energy	520
Projected Range	521
Implantation Masking	522
Ion Channeling	527
Modeling Implanted Dopants in Silicon	528
References	538
8. MICROLITHOGRAPHY FOR VLSI	541
<i>R. Fabian Pease</i>	
Introduction	541
Forming the Resist Film	542
Generation of the Aerial Image for Electron Beam Mask Making	543
Interaction of Electrons with the Workpiece	545

Exposure and Development of Photoresist on Semiconductor	
Wafers	551
Formation of the Aerial Image in Projection Mask Aligners	556
Interaction of Ultra-Violet Light with Photoresist	559
Exposure and Development of Photoresist Films on Reflective	
Substrates	564
Emerging New Technologies	567
Multiple Level Resist	568
Electron-Beam Direct Write	570
X-Ray Lithography	571
Vote Taking Lithography	573
References	574
9. METALLIZATION FOR VLSI INTERCONNECT AND PACKAGING	575
<i>Paul S. Ho</i>	
Introduction	575
Wiring Structure	581
Impact of Device Scaling	587
Electrical Characteristics	589
Material Reaction	595
Metallization Reliability	598
Junction and Gate Contacts	598
Electromigration	604
Summary	607
References	607
10. CHARACTERIZATION OF SEMICONDUCTOR MATERIALS	610
<i>Gary E. McGuire</i>	
Introduction	610
Surface Analysis Techniques	611
Auger Electron Spectroscopy	611
Photoelectron Spectroscopy	618
Secondary Ion Mass Spectroscopy	624
Resonance Ionization Spectroscopy	632
Rutherford Backscattering Spectroscopy	635
Summary	640
Imaging Analysis Techniques	640
Scanning Electron Microscopy	640
Scanning Transmission Electron Microscopy	647
X-Ray Topography	653
Bulk Analysis Techniques	654
Fourier Transform Infrared Spectroscopy	654
Deep Level Transient Spectroscopy	659
Photoluminescence Spectroscopy	660
Neutron Activation Analysis	664
References	666
INDEX	669

Introduction

Gary E. McGuire

*Tektronix, Inc.
Beaverton, Oregon*

There has been a major thrust throughout the semiconductor industry to establish the capability to process Very Large Scale (VLSIC) and Ultra Large Scale (ULSIC) Integrated Circuits, as well as, Very High Speed Integrated Circuits (VHSIC). The generally accepted goals of VLSI technology are to produce devices with 10^6 gates or memory bits per circuit with geometries of less than 1 micrometer. The goal of ULSI is to produce devices with 10^7 - 10^9 gates or memory bits per circuit.¹ The goals of VHSIC are to develop the technology to produce devices with 1 micrometer features that have an equivalent gate clock frequency product exceeding 5×10^{11} gate Hz/cm² and a minimum clock rate of 25 MHz. The goals of ULSI and VHSIC are a natural evolution of current IC technology since the number of devices on a single IC have nearly doubled every year for the past twenty years. Similarly, as shown in Figure 1, the minimum horizontal dimension has been reduced throughout the past decade by a factor of two about every seven years. Even though the minimum horizontal dimension has been reduced the overall chip length has increased as illustrated in Figure 2.

The device parameters of gate length, junction depth and gate oxide thickness have all decreased with each succeeding generation of product. There is significant interaction between these parameters and device performance.² As shown in Figure 3, the gate oxide thickness has decreased with scaling of the active channel length. The device speed, minimum gate delay, has also decreased with gate length, Figure 4, yielding improvements in device performance. Part of the improved performance is a direct result of shorter distances that signals must travel

2 Semiconductor Materials

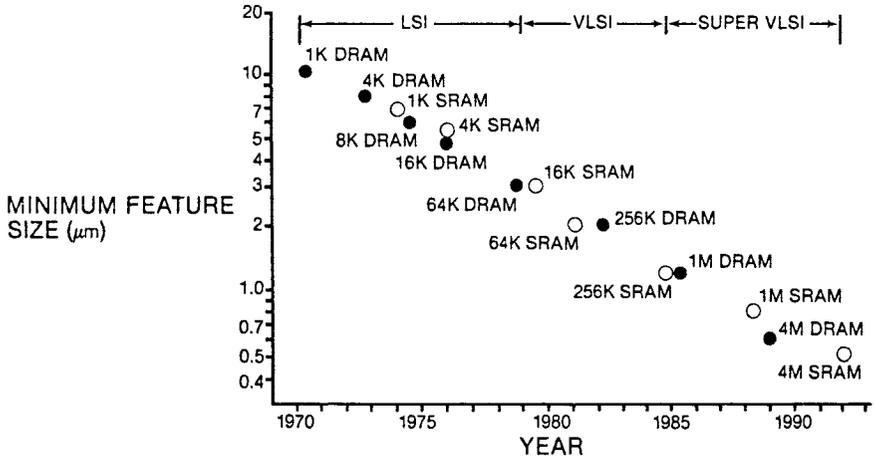


Figure 1: Projected decrease in feature size for dynamic and static random access memories as a function of the year of introduction.

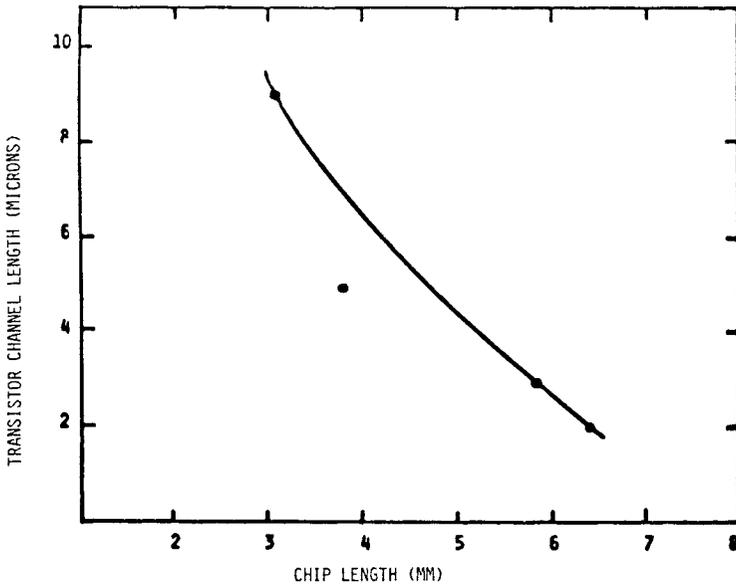


Figure 2: Plot of transistor channel length versus overall chip size showing the trend toward larger chips even though the feature sizes have decreased.

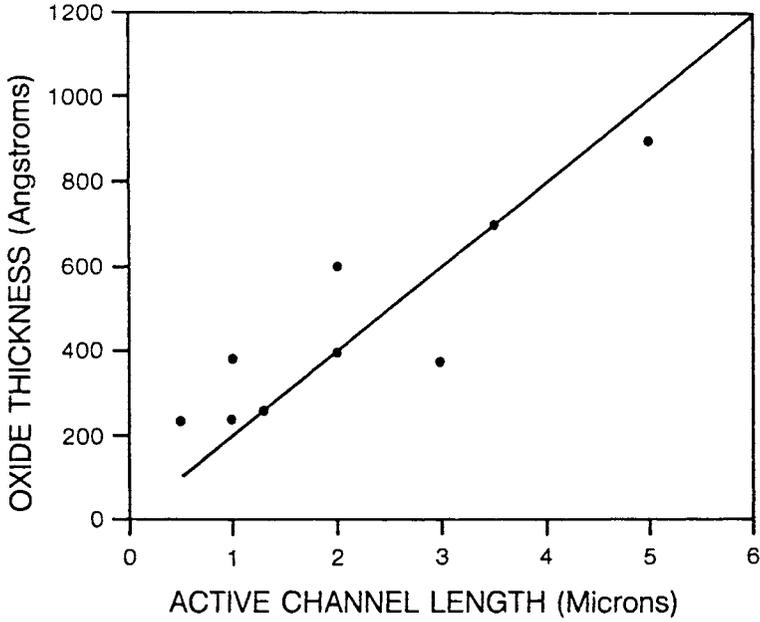


Figure 3: Plot of gate oxide thickness as a function of the active channel length.

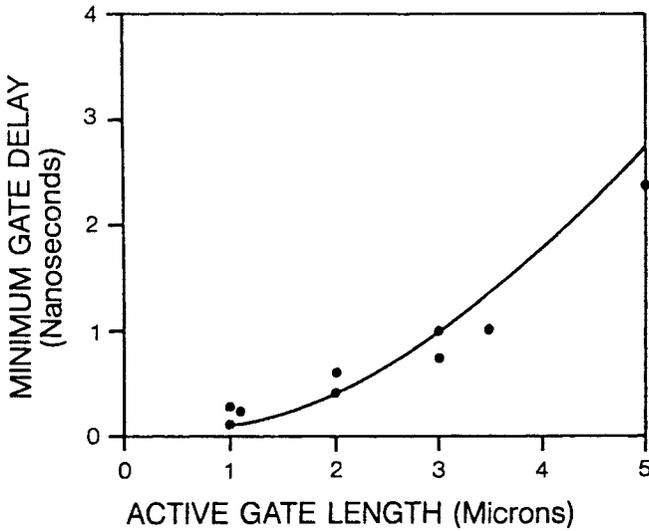


Figure 4: Plot of minimum gate delay as a function of the active gate length.

4 Semiconductor Materials

between circuits. Miniaturization also plays a large role in the steady decrease in the energy utilized per switching operation. The trend in this parameter, the product of power per circuit and logic delay is shown in Figure 5.³

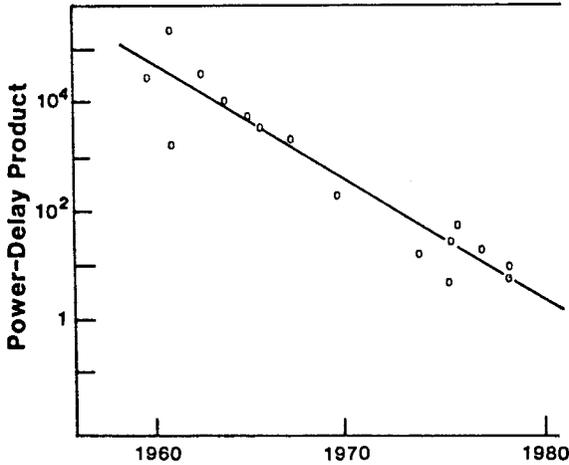


Figure 5: Plot of the decreasing power-delay time product as a function of the year of component introduction.

Additional device related concerns arise as a result of device scaling. The subthreshold scaling problem is one that poses significant technological barriers.⁴ A plot of the drain current versus gate voltage for a reference and scaled device, Figure 6, shows that when the gate voltage is decreased to below the threshold value, the channel current does not drop linearly to zero but decreases exponentially. This rate of decrease in the channel current in the subthreshold region is dependent on the gate voltage and independent of the channel length at a given temperature. If the designed threshold of the scaled device is too low, a significant amount of current continues to flow in the channel when the gate voltage is reduced to zero. As a result, the stored charge of a capacitor on a dynamic random access memory (DRAM) will leak off between refresh cycles. This phenomenon precludes scaling of DRAM threshold voltages according to the dictates of scaling theory, and therefore precludes accurate scaling of power supply voltages. As DRAM's are made smaller and gate insulators are made thinner, it may be necessary to operate at higher fields across the oxide, increase the storage capacitor area to increase the total charge stored or replace the SiO₂ with a higher dielectric constant material in order to minimize the effects of leakage current across the channel.² The current leakage in the channel is also complicated by current leakage around the periphery of the device. The leakage rate of DRAM cells around the periphery has increased with scaling as a result of the change in the perimeter-to-area ratio of the cell, Figure 7.⁵

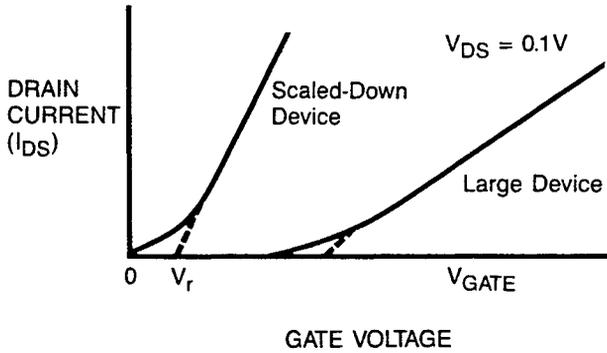


Figure 6: Plot of the drain current versus gate voltage illustrating the lack of a sharp cut-off voltage for VLSI devices. Reference: VLSI Technology and Design, *IEEE*, O. Folberth and W. Grobman (1984).

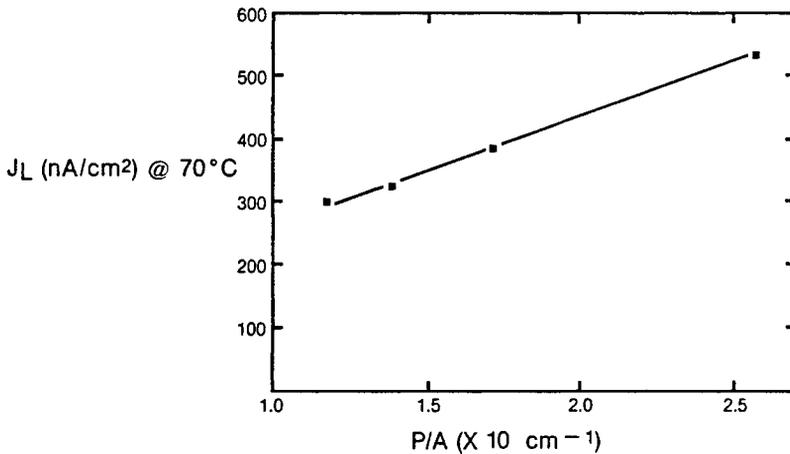


Figure 7: Dependence of the current leakage rate of a dynamic random access memory cell as a function of the periphery-to-area ratio. Reference: P.K. Chatterjee et al., *IEEE Trans. Electron Devices* ED-26:486 (1979).

The sheer complexity of the design of chips containing many circuits impedes progress. Interconnecting and packaging these devices is another area of device technology that requires significant development to circumvent the problems that arise as a result of device scaling. For example, the resistance of interconnecting wires increases as dimensions are reduced. The increasing length of interconnections on a chip as a result of the increasing number of circuits accentuates this problem.³

The methodology necessary for achieving ULSI and VHSI circuit densities and dimensions are related to traditional horizontal scaling, as well as vertical scaling. Feature sizes have historically been limited by the

6 Semiconductor Materials

lithographic techniques. However, all the available lithographies (optical, x-ray, ion and electron beam) are capable of patterning geometries less than 1 micrometer while maintaining acceptable alignment tolerances and defect levels.⁶ The shrinking feature size has also necessitated a reduction in film thickness in order to achieve the desired physical dimensions and electrical parameters. This has generated new materials and process technologies to meet these changing requirements.

There are several common themes behind the emergence of these particular technologies. Foremost among these are small area pattern definition, registration and replication as driving forces for the various lithographies and etching techniques. For example, anisotropic etching removes material in the vertical direction with minimal or no etching in the horizontal direction. This provides precise replication of the exposed pattern. Implementation of plasma or reactive ion etching requires thorough characterization of the etch rate selectivity, anisotropy, uniformity and process reliability. Once the desired material has been deposited and patterned, there is a need to minimize any subsequent process conditions that will alter these properties. One approach that has been taken is to use lower temperature processes which minimize diffusion of dopants, impurities and contact metallization or reduce the nucleation of stacking faults, dislocations and precipitates and dimensional changes in the substrate. Implementation of ion implantation, chemical vapor and plasma deposition have lowered the temperature required for many process steps. An alternate approach which reduces the solid state interaction of materials is rapid thermal processing. The process temperature is ramped up and down quickly so that the substrate is exposed to an elevated temperature for only a short period of time.

In addition to the new process technologies, there has been increasing demands placed on materials. The increasing demands have pushed some materials beyond their fundamental limits and created a search for new or improved materials. For example, the electrical conductivity and associated electromigration problems of Cu or Si doped Al contacts was clearly unacceptable for VLSI circuits. This led to the development of low resistivity refractory silicide and doped polysilicon interconnects. Smaller geometries also dictate higher sensitivity substrates, with fewer defects and better dimensional control.

In an ultra-small electronic structure, the device is approaching the dimension of long-range order in the material. For this reason one must be concerned with diffusion, microstructure and phase transitions within the host material. The nucleation of thin films is generally governed by non-equilibrium thermodynamics resulting in questions about solid phase reactions, segregation and agglomeration. Our lack of understanding of physical phenomena on the microscopic scale has created the need for extensive characterization.

At the same time device geometries diminished, the analytical tools with high spatial resolution flourished. The surface and thin film analysis techniques x-ray photoelectron spectroscopy, Auger electron spectroscopy, secondary ion mass spectroscopy and Rutherford backscattering spectroscopy have grown in popularity in parallel to the decrease in the

vertical dimension of IC's. The new trace analyses technique Fourier transform infrared, photoluminescence, deep level transient and resonance ionization spectroscopy have emerged to compliment the more traditional trace analysis technique, neutron activation analysis, reflecting the need to characterize the higher purity materials. The low defect density requirements for IC materials has given a boost to the defect imaging techniques, transmission electron microscopy and x-ray topography. The increased complexity of ULSI and VHSL circuits has been an incentive in the development of the electrical evaluation technique voltage contrast.

At each new stage of miniaturization physical effects and phenomena are encountered which were previously unknown or could be neglected. It is easy to imagine that this tendency will continue with each new level of miniaturization. One might expect that the obstacles will be more difficult to overcome the closer one gets to the ultimate limitations of the technology. Although the majority of the material in this text does not directly address these fundamental barriers, it does review the present state-of-the-art and future directions which is in itself a reflection of the barriers that have already been overcome.

REFERENCES

1. J.D. Meindl, *IEEE Transactions on Electron Devices*, ED-31 #11, 1555 (1984).
2. A. Reisman, in *VLSI: Technology and Design*, Otto G. Folberth and Warren D. Grobman, eds, IEEE Press, New York (1984).
3. R.W. Keyes, *IEEE Transactions on Electron Devices*, ED-26 #4, 271 (1979).
4. F.H. Gaenssien, V.L. Rideout, E.J. Walker and J.J. Walker, *IEEE Transactions on Electron Devices*, ED-24, 218 (1977).
5. P.K. Chatterjee, G.W. Taylor, A.F. Tasch, Jr., and H-S Fu, *IEEE Transactions on Electron Devices*, ED-26, 564 (1979).
6. A.N. Broers, *IEEE Transactions on Electron Devices*, ED-28, 1268 (1981).

Silicon Materials Technology

William C. O'Mara

*Aeolus Laboratory
Palo Alto, California*

1. INTRODUCTION

This is the silicon age. Just as previous historical periods were named by the characteristic material, we may assume that silicon will be seen as paramount in importance to the era to which some refer as the second industrial revolution. Certainly the transformation in the way people both work and relax is being changed significantly by electronic devices such as computers, control systems, and audio and video products. These electronic devices are all based on silicon, especially as used for integrated circuits.

Although a large amount of technical literature exists on silicon devices, comparatively little has been written on the material itself. This is especially true of material of an introductory nature. This chapter attempts a survey of the way silicon is made, and includes information on material properties, especially as modified by the presence of small amounts of oxygen. Silicon turns out to be a fascinating substance, and readers of this view are invited to turn to the references for further information.

2. SILICON CRYSTAL GROWTH

Pure silicon crystallizes from the melt in an open network of atoms termed the diamond structure. It occupies the Group IV position below carbon in the periodic table and has the same arrangement of tetrahedral bonds found in the crystalline form of that element. Fortunately, synthetic silicon crystals are much easier to prepare than those of carbon. They also form the basis for the \$30 billion, in 1984, semiconductor-device industry.

The method for single-crystal silicon growth was invented by Teal and Beuhler in 1951.¹ It was an extension of earlier work by Teal and Little in which single crystals of germanium were prepared for transistor manufacturing. Germanium, like silicon, occupies a Group IV position in the periodic table, and has semiconductor properties that facilitated the initial device manufacturing. However, the superior properties of silicon were soon realized and the method was extended to this element. Pure material was melted in a quartz crucible in an inert ambient, and a seed crystal was lowered to begin controlled freezing of the melt. Dopants were added as needed to control the electrical properties of the material. Reference 1 describes this early work in detail.

Production of silicon ingots today follows this same method, although extensive improvements have been made in equipment, starting material and process control. This section describes current practice, with emphasis on material perfection and controlled impurity incorporation. Some people refer to this method as Czochralski silicon growth, after an earlier experimental method. However, this method was not designed for, nor did it produce, single-crystal material.² The ability of Teal and Little to make single crystals repeatably was crucial to the growth of the solid-state-device industry.

Other processes have been developed for silicon-crystal growth. The most important of these is the floating-zone method, in which the crystal is solidified from a small molten zone resting on the crystal itself. A polycrystalline feed rod is lowered from above into an RF induction coil which melts its lower end. The rate of lowering is matched to the rate of withdrawal of the crystal from below in order to maintain a constant melt volume. This method differs from that of Teal and Little in that no crucible is employed; the melt is suspended and maintained by surface tension. Because the melt is not in contact with quartz, no oxygen is incorporated into the silicon. Some devices, such as high-voltage, high-power transistors, rectifiers and thyristors require oxygen-free starting material. The majority of devices, however, including virtually all integrated circuits, benefit from the presence of dissolved oxygen and therefore require silicon grown from a quartz crucible.

Several other processes have been investigated for the manufacture of silicon devices.³ These include sheet-growth methods using dendrites or fast-growing silicon-crystal forms, growth from dies or free-form crystallization. Casting has also been employed to prepare ingots of large-grain polycrystalline material. One product of these novel growth methods is infrared window blanks for various applications. The main thrust of work in this area, however, has been to prepare low-cost starting material for photovoltaic applications. Currently, casting of polycrystalline ingots is the most common way to prepare photovoltaic substrates.

2.1 Crystal Growth Equipment and Process

An example of the equipment used for silicon-crystal growth is shown in Figure 1, and is represented schematically in Figure 2. The photograph in Figure 1 shows a large water-cooled chamber which contains the

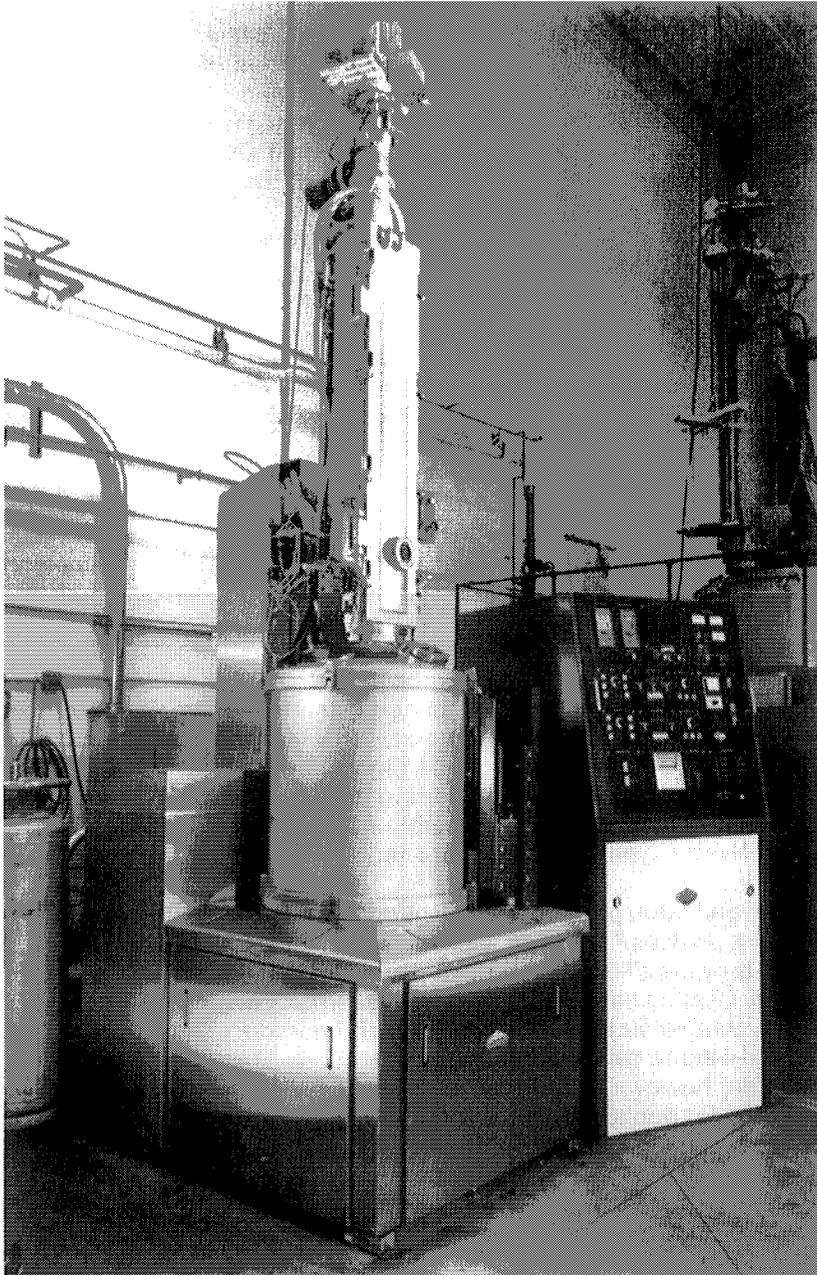


Figure 1: Photograph of a modern silicon-crystal puller (courtesy Kayex-Hamco).

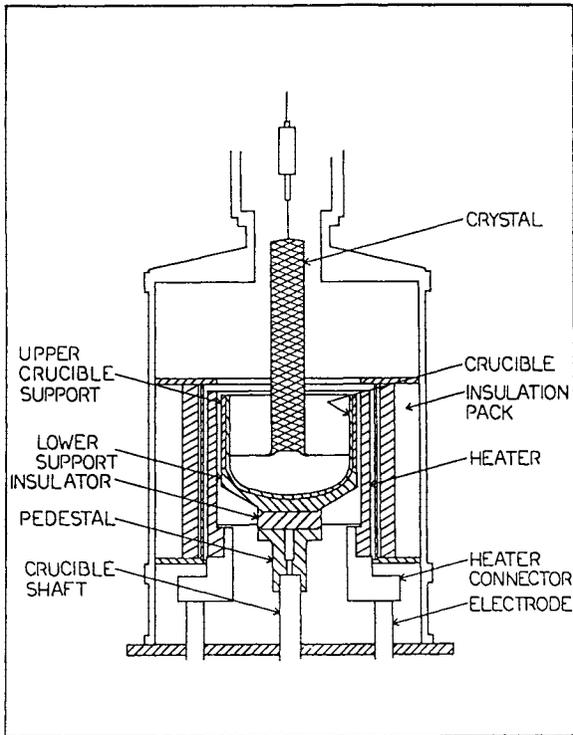


Figure 2: Schematic cross-section of crystal puller.

heater, susceptor and molten-silicon charge. Mounted above the "hot zone" chamber is the chamber into which the ingot is pulled. It is raised by means of a chain mechanism which lifts the seed to which the ingot is attached. The chain mechanism sits at the top of the crystal-growth apparatus. To the right of the growth chamber, an electronic console provides controls for heater power, seed lift, crucible and seed rotation, as well as means for recording these for reference.

Modern crystal pullers produce cylindrical silicon ingots ranging from 100 mm to 150 mm in diameter, from polysilicon charges ranging from 20 kgm to 40 kgm, with an ingot length of approximately one meter. Purified polycrystalline starting material is loaded into a quartz crucible of 12"-14" in diameter, along with small amounts of dopant needed to give the desired electrical properties to the finished material. Because quartz is quite soft at the melting point of silicon, 1425°C, the crucible is supported by a graphite susceptor, which is mounted on a graphite support which can be rotated. This rotation helps to minimize the effects of temperature fluctuations which arise from non-uniformities in heater resistance. The heater itself is graphite formed into a cylinder and machined into a "picket fence" which surrounds the susceptor. This forms a continuous resistive path for current which heats it to greater than 1500°C.

In order to accomplish single-crystal growth, three things are needed. First, a seed crystal must be used in order that the desired atomic arrangement will be achieved. The seed crystal is suspended from the chain and lowered into the melt. It is typical to rotate the seed in a counter sense to the crucible rotation. This promotes a homogeneous solid-liquid interface, needed for microscopic uniformity of dopant distribution. The second requirement for crystal growth is to locate the melt surface with respect to the heater so that the proper temperature gradient is achieved along the growing ingot. This is done by raising or lowering the susceptor support. Initially this is somewhat of a trial-and-error procedure, but once the start position is established it remains constant for a given puller. During ingot growth, the crucible and susceptor are raised to maintain a more or less constant melt position with respect to the heater as the melt volume decreases. The third requirement for crystal growth is that the central portion of the melt be cooler than the outer portion, so that freezing can occur locally while the melt remains in a liquid state. Because heating occurs from the outside, this is automatically accomplished. However, very precise control of heater current is required to bring the temperature at the center to just the freezing point and not lower.

Molten silicon must be contained in a non-oxidizing atmosphere so that SiO_2 formation is avoided. Nitrogen cannot be used because of silicon nitride formation. Argon is most commonly used as the gaseous ambient, although helium can be employed. Hydrogen, used originally by Teal and Little, significantly modifies the properties of the material. As the molten silicon continuously erodes the quartz crucible, silicon monoxide vapor is evolved from the melt surface. Because of this continual vaporization, crystal yields can be improved by reducing the pressure over the melt surface. The reduced pressure, on the order of 30 torr, allows SiO to be swept away from the furnace into a suitable trap where it cannot interfere with crystal growth.

Automatic power supplies bring the heater temperature to a value sufficient to melt the silicon, 1425°C . The seed is lowered into contact with the melt, and the melt temperature is reduced slightly so that freezing can begin onto the seed. Freezing proceeds laterally from the 5-10 mm diameter seed until the final ingot diameter of 100-150 mm diameter is reached. At this point the seed lift is engaged, and further freezing adds to the ingot height, but does not increase the diameter. Seed-lift rates of 50-100 mm per hour are common for these ingot diameters. The upper portion of the crystal-growth chamber contains an infrared sensor and lens which monitors the bright edge of the solid-liquid interface during growth. Any deviations in the diameter of the crystal are translated into changes in seed lift rate. These changes are automatically made in a way that brings the dimensions of the crystal back within prescribed limits. If the diameter increases beyond the control limit, seed lift is increased to reduce it and vice versa for a diameter decrease. Crystal growth proceeds over the course of several hours until 80-90% of the melt has solidified. At this point the ingot diameter is reduced by raising the temperature. Diameter reduction continues until the crystal tail resembles an inverted cone with a sharp point. This practice prevents the thermal shock of furnace shutdown from

introducing dislocations into the lower end or tang of the crystal. These dislocations, if introduced, could propagate upward and destroy crystal perfection in much of the ingot.

An important improvement in crystal growth was made in the late 1950s by Dash,⁴ which allowed the production of ingots free of dislocations, termed zero-D growth. The process, represented schematically in Figure 3, involves special growth conditions during the initial seeding process. The seed is a single crystal of silicon, usually oriented along a $\langle 100 \rangle$ or $\langle 111 \rangle$ direction. Although it is a single crystal, in general it will contain dislocations or extended disruptions of the lattice. As material is added to the seed by freezing, the dislocations will propagate. By reducing the seed diameter to 5 mm or half the initial diameter, and making use of the fact that dislocations virtually always make at least a small angle with respect to the vertical axis, the seed can be grown to the point at which all dislocations have reached its surface. Once a dislocation is at the surface, it is “pinned”, and substantial energy is required to initiate a new one. Subsequent growth of the crystal is routinely maintained in the dislocation-free condition, and all silicon substrates are supplied in this state. The method of Dash was crucial to the production of ingots of three-inch diameter and larger, avoiding the tendency of large dislocated crystals to become polycrystalline. Figure 4 shows the stages of crystal growth in a production puller, while Figure 5 shows a completed ingot ready for further processing.

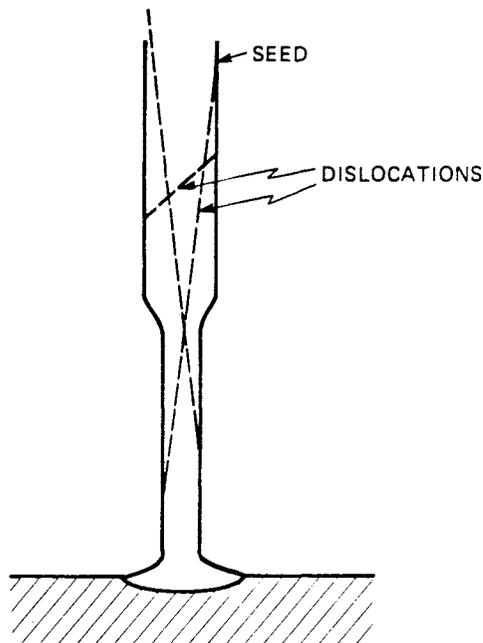


Figure 3: Method of Dash for dislocation-free crystal growth. Any dislocations in the seed are allowed to grow to the surface and are pinned.⁴

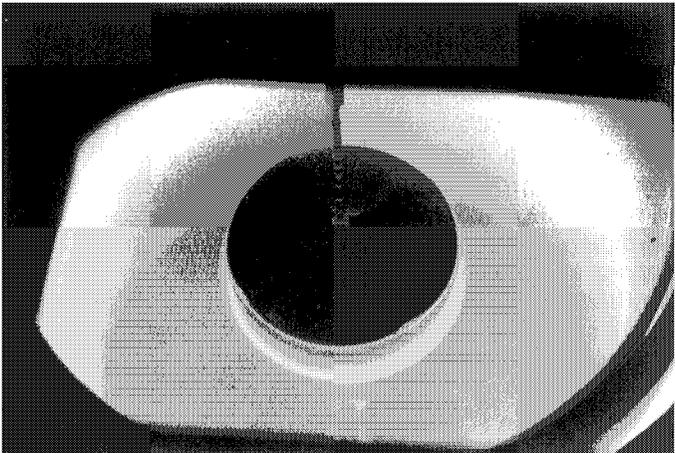


Figure 4: View of silicon crystal being solidified from the melt.

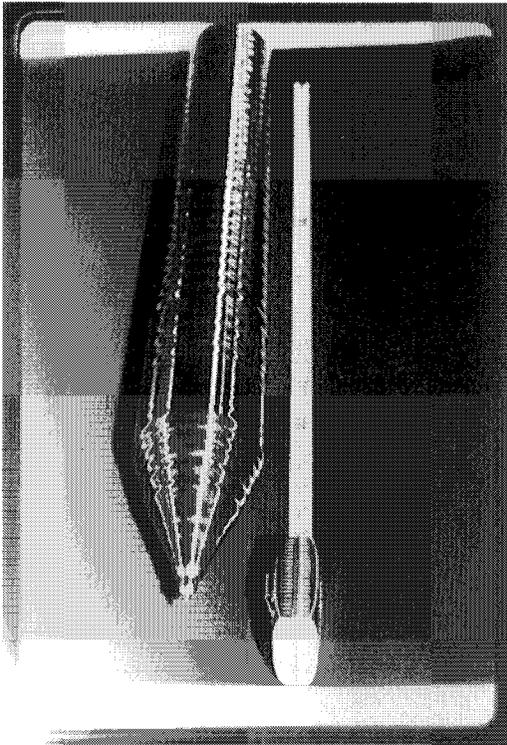


Figure 5: Photograph of finished ingot.

2.2 Dopant and Impurity Incorporation

Silicon is a good electrical insulator at room temperature in the pure state. Dopants are intentionally added to lower the resistivity to values that approach metallic conductivity. Silicon also possesses the property of conducting electricity by free electrons, as in metals, or by “holes”—the absence of a valence electron. The hole represents a well-defined conductor because the crystal lattice is essentially perfect throughout the specimen.

Excess holes or electrons for the appropriate conductivity can be introduced into the lattice by adding specific dopants or impurities during crystal growth. Elements of Group III of the periodic table cause silicon to be p-type, or positively conducting via a hole mechanism. Group VI elements add free electrons to the material and result in negatively conducting, or n-type silicon. Boron is used for p-type doping, while phosphorus, arsenic and antimony can be used for n-type silicon. The incorporation of these elements into the melt is a function primarily of the segregation coefficient, a number unique to the material (silicon) and the impurity or dopant in question.

When a relatively pure material freezes, any impurity is preferentially rejected. The amount rejected is expressed by the segregation coefficient,

$$\begin{aligned} \text{where } k &= \text{segregation coefficient} = C_s/C_l \\ C_s &= \text{concentration of impurity in the solid} \\ C_l &= \text{concentration of impurity in the liquid.} \end{aligned}$$

The segregation coefficient, k , is less than one for virtually all impurities in silicon except for oxygen. Some common elements and their segregation coefficients are listed in Table 1. Because the impurity is rejected by the freezing solid, the concentration in the liquid grows as the ingot is withdrawn. This is expressed by the following relation:

$$C_s = C_o k(1 - g)^{k-1}$$

$$\begin{aligned} \text{where } C_s &= \text{concentration of impurity in the solid being frozen} \\ C_o &= \text{initial impurity concentration in the liquid} \\ k &= \text{segregation coefficient} \\ g &= \text{fraction of melt solidified.} \end{aligned}$$

This relation is shown graphically for a number of different values of the segregation coefficient in Figure 6. The implication is that the seed end of the ingot is less heavily doped with impurity than the tang. Because of this a resistivity variation will occur along the length of the ingot. For boron-doped ingots this variation is a factor of two, while for phosphorus it is a factor of three.

Figure 7 shows the resistivity variation from seed to tail of a boron-doped ingot. The first part of the ingot contains relatively little boron, so the resistivity is high. As the melt freezes, the boron concentration builds in the melt; none is lost by evaporation. Subsequent portions of the ingot contain ever-increasing amounts of boron, so the resistivity decreases smoothly from seed to tang. Figure 8 shows the axial resistivity variation of a

Table 1: Segregation Coefficients and Solid Solubilities of Some Elements in Silicon

<u>Element</u>	<u>k_0</u>	<u>Solubility (atoms/cm³)</u>	
<u>Electrical Dopants</u>			
Boron	0.8	6×10^{20}	p-type
Phosphorus	0.35	1.5×10^{21}	n-type
Arsenic	0.35	1.9×10^{21}	"
Antimony	0.023	6.8×10^{19}	"
<u>Ubiquitous Impurities</u>			
Oxygen	1.25	1.2×10^{18}	
Carbon	0.07	5×10^{17}	
Nitrogen	7×10^{-4}	4.5×10^{15}	
<u>Metals</u>			
Iron	8×10^{-6}	2×10^{16}	
Nickel	2.7×10^{-6}	8×10^{17}	
Copper	4×10^{-4}	1.1×10^{18}	
Gold	6×10^{-6}	1×10^{17}	
Aluminum	2×10^{-3}	2×10^{18}	p-type

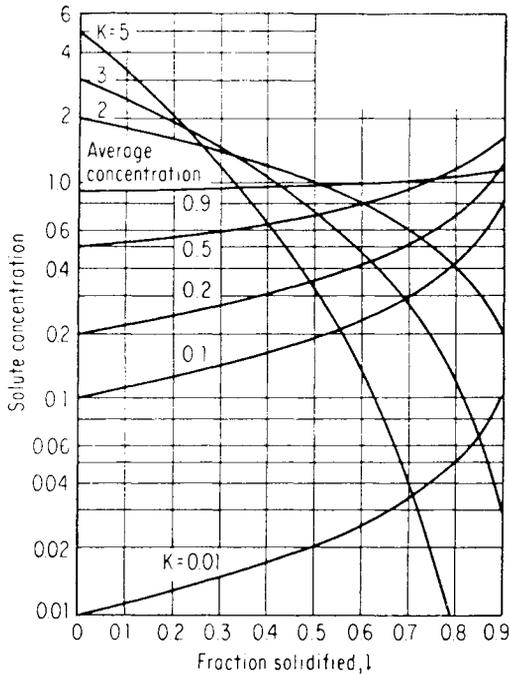


Figure 6: Incorporation of impurities as a function of melt fraction solidified. Curves are shown for various values of the segregation coefficient.⁵

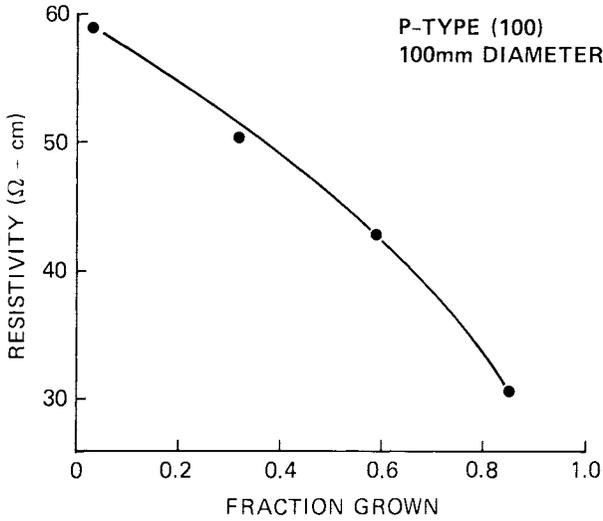


Figure 7: Plot of resistivity of boron-doped silicon as a function of distance from seed end. Boron segregation during growth results in a decrease in resistivity from seed to tang end of the crystal.

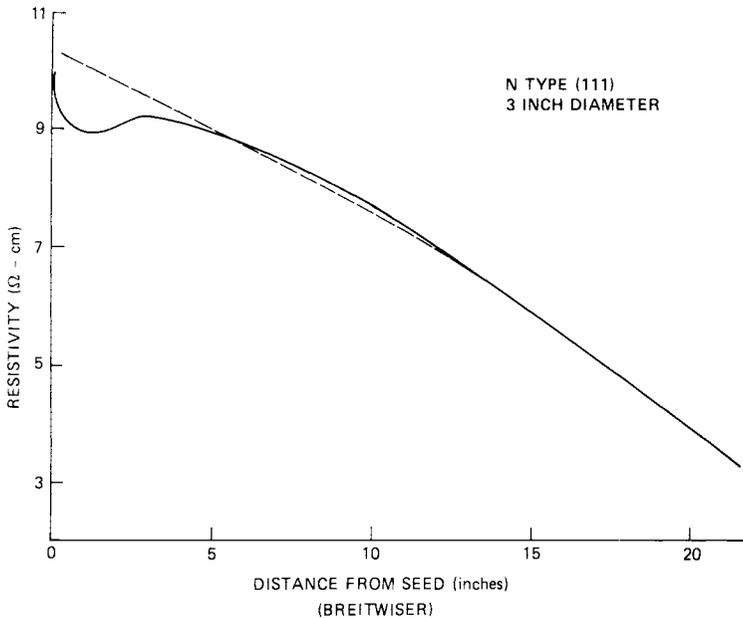


Figure 8: Plot of resistivity of phosphorus-doped silicon as a function of distance from seed end. Resistivity dip near seed end is a result of fast growth rate here and an effective segregation coefficient greater than the equilibrium value.⁶

phosphorus-doped ingot.⁶ The lower segregation coefficient of 0.35, compared to 0.8 for boron, leads to a steeper resistivity gradient along the ingot. In addition, this particular ingot shows a resistivity variation near the seed end which does not follow the impurity incorporation expression for normal freezing. After the ingot has reached the desired diameter, the initial portion of the ingot is often withdrawn at a greater rate than the lower portion. This fast freezing causes impurities, especially those with low segregation coefficients, to be incorporated in a non-equilibrium fashion. The deviation is always towards an “effective” segregation coefficient which is larger than the equilibrium value, which explains the high phosphorus concentration (low resistivity) in the seed end of the ingot of Figure 8.

In addition, local fluctuations of dopant concentration can occur because of changes in growth rate or temperature. These can cause a sudden increase in dopant level as the process deviates from equilibrium. As a result, the uniformity of resistivity across a silicon wafer may vary by $\pm 20\%$, for the case of phosphorus. This variation is shown on a local scale by means of spreading resistance measurements in Figure 9. Local variations of boron are typically less than $\pm 10\%$.

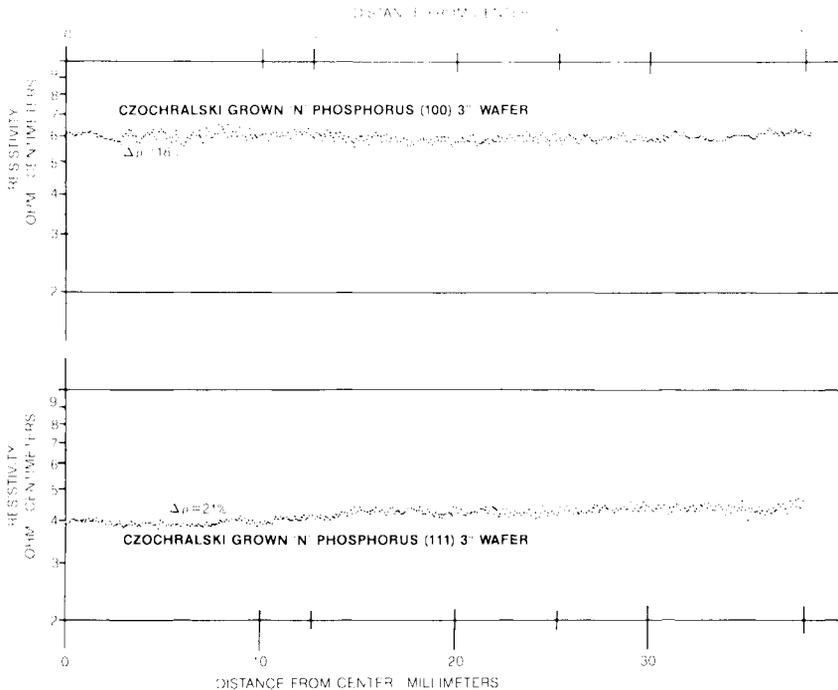


Figure 9: Fluctuations in resistivity from center to edge of a phosphorus-doped wafer. Instantaneous growth-rate fluctuations also change the effective segregation coefficient.

2.3 Incorporation of Oxygen

In addition to intentionally added dopants, oxygen is incorporated into silicon crystals due to the dissolution of the quartz crucible. Any material in contact with molten silicon will dissolve, with vitreous silica dissolving at one of the lowest rates. It is the only material suitable to contain the melt, and provides a source of oxygen to the melt throughout the growth process. The rate of dissolution of the crucible is a function of the temperature at the crucible wall, and the stirring currents which sweep the oxygen into the interior. An added factor is the escape of oxygen from the melt surface in the form of SiO. This latter factor is a constant while the oxygen source diminishes throughout the growth process. The source of oxygen is a function of the surface area of the crucible wetted by the melt which decreases throughout the growth process.

This means that the oxygen incorporated into the crystal at the seed end is the maximum amount possible for a given set of growth conditions. Given fixed seed and crucible rotations, the oxygen level will decrease from seed to tang end of the crystal. This is shown in Figure 10 for a 3-inch-diameter ingot. Modeling of this behavior has been presented by Carlberg et al.⁷ Because SiO is escaping from the edge of the solid-liquid interface, the oxygen level is lower at the edge than the center. This fall-off in oxygen level is shown in Figure 11. This data was obtained from the same 3" diameter ingot as Figure 10.

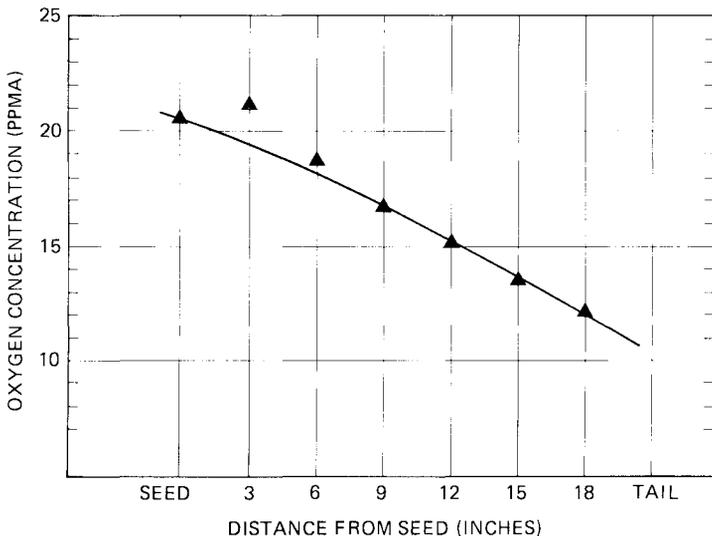


Figure 10: Axial gradient of oxygen in a silicon crystal. Oxygen measured according to ASTM F121-80.

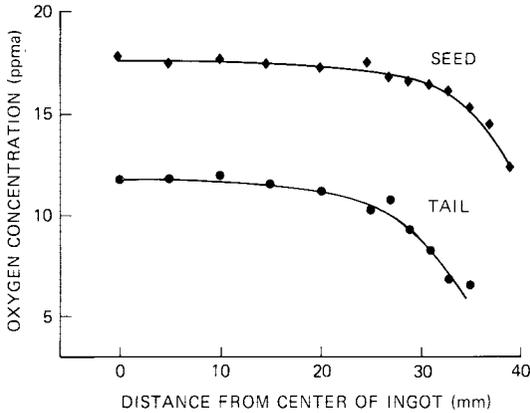


Figure 11: Radial variation in oxygen level.

Ingot manufacturers have attempted to develop ways to control the level of oxygen in crystals. This is because the oxygen level affects wafer performance in several ways, some beneficial, others harmful to device yield and performance. Oxygen at some level is essential for the majority of devices, and is the reason why float-zone silicon, which is oxygen-free, cannot be used in the vast majority of device applications. In the standard crystal-growth method of Teal and Little, the control of oxygen level is primarily obtained by the variation of seed and crucible rotations.⁸ In this way, the stirring currents that sweep oxygen into the melt and past the solid-liquid interface are controlled. The limits on this control are due to the fact that seed and crucible rotations also control the uniformity of dopant incorporation for electrical resistivity. This limits the oxygen levels to a range of 10-20 ppmA for standard crystal-growth processes.

The direct determination of the segregation coefficient of oxygen in silicon by Yatsurugi et al. indicated a value of $k = 1.25$.⁹ This value has been the subject of recent dispute, with indirect determinations or estimates of $k = 0.3$ to $k = 1.4$.¹⁰ A value of k greater than one suggests two distinct oxygen species, since the "interstitial" oxygen incorporation was shown by Yatsurugi to have $k = 1.0$. The predominant form of oxygen in silicon is this interstitial species, which actually consists of oxygen lying between two near-neighbor silicon atoms, forming a nearly linear Si-O-Si structure in place of the Si-Si bond. It is not a true interstitial species. If the segregation coefficient of this species is 1.0, and the overall segregation coefficient is 1.25, then a second form of the atom in the lattice is indicated. The original suggestion for the form of this second species was small oxygen clusters⁹. I have suggested that the second oxygen species exists in the form of a fully substitutional atom, which is incorporated into the lattice with a segregation coefficient of 0.25.¹¹ This idea can help to explain many of the puzzling features of oxygen in silicon, such as the existence of oxygen striations in silicon,¹² and the striated distribution of the oxygen donor in the material.¹³ Effects related to precipitation of oxygen from solution are discussed in a later section.

2.4 Incorporation of Carbon

Carbon has a low segregation coefficient of 0.07, which means it is strongly rejected by the freezing silicon. It is never intentionally added as a dopant, but may be present from one of three sources. The polysilicon charge itself may contain carbon, but this is unusual. At times, the silicon charge may contain previously melted silicon, or "remelt." This remelted silicon is enriched in carbon from the previous segregation. Finally, the melt may entrain carbon from CO gas passing over it. The source of this gas is a water or oxygen leak in the furnace which reacts with hot graphite.

The three situations are depicted in Figure 12. The maximum level of carbon is due to remelt in the original charge. The next-lower level of carbon is due to standard polysilicon in a furnace operated at one atmosphere pressure. When the overpressure is reduced to 30 torr, the CO entrained in the melt is much reduced, and the lower curve of carbon incorporation is obtained. As will be shown later, high levels of carbon are uniformly detrimental to material and device performance.

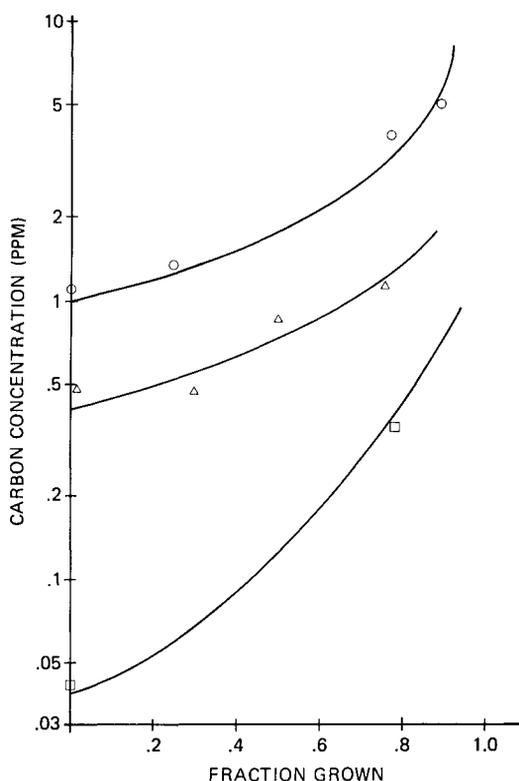


Figure 12: Carbon concentration versus fraction of melt solidified for three conditions of crystal growth: circles represent the maximum level of carbon due to remelt in the original charge, triangles represent crystals grown with standard polysilicon at atmospheric pressure and squares present data for material grown with standard polysilicon but a reduced pressure of 30 torr.

3.0 WAFER PREPARATION

3.1 Mechanical Shaping Procedures

As-grown silicon ingots are subjected to a number of mechanical and chemical operations to prepare slices or wafers ready for device manufacture. The mechanical steps begin with grinding of the ingot to make it perfectly cylindrical, followed by the grinding of one or more flats along its length. The flats define specific crystal planes in the material such as the (100) or (110) planes, and serve to identify wafer orientation and type. The flattening procedure is guided by X-ray orientation of the ingot. Diamond-tipped grinding tools similar to those in many machine shops are used for this procedure. Grinding and flattening operations, as well as the subsequent mechanical operations of slicing, lapping and edge rounding, introduce subsurface work damage into the material. Great care is taken to avoid cracks and fractures, and the remaining damage is subsequently removed by etching.

Silicon ingots are produced with flats for two purposes. Originally, finished die on a wafer were separated for packaging by scribing along the directions of crystal planes and breaking the wafer into squares along directions of easy cleavage. The orientation flat at one side of the wafer served to guide the alignment for scribing. Although die separation is now more commonly done by means of a diamond saw, the crystal planes are still chosen for sawing. In addition, automatic handling equipment makes use of the flat for wafer orientation in photolithographic mask alignment and other operations.

An additional function of wafer flats is the use of a second, minor flat in combination with the primary (110) flat to identify wafer type and orientation. This function has been standardized by the Semiconductor Equipment and Materials Institute (SEMI), a trade organization of suppliers to the IC industry. Figure 13 shows schematically the four combinations of flats that identify p- and n-type, and (111) and (100) orientations. This allows visual verification of these parameters and helps to avoid mixing of wafer lots.

After grinding, the ingot is mounted on a graphite beam for slicing. Again, X-ray orientation serves to ensure that cuts are made on the correct crystal orientation. Most (100) oriented ingots are cut parallel to the (100) plane. (111) oriented crystals can either be sliced on orientation, or 3-5° off orientation towards (110). The latter choice is made when an epitaxial silicon layer will be grown on the wafer in order to improve the growth kinetics of the layer. Figure 14 shows a schematic diagram of the slicing process. The mounted crystal is held near a thin stainless-steel blade which is coated with diamond grit on its inner surface. The blade resembles a drum head with a central hole, and is rotated in a spindle at several thousand RPM. The blade is lowered at a rate of 1-3 inches per minute through the silicon, cutting a wafer or disk from the ingot. Indexing equipment allows the setting of wafer thickness. If the blade does not cut true, either bow or taper can be introduced into the slice as shown in the figure. These undesirable deviations from flatness can be detected by an eddy current sensor, shown mounted above the blade near the crystal. The

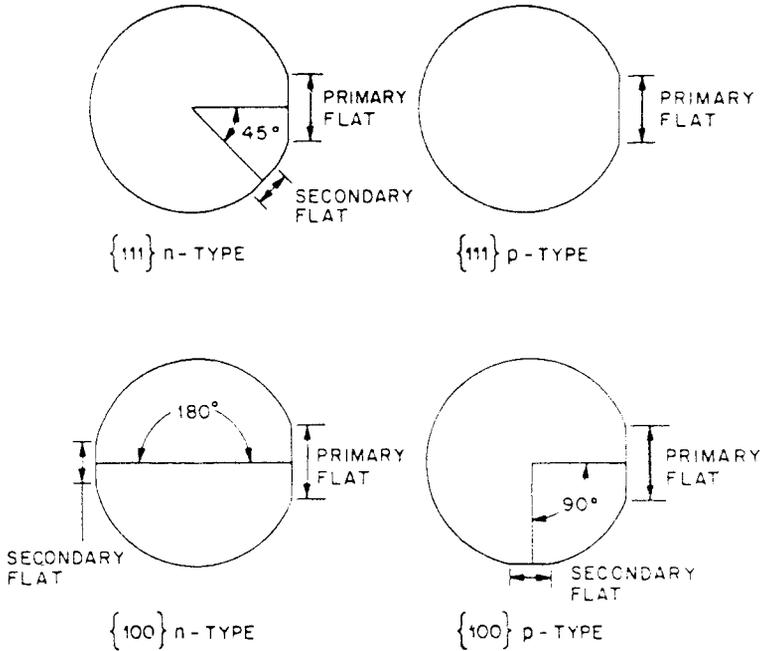


Figure 13: Semiconductor Equipment and Materials Institute (SEMI) standard flat locations for silicon wafers.

sensor indicates blade deviation which can be corrected by appropriate dressing of the grit.

Sliced wafers receive two additional mechanical operations. The wafer edge is rounded or contoured to reduce edge chipping in subsequent use. In addition, the front and rear surfaces are made flat and parallel by planetary lapping. In this operation, a SiC or Al_2O_3 slurry is used to remove a small amount of silicon while the wafers are rotated between two steel plates. This operation leaves wafers flat to as little as $1 \mu\text{m}$ deviation of flatness over the entire wafer diameter.

3.2 Wafer Etching

Mechanical machining to produce a wafer of the desired thickness and flatness produces a layer or skin on the wafer which is work damaged, containing numerous dislocations. The damaged layer is removed by a chemical etching step, which may also reduce work-induced stress and remove contaminants introduced into the material. Etching can be done using a mixture of hydrofluoric, nitric and acetic acids, or by using a caustic KOH bath. The former technique produces a smooth, featureless surface, whereas the latter etch leaves the surface with microscopic pits and a specular appearance. KOH etching is finding favor because of its ease of handling and disposal.

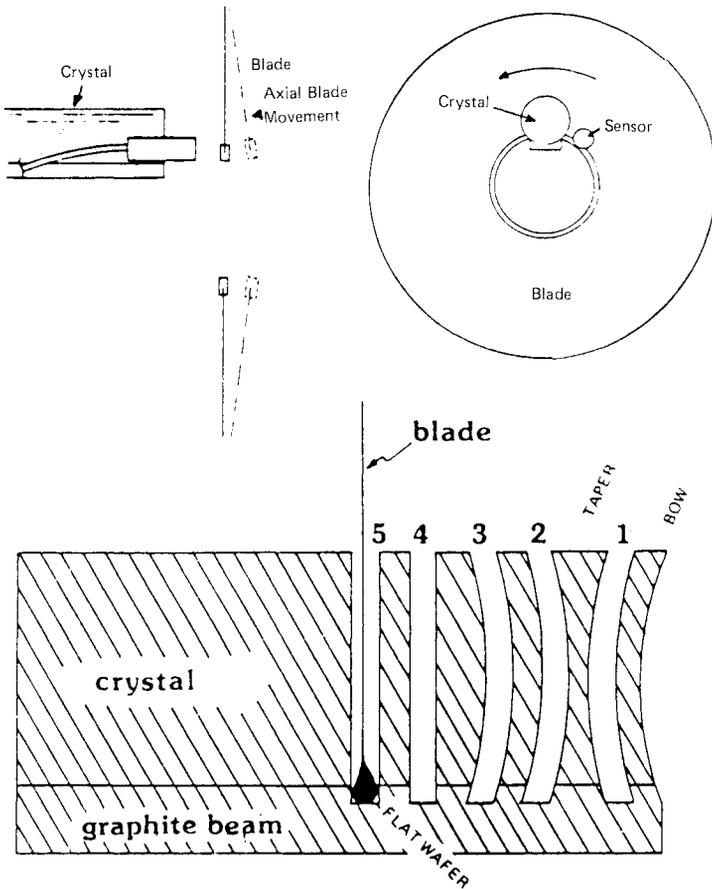


Figure 14: Schematic of ingot-slicing operation. Blade mounting and conditioning are done carefully to allow for straight cuts. Blade deviation can result in either bow (cuts 1 and 3) or taper (cuts 1 and 2) in the slice.

Chemical etching is sometimes followed by a re-introduction of damage in a controlled fashion on the backside of the wafer. Damage is introduced by sandblasting with fine quartz spheres, or by abrasion with sandpaper. A shallow network of dislocations is introduced into the wafer surface. Some users find this damaged layer helpful in preventing haze formation on the front side of the wafer during initial oxidation. As a result, wafers treated in this way are supplied at customer request. A recent alternative approach is to deposit a layer of polysilicon on the wafer backside, which functions in a similar way to abrasion.

A new operation is being introduced into wafer manufacturing, often after the etching step, in which a laser is used to mark the wafer surface with an identification code. This code contains pertinent information about

the wafer characteristics, and can be read visually or with an automated reader. The code number is generally placed on the front surface near the major flat.

3.3 Polishing

Front-surface polishing of the wafer leaves it with a mirror finish needed for device fabrication. Wafers are mounted on carriers for this operation, held either by a thin wax layer or by a friction bond to the carrier. These carriers are pressed onto a rotating polishing pad of polymeric material while a polishing slurry is applied. Generally the process consists of two steps on two different polishing machines. In the first step, perhaps 0.001" or one mil of silicon is removed from the surface in a process termed stock removal. The carrier is then moved to a machine with a smoother pad surface for final, mirror-finish polishing.

The polishing slurry consists of a solution of colloidal silica maintained at a pH of 11 for stock removal and a pH of 9 for finish polish. The function of the silica in the polishing process is not understood. It is probable that the product of the polishing process is silica itself. That is, the surface of the wafer is oxidized and hydrated to form a silica that can be wiped away by the pad. It is curious that one must add the reaction product to initiate the polishing process. The silica acts as some sort of catalyst, and the primary polishing agent is probably water!¹⁵

3.4 Cleaning

The final step in wafer preparation is a careful cleaning of the surface. A sequence of acids and bases is used to remove any contaminants, including wax residues, if any, and metallic contaminants in the polishing medium. The cleaning process is assisted by adding hydrogen peroxide to oxidize these materials, and by heating the liquid baths. The basis for chemical cleaning has been established by Kern.¹⁶

The final bath in chemical cleaning process is usually chosen so that the wafer surface is hydrophilic, or water-loving. This means that the water wets the surface as the wafer is withdrawn from the bath, and can be removed uniformly. A combination of ammonium hydroxide and hydrogen peroxide is an example of such a bath. On the other hand, a wafer withdrawn from a hydrofluoric acid bath will be hydrophobic, and water will bead on its surface. This beading can lead to spots on the surface after the water dries. The difference between a hydrophilic and hydrophobic surface consists of the difference in the native oxide thickness. Silicon will instantly grow a thin oxide upon emerging from a chemical bath. In the hydrophobic case, the oxide thickness is 10Å or so, while for the hydrophilic case the oxide thickness can range from 15-50Å, depending on bath characteristics. This can lead to variations in thickness of a subsequent thermal oxide layer. This variation is important for thin gate-oxide-layer growth in MOS device fabrication.

Throughout the fabrication process, wafers are inspected for a variety of parameters including physical dimensions, electrical resistivity, flatness and surface perfection. After final inspection, most wafers are packaged in

Class 100 clean rooms in such a fashion that the cleanliness will be maintained until use by the customer.

4. MATERIAL PROPERTIES

4.1 Crystal Structure

Silicon crystallizes in an open, low-density structure termed the diamond lattice structure, which it shares with the gemstone modification of carbon, the single-crystal form of germanium, and one of the crystalline forms of tin. In this structure, each atom is bonded to four near neighbors arranged tetrahedrally about the central atom. The local arrangement is repeated throughout the crystal, giving rise to the unique properties of the material. Atoms are joined to each other by sp^3 hybrid covalent bonds, which are quite stable. The bond strength is 25 kcal per mole for silicon, and the near-neighbor spacing is 2.35\AA . If one looks at a silicon lattice containing several atoms, he can see that a cubic structure is formed. A diagram of this structure is shown in Figure 15. There are actually two interpenetrating cubic structures, with an atom at each cube face. The cube side has a dimension of 5.43\AA .

The regular crystal structure of covalent bonds is quite inert. Very pure silicon is a good electrical insulator with an energy of 1.11 eV required to

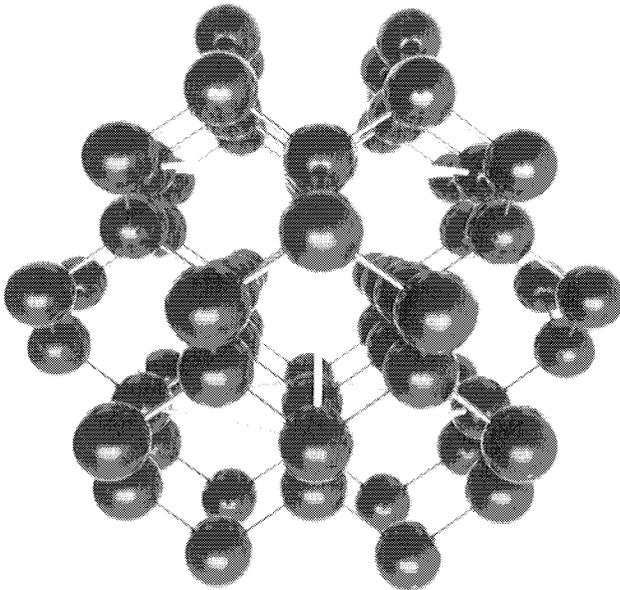


Figure 15: Silicon lattice structure.¹⁷

detach a valence electron from one of the covalent bonds. This property, as well as some other important properties of silicon, are listed in Table 2. Silicon is useful for electronic devices because the electrical properties can be modified by the addition of impurities or dopants. For example, phosphorus or other Group V atoms substitute for silicon atoms at lattice positions when added during crystal growth.¹³ However, phosphorus has an extra valence electron, which is easily removed from the vicinity of the donor atom, and contributes to electrical conduction as in a metal.

On the other hand, if a Group III element such as boron is added as an impurity, a property unique to semiconductors is manifest. Boron also substitutes for silicon, maintaining the perfect lattice structure. But boron has only three of the required four valence electrons, so that one of the valence bonds is “short” an electron. This electron deficit is termed a hole. The hole can migrate when an adjacent electron jumps into the bond lacking one electron. Under the influence of an electric field in boron-doped silicon, the electrons jump from valence bond to valence bond. The effect is as if the “hole” migrates in the opposite direction.

Because silicon can be doped both with donors (such as phosphorus) and acceptors (such as boron), two types of electrical conduction are possible in the material. Because of the perfect crystal structure, the regions of different doping can be maintained permanently separated, allowing fabrication of resistors, diodes, and transistors in different portions of the material. The existence of a native oxide allows metallic interconnects and capacitors to be made on the surface of the wafer, the sum total constituting an integrated circuit of great complexity.

4.2 Electrical Properties

Figure 16 shows the resistivity versus impurity density for n-type (negative or electronic conduction due to donors) and p-type (positive or hole conduction due to acceptors) silicon as determined by Irwin. After a donor such as phosphorus or arsenic is added to the lattice, it loses its identity, uniformly adding one carrier per impurity atom. This is true for the commonly used donors and acceptors whose carriers are thermally excited at room temperature, and thus are fully ionized. Polynomial fits to these curves have recently been published in a form useful for small-computer data reduction.^{19,20}

Although each dopant atom adds one carrier to the lattice, the electrical properties differ depending on whether the carrier concentration is high or low. This is expressed in terms of carrier mobility, which is the average velocity of excess carriers per unit electric field. When carrier concentration is low, mobility is relatively high, since the carriers “see” only a perfect silicon lattice and scattering is minimized. However, when carrier concentrations approach $10^{17}/\text{cm}^3$, the quantity of impurity atoms is great enough to perturb the electrons or holes. The dopant atoms scatter or screen the carriers, reducing the mobility. Figures 17 and 18 show this effect as a function of carrier concentration for n- and p-type silicon. It is for this reason that high-speed devices are designed using lightly doped silicon.

The velocity for carriers in silicon is shown in Figure 19 as a function of electric field. The maximum velocity is a function of the material itself, and

Table 2: Some Properties of Silicon

Atomic Weight	28.09
Atoms/cm ³	4.995 x 10 ²²
Crystal Structure	Diamond
Lattice Constant (Å)	5.43
Density (grams/cm ³)	2.33
Melting Point (°C)	1420
Density of Surface (atoms/cm ²)	
(100)	6.78 x 10 ¹⁴
(110)	9.59 x 10 ¹⁴
(111)	7.83 x 10 ¹⁴
Energy Gap (eV)	1.11
Density of States/cm ³	
Conduction Band	2.8 x 10 ¹⁹
Valence Band	1.04 x 10 ¹⁹
Intrinsic Carrier Concentration	
n _i (/cm ³)	1.45 x 10 ¹⁰
n _i ² (/cm ⁶)	2.103 x 10 ²⁰
Intrinsic Resistivity (Ω-cm)	2.3 x 10 ⁵
Dielectric Constant	11.8
Refractive Index	3.4
Hardness (Moh)	7
Elastic Constants (dynes/cm ²)	
C ₁₁	1.67 x 10 ¹²
C ₁₂	0.65 x 10 ¹²
C ₄₄	0.79 x 10 ¹²
Young's Modulus (dynes/cm ²)	
(111) direction	1.9 x 10 ¹²
Bulk Modulus (dynes/cm ²)	7.7 x 10 ¹¹
Heat of Fusion (k cal/mole)	12.1
Thermal Conductivity (cal/sec/cm/°C) at 20°C	0.3
Expansion on Freezing (volume increase)	9%
Linear thermal coefficient of expansion (/°C) at 25°C	2.33 x 10 ⁻⁶

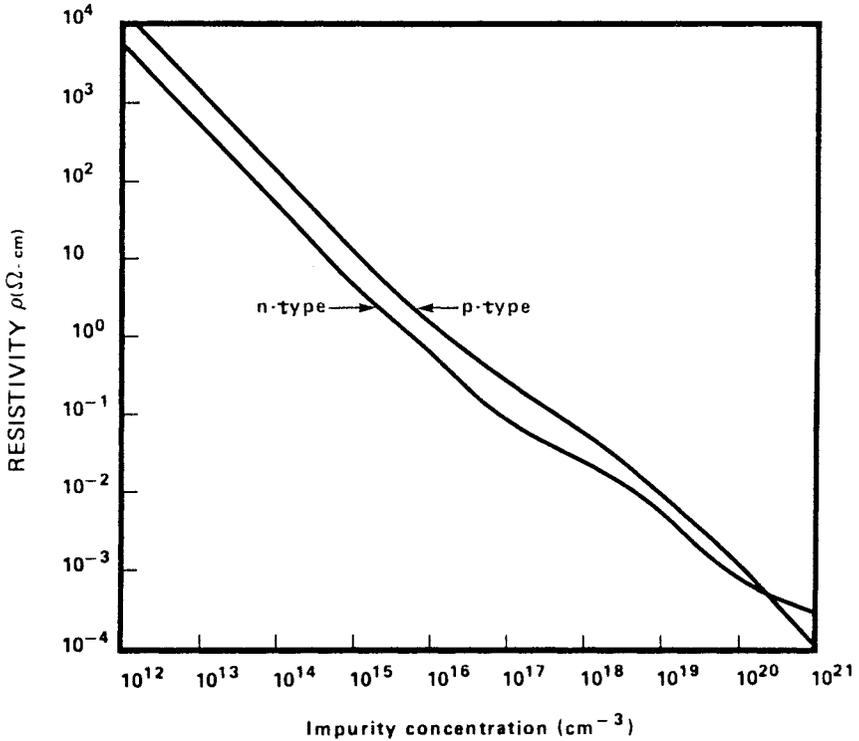


Figure 16: Resistivity of silicon as a function of n- and p-type dopant concentration, for carrier concentrations 10^{14} - $10^{21}/\text{cm}^3$.¹⁸

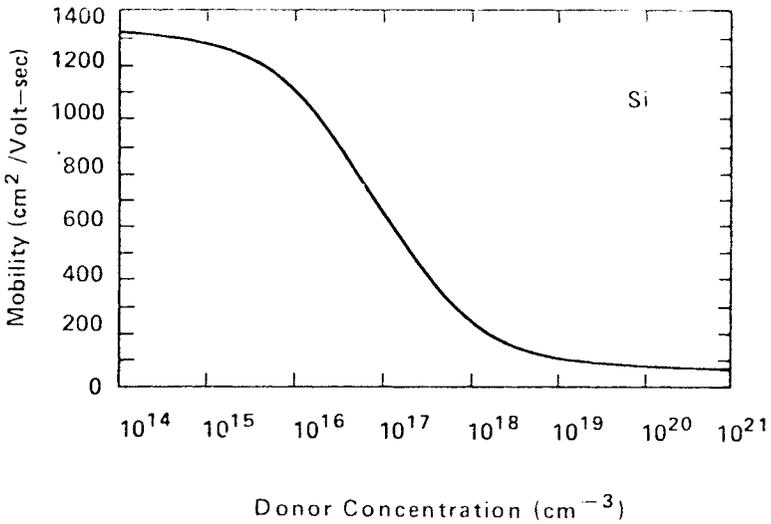


Figure 17: Electron mobility in silicon as a function of carrier concentration.²¹

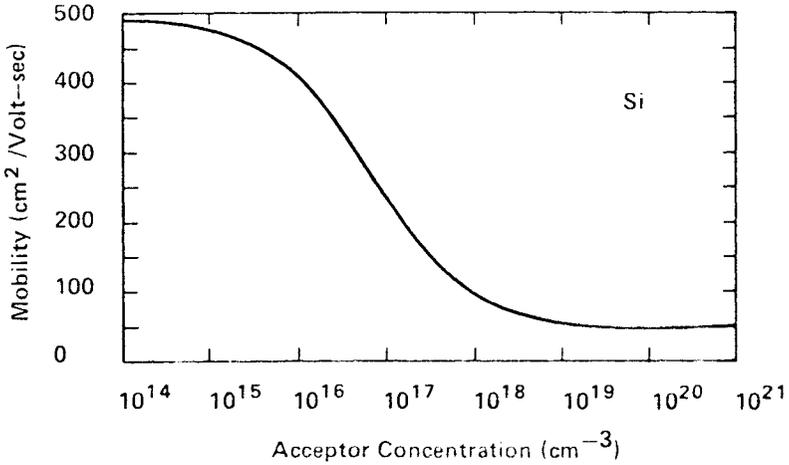


Figure 18: Hole mobility in silicon as a function of carrier concentration.²¹

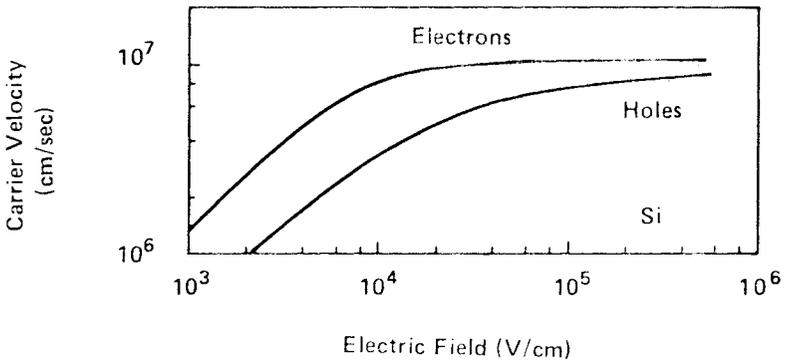


Figure 19: Carrier velocity in silicon as a function of electric-field strength.²¹

so is about the same for electrons and holes. Because of the lower hole mobility, the field at which velocity saturation is reached is greater than for electrons.

4.3 Optical Properties

At wavelengths shorter than the band gap energy of 1.11 eV, silicon is an excellent absorber of optical energy. Since this includes the visible and near-infrared portions of the spectrum, silicon can be used as a solar cell, with collection efficiencies near 20% of the impinging photons.

For the mid and far infrared portions of the optical spectrum, undoped silicon is transparent. As free carriers are added, the transmission diminishes. The absorption spectrum for n- or p-type silicon possesses a characteristic

minimum, whose wavelength changes as a function of doping. This provides a basis for resistivity measurement using optical absorption in the wavelength region of 1-100 μm . Figures 20 and 21 show plots of the plasma minimum wavelength versus carrier concentration for n- and p-type silicon. This technique for resistivity determination has recently been extended to doped polysilicon films as well as single-crystal material.²³

There are a variety of impurities in the silicon lattice which give rise to local mode absorption in the infrared. This is absorption of optical energy by the impurity nucleus, which vibrates with characteristic frequencies against the surrounding silicon lattice. The impurity vibrations lie in the infrared region of the spectrum, and can be used for quantitative analysis. This is especially important in determining oxygen and carbon levels, since they are normally electrically inactive. Other types of quantitative determination of carbon and oxygen are slow and tedious compared to infrared. Reviews and recommended procedures are available both for oxygen²⁴ and carbon analysis.²⁵

4.4 Mechanical Properties

It is only recently that the mechanical properties of silicon have been studied. Peterson²⁶ reviewed the materials properties with a view towards the manufacture of micromechanical devices. Sumino et al.²⁷ described plastic deformation. These properties are important in terms of wafer

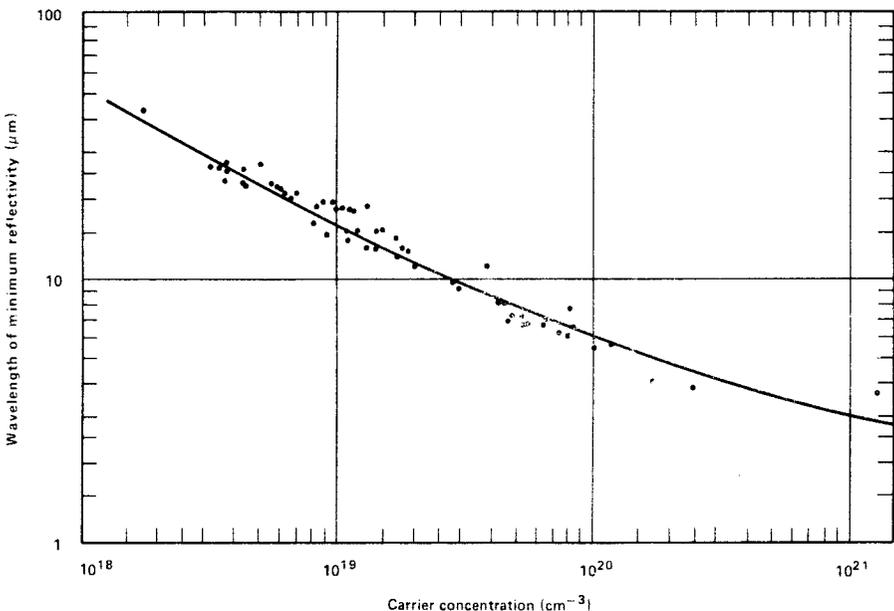


Figure 20: Wavelength of the plasma minimum for electrons in silicon as a function of carrier concentration.²²

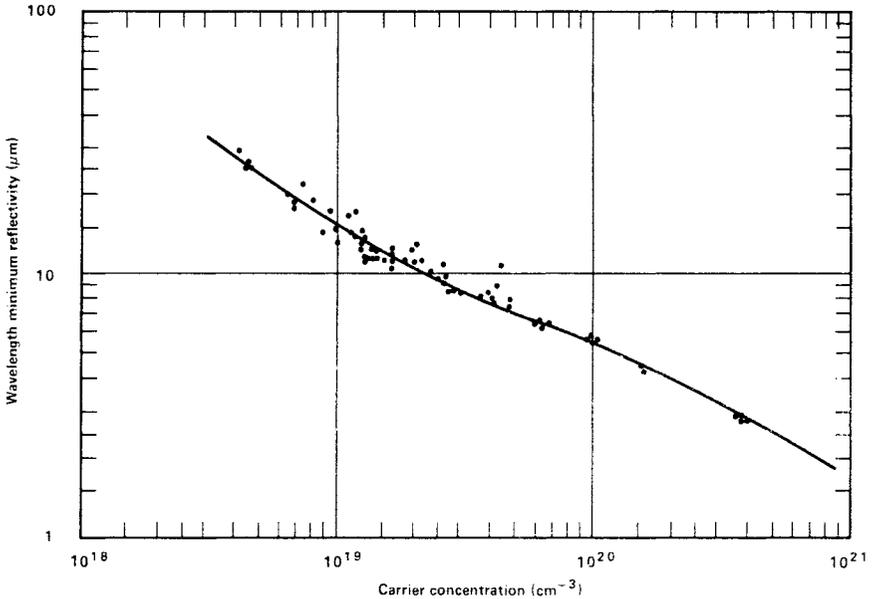


Figure 21: Wavelength of the plasma minimum for holes in silicon as a function of carrier concentration.²²

warpage induced by thermal shock in semiconductor-device processing. The presence of oxygen in the lattice plays an important role in minimizing warpage induced by thermal shock in semiconductor-device processing. Small precipitates,²⁹ can prevent the propagation of dislocations in silicon. Oxygen thus plays a role in silicon analogous to that of carbon in iron; small amounts of impurity provide mechanical strength that is much greater than for the pure material. Nitrogen can also serve to strengthen silicon, although it is much more difficult to incorporate during crystal growth.

5. PROCESS-INDUCED DEFECTS

5.1 Oxidation-Induced Stacking Faults

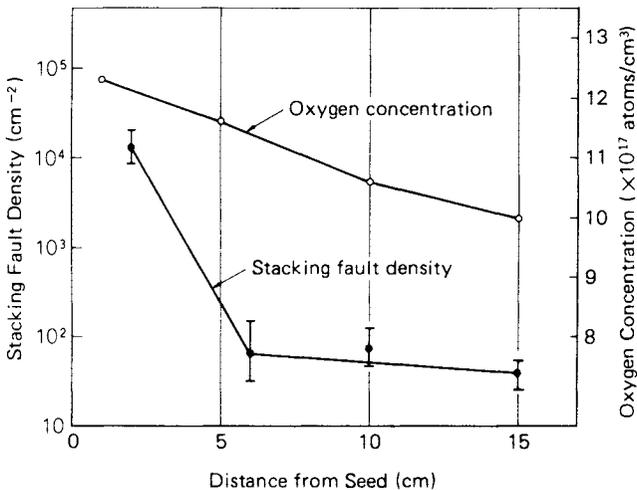
Stacking faults are excess silicon atoms that coalesce to form a “platelet” which lies between two adjacent (111) planes. Because the platelet or stacking fault fits between planes, the surrounding lattice is strained. After a certain strain is reached, the lattice will rupture, producing a dislocation loop which bounds the fault. Heat treatments such as oxidation can cause the nucleation and growth of stacking faults.³⁰ Normally they are electrically inactive, but they can serve as diffusion pipes³¹ and are generally undesirable near the wafer surface where the active-device regions are located. The defects are produced in silicon in quantities which depend on crystal-growth conditions, oxygen level, and electrical dopant

(boron or phosphorus).³² Generally, material near the seed end, where growth conditions vary the most, contains a higher level of stacking faults after a standard heat cycle. While the level of oxygen also affects the stacking-fault density, it is of secondary importance. Figure 22 shows the strong variation in stacking-fault (SF) density and oxygen level as a function of crystal position in boron-doped silicon. A phosphorus-doped ingot grown by the same experimenters showed stacking-fault levels nearly three orders of magnitude higher than for the boron-doped case. While the exact cause for this is not known, it is possibly due to fluctuations in doping levels, or striations, that are much greater in the case of phosphorus than for boron.

The bulk stacking faults discussed here are to be distinguished from surface stacking faults which arise from other causes. As first mentioned by Rozgonyi,³³ bulk and surface stacking faults can be distinguished from each other by etching and optical microscopy. The surface variety are due to impurities, scratches, abrasions and so forth. This is another indication of the importance of the cleaning process prior to each high-temperature step in order to avoid formation of surface stacking faults.

5.2 Saucer Pits

Certain defects are seen on the silicon-wafer surface after oxidation, oxide removal, and defect etching.³⁴ These are termed saucer pits or shallow pits due to their smooth, featureless appearance. Their appearance is associated with epi stacking faults, if an epitaxial layer is subsequently grown. In addition, most capacitors built on regions of high saucer-pit



(Diado et al.)

Figure 22: Stacking-fault concentration as a function of wafer position in the ingot. Faults are seen after an oxidation procedure. Oxygen level is also shown.³²

density will show undesirably low minority-carrier lifetimes.³⁵ Recent experiments have determined the presence of fast-diffusing metallic impurities as the source of these defects.³⁶ Their elimination is accomplished by a variety of “gettering” techniques, including back-side abrasion or polysilicon deposition, mentioned in a previous section. In addition, oxygen precipitation in the bulk of the wafer can constitute an “internal gettering” mechanism which also prevents the formation of shallow pit defects. This is discussed in the final section on oxygen-related effects in wafer processing.

5.3 Carbon-Related Defects

Commercially available silicon wafers generally contain carbon at levels of 5×10^{16} atoms/cm³ or less. Both producers of polysilicon and wafer manufacturers routinely inspect material to ensure low levels of this impurity. This is because carbon in concentrations above $\sim 1 \times 10^{17}$ atoms/cm³ can produce defects which degrade device performance. The section on crystal growth mentioned the incorporation of carbon as a function of the low segregation coefficient ($k = 0.07$), and its incorporation from the gas ambient in the growth furnace. Figure 12 shows carbon incorporation at various levels depending on starting material purity and growth conditions. Ordinarily, the curve corresponding to the lowest carbon levels applies to commercial material. When the carbon level approaches 7-10 ppma in the growing crystal (corresponding to a melt concentration of 100 ppma or so), silicon carbide will form at the solid-liquid interface, terminating single-crystal growth.

At levels below this, carbon is incorporated as individual substitutional atoms in the lattice. If the dissolved carbon level is greater than 2-4 ppma, subsequent heat treatment may cause it to precipitate. In one case, silicon wafers with a carbon concentration of 4 ppma (2×10^{17} atoms/cm³) were heat-treated according to an MOS process cycle.¹⁴ Various platelet-like defects were formed. A Secco etch showed characteristic teardrop-shaped etch pits, as shown in Figure 23. Transmission-electron microscopy showed

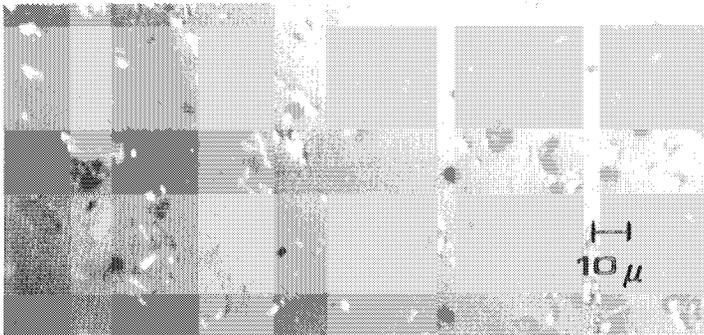


Figure 23: Optical micrograph of etch pits due to carbon-related defects in silicon. Features delineated by Secco etch on 100 surface.¹⁴

platelet formation due to precipitation from solid solution. The precipitates caused lattice strain which was relieved by dislocations in the surrounding silicon. One of the micrographs showed a platelet morphology characteristic of β -SiC, as shown in Figure 24. In addition to minimizing carbon levels so that such defects do not form during processing, gettering processes for this element have recently been investigated.³⁷

6. OXYGEN IN SILICON

6.1 Oxygen in As-Grown Silicon

The properties in oxygen in silicon are both a subject of great interest and great dispute. Experiments designed to measure a certain property often give apparently different results when done by different investigators. The problems begin with the determination of the segregation coefficient, and extend through the precipitation of oxygen from solid solution during device fabrication. The reason for conflicting results may be due to the lack of a reference condition for the silicon itself. That is, a given wafer used for an experiment may have oxygen in a different state than another one used for the same experiment, even if the oxygen level as measured by infrared absorption is the same.

6.1.1 Quantitative Analysis of Oxygen in Silicon. Oxygen was discovered to exist in crucible-grown silicon by Kaiser and Keck in 1957.³⁸ Since then, considerable effort has been expended in studying its incorporation into the lattice, since the properties of the material are so greatly influenced by the presence of small amounts of this element. The amount of oxygen incorporated during crystal growth is a strong function of the growth conditions themselves, as mentioned previously. In addition, quanti-

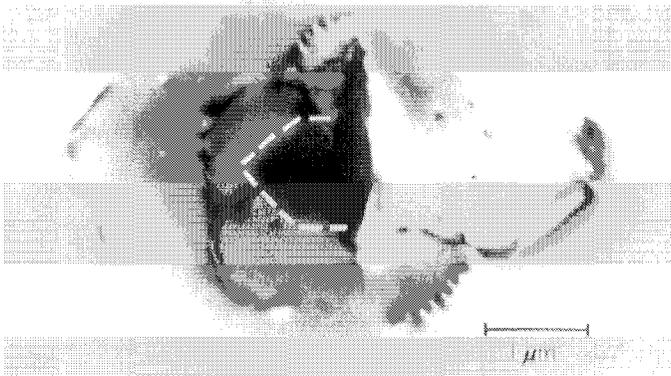


Figure 24: Transmission electron micrograph of β -Si precipitate formed by heat-treating carbon-rich silicon.¹⁴ (Photograph courtesy of J. Peng.)

tative measurement of oxygen levels has proven to be difficult, and results from different laboratories have varied widely. In order to measure oxygen levels, direct chemical analysis of some sort is necessary. Trace levels of this element exist in most environments, including equipment used for the analysis, so different groups have obtained widely varying results. Vacuum fusion has been used to calibrate the infrared absorption at $9\ \mu\text{m}$ due to oxygen, which is shown in Figure 25.³⁸ Once the calibration is done, the intensity of absorption can be used for non-destructive measurement of the oxygen level, in the range of 5×10^{16} - 10×10^{17} atoms/cm³. Most silicon contains oxygen at a level of 5 - 10×10^{17} /cm³.

The varying calibration results obtained by different groups were reviewed by Patel.³⁹ This has resulted in confusion in quoting oxygen levels when only infrared is used, and reference is made to only one of the experimental procedures. Recently an extensive re-evaluation of the measurement was performed by the Japan Electronic Industries Development Association,⁴⁰ and this calibration agrees closely with the original one of Kaiser and Keck, as well as previous work by Japanese researchers.⁴¹ According to this work, the concentration of oxygen is obtained by multiplying the absorption coefficient of the $9\ \mu\text{m}$ line by the value $3.03 \pm 0.02 \times 10^{17}$ atoms/cm².

6.1.2 Interpretation of Infrared Absorption Spectra. The absorption at $9\ \mu\text{m}$ is due to a well-defined species termed interstitial oxygen. This is not a true interstitial in the sense of a free atom which resides in the openings between silicon bonds; rather, the oxygen forms two strong

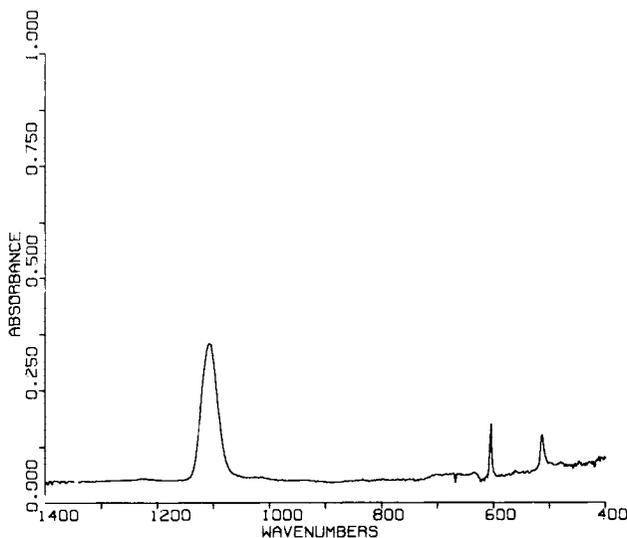


Figure 25: Infrared spectrum of silicon containing oxygen and carbon. Two peaks due to oxygen are seen.

bonds with near-neighbor silicon atoms, with a nearly linear Si_2O structure. The bond angle is 150° - 160° .⁴² The absorption spectrum can be analyzed as if the Si_2O species existed as a gas, or in other words, as if the silicon lattice wasn't there at all. The $9\ \mu\text{m}$ band is termed the asymmetric stretching mode of this species. Such a species should show other infrared-active modes as well.

Figure 25 shows a line at $513\ \text{cm}^{-1}$ or $19.5\ \mu\text{m}$, which is also due to oxygen. In early studies, this line was thought to be another vibrational mode of the interstitial oxygen species. However, Bosomworth and co-workers showed that this could not be so,⁴² and the line has remained unassigned and a source of some controversy. In 1980 I first proposed that this line was due to oxygen in another site, most probably substitutional.⁴³ The evidence for this assignment came from examining the low temperature absorption frequencies of isotopes of boron, carbon, and oxygen in silicon which are shown in Table 3.

The frequencies of Table 3 are plotted versus $1/\sqrt{m}$ and are shown in Figure 26. The straight-line relationship indicates that a simple harmonic oscillator describes the motion for all the isotopes, indicating that all the impurities occupy a lattice site with the full T_d symmetry of the "diamond" structure. This means either a true interstitial, or a fully substitutional site. Baker and co-workers have shown that addition of carbon to pure silicon causes the lattice to contract in a monotonic fashion until SiC precipitates from solution. This is strong evidence that carbon exists in a substitutional site,⁴⁷ and contrasts strongly with the effect of "interstitial" oxygen, which expands the lattice.⁴⁸ Taken together with the assumption that boron is also substitutional, this implies that oxygen exists as a substitutional as well as an interstitial species in oxygen.

Table 3: Low Temperature Absorption Frequencies for Light Elements in Silicon

<u>Isotope</u>	<u>Frequency (cm^{-1})</u>	<u>Reference</u>
O^{18}	506	43
O^{16}	517	43
C^{14}	573	44
C^{13}	590	44
C^{12}	611	44
B^{11}	623	45
B^{10}	646	45

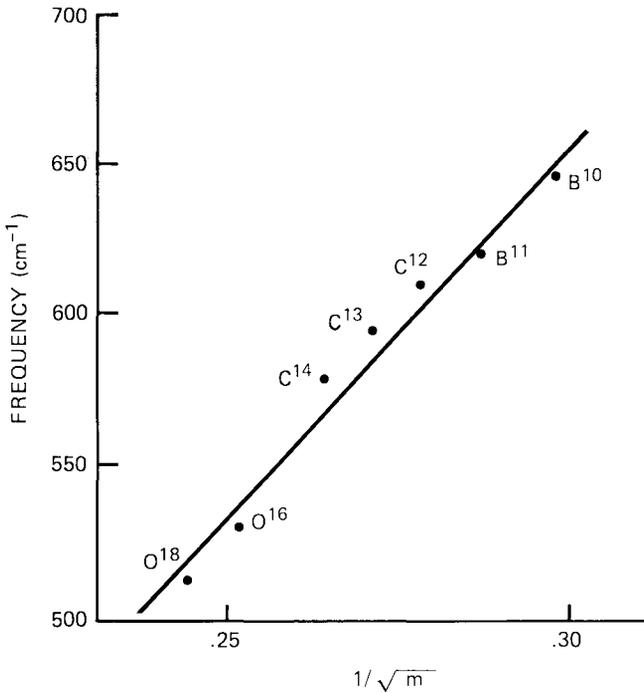


Figure 26: Plot of infrared absorption frequency of light elements in silicon versus inverse square root of atomic mass. The oxygen absorption near $19.5 \mu\text{m}$ is related to carbon and boron absorptions.

The existence of two oxygen species can explain many puzzling aspects of the behavior of this element in silicon. First of all, it explains how the segregation coefficient can be greater than one.¹¹ Incorporation into two sites in the lattice would proceed independently, but since the same atom is involved, the individual segregation coefficients would be additive. It has been shown⁹ that the segregation coefficient of the interstitial species is 1.0, and that the total segregation coefficient of oxygen in silicon is 1.25. Therefore the substitutional species has a value of $k=0.25$. Such a low value is consistent with other substitutional atoms with excess electrons, such as phosphorus with $k = 0.35$. It can explain other phenomena associated with oxygen which are discussed in succeeding paragraphs.

6.1.3 Solid Solubility. The solubility of oxygen in silicon is a strong function of temperature. Figure 27 shows solubility versus temperature over the range commonly employed for oxidation and diffusion. The data are taken from an article by Craven,⁴⁹ and are plotted according to the JEIDA calibration for infrared absorption. The solid solubility at the melting point is $\sim 2 \times 10^{18}$ atoms/ cm^3 , but normal crystal-growth processes give oxygen levels of $5\text{-}10 \times 10^{17}/\text{cm}^3$. It is important to note that at lower temperatures the solubility will correspond to a value less than the as-

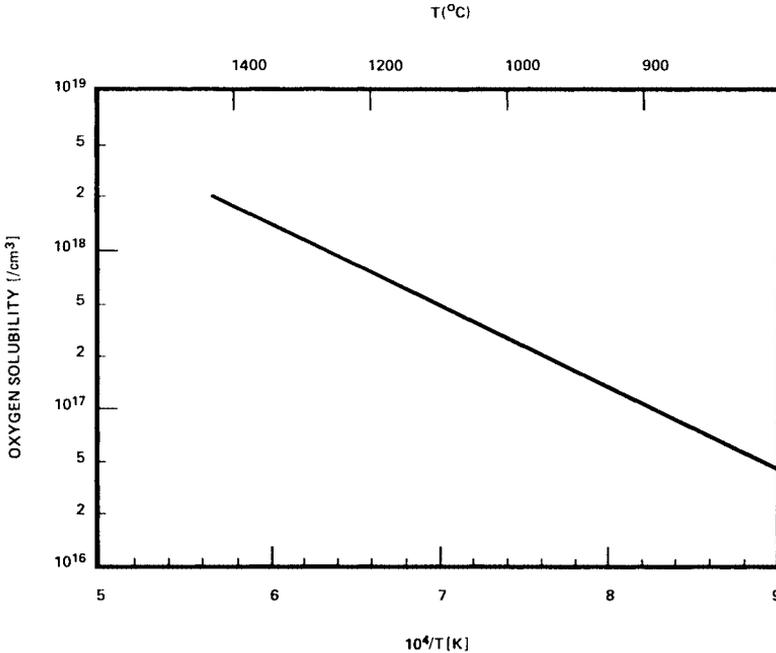


Figure 27: Oxygen solubility versus inverse temperature. Data of Reference 49 are plotted according to the infrared calibration of Reference 40.

grown level. This provides a driving force for precipitation from solution when the wafer is heated to lower temperatures for oxidation or other processes. This is the basis for much recent work on controlling such precipitation for purposes of internal gettering, discussed below.

As is true with other physical properties of oxygen in silicon, the solubility as a function of temperature is apparently different when measured by different researchers. The results determined by Craven agree well with those determined by Hrostowski and Kaiser,⁵⁰ and Takano and Maki,⁵¹ but other researchers have obtained quite different results (see the review by Patel³⁹ for references).

6.1.4 Diffusion Coefficient. The diffusion of oxygen plays an important role in the creation of defect-free surface layers during device manufacture, the formation of precipitates, and the formation of donors in the 450°C temperature range.

Oxygen is a very fast diffuser in silicon, and the diffusion properties have been measured over a wide temperature range by Mikkelsen.⁵² The results can be expressed in the relation

$$D = 0.07 \exp(-2.44\text{eV}/kT) \text{ cm}^2/\text{sec.}$$

which gives the diffusion coefficient as a function of temperature. Standard complementary error-function solutions to the diffusion equation apply to

the diffusion of oxygen in silicon. These results agree quite well with those obtained by Takano and Maki.⁵¹

At low temperatures, in the range 300°-500°C, the situation is more complicated, and again oxygen demonstrates its duality. An optical experiment was performed which gave an indirect way of obtaining the diffusion coefficient in this temperature range.⁵³ The results of the experiment yielded two quite different diffusion coefficients, depending on the prior heat treatment of the material. In one case, the diffusion coefficient had the same value as extrapolated from the high-temperature work cited previously.^{51,52} In this case, the material had been heated above 1200°C prior to measuring the diffusion coefficient. It is quite possible that the heating step produced an equilibrium distribution of oxygen among different sites which does not exist even in the as-grown case. The second diffusion coefficient was measured to be one hundred times faster than the first one, and was seen after extended heating at 900°C. At this temperature, departure from equilibrium is quite probable. The two values of the diffusion coefficient can be due to the predominance of the interstitial or substitutional species proposed earlier.⁴³

6.1.5 Donor Formation. A property unique to oxygen in silicon is its ability to form electrically active complexes when heated at 450°C. This phenomenon was reported prior to the discovery that crucible-grown material contained oxygen.⁵⁴ The temperature range of donor formation is quite narrow ($\leq \pm 50^\circ\text{C}$), indicating a specific mechanism for the process. The rate of formation and maximum concentration depend on the oxygen level. The formation rate is proportional to the fourth power of the (interstitial) oxygen concentration, while the maximum concentration is proportional to the third power of the oxygen level. These and other properties of the donor have been reviewed by Gosele and Tan.⁵⁵ The formation rate can be as high as 10^{13} carriers/cm³. sec, and donors can reach a maximum level of $\sim 5 \times 10^{16}$ /cm³. Evidently some electrically active complex of more than one oxygen atom is responsible for the donor behavior. Based on the formation kinetics, an SiO₄ complex was suggested by Kaiser, Frisch and Reiss,⁵⁶ although no reason for the electrical activity was suggested.

Oxygen is a double donor, with levels at 60 and 120 milli-electron volts. Low-temperature infrared studies show that more than one donor state is formed, depending on the time of heating at 450° C.⁵⁷ Heating at temperatures in excess of 500°C causes the donors to disappear. One practical consequence of this phenomenon is the appearance of donors in as-grown crystals, due to cooling of the ingot in the growth chamber. Most wafer manufacturers use a heating step in the range 650-700°C to eliminate donors formed during ingot cool-down.⁵⁸ Otherwise, the resistivity due to intentionally added dopant would be impossible to measure. Both the process of donor formation and annihilation appear to be due to some sort of polymerization behavior which can lead to rods, platelets or other shapes of SiO_x precipitates at higher temperatures.

Recently, a convincing model for the oxygen donors was proposed by Keller.⁵⁹ In this model, three interstitial oxygen atoms surround a substitutional oxygen species in next-neighbor locations. Nine configurations are possible for a strain-free four-oxygen complex. The substitutional

oxygen can release two valence electrons to the conduction band once the strain is compensated. This is the first model to effectively combine all the observations about the oxygen donor and lends additional credence to the idea of interstitial and substitutional species existing independently in the lattice.

One final point about the oxygen donor concerns its formation during IC-chip fabrication. Device manufacturers often anneal or alloy wafers to improve ohmic contact of aluminum to silicon, or to remove silicon-silicon dioxide surface states in MOS gates or capacitors. If the annealing temperature is chosen to be 450°C, as is sometimes done, the resistivity of the substrate can be substantially changed. A slightly lower temperature, such as 400°C, will avoid donor formation during this anneal step.

6.2 Precipitation from Solid Solution

In the as-grown state, oxygen is ordinarily well dispersed throughout the material. In device manufacturing, wafers are heated to temperatures in the range 900-1200°C for oxidation, diffusion and insulating film deposition. The oxygen incorporated into the lattice during crystal growth is in a condition of supersaturation at these temperatures. It will precipitate as SiO_x given enough time and the proper nucleation. When precipitation does occur, it can happen uniformly throughout the material or non-uniformly in circular or ring-shaped patterns. These latter are sometimes referred to as “oxygen swirl,” but are distinct from swirl defects identified in float-zone silicon. Abe has shown that non-uniform oxygen precipitation results from growth-rate fluctuations during the crystal-growth process.¹² Manufacturers attempt to minimize these fluctuations and provide wafers in which oxygen precipitates uniformly.

Uniform precipitation is termed homogeneous, and can be predicted and controlled. The number of precipitates formed is determined by a relatively low-temperature “nucleation” process, which consists of heating for several hours at a temperature in the range 600-900°C. Note that some wafers may previously have received such an annealing step for purposes of donor annihilation, which must be taken into account when nucleation is performed as a separate step in wafer fab. Nucleation causes the formation of very small clusters of oxygen atoms, on the order of 10Å in size.⁶⁰ These nuclei serve as sites for precipitation during subsequent high-temperature processing.

Intentional precipitation of oxygen is carried out as a means of “gettering” trace metallic impurities. Assuming that the wafer receives a thorough cleaning prior to a high-temperature process, the main source of these impurities is the furnace heating elements.⁶¹ Metals diffuse through the quartz furnace liner, and are deposited on the wafers. Metals such as iron, copper, and nickel are very fast diffusers in silicon. Although the gettering process is not completely understood, one result suggests that metal atoms diffusing through the wafer can be entrained in a growing SiO_x precipitate.⁶² Once in the precipitate, they are trapped and electrically inactive. If metals remain free, however, they can seriously degrade device performance. The starting oxygen level and the nucleation cycle are

chosen so that precipitation occurs throughout the device-manufacturing process. In general, these will differ significantly for bipolar and MOS processes, and the sequence is usually tailored for each process line. It is important to leave enough unprecipitated oxygen so that the wafer retains its mechanical strength and warpage is prevented.²⁸

6.3 Denuded Zone Formation

In order to employ controlled precipitation of oxygen in wafer fab, a surface-denuding step is generally required. In this operation, oxygen is out-diffused from the surface so that its concentration falls to a low value. Then, since the surface oxygen level is below the solubility level during subsequent high-temperature treatments, precipitation will not occur and the active-device region remains precipitate-free. The denuding temperature and time can be selected based on the initial oxygen level.⁶³ Optimum temperature is in the range of 1000-1100°C. The denuding step can be performed before or after nucleation, and may coincide with initial oxidation. Figure 28 shows a wafer cross-section after angle lap and defect etch. The wafer has received the three-step sequence of denuding, nucleation and precipitation. A high density of precipitates is seen in the interior of the wafer, while the surface is defect-free.

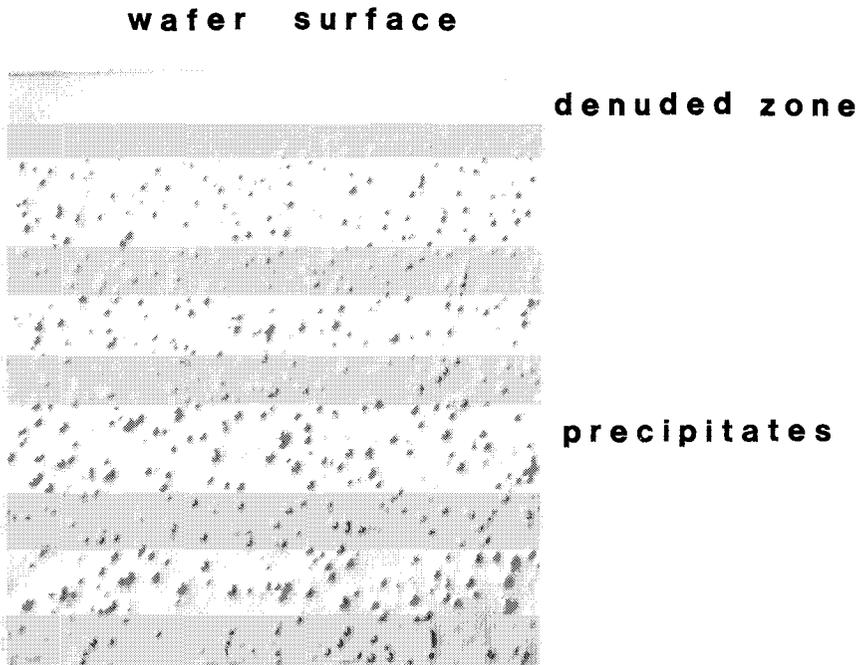


Figure 28: Optical micrograph of angle-lapped and etched silicon wafer. Denuded zone and bulk oxygen precipitates and stacking faults are seen. (Photograph courtesy of W.M. Bullis.)

6.4 Device Application

Controlled oxygen precipitation for impurity removal is becoming common practice for many device lines. MOS memories constitute one class of devices whose yield depends on in-process precipitation. Metallic impurity removal results in very high minority-carrier lifetimes near the surface, which translates to long refresh times, very desirable in memories.⁶² Bipolar devices can also show improved yields due to "gettering."

Other applications for oxygen precipitates are also being explored. High precipitate density in the bulk of the wafer reduces minority-carrier lifetimes in this region, thereby suppressing the flow of lateral currents. This effect has been applied to the suppression of crosstalk in memories,⁶⁴ and in optical sensor arrays.⁶⁵ In the latter example, high precipitate density also resulted in improved immunity to latchup in CMOS devices on the same chip. It is quite likely that other ways to control the electrical properties of silicon will be found based on effects due to oxygen.

REFERENCES

1. G.K. Teal, *IEEE Transac. Electron. Dev.* 23: 621 (1976).
2. W.R. Runyan, *Silicon Semiconductor Technology*, McGraw Hill, New York, 1965.
3. For a review of some of these techniques, see *Electronic and Optical Properties of Polycrystalline or Impure Semiconductors and Novel Silicon Growth Methods*, K.V. Ravi and W. O'Mara, eds., Electrochem. Soc., New York, 1980.
4. W.C. Dash, *J. Appl. Phys.*, 29: 228 (1958), *J. Appl. Phys.* 29: 736 (1958).
5. W.G. Pfann, *Trans. AIME* 194: 747 (1952).
6. B. Breitwieser, private communication.
7. T. Carlberg, T.B. King and A.F. Witt, *J. Electrochem. Soc.* 129: 189 (1982)
8. K. Hoshikawa, H. Hirata, H. Nakanishi and K. Ikuta, in "*Semiconductor Silicon 1981*," H.R. Huff, R.J. Kriegler, and Y. Takeishi, ed., Electrochem. Soc., New York, 1981, p. 101.
9. Y. Yatsurugi, N. Akiyana, Y. Endo, and T. Nozaki, *J. Electrochem. Soc.* 120: 1975 (1973).
10. R.J. Jaccodine and C.W. Pearce, in "*Defects in Silicon*," W.M. Bullis and L.C. Kimerling ed., Electrochem. Soc. New York, 1983, p. 115.
11. W.C. O'Mara, in "*Defects in Silicon*," op cit., p. 120.
12. T. Abe, K. Kikuchi, and S. Shirai, in "*Semiconductor Silicon 1977*," R.H. Huff and E. Sirtl, ed., Electrochem. Soc. New York, 1977, p. 95.
13. T. Abe, K. Kikuchi, S. Shirai, and S. Muraoka, in "*Semiconductor Silicon 1981*," H.R. Huff, R.J. Kriegler, and Y. Takeishi, ed., Electrochem. Soc., New York, 1981, p. 54.
14. W.C. O'Mara and D. Guidici, *Electrochem. Soc. Extended Abstracts*, Spring 1979.
15. R. Walsh, private communication.
16. W. Kern and D. Puotinen, *RCA Review* 31: 1987 (1970).
17. W. Shockley, "*Electrons and Holes in Semiconductors*," Van Nostrand, New York, 1950, p. 6.
18. J.C. Irvin, *Bell Sys. Tech. J.* 41: 387 (1962).
19. W.R. Thurber, R.L. Mattis, Y.M. Liu, and J.J. Filliben, *J. Electrochem. Soc.* 127: 1807 (1980).

20. W.R. Thurber, R.L. Mattis, Y.M. Liu, J.J. Filliben, *ibid*, 127: 2291 (1980).
21. D.M. Caughery and R.E. Thomas, *Proc. IEEE* 55: 219Z (1967).
22. P.A. Schumann, Jr., *Solid State Technology*, 13: 50, (January 1970).
23. J.P. Lavine, F.C. Lo, F. Moser, B.C. Barkey, and F.T.J. Smith, in *Electronic and Optical Properties of Polycrystalline Semiconductors*, K.V. Ravi and W. O'Mara, ed., Electrochem. Soc., New York, 1980, p. 96.
24. ASTM Procedure F120-75 and F121-80, Amer. Soc. Test. Mat., Philadelphia, 1985.
25. ASTM Procedure F120-75 and F123-81, Amer. Soc. Test. Mat., Philadelphia, 1985
26. Kurt Peterson, *Proc. IEEE* 70: 429 (1982).
27. K.Sumino, in *Semiconductor Silicon 1981*, op cit. 208.
28. Y. Kondo, in "Semiconductor Silicon 1981" op cit., p. 220.
29. H.D. Chiou, J. Moody, R. Sandford, and F. Shimura, in *VLSI Science and Technology/1984*, K.E. Bean and G.A. Rozgonyi, eds., Electrochem. Soc., New York, 1984, p. 59.
30. C.W. Pearce and G.A. Rozgonyi, in *Semiconductor Silicon 1977*, H.R. Huff and E. Sirtl, ed., Electrochem. Soc., New York, 1977, p. 606.
31. C.J. Varker and K.V. Ravi, in *Semiconductor Silicon, 1977*, op cit., p. 785.
32. K. Daido, S. Shinoyama, and N. Inoue, in *Rev. Elec. Comm. Labs (Japan)* 27: 33 (1979).
33. G.A. Rozgonyi and T.E. Seidel, in *Semiconductor Silicon, 1977*, op cit., p. 616.
34. A list of etchants for revealing process-induced defects may be found in *Solid State Technol.*, Nov. 1983, p. 111.
35. G.A. Rozgonyi and C.W. Pearce, *Appl. Phys. Lett.* 32: 747 (1978).
36. R.A. Craven, T. Shimura, R.S. Hockett, L.W. Shive, P.W. Fraundorf, and G. Keefe-Fraundorf, in *VLSI Science and Technology/1984*, K.E. Bean and G.A. Rozgonyi, ed., Electrochem. Soc., New York, 1984, p. 20.
37. W.E. Bailey, R.A. Bowling, and K.E. Bean, in *Defects in Silicon*, op cit., p. 204.
38. W. Kaiser and P.H. Keck, *J. Appl. Phys.* 28: 822 (1957).
39. J.R. Patel, in *Semiconductor Silicon 1977*, op cit., p. 521.
40. T. Iizuka, S. Takasu, M. Tajima, T. Arai, T. Nozaki, N. Inoue, and M. Watanabe, in *Defects in Silicon*, op cit., p. 265.
41. T. Nozaki, Y. Yatsurugi, and N. Akiyama, *J. Radioanalyt. Chem.* 4: 87 (1970).
42. D.R. Bosomworth, W. Hayes, A.R.L. Spray, and G.D. Watkins, *Proc. Roy. Soc. A317*: 133 (1970).
43. W.C. O'Mara, *Electrochem. Soc. Extended Abstracts* 80-1: 475 (May 1980).
44. H.J. Hrostowski and R.H. Kaiser, *Phys. Rev.* 107: 966 (1957).
45. R.C. Newman and J.B. Willis, *J. Phys. Chem. Solids* 26: 373 (1965).
46. S.D. Smith and J.F. Angress, *Phys. Lett.* 6: 131 (1963).
47. J.A. Baker, T.N. Tucker, N.E. Moyer, and R.C. Buschert, *J. Appl. Phys.* 39: 4365 (1968).
48. W.L. Bond and W. Kaiser, *J. Phys. Chem. Solids* 16: 44 (1960).
49. R.A. Craven, in *Semiconductor Silicon 1981*, op cit., p. 254.
50. H.R. Hrostowski and W. Kaiser, *J. Phys. Chem. Solids* 9: 214 (1959).
51. Y. Takano and M. Maki, in *Semiconductor Silicon 1973*, H.R. Huff and R.R. Burgess, ed., Electrochem. Soc., New York, 1973, p. 469.
52. J. Mikkelsen, *Appl. Phys. Lett.* 40: 336 (1982).
53. M. Stavola, J.R. Patel, L.C. Kimerling, and P.E. Freeland, *Appl. Phys. Lett.* 42: 73 (1983).
54. C.S. Fuller and R.A. Logan, *J. Appl. Phys.* 28: 1427 (1957).
55. U. Gösele and T.Y. Tan, *Appl. Phys.* A28: 79 (1982).
56. W. Kaiser, H.L. Frisch and H. Reiss, *Phys. Rev.* 112: 1546 (1958).
57. D. Wruck and P. Gaworzewski, *Phys. Stat. Solidi a* 56: 557 (1979).
58. W.C. O'Mara, J.E. Parker, P. Butler, and A. Gat, *Appl. Phys. Lett.* 46: 299 (1985).
59. W.W. Keller, *J. Appl. Phys.* 55: 3471 (1984).
60. N. Inoue, J. Osaka, and K. Wada, *J. Electrochem. Soc.* 129: 2780 (1982).
61. P.F. Schmidt and C.W. Pearce, *J. Electrochem. Soc.* 128: 630 (1981).

62. R. Huff, H.F. Schaake, J.T. Robinson, S.C. Baber, and D. Wong, *J. Electrochem. Soc.* 130: 1551 (1983).
63. J.M. Andrews, in *Defects in Silicon*, op cit., p. 133.
64. H. Otsuka, K. Watanabe, H. Nishimura, H. Iwai and H. Nihira, *IEEE-ED Lett* 3: 1982 (1982).
65. C.N. Anagnostopoulos, E.T. Nelson, J.P. Lavine, K.Y. Wong and D.N. Nichols, *IEEE Trans.* ED-31: 225 (1984).

The Thermal Oxidation of Silicon and Other Semiconductor Materials

Bruce E. Deal
Research Center
Fairchild Semiconductor Corporation
Palo Alto, California

1. INTRODUCTION AND BACKGROUND

The thermal oxidation process associated with semiconductor technology has been used primarily in conjunction with silicon and silicon containing materials. Attempts to thermally oxidize germanium and compound semiconductors have been generally unsuccessful, except by employing a field-assisted process such as anodizing. More often, deposited oxides and dielectrics have been used to passivate compound semiconductors, while germanium is not a significant factor in today's semiconductor industry. A description of CVD (chemical vapor deposition) and other deposition processes used for dielectric films in semiconductor applications is included in other chapters of this volume.

Silicon semiconductor technology has depended heavily on the thermal oxidation process since the 1950's, when silicon devices were first developed. Initially, thermal oxides were used to selectively mask dopants during the fabrication of diffused transistors.¹ Additional investigations at that time by Atalla, Liginza, Spitzer, and other workers at Bell Laboratories provided considerable information about silicon thermal oxides, especially their passivation properties.^{2,3} These investigations led in 1960 to two of the most important developments of semiconductor technology: the Planar process invented by Hoerni,⁴ and the MOS transistor which was first disclosed by Kahng and Atalla.⁵

Many investigations related to silicon thermal oxidation and other types of dielectric films have been undertaken since 1960.^{6,7} These have resulted in a number of technological developments and have helped to make possible the amazing growth of the semiconductor industry. Some of the uses of thermal oxides and dielectric films in today's semiconductor technology are listed in Table 1. Thermal silicon dioxide is and will most likely continue to be the mainstay of silicon device technology, even as we move into the realm of sub-micrometer VLSI and beyond.

Table 1: Uses of Dielectric Films in Semiconductor Technology

- COMPONENTS IN DEVICES
- CORROSION PROTECTION
- DEVICE ISOLATION
- DOPANT DIFFUSION SOURCE
- GETTER IMPURITIES
- INCREASE BREAKDOWN VOLTAGE
- INSULATE METAL LAYERS
- MASK AGAINST DOPANTS
- MASK AGAINST IMPURITIES
- MASK AGAINST OXIDATION
- MECHANICAL PROTECTION
- PASSIVATE JUNCTIONS
- SMOOTH OUT TOPOGRAPHY

This chapter deals primarily with the thermal oxidation of silicon. The kinetics of silicon thermal oxidation is first reviewed, with emphasis on the general oxidation relationship and the thin oxide regime. Important oxide properties are then summarized. Following is the most important part of the chapter, a description of process variable-oxidation reaction interdependencies. Some of the variables included are effects on silicon surface properties (dopant redistribution, charges), surface property effects on the oxidation process (orientation, doping, cleaning), and ambient effects (type, chlorine addition, pressure). After a discussion of oxidation mechanisms, the chapter concludes with a section on other oxidation processes and an indication of thermal oxidation applications and trends with respect to future semiconductor devices.

2. SILICON THERMAL OXIDATION KINETICS

The thermal oxidation process provides superior passivating charac-

teristics for silicon devices as compared to the various deposited dielectric processes. Used in combination, however, with other dielectrics (e.g. deposited silicon nitride over thermal oxide), the most stringent requirements can be met. In the thermal oxidation process, silicon reacts with either oxygen or water vapor (steam) at temperatures between 600° and 1250°C to form silicon dioxide. The oxidation reaction may be represented by the following two reactions:



Special marker experiments⁸ have demonstrated that oxidation proceeds by the diffusion of either an oxygen or water species through the oxide already formed which then reacts with the silicon at the Si-SiO₂ interface. As oxidation continues, the interface moves into the silicon and a new, clean silicon surface is produced. As a result, original silicon surface states (unsatisfied bonds) and contamination are consumed and optimized device passivation is achieved. From the densities and molecular weights of silicon and amorphous silicon dioxide, it can be shown that for every thickness x_o of oxide formed, 0.45 x_o of silicon is consumed. The exact nature and charge of the diffusing oxidation species (O₂, O, O₂⁻, O⁻, H₂O, H₃O⁺, OH⁻, etc.) have not yet been identified. It is known, however, that for steam oxidation, considerable exchange occurs between the already formed silicon dioxide and the diffusing water species. On the other hand, very little exchange takes place between oxygen and the oxide network.

Thermal oxidation of silicon is normally carried out in a fused quartz tube in a resistance heated furnace. The silicon wafers are placed vertically in slots in a flat quartz "boat," most present day furnaces accommodating up to 200 four to six-inch diameter wafers. For dry O₂ oxidation, high purity oxygen from a liquid source is transported into the furnace tube through suitable regulators, valves, traps, filters, and flowmeters. For a number of years, water or steam oxidation was carried out by bubbling O₂ or N₂ through a flask of deionized water maintained at a particular temperature. Thus, a specified vapor pressure of water could be provided in the oxidizing ambient. More recently, however, pyrogenic systems⁹ have been employed which permit H₂ to react with O₂ at the inlet end of the oxidation tube, thus providing water vapor of much higher purity and control.

Silicon oxidation data are obtained by determining oxide thickness (x_o) as a function of oxidation time (t) and other variables such as oxidation temperature and silicon orientation. Typical results are given in Figure 1 and 2, which contain plots of $\log x_o$ vs $\log t$ for silicon oxidation at various temperatures in dry O₂¹⁰ and pyrogenic steam.¹¹

2.1 General Relationship

As indicated earlier, silicon thermal oxidation proceeds by the diffusion

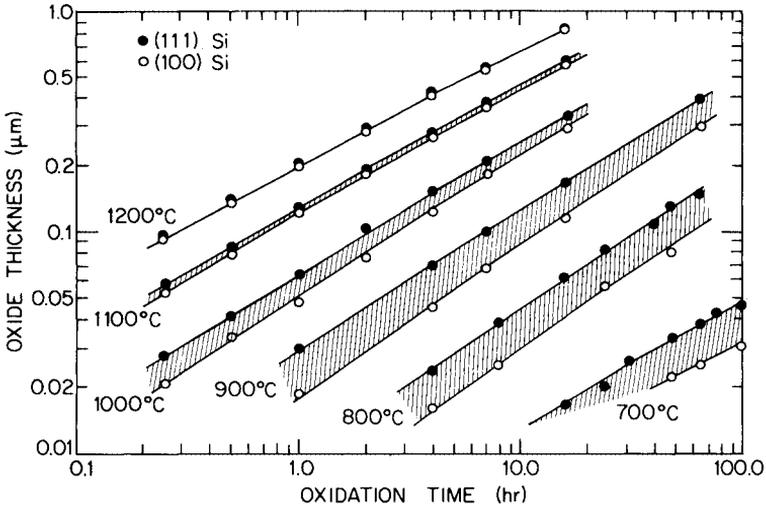


Figure 1: Oxide thickness vs oxidation time for silicon oxidation in dry oxygen at various temperatures (after Hess and Deal¹⁰).

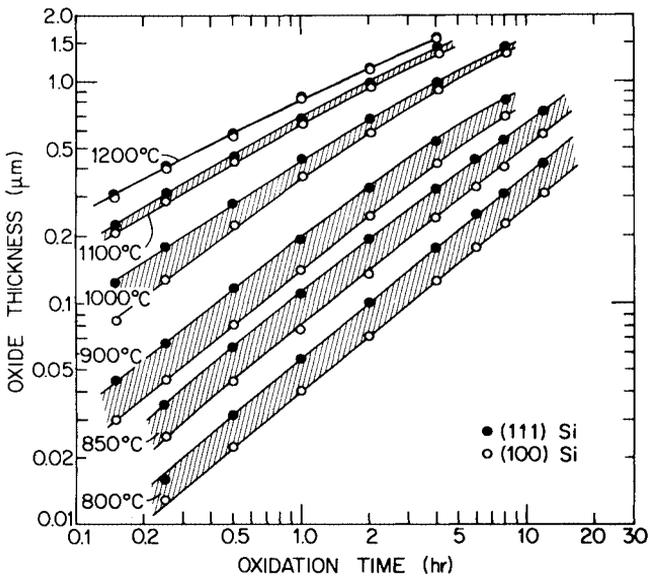


Figure 2: Oxide thickness vs oxidation time for silicon oxidation in pyrogenic steam (~ 640 Torr) at various temperatures (after Deal¹¹). Reprinted by permission of the publisher, The Electrochemical Society, Inc.

of an oxidizing species through the oxide already formed. This process is indicated in Figure 3. It has been proposed by Deal and Grove¹² that during thermal oxidation three consecutive reactions occur whose fluxes are equal under steady-state conditions. These are denoted in Figure 3 and are summarized below, along with the expressions which represent the fluxes (flux (F) is defined as the number of molecules—in this case oxidant—crossing a unit surface area in a unit time).

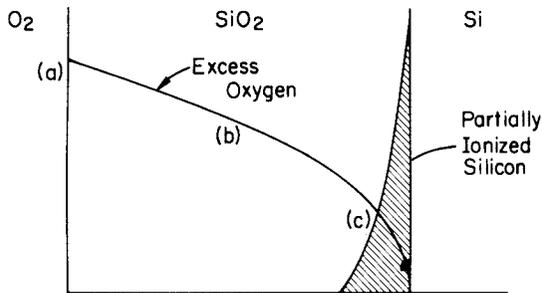


Figure 3: Schematic illustration of the silicon thermal oxidation process.

- (a) Transfer of the oxidant from the gas phase to the oxide outer surface:

$$F_1 = h (C^* - C_o) \quad [3]$$

where h is a gas-phase transport coefficient, C_o is the concentration of the oxidant in the outer oxide surface, and C^* is the equilibrium concentration of the oxidant in the oxide, and assumed to be proportional to the partial pressure of the oxidant in the gas ambient.

- (b) Diffusion of the oxidizing species through the oxide to the silicon:

$$F_2 = -D_{eff} \frac{C_o - C_i}{x_o} \quad [4]$$

where D_{eff} is the effective diffusion coefficient of the oxidizing species in the oxide, C_i is the oxidant concentration in the oxide near the Si-SiO₂ interface, and x_o is the oxide thickness.

- (c) Reaction of oxidizing species with silicon at the Si-SiO₂ interface to form SiO₂:

$$F_3 = k C_1 \quad [5]$$

where k is the interface reaction rate constant.

By assuming all the above fluxes to be equal, the following general oxidation relationship has been derived.¹²

$$x_o^2 + Ax_o = B(t+\tau) \quad [6a]$$

also written in the form

$$\frac{x_o^2 - x_i^2}{B} + \frac{x_o - x_i}{B/A} = t \quad [6b]$$

where x_o = oxide thickness, t = oxidation time, and A , B , τ , and x_i are constants as defined below:

$$A = 2 D_{eff} (1/k + 1/h) \quad [7]$$

$$B = 2 D_{eff} C^* / N_1 \quad [8]$$

$$\tau = (x_i^2 + Ax_i) / B \quad [9]$$

where D_{eff} = effective oxidant diffusion constant in the oxide;
 k, h = rate constants at the Si-SiO₂ and gas-oxide interfaces;
 C^* = equilibrium concentration of the oxide species in oxide;
 N_1 = number of oxidant molecules in the oxide unit volume; and
 x_i = oxide thickness at the start of oxidation.

Two limiting forms of Equation 6 can be noted. At large oxidation "times," i.e. $t \gg A^2/4B$ and $t \gg \tau$

$$x_o^2 = Bt \quad [10]$$

This equation represents a parabolic oxidation and B is the parabolic rate constant.

For small oxidation "times," $t \ll A^2/4B$ and the linear oxidation expression is obtained:

$$x_o = B/A(t+\tau) \quad [11]$$

B/A is the linear rate constant.

From Equations 7 and 8 it can be noted that when the oxidation process is controlled primarily by the parabolic rate constant (at high temperatures or thick oxides), the kinetics are affected by changes in the diffusion process or oxidant solubility in the oxide. The latter is proportional to ambient pressure. On the other hand, at low temperatures or for thin oxides, where the linear rate constant predominates, the oxidation is also sensitive to oxidant solubility in the oxide (and ambient pressure) but depends on those factors affecting the interface rate constants h and k . These effects will be discussed in more detail later.

Special mention should be made of the correction factor τ which is related to initial oxide thickness x_i in Equation 9. It has been noted that for oxidation in dry O₂, an initial thickness region does not appear to be satisfied by the general relationship Equation 6. Rather, the plot of oxide thickness versus oxidation time tends to extrapolate through the thickness

axis at about $x = 150\text{\AA}$. Thus, in the absence of a model for oxidation in this region, the practice has been to assign a value of τ corresponding to $x_1 = 150\text{\AA}$. Further discussion on the mechanism of thermal oxidation for these very thin oxides is presented in the next section.

The thermal oxidation of silicon can be represented by Equation 6 for a wide range of temperatures, oxide thicknesses, orientations, and oxidation ambients, provided the dependence of the rate constants B and B/A as a function of these variables is known. Values of the rate constants have been determined by rearranging the general relationship Equation 6 into a linear expression, plotting x_0 vs $(t+\tau)/x_0$ and extracting B as the slope and $-A$ as the intercept from the resulting plots. Arrhenius expressions of the form

$$C = c_1 e^{-E/kT} \quad [12]$$

have been used in plotting $\log B$ and $\log B/A$ vs $1/T$. Such plots are presented in Figures 4 and 5, and values of the constants for the Arrhenius expression are tabulated in Table 2. These data can be used to determine any thickness-time relationship for a given set of oxidation conditions. Similar data are incorporated in the SUPREM program¹³ and related computer process modeling programs.

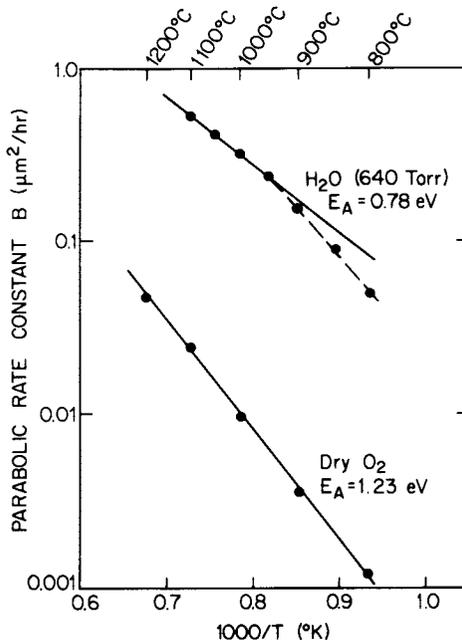


Figure 4: Dependence of the parabolic rate constant B on temperature for the thermal oxidation of silicon in pyrogenic steam (~ 640 Torr) and dry oxygen (after Deal¹¹). Reprinted by permission of the publisher, The Electrochemical Society, Inc.

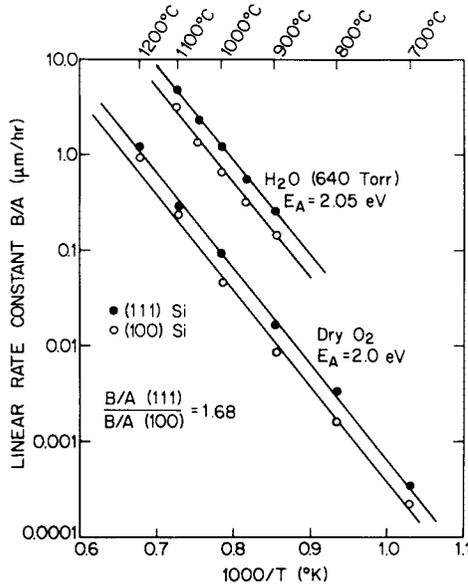


Figure 5: Dependence of the linear rate constant B/A on temperature for the thermal oxidation of silicon in pyrogenic steam (~640 Torr) and dry oxygen (after Deal¹¹). Reprinted by permission of the publisher, The Electrochemical Society, Inc.

Table 2: Dependence of Silicon Oxidation Rate Constants on Temperature

PARABOLIC	$B = C_1 e^{-E_1/kT}$
LINEAR	$B/A = C_2 e^{-E_2/kT}$
(111) SILICON	
DRY O ₂	$C_1 = 7.72 \times 10^2 \mu\text{m}^2/\text{hr}$ $C_2 = 6.23 \times 10^6 \mu\text{m}/\text{hr}$ $E_1 = 1.23 \text{ eV}$ $E_2 = 2.0 \text{ eV}$
STEAM (PYROGENIC)	$C_1 = 3.86 \times 10^2 \mu\text{m}^2/\text{hr}$ $C_2 = 1.63 \times 10^8 \mu\text{m}/\text{hr}$ $E_1 = 0.78 \text{ eV}$ $E_2 = 2.05 \text{ eV}$
(100) SILICON	$C_2 (100) = C_2 (111)/1.7$

2.2 Thin Oxide Formation

As indicated above, considerable experimental evidence is available which suggests a different or modified reaction mechanism for dry O_2 thermal oxide formation below 200\AA . This difference has resulted in the requirement (in the case of dry O_2 oxidation) for the use of a constant, τ , in the general relationship, Equation 6. A number of investigators have attempted to model the thermal oxidation process in the very thin regime, but as yet no one satisfactory relationship has been established. Likewise, the exact characterization of thin oxide properties, electrical, physical, etc., has not yet been accomplished. These properties are undoubtedly related to the kinetics of thin oxide formation and also make accurate thickness measurements quite difficult.

Among the first to investigate the mechanisms involved in the initial stages of silicon thermal oxidation were van der Meulen and Ghez.^{14,15} They proposed complex reactions at the Si-SiO₂ interface involving both molecular and atomic reactions with silicon. Thus, various reported pressure dependencies of the oxidation reaction on oxygen partial pressures could be explained, with $P^{1.0}$ dominating for thicker oxides and $P^{0.5}$ being prevalent in the thinner oxide regime. Blanc¹⁶ has proposed a similar model but with only a $P^{0.5}$ dependence for thin oxide formation. This, however, does not satisfy pressure dependencies observed for thicker oxides.

More recently Massoud and Plummer¹⁷ have suggested that the initial stages of silicon oxidation in dry O_2 may be represented by the following equation:

$$\frac{dx_o}{dt} = \frac{B}{2x_o + A} + c_1 \exp\left(\frac{-x_o}{L_1}\right) + c_2 \exp\left(\frac{-x_o}{L_2}\right) \quad [13]$$

In this expression, the first term on the right-hand side is the contribution from the original linear-parabolic model. The second term incorporating L_1 is possibly related to effects of residue left on the silicon surface from the cleaning treatment ($X_o \leq 15\text{\AA}$). The contribution of L_2 has not yet been explained. Subsequently, Han and Helms¹⁸ have proposed that a mechanism based on parallel diffusion reactions provide even a better fit to oxidation data over the entire thickness range.

It is important for future applications of devices having sub-micrometer feature sizes and film thicknesses in the nanometer range (especially MOS gate and capacitor oxides), that reaction mechanisms be understood and characterized for oxides in the very initial stages of formation. The investigations described are a good step in that direction.

3. PROPERTIES OF THERMAL OXIDES

Thermal silicon dioxide (SiO₂) produced by the oxidation of silicon in O_2 or H_2O at elevated temperatures is essentially amorphous silica. Its properties are almost identical to those of the fused quartz tubes in which most oxidations are carried out. The molecular structure is a random version of

crystalline quartz with each silicon atom tetrahedrally surrounded by four oxygen atoms. In turn each oxygen atom is bonded to two silicon atoms. Thermal oxides act as somewhat of a barrier to high-temperature diffusion of the common dopants B, P, As, and Sb and can be used as a mask against ion implantation. For these cases, masking is limited and suitable masking data curves must be used to ensure that the species are prevented from penetrating into the substrate. On the other hand, alkali impurity ions, such as Li^+ , Na^+ , K^+ , and even H^+ or H_2O can diffuse rapidly through thermal oxide, even at relatively low temperature. In general, other more dense dielectric films, such as silicon nitride or phosphosilicate glass, are used in combination with thermal oxide to passivate against these impurities.

Thermal SiO_2 is normally quite stable chemically. The most common etchant is an HF-based solution. Optical and electrical properties are similar to those of fused quartz.

Electrical properties of thermal oxides are extremely critical with respect to device performance and reliability. Such parameters, as conductivity, carrier trapping characteristics, and oxide charges, can have an appreciable effect on today's small-geometry integrated circuits. These properties have been evaluated by appropriate techniques for varying preparation conditions.¹⁹ Also critical to devices is the defect density of passivating oxides. Many investigations have been reported that are related to dielectric strength, pinhole density, and other oxide properties which result primarily from impurity incorporation during processing.²⁰ A number of common properties of thermal SiO_2 are listed in Table 3.

Various techniques are used to measure properties of thermal oxide films. One of the most important properties, thickness, is determined primarily by ellipsometry or other spectrophotometric or interferometric methods. Typical film thicknesses range from over 1 μm to below 100 \AA in special devices. Since structural and optical properties of the thinner films may differ from bulk properties, some difficulties arise in thin oxide measurements.

4. PROCESS VARIABLE/OXIDATION REACTION DEPENDENCIES

The silicon thermal oxidation process has been found to be a direct function of a number of process variables, including silicon surface properties. Conversely, some important silicon surface properties are dependent on the oxidation process. In Figure 6, one form of the general oxidation relationship (Equation 6b) is re-presented, with an indication of some of the physical variables which contribute to the oxidation reaction through the individual components of the rate constants. In this section, some of the important inter-relationships between process variables and the oxidation process are discussed.

4.1 Effects of the Oxidation Reaction on Surface Properties

4.1.1 Oxide Charges. At least four general types of electrical charges have been observed to be associated with the thermally oxidized silicon

Table 3: Properties of Thermal Silicon Oxides (25°C unless indicated)

Physical Properties	
Formula	SiO ₂
Molecular weight	60.08
Molecules/cm ³	2.3×10^{22}
Density	2.27 gm/cm ³
Melting point	~1700°C
Specific heat	1.0 Joule/g°C
Vapor pressure	10^{-3} Torr at 1450°C
Thermal conductivity	0.014 watt/cm°C
Linear coefficient of expansion	0.5×10^{-6} °C ⁻¹
Young's modulus	1×10^7 psi
Electrical Properties	
Resistivity	5×10^{15} Ω-cm
Dielectric constant	3.9
Dielectric strength	$\sim 1 \times 10^7$ V/cm
Energy gap	~8 eV
Optical Properties	
Refractive index	1.462 at 5459 Å
Absorption coefficient	≈ 0 for E < 8 eV
Chemical Properties	
Etch rate (1:10 HF:H ₂ O)	5 Å/sec
Oxygen solubility	5.5×10^{16} cm ⁻³ at 1000°C
Water solubility	3.4×10^{19} cm ⁻³ at 1000°C

structure.^{21,22} These charges and their locations are indicated in Figure 7, which is a representation of an oxide cross section similar to that shown in Figure 3. The symbols selected to denote these charges²³ are based on the following:

Q = effective net charge per unit area at the Si-SiO₂ interface (C/cm²),

N = Q/q = net number of charges per unit area at the Si-SiO₂ interface (number/cm²),

D = net density of interface trapped charges per unit area and energy (number/cm²-eV),

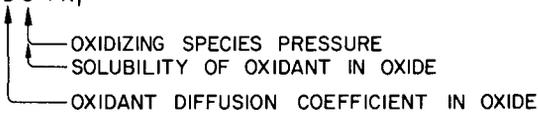
q = charge of an electron.

GENERAL RELATIONSHIP

$$\frac{(x_0^2 - x_i^2)}{B} + \frac{(x_0 - x_i)}{B/A} = t$$

PARABOLIC RATE CONSTANT

$$B = 2 D C^*/N_1$$



LINEAR RATE CONSTANT

$$B/A =$$

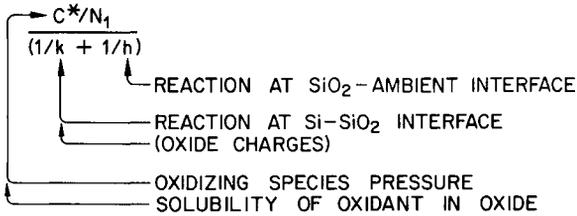


Figure 6: Relationship of process variables to oxidation general relationship.

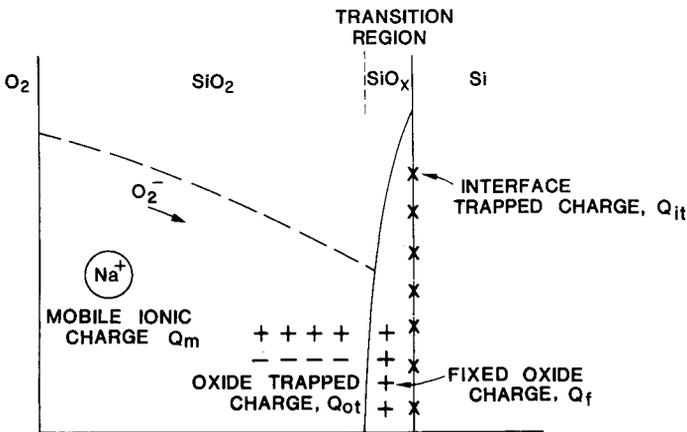


Figure 7: Names and location of charges associated with the thermally oxidized silicon structure.

A brief description of the four types of charges is presented. As may be noted, all of them are directly related to the oxidation and associated processing. Techniques for measuring the charge densities are also indicated.

Fixed Oxide Charge; Q_f, N_f . As indicated in Figure 7, the fixed oxide charge is positive and located in the oxide very close to the Si-SiO₂ interface (<20 Å). It is due primarily to structural defects (ionized silicon) in the SiO₂ lattice and directly dependent on conditions of oxidation. For instance, its density which ranges from 10¹⁰ to 10¹² cm⁻² depends on oxidation ambient and temperature, anneal/cooling conditions, and silicon orientation.²² Its density normally does not vary with surface potential which distinguishes it from interface trapped charge—hence the name *fixed oxide charge*.

An important Q_f process relationship is that the density of Q_f for either steam or dry O₂ oxidation increases with decreasing temperature.^{21,22,24} However, a subsequent anneal in an inert ambient such as argon will decrease the density of Q_f to a minimum equilibrium value, giving rise to the “ Q_f triangle” effect.²¹

Another important property of fixed oxide charge is that its effective density can be increased by the application of high negative fields to fieldplates of an MOS structure at moderate temperatures (100°–400°C).²¹ This increase is proportional to the applied field as well as the initial Q_f density. The interface trapped charge density also increases as a result of negative field application. Such an effect can lead to instabilities in p-channel MOS devices.

Mobile Ionic Charge; Q_m, N_m . The mobile ionic charge is primarily due to the positive alkali ions, Li⁺, Na⁺, K⁺, and also possibly H⁺. In addition, it is possible to observe charge effects due to the larger negative ions such as F⁻, Cl⁻, and also Cs⁺, Au⁺, and the like. These latter ions normally do not migrate at typical device temperatures, however, and will not lead to instabilities. Likewise, their presence is more difficult to detect. The field-induced “drift” of the alkali ions is the leading cause of instabilities in MOS devices²⁵ and the rate of drift is inversely proportional to ion size (Li⁺ > Na⁺ > K⁺).

Almost every semiconductor device fabrication step can be a source of ionic contamination. Elaborate steps have been established to minimize their effect. Since this is impossible on an absolute basis, special “gettering” processes have been developed.²⁶ These involve the use of phosphorus glass, chlorine species, or other materials which can complex or getter impurity ions from the oxide and render them inactive. Dense dielectric films such as silicon nitride are also used to mask ionic impurities from entering the oxide.

Interface Trapped Charge; Q_{it}, N_{it}, D_{it} . Closely related in physical origin to the fixed oxide charge is the main form of interface trapped charge. Both charges arise from the formation of partially ionized silicon species during the thermal oxidation process. The main difference is that Q_{it} may be charged or discharged as a function of surface potential, while Q_f is not in electrical communication with the silicon and remains charged. Interface trapped charge does have many of the same process dependencies as Q_f ,

such as oxidation temperature, silicon orientation, annealing conditions, etc. One significant difference between the two charges, however, is that interface traps can be complexed at low temperatures (300°-500°C) with an active hydrogen species and thus their effective density reduced significantly. As-oxidized N_{it} densities are normally in the 10^{12}cm^{-2} range, while after a 400°C forming gas anneal their values drop to below 10^{10}cm^{-2} .

Other types of interface trapped charges result from ionizing radiation, and the presence of heavy metals (Fe, Cu) at the Si-SiO₂ interface. Both result in the same kind of charge formation with respect to the silicon band gap.

Oxide Trapped Charge; Q_{ot} , N_{ot} . The fourth type of oxide charge is due to the presence or generation of trapped holes or electrons in the oxide. Generally these are produced by ionizing radiation, avalanched junctions, high currents through the oxide, or other reactions which either tend to break Si-O bonds in the oxide or otherwise lead to carrier trapping on sites or traps already present in the oxide.²⁷ Charge trapping, either due to ionizing radiation or the presence of high fields leading to avalanching, have been cause for concern in the past and will be even more of a problem as device geometries shrink and radiation producing processes are employed. Some of these, such as sputtering, plasma deposition and etching, and electron beam/x-ray lithography, result in considerable electron-hole trapping and these trapped charges often are not easily removed. It is possible, however, to modify oxidation processes such that oxide charge trapping is minimized. Several studies have been reported which indicate process effects on oxide trapping.^{28,29}

Measurement and Control of Oxide Charges. A number of methods have been developed for measuring the effective density of charges associated with thermally oxidized silicon. It is beyond the scope of this chapter to discuss them all but the reader is referred to reviews by Nicollian and Brews,⁷ Bartelink³⁰ and Goetzberger et al.³¹ These and other reviews describe measurements particularly of the interface trapped charge density N_{it} and include quasistatic capacitance-voltage analysis, deep level transient spectroscopy, conductance-voltage analysis, and others. These methods generally permit measurement of interface trap charge density as a function of energy across the middle portion of the silicon band gap.

The high frequency capacitance-voltage (C-V) technique, however, is the most suitable for on-line measurement and control of charges in oxidized silicon structures.^{32,33} It generally involves a high frequency (1 MHz) capacitance measurement as a function of dc bias across an MOS capacitor. The latter consists of an evaporated metal field plate (normally an aluminum dot evaporated through a metal mask) over a 0.1 to 0.2 μm thermal silicon oxide. The substrate should be medium doped (10^{15} to 10^{16}cm^{-3}) silicon of the appropriate orientation. Fixed oxide charge N_f , mobile ionic charge N_m , and oxide trapped charge N_{ot} can all be determined rapidly using C-V analysis. N_f is determined by the expression

$$N_f = (-V_{FB} + \phi_{MS}) \frac{C_o}{q} = (-V_{FB} + \phi_{MS}) \frac{k_o \epsilon_o}{qx_o} = (-V_{FB} + \phi_{MS}) \frac{2.13 \times 10^{10}}{x_o (\mu\text{m})} \quad [14]$$

where V_{FB} = flatband voltage

ϕ_{MS} = metal-semiconductor work function difference

C_o = oxide capacitance

q = electronic charge

k_o = oxide dielectric constant

ϵ_o = permittivity of free space

x_o = oxide thickness in micrometers

In the case of N_m , a bias-temperature stress test is used to measure total impurity ion content. N_m is determined by the following expression which is based on Equation 14:

$$N_m = (\Delta V) \frac{2.13 \times 10^{10}}{x_o (\mu m)} \quad [15]$$

where ΔV is the difference between flatband voltages of C-V plots after positive and negative stress tests. The conditions for the test are normally +50 and -25 V/ μm at 300°C for 2 minutes. An example of N_f and N_m measurement using C-V analysis is presented in Figure 8, where C-V plots before and after bias-temperature stress tests are shown.

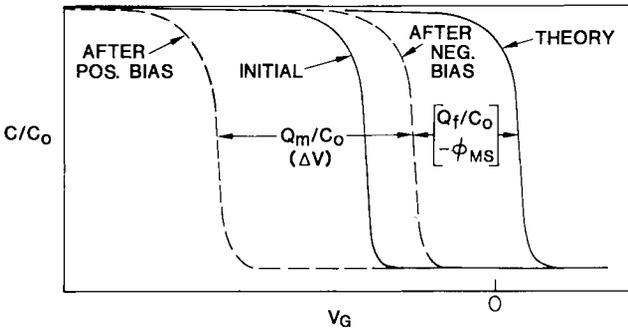


Figure 8: The determination of fixed oxide charge density Q_f and mobile ionic charge density Q_m in thermal silicon dioxide using the MOS capacitance-voltage technique.

Oxide trapped charge density N_{ot} is normally determined following a procedure similar to that used for ionic contamination. In the case of N_{ot} , however, the C-V plots are compared before and after the charge trapping process and no elevated temperature is employed.

As gate oxides for VLSI MOS devices approach the 200Å thickness range, the selection of a proper value of work function difference ϕ_{MS} in Equation 14 becomes quite critical. In fact, it has been determined that the effective value of ϕ_{MS} can depend on the processing sequence³⁴ so that the proper choice of ϕ_{MS} becomes especially difficult.

4.1.2 Dopant Redistribution. It has been known for many years that silicon dopants segregate preferentially on one or the other side of the Si-SiO₂ interface during thermal oxidation. This dopant pile-up or depletion at the silicon surface can affect several critical device characteristics, such as MOS threshold voltage, junction breakdown voltage, and others. The amount and nature of the dopant segregation depends on several factors, the most important being the dopant segregation coefficient m (m = ratio of equilibrium concentration of dopant in the silicon to that in the oxide), the oxidation rate, and the relative dopant diffusion rates in the oxide and silicon.³⁵ In general, n-type dopants such as boron pile up ($m > 1$) and p-type dopants such as phosphorus ($m < 1$) and p-type dopants such as boron deplete ($m < 1$). Four possible redistribution effects are shown in Figure 9.³⁵ Cases A and C represent normal p-type depletion and n-type pile-up. However, the presence of hydrogen causes increased boron diffusion in the oxide (case B) and greater depletion. In case D, increased gallium diffusion in the oxide leads to depletion, even though $m > 1$ for this p-type dopant.

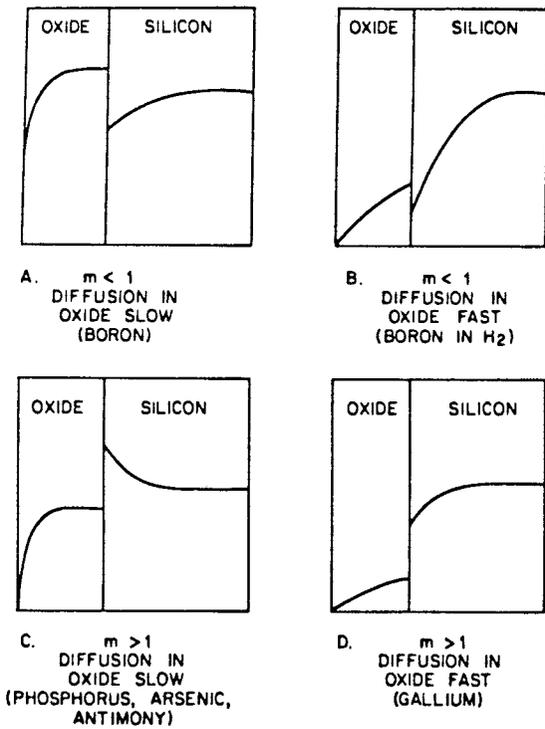


Figure 9: Schematic illustration of the dopant distribution as a function of position in the SiO₂/Si structure indicating the redistribution and segregation of dopants during silicon thermal oxidation (after Grove et al³⁵).

Because of the various factors affecting redistribution mentioned above (segregation coefficient, oxidation rate, dopant diffusion rates), oxidation variables such as temperature, ambient type, and others will determine the amount of depletion or pile-up. For instance, steam ambients and/or lower oxidation temperatures generally result in more redistribution. Furthermore, oxidation pressures which do not allow equilibrium to be established also result in modified redistribution profiles. The overall mechanism of dopant redistribution remains unclear due to its complexity, and computer techniques will be required to properly model the process.

4.2 Effects of Surface Properties on the Oxidation Reaction

4.2.1 Silicon Orientation. It has been observed for a number of years that the silicon thermal oxidation process is dependent on the substrate silicon orientation (see Figures 1 and 2), and that this dependence is more pronounced at lower temperatures (900°C). The effect was first reported by Ligenza³⁶ and later verified by Pliskin.³⁷ It is now established that the orientation dependence is reflected primarily through the linear rate constant and is undoubtedly associated with the Si-SiO₂ interface reaction (see Figure 6). While a quantitative mechanism has still not been established, it is assumed that the effect is based on the bond density or "availability," and it is observed that the (111) orientation results in the fastest interface reaction while (100) silicon is the slowest. As the oxidation temperature decreases and the oxidation process becomes more surface reaction controlled, the orientation effect thus becomes more pronounced.

An orientation effect related to oxide charge density (Q_f and Q_{it}) similar to that for the interface oxidation reaction has also been observed. The similarity of these two effects leads to the conclusion that oxide charge origin (Q_f and Q_{it}) is directly related to the oxidation process at the Si-SiO₂ interface. A comparison of these two orientation dependent effects is shown in Table 4.

Table 4: Dependence of Oxidation Linear Rate Constant and Oxide Fixed Charge Density on Silicon Orientation

OXIDE TYPE	OXIDATION TEMPERATURE (°C)	SILICON ORIENTATION	B/A (μm/hr)	Q_f/q (cm ⁻²)
DRY O ₂	1200	(111)	1.12	1.7 X 10 ¹¹
		(110)	0.90	0.6
		(100)	0.56	0.2
WET O ₂ (95°C H ₂ O)	1200	(111)	14.40	4.0
		(110)	12.0	1.7
		(100)	7.2	1.2

As predicted from the general relationship, the parabolic rate constant is relatively independent of silicon orientation. However, there is some indication that for lower temperatures or thinner oxides the silicon substrate may cause some oxide structural effect, which in turn might result in an orientation dependence of the parabolic rate constant.

4.2.2 Dopant Concentration. It was observed some time ago that areas of semiconductor devices having surface dopant concentrations greater than 10^{19}cm^{-3} , i.e. emitter regions in bipolar transistors, exhibit higher oxidation rates than adjoining lightly doped silicon.^{37,38} Experiments indicated that the effect for n-type dopants is more pronounced at lower temperatures or thinner oxides, while for boron doping, an oxidation increase is noted to some extent over the entire temperature range. These results implied that high concentrations of phosphorus affect the oxidation process primarily through the surface reaction rate constant B/A , while for boron both B and B/A contribute to the increase in oxidation rate.

More recently Ho and co-workers³⁹ investigated in more detail the effect of phosphorus concentration on the thermal oxidation process. They have attributed this effect to vacancy generation resulting from high phosphorus concentrations in silicon. These vacancies provide a driving force for increased interface reaction rates and relate to the more recently proposed atomic model of oxidation to be discussed in a later section of this chapter. Boron tends to segregate into the oxide and will therefore tend to affect the parabolic rate constant which is more important at higher temperatures. As boron concentration increases, however, that at the silicon surface will also increase and therefore raise the interface oxidation reaction rate. Typical oxide thickness - time data are shown in Figure 10 for phosphorus.³⁹ Note that the increase of oxidation rate is more pronounced at lower temperature (800°C) and for thinner oxides, which reflects the greater contribution of the linear rate constant B/A under these conditions.

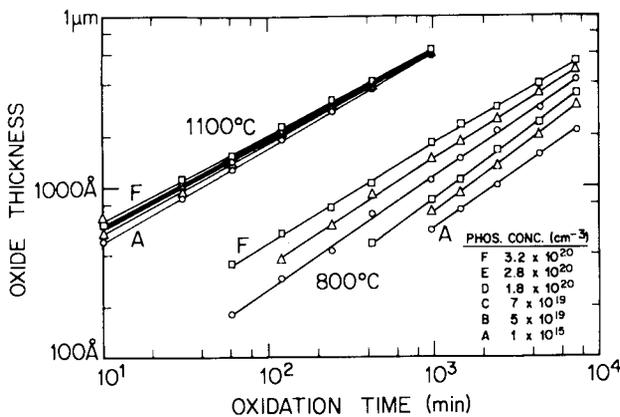


Figure 10: Oxide thickness vs oxidation time for oxidation time for silicon oxidation in dry oxygen at 800° and 1100°C using (111) silicon substrates doped with phosphorus up to solid solubility (after Ho et al³⁹).

4.2.3 Surface Preparation. Some of the effects mentioned above relating silicon surface properties to oxidation rate indicate that such effects are more prevalent at lower temperatures. A similar effect has been noted for the surface condition of the silicon prior to oxidation. If the silicon surface is not cleaned uniformly it is much more likely that resulting patches of non-uniform thickness of oxide can be observed at 1000°C or less as opposed to 1200°C. Likewise, if differences in oxidation rate due to variations in physical treatments such as polishing, lapping, etc., occur, they will more likely occur at lower temperatures or in the thinner oxide regime. This again reflects the fact that surface related effects are associated with the linear rate constant which is much more oxidation rate controlling at lower temperatures.

Somewhat related to these effects is the observation made by Schwettmann and others that the type of pre-oxidation cleaning treatment can affect the subsequent oxidation rate. An example of this effect is indicated in Figure 11, where oxidation thickness-time data are shown for three different cleaning treatments.⁴⁰ Since the treatment incorporating ammonium hydroxide results in the slowest oxidation, it is postulated that an inhibiting nitride layer is formed on the silicon which retards the subsequent oxidation process. More work is necessary to better understand the mechanisms involved in these effects, since the use of very thin oxides for VLSI applications will require more stringent control of thin oxide thickness and uniformity.

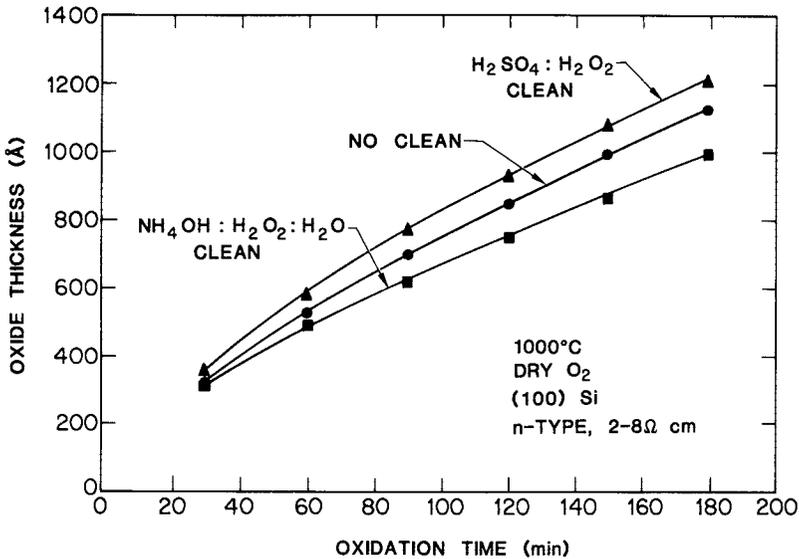


Figure 11: Effect of pre-oxidation cleaning process on oxide growth rate (after Schwettmann et al⁴⁰).

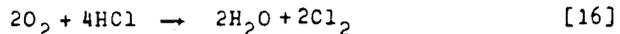
4.3 Effects of Ambients on the Oxidation Reaction

4.3.1 Ambient Type. The data presented in Figures 1 and 2 and in Table 2 demonstrate the considerable difference in silicon oxidation rates between dry oxygen and steam ambients. While several physical factors related to Equation 6 can contribute to such differences, in this case the primary effect appears to be oxidant solubility in the oxide. The solubility of water in thermal SiO_2 is three orders of magnitude greater than that of oxygen.¹² As indicated earlier, investigations are now being conducted to determine mechanisms of oxidant diffusion through the oxide and other reactions which occur during silicon thermal oxidation. From a practical consideration, dry oxygen is more commonly used for preparing thinner, reproducible oxides, such as MOS gates, while steam is employed for thicker oxides normally used for isolation outside the active junction area. Variations in O_2 or H_2O partial pressures are employed for optimizing thickness control in certain applications, while small amounts of chlorine are sometimes added for impurity gettering. Otherwise, no other types of oxidation ambients have been reported.

4.3.2 Chlorine Additions. In the early 1970's it was reported that the addition of a small amount of a chlorine species, either HCl , Cl_2 , or some organochloro component, to the oxidation ambient can provide several beneficial effects with respect to the resulting oxide.^{41,42} These benefits include improved oxide charge stability, fewer pinholes, reduced interface trapped charge density, and better device performance in general. It is now quite common to add a few percent (1-5%) of a chlorine-containing compound, such as HCl , to oxidation ambients used for both MOS and bipolar devices. It is also common to "clean" oxidation tubes with oxygen- or nitrogen-chlorine mixtures prior to oxidation.

It has been determined that a certain amount of chlorine remains in the oxide after oxidation in a chlorine-containing ambient, and that this chlorine residing very near the Si-SiO_2 interface can provide improved passivation. If too much chlorine is present, however, device properties become degraded and the oxide may blister and peel off the silicon. In most cases the oxidation rate increases due to chlorine additions—the greater the chlorine content and the higher the oxidation temperature, the greater the rate increase.

The mechanism for chlorine oxidation is not completely understood, but it is believed that the reaction:



occurs. It follows that Cl_2 must be the primary chlorine species incorporated in the oxide and is driven to the Si-SiO_2 interface by a field in the oxide during oxidation. If water is added to the ambient, the above chemical reaction is driven to the left and less chlorine is incorporated in the oxide. This is supported by the fact that little chlorine is observed in steam produced oxides. Figure 12 includes an Auger profile of an 1100°C chlorine-

containing oxide as well as a plot of chlorine concentration at the Si-SiO₂ interface as a function of oxide thickness.⁴³ The latter demonstrates that chlorine content in the oxide increases with oxidation time.

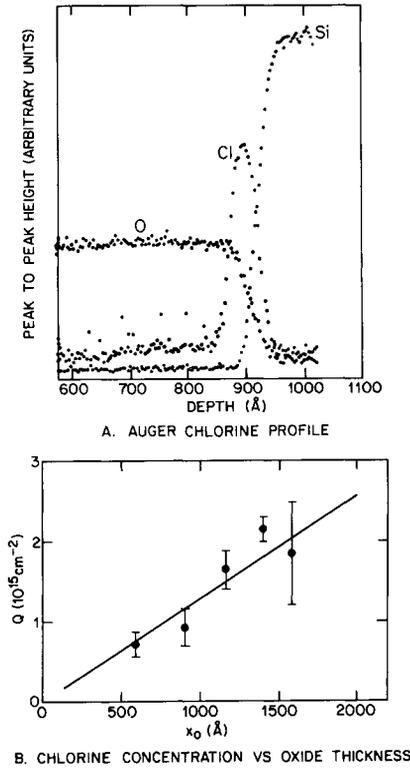


Figure 12: Auger sputter profile (A) and chlorine concentration vs oxide thickness (B) for thermal oxide prepared in 5% HCl/O₂ ambi at 1100°C using (100) silicon (after Rouse et al⁴³).

4.3.3 Nitridation. So-called inert gases, such as nitrogen, argon, and others have been used for many years to dilute oxygen ambients, reduce charge densities, promote dopant diffusion, and to provide an oxygen-free ambient for cooling and pulling wafers. It was determined, however, that while argon and helium are inert, nitrogen will react at elevated temperatures with silicon. In fact, Raider found that a silicon-nitrogen reaction occurs at the Si-SiO₂ interface even in the presence of appreciable thicknesses of thermal oxides.⁴⁴ In general, however, attempts to produce silicon nitride films by direct thermal reaction were not successful.

More recently, improved gas purity and techniques have permitted

reasonably good silicon nitride films to be produced by reacting N_2 or NH_3 directly with silicon at elevated temperatures (1000° - $1300^\circ C$), with or without the use of plasma excitation.^{45,46} These thin films exhibit improved properties over thermal oxides (increased breakdown, fewer pinholes, higher dielectric constant) and are in the thickness range, 50-100Å, which is required for advanced VLSI MOS structures. However, continuing process and reproducibility problems appear to preclude their use in actual devices. On the other hand, it has been subsequently reported⁴⁷ that advantages of the silicon nitride might be achieved while still maintaining the superior interface properties of thermal SiO_2 by converting the outer portion of the oxide to nitride using an NH_3 anneal at temperatures greater than $900^\circ C$. The resulting oxy-nitride structures exhibit greatly improved dielectric breakdowns, resistance to subsequent oxidation impurity, and dopant diffusion, and improved integrity—all of which make them much more suitable for submicrometer MOS device application.⁴⁸ Auger analysis has indicated that the actual amount of nitrogen in these nitrified oxide films is fairly low, most of it concentrated at the outer surface of the oxide or near the Si- SiO_2 interface.⁴⁸

4.3.4 Oxidant Pressure. High pressure oxidation of silicon was first employed more than twenty years ago by Ligenza and Spitzer³ in order to accelerate the oxidation process at lower temperatures. They employed a stainless steel “bomb” which contained the silicon wafer in a steam ambient. While devices were passivated by this method, it did not become widely accepted in the industry. More recently, Panousis and Schneider⁴⁹ reported a high pressure oxidation system more suitable for production. This system allowed continuous flow of the pressurized ambient through a quartz tube and was the basis for today’s commercial systems.⁵⁰⁻⁵² These systems can be used for dry O_2 or steam up to 25 atm, and have capacities of up to 200 four-inch diameter wafers. Other experimental systems, similar to the closed bombs of Ligenza and Spitzer, employing dry O_2 up to 750 atm, have been reported,⁵² but are not used commercially.

As indicated in Fig. 6, the general relationship predicts that both the parabolic and linear rate constants should be directly proportional proportional to ambient pressure through C^* , the equilibrium concentration of the oxidant in the oxide. Thus, the time required to produce a given oxide thickness should be inversely proportional to pressure. This will provide several advantages, especially with regard to today’s small geometry devices having very shallow junctions. Junction movement during oxidation as well as dopant redistribution will be minimized. It has also been found that defect levels are reduced. These improvements are either due to the shorter times and lower temperatures employed for the high pressure oxidation process, or because of the increased oxidation rate.

Recent kinetic studies of silicon oxidation in steam up to 20 atm have indicated that both the linear and parabolic rate constants are directly proportional to steam pressure from 800° to $1000^\circ C$, as predicted above.⁵³ For dry O_2 , however, while the parabolic rate constant has a linear ($B \propto P$) dependence, the linear rate constant falls off with pressure⁵⁴ ($B/A \propto P^{0.7}$). Typical thickness-time oxidation data for high pressure steam are shown in Fig. 13.

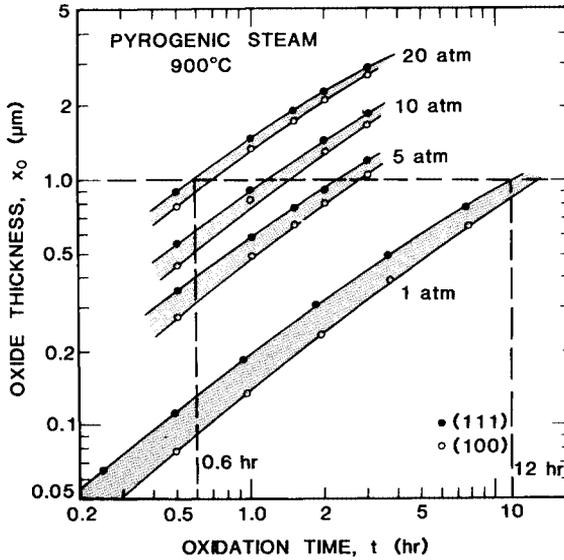


Figure 13: Oxide thickness vs oxidation time for silicon oxidation in pyrogenic steam (~640 Torr) at 900°C and various pressures (after Razouk et al⁵³). Reprinted by permission of the publisher, The Electrochemical Society, Inc.

Reduced partial pressures of O₂ and H₂O in inert carrier gases have been used for producing thin, controlled oxide films. In general, rate constants are proportional to oxidant pressures down to about 0.1 atm. However, as film thicknesses approach 200Å, the deviation in oxidation kinetics mentioned earlier occurs and pressure dependencies are not clearly understood.

5. OXIDATION MECHANISM

5.1 Atomic Reactions

As indicated in Section 2 above, the thermal oxidation of silicon in either dry oxygen or steam can be characterized by the general relationship Eq. 6. However, the actual atomic reactions at the Si-SiO₂ interface during thermal oxidation have not been well understood in the past. More recently efforts have been made to better characterize these reactions, especially the mechanisms associated with the rate constant k in Eq. 6. It has been proposed^{55,56} that at least three individual phenomena occur at the Si-SiO₂ interface as oxidation proceeds. These are shown in Fig. 14, and are also proposed to contribute to other effects observed during thermal oxidation.

First, each SiO₂ molecule produced occupies considerably more volume than that of the silicon reacted. Thus, appreciable strain results at the interface region as is indicated in the upper box of Fig. 14. This compressive

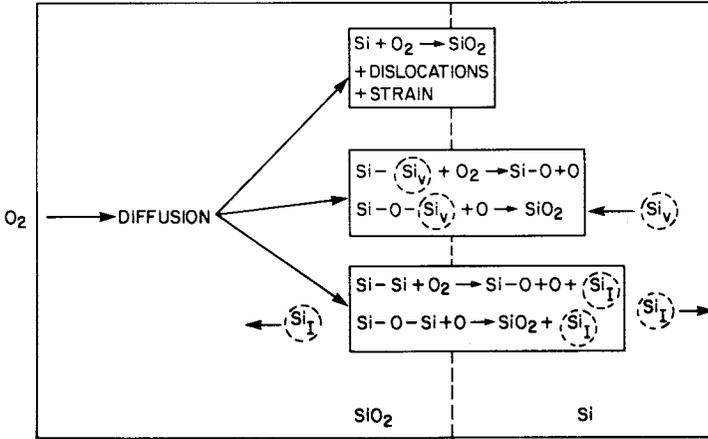


Figure 14: Proposed mechanisms occurring at the Si-SiO₂ interface during silicon thermal oxidation (after Plummer⁵⁵). These figures were originally presented at the Spring 1981 Meeting of The Electrochemical Society, Inc., held in Minneapolis Minnesota.

stress accounts in part for the excellent passivation property of thermal silicon dioxide, but can also lead to lattice mismatch and other defects, if some mechanism does not permit the stress to be relieved.

One of the ways of relieving this stress is shown in the lower box of Figure 14, which is the generation and diffusion away from the interface of silicon interstitials. Silicon interstitials produced by the oxidation process have been proposed to cause enhanced diffusion of dopants in the silicon during thermal oxidation—OED,⁵⁷ as well as to promote stacking fault formation—OISF.^{58,59} It has also been proposed that they contribute to charges such as Q_f or Q_{it} in the oxide.^{55,60} For the enhanced dopant diffusion effect, a relationship relating the oxidation rate and effective diffusion coefficient has been developed⁵⁷ which agrees reasonably well with experimental data:

$$D_{eff} = D_i + K \left(\frac{dx}{dt} \right)^n \quad [17]$$

where D_i = the normal diffusion coefficient due to vacancy mechanisms and $K(dx/dt)^n$ = the silicon interstitial contribution. The value of n has been determined to range from 0.2 to 1.0. An expression for stacking fault generation and retrogrowth has also been proposed:⁵⁹

$$\frac{dl}{dt} = K_1 \left(\frac{dx}{dt} \right)^n - K_2 \quad [18]$$

where dl/dt = the growth/retrogrowth rate, K_2 = the shrinkage rate in the absence of oxidation, and $K_1(dx/dt)^n$ = the interstitial contribution to the

growth rate. The best value of n appears to be 0.4. A number of process variables in addition to oxidation rate dx/dt , such as ambient type, silicon orientation, HCl presence, and mechanical damage, have been shown to affect oxidation-enhanced diffusion and stacking fault generation through the formation of silicon interstitials during oxidation.⁵⁵

The third reaction or mechanism proposed to occur at the Si-SiO₂ interface indicated by the middle portion of Fig. 14, involves the possible contribution of silicon vacancies to the oxidation reaction. Under normal conditions (lightly doped silicon) the vacancy concentration is reasonably low and oxidation proceeds, giving rise to the mechanisms already discussed (strain generation and silicon interstitial effects). However, for heavier dopant concentrations ($C_B = \geq 10^{20} \text{cm}^{-3}$) enough silicon vacancies are present so as to provide additional free volume which can accommodate additional interstitials and as a result, the oxidation rate increases. This mechanism has been discussed and modeled by Ho and Plummer³⁹ which helps to explain the well-known heavily doped oxidation effect discussed earlier.

An understanding of the detailed mechanisms of reactions occurring at the Si-SiO₂ interface during thermal oxidation, such as those described above, and the relationship to associated reactions occurring during the oxidation process, should provide the basis for producing and controlling the thin oxides required for VLSI circuits. It will also be necessary, however to understand the details of the diffusion of oxidizing species through the oxide. Various techniques, such as radioactive exchange,^{61,62} are being used for this purpose. Finally, an actual physical "picture" of the oxidized silicon interface region will be required if all the observed effects and mechanisms of oxidation are to be explained. The current status of the clarity of the picture is presented in the next section.

5.2 Structure of the Si-SiO₂ Interface

For at least twenty-five years, investigators have speculated about the nature and physical structure of the Si-SiO₂ interface region associated with thermally oxidized silicon. Most of the earlier speculations were based on electrical characteristics of the interface and MOS devices. It was initially believed that the interface region depth was less than 200 Å but lack of sensitive instrumentation prevented any detailed knowledge about how much less it might be. It was also believed that some of the oxide charges which resided in this region were due to disrupted Si-Si or Si-O bonds or other similar defects. Not much more was known about the Si-SiO₂ interface.

In the 1970s and early 1980s, the sophistication and resolution of analytical tools have improved considerably. As a result, and because of the increased emphasis of complex device structures, many investigations regarding the structure of the Si-SiO₂ interface have been carried out. These have involved Auger spectroscopy, x-ray diffraction, electron microscopy, x-ray photoelectron spectroscopy, Rutherford backscattering, secondary ion mass spectrometry, electron spin resonance, ellipsometry, photo emission, and numerous others. In addition, theoretical computer modeling has been used to predict interface properties. The results obtained

from these various types of analysis have been correlated with those obtained using improved electrical techniques such as quasistatic C-V, DLTS, and conductance-voltage measurements. All of these plus actual device measurements have provided considerable insight into the exact nature of the Si-SiO₂ interface. It is not possible here to reference even a small number of the papers concerning the evaluation of the Si-SiO₂ interface; however, some of the more comprehensive reviews on the subject of surface and interface analysis may be consulted.⁶³⁻⁶⁵

The current understanding of the nature of the Si-SiO₂ interface in thermally oxidized silicon may be summarized as follows. First, it is generally agreed that the transition region between silicon and the bulk oxide is no more than 10Å or even one or two monolayers. In this region, the composition changes rapidly from Si to SiO₂; and the oxide is apparently crystalline in nature immediately adjacent to the silicon. As a result, the physical, electrical, and chemical properties of the oxide in this transition region are markedly different from those of amorphous SiO₂ and affect the net properties of oxides up to 200Å or more. There is also a good possibility that the Si-O bond angles in the transition region and beyond (up to 50Å) are strained, which can also affect oxide properties.

Depending on the oxidation conditions and the silicon orientation, a limited number of silicon atoms at the silicon surface (as few as one in 10⁵) might not be bonded to oxygen and thus could act as trapping sites (Q_{it}). Similarly, some of the silicon atoms on the oxide side of the interface might be disconnected from adjacent oxygen ions (or certain oxygen atoms might be missing) and these silicon species could also act as charge or trapping sites (Q_f). Although these specific defects or trapping sites have yet to be positively observed, the evidence for their presence is fairly conclusive as a result of recent investigations.^{18,19,66,69} A proposed cross section structure of the Si-SiO₂ interface region is presented in Fig. 15, which includes the possible origin of the four types of oxide charges.

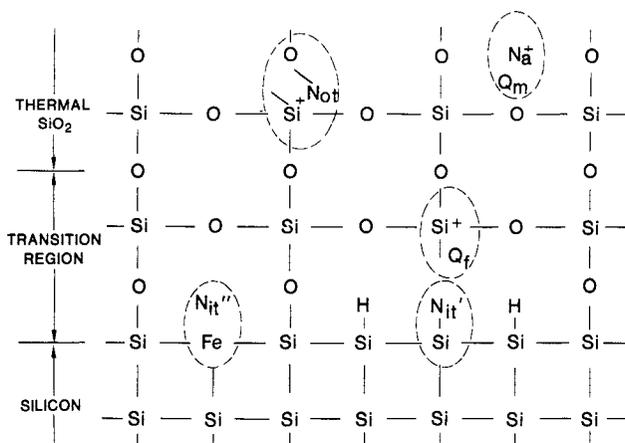


Figure 15: Idealized structure of thermally oxidized silicon interface region showing possible origins of four types of oxide charge.

concept of the Si-SiO₂ interface structure is the most likely to date. As more sophisticated analysis equipment is developed, it is reasonable to assume that a more accurate description of the Si-SiO₂ interface will emerge. Under any consideration, this interface will play a most important role in future semiconductor devices.

6. OTHER OXIDATION PROCESSES

6.1 Assisted Oxidation

It is known that various types of radiation can affect the thermal oxidation process. The radiation can involve electrons, protons, x-rays, photons (including laser, UV or IR radiation), gamma rays, ions, microwave, and various types of plasmas. Generally, the net effect of the radiation is to increase the rate of oxidation. This can occur by either the activation of the oxidizing species in the gas phase, whereby both the parabolic and linear rate constants are increased, or by breaking Si-O bonds in the oxide which could cause increased diffusion of the oxidant through the oxide. It is also possible that the interface reaction might be increased. Radiation processes may therefore be employed to increase the rate of thermal oxidation at lower temperatures. An example of the use of plasma to form thermal oxides of reasonable thickness at very low temperatures is demonstrated in Figure 16.⁷⁰

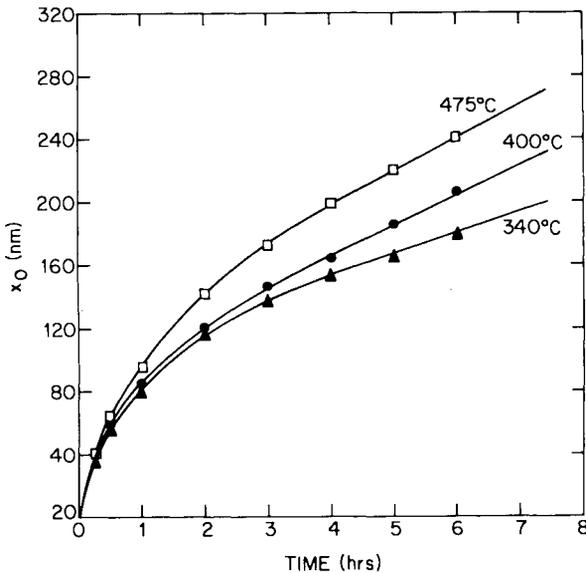


Figure 16: Oxide thickness vs oxidation time for silicon oxidation in dry oxygen plasma (30 mTorr, 1 kW, 0.5 MHz) (after Ray and Reisman⁷⁰). Reprinted by permission of the publisher, The Electrochemical Society, Inc.

On the other hand, these radiation effects might lead to undesirable results, such as excess oxide charge formation and non-uniform oxide growth. In fact the use of many radiation-producing processes in VLSI device fabrication has been the basis for considerable concern by investigators studying the effects of plasma etching, sputtering, electron beam and x-ray lithography, and other advanced types of processes.⁷¹

Electric fields similar to radiation influence the oxidation mechanism. Electric fields have been used to both affect the oxidation reaction and to study its mechanism. It has been known since the original experiment of Jorgensen⁸ that an electric field is probably present in the oxidizing silicon system. Furthermore, the application of an external field on the structure can accelerate or retard the oxidation reaction. The complete picture is still not clear, although some of the radiation processes described above (plasma, ions, etc.) include the effects of an electric field as part of the means of accelerating oxidation. In addition, the anodization process has also been reported to be effective in oxidizing silicon and compound semiconductors at low temperatures.⁷²

6.2 Silicon-Containing Materials

Typically, over the past twenty-five years, the chemical elements used to fabricate silicon semiconductor devices have consisted of silicon, oxygen, and aluminum, with added dopants. More recently, silicon-containing materials such as silicon nitrides and refractory metal silicides have been employed as device components. As it turns out, these alternate materials can be thermally oxidized to form SiO_2 in a manner similar to silicon oxidation. In fact, the oxide formed is essentially identical to SiO_2 associated with silicon.

The mechanisms involved in three different silicon-containing materials used in devices are illustrated in Figure 17. The oxidation of polycrystalline silicon (used for MOS gates and interconnects) most closely resembles single crystal silicon oxidation, except for the presence of grains and grain boundaries in the poly-Si (Figure 17A). This tends to make the process more complex and less controllable.^{73,74} Nevertheless, thermal oxides are used successfully to passivate and insulate polycrystalline silicon in today's device structures. In the case of silicon nitride (Si_3N_4) (used to mask oxidation and impurity ions) the mechanism of conversion to SiO_2 is similar to that of silicon, but nitrogen produced by the reaction must diffuse out of the oxide (Figure 17B). The kinetics of Si_3N_4 oxidation appears to follow the general oxidation relationship but the rate is at least an order of magnitude less than for silicon.^{75,76} The SiO_2 produced is identical to that obtained by silicon oxidation.

The refractory metal silicides (WSi_2 , TaSi_2 , TiSi_2 , MoSi_2) are currently used in devices in combination with polycrystalline silicon as interconnects and MOS gates. An oxidized TaSi_2 /poly-Si structure is shown in Figure 17C. Thermal oxidation of the silicide apparently proceeds by the diffusion of silicon atoms from the underlying poly-Si layer up through the metal silicide.⁷⁷⁻⁷⁹ Assuming a sufficient supply of substrate silicon, stoichiometric SiO_2 is produced over the silicide, the latter not being consumed by the

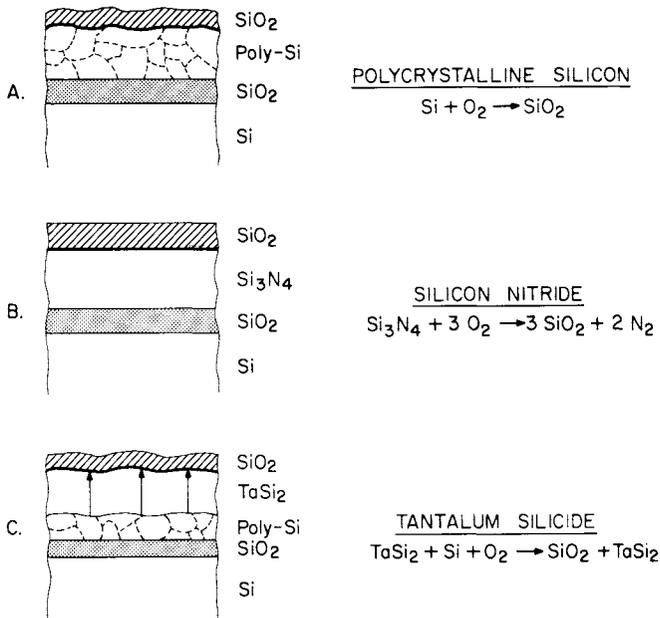


Figure 17: Schematic illustration of thermal oxidation of silicon-containing materials in which silicon dioxide is reaction product: (A) polycrystalline silicon, (B) silicon nitride, (C) refractory metal silicide.

reaction. Analysis of the data indicates that the rate determining step is primarily diffusion of the oxidizing species through the oxide. Values of B, the parabolic rate constant are almost identical to those obtained for conventional silicon oxidation, while B/A values are much higher.⁷⁷⁻⁷⁹ If no silicon is present beneath the silicide, the resulting oxides are mixtures of refractory metal and silicon oxides and are generally not stable or reproducible.

Typical thickness-time data for the thermal oxidation of tantalum silicide deposited over polycrystalline silicon are shown in Figure 18.⁷⁷ Single crystal silicon oxidation data are included in the figure for comparison.

6.3 Other Semiconductors

One of the reasons that germanium did not become a significant factor in semiconductor device technology was that it could not be passivated through thermal oxidation. Thermal germanium oxides are generally unstable and decompose during subsequent high temperature processing. Alternate approaches, such as anodizing or deposition of silicon oxides, have been attempted, but the almost ideal passivating properties of the thermally oxidized silicon have helped to make this semiconductor the main device material over the past twenty-five years.

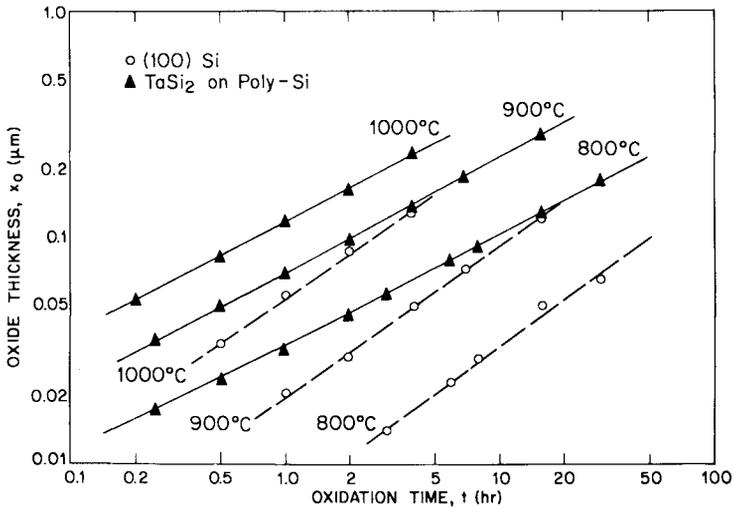


Figure 18: Oxide thickness vs oxidation time for thermal oxidation of TaSi₂/poly-Si structure in dry oxygen at various temperatures (after Razouk et al⁷⁷).

Compound semiconductors, such as gallium arsenide, indium antimonide, gallium phosphide, and others, have also been difficult to passivate by thermal oxidation. Like germanium, the oxides formed are not generally stable. In addition, the presence of two or more species in the semiconductor causes competing reactions during the oxidation process and leads to non-uniform films which depend on such factors as composition, orientation and other variables. Some success has been achieved using anodic oxidation for passivating some of the compound semiconductors. In addition, chemical vapor deposited SiO₂ or Si₃N₄ provide satisfactory diffusion masks and passivating layers for compound semiconductors used for devices of various kinds. Because the nature of the interfaces associated with these passivated structures is so complex, much less is known about their chemical and electrical properties. Perhaps as more commercial applications are developed involving compound semiconductors, more efforts can be devoted to understanding their surface and interface properties. In certain cases it might even be possible to develop a suitable thermal oxidation process. Several informative reviews which deal with thermal and anodic oxidation of compound semiconductors are available.^{72,80,81}

7. FUTURE TRENDS

Considerably more stringent requirements will be placed on thermal oxides as silicon integrated circuit feature sizes move from micrometer

into the submicrometer region. Continuing device scaling will require even thinner, more reproducible oxides for gates in MOS structures. This implies improvements or modifications in several areas. First, oxide thickness will have to be controlled to even closer dimensions than it is now. Because of differences in optical and electrical properties of thin thermal oxides, improved or new thickness measurement techniques will have to be developed. Equally important will be the need to better understand the oxidation kinetics in the thin region so that control and reproducibility of the oxide thickness can be achieved.

It will also be necessary to control and understand the effects of process variables on other oxide properties such as electrical conductivity and oxide charge formation. Equally important will be minimization of defects, pinholes, and the like in these ultrathin films. Control of all these oxide properties implies a better understanding of the Si-SiO₂ interface region. As device dimensions reach a critical minimum size, statistical variations in individual oxide charge densities may not permit specific bits of the device to function.

With respect to thicker oxides used in future device structures, procedures will have to be developed to minimize or even eliminate oxide encroachment. Up to now, oxides used to isolate the individual devices have exhibited some form of "birdsbeak." Since this encroachment can be of the order of a micrometer, it is obvious that this much "lost" area cannot be tolerated in submicrometer structures. These oxide isolation problems are being solved in part by (1) the use of other types of nitride masking procedures which retard encroachment,⁸² (2) the fabrication of etched trench structures which can be filled by various types of dielectric materials,⁸³ or (3) selective epitaxial growth of silicon within insulating walls of silicon oxide.⁸⁴

The final answer to controlling all the above properties may lie in our ability to properly model the oxidation process itself and the resulting effects on the oxide properties, and ultimately, the device parameters. Since most future devices will involve very complex, three dimensional configurations, our ability to model multidimensional aspects (two and three dimensional) of oxide formation must be greatly improved. Such modeling will of course be based on advanced computer techniques.

The trend for all semiconductor processing of the future includes lower temperatures and shorter times—required for maintaining the extremely small structures in VLSI devices. This may be accomplished by the use of high pressure and/or plasma-assisted oxidation. More reliable devices with better performance and tighter specifications have been produced by the use of chlorine in the oxidation ambient. More of these types of improved oxidation techniques can be expected in the future.

Finally, although thermal oxides, and silicon semiconductors, have been the mainstay of device technology for more than twenty-five years, SiO₂ will only be used in the future if it continues to satisfy the technological requirements. For specific applications involving MOS gate structures, other dielectrics such as thermal silicon nitride are being investigated. Whether these or other materials replace thermal oxides in certain cases remains to be seen. Similarly, it is reasonably certain that other semi-

conductors, e.g. GaAs, GaAlP, etc., will be used for various applications, including integrated circuits. It is reasonably certain, however, that both silicon and thermal silicon dioxide will continue to play major roles in semiconductor technology.

REFERENCES

1. C.J. Frosch, and L. Derick, *J. Electrochem. Soc.* 104: 547-52 (1957).
2. M.M. Atalla, E. Tannenbaum and E.J. Scheibner, *Bell Sys. Tech. J.* 38: 749-84 (1959).
3. J.R. Ligenza and W.G. Spitzer, *J. Phys. Chem. Solids* 14: 131-36 (1960).
4. J.A. Hoerni, paper presented at the IRE Electron Devices meeting, Washington, D.C., Oct. 1960; U.S. Patents 3,025,589 (1962) and 3,064,167 (1962).
5. D. Kahng and M.M. Atalla, paper presented at the IRE Solid State Device Research Conference, Pittsburgh, PA, June 1960.
6. W.A. Pliskin and R.A. Gdula, *Handbook on Semiconductors* (T.S. Moss, ed.), vol. 3, *Materials, Properties, and Preparation* (S.P. Keller, ed.), pp 641-687, North-Holland Publishing Co., Amsterdam (1980).
7. E.H. Nicollian and J.R. Brews, *MOS Physics and Technology*, New York: John Wiley (1981); S.A. Schwarz and M.J. Schulz, in: *VLSI Electronics Microstructure, Vol. 10* (N.G. Einsbruch and R.S. Bauer, eds) pp 29-77, Academic Press, Orlando (1985); G. Barbottin and Vapaille, *Instabilities in Silicon Devices Vols. 1 and 2*, Amsterdam: Elsevier (1986); F.P. Fehlner, *Low Temperature Oxidation*, pp 211-247, Wiley-Interscience, New York (1985).
8. P.J. Jorgensen, *J. Chem. Phys.* 37: 874-77 (1962).
9. R.M. McLouski, Paper No. 177 presented at The Fall Meeting of The Electrochemical Society, Chicago, IL., Oct. 15-20, 1967.
10. D.W. Hess and B.E. Deal, *J. Electrochem. Soc.* 124: 735-39 (1977).
11. B.E. Deal, *J. Electrochem. Soc.* 125: 576-79 (1978)
12. B.E. Deal and A.S. Grove, *J. Appl. Phys.* 36: 3770-78 (1965).
13. R.W. Dutton, et al., *IEEE J. Solid State Circuits* SC-12: 349-55 (1977); *Proc. IEEE* 69: 1305-20 (1981).
14. Y.J. Van der Meulen, *J. Electrochem. Soc.* 119: 530-34 (1972).
15. R. Ghez and Y.J. Van der Meulen, *J. Electrochem. Soc.* 119: 1100-06 (1972)
16. J. Blanc, *Appl. Phys. Lett.* 33: 424-6 (1978).
17. H.Z. Massoud, J.D. Plummer and E.A. Irene, *J. Electrochem Soc.*, 132:2685-2700 (1985).
18. C.J. Han and C.R. Helms, *J. Electrochem. Soc.*, to be published.
19. S.T. Pantelides, (ed.), *The Physics of SiO₂ and its Interfaces*, New York: Pergamon Press (1978); G. Lucovsky, S.T. Pantelides and F.L. Galeener, (eds), *The Physics of MOS Insulators*, New York: Pergamon Press (1980).
20. See for instance C.M. Osburn and D.W. Ormond, *J. Electrochem. Soc.* 119: 591-97, 597-603 (1972); P. Solomon, *J. Vac. Sci. Tech.* 14: 1122-30 (1977).
21. B.E. Deal, M. Sklar, A.S. Grove and E.H. Snow, *J. Electrochem. Soc.* 114: 266-74 (1967).
22. Y.C. Cheng, *Prog. Surface Science* 8: 181-218 (1977).
23. B.E. Deal, *J. Electrochem. Soc.* 127: 979-81 (1980); *IEEE Trans. Electron Devices* ED-27: 606-8 (1980).
24. B.E. Deal, *J. Electrochem. Soc.* 121: 198C-205C (1974).
25. J.R. Davis, *Instabilities in MOS Devices*, New York: Gordon and Breach Science Publishers (1981).

26. P. Balk and J.M. Eldridge, *Proc. IEEE* 57: 1558-63 (1969).
27. For a discussion of trapping in silicon oxides, see: special issues on device radiation effects, *IEEE Trans. Nucl. Sci.*, Dec. issues, Vols. NS21-28, 1974-81; C.T. Sah, *IEEE Trans. Nucl. Sci.* NS-23: 1563-68 (1976); W.R. Dawes, Jr., G.F. Derbenwick and B.L. Gregory, *IEEE J. Solid-State Circuits* SC-11: 459-66 (1976).
28. R.R. Razouk and B.E. Deal, DARPA Final Technical Report No. NR 322-080, Contract No. N00014-79-C-0297, April 1982.
29. D.R. Young, *J. Appl. Phys.* 52: 4090-4 (1981).
30. D.J. Bartelink, in: *Integrated Circuit Process Models* (J.D. Meindl and K.C. Saraswat, eds.), Chapt. 15, Englewood Cliffs, NJ: Prentice-Hall, Inc., to be published.
31. A. Goetzberger, E. Klausmann and M.J. Schulz, *CRC Crit. Reviews Solid State Science* 6: 1-43 (1976).
32. A.S. Grove, B.E. Deal, E.H. Snow and C.T. Sah, *Solid-State Electronics* 8: 145-63 (1965).
33. K.H. Zaininger and F.P. Heiman, *Solid-State Tech.* 13 (5): 49-55 (1970); 13 (6): 46-55 (1970).
34. R.R. Razouk and B.E. Deal, *J. Electrochem. Soc.* 129: 806-810 (1982).
35. A.S. Grove, O. Leistiko, Jr. and C.T. Sah, *J. Appl. Phys.* 35: 2695-701 (1964).
36. J.R. Ligenza, *J. Phys. Chem.* 65: 2011-14 (1961).
37. W.A. Pliskin, *IBM J. Rsch. Dev.* 10: 198-206 (1966).
38. B.E. Deal, and M. Sklar, *J. Electrochem. Soc.* 112: 430-35 (1965).
39. C.P. Ho and J.D. Plummer, *J. Electrochem. Soc.* 125: 665-71 (1978); 126: 1516-30 (1979).
40. F.N. Schwettmann, K.L. Chaing and W.A. Brown, Paper No. 276 in The Spring Meeting of The Electrochemical Society, Seattle, WA, May 21-26, 1978.
41. R.J. Kriegler, Y.G. Cheng and D.R. Colton, *J. Electrochem. Soc.* 119: 388-96 (1972).
42. B.R. Singh and P. Balk, *J. Electrochem. Soc.* 125: 453-61 (1978).
43. J.W. Rouse, C.R. Helms, B.E. Deal and R.R. Razouk, *J. Electrochem. Soc.*, 131: 887-894 (1984).
44. S.I. Raider, R.A. Gdula and J.R. Petrak, *Appl. Phys. Lett.* 27: 150-52 (1975).
45. T. Ito, S. Hijiya, T. Nozaki, H. Arakawa, M. Shinoda and Y. Fukukawa, *J. Electrochem. Soc.* 125: 448-52 (1978).
46. T. Ito, I. Kato, T. Nozaki, T. Nakamura and H. Ishikawa, *Appl. Phys. Lett.* 38: 370-72 (1981).
47. T. Ito, T. Nozaki and H. Ishikawa, *J. Electrochem. Soc.* 127: 2053-57 (1980).
48. S.S. Wong, C.G. Sadini, T.W. Ekstedt, H.R. Grinolds, K.H. Jackson, S.H. Kwan and W.G. Oldham, *J. Electrochem. Soc.* 130: 1139-44 (1983).
49. P.T. Panousis and M. Schneider, Paper No. 53 presented at The Spring Meeting of The Electrochemical Society, Chicago, IL, May 13-18, 1973.
50. R. Champagne and M. Toole, *Solid State Tech.* 20(12): 61-63 (1977).
51. N. Tsubouchi, H. Miyoshi, A. Nishimoto and H. Abe, *Jap. J. Appl. Phys.* 16: 855-56 (1977).
52. R.J. Zeto, N.O. Korolkoff and S. Marshall, *Solid State Tech.* 22(7): 62-69 (1979).
53. R.R. Razouk, L.N. Lie and B.E. Deal, *J. Electrochem. Soc.* 128: 2214-20 (1981).
54. L.N. Lie, R.R. Razouk and B.E. Deal, *J. Electrochem. Soc.* 129: 2828-34 (1982).
55. J.D. Plummer, in *Semiconductor Silicon 1981* (H.R. Huff, R.V. Kriegler and Y. Takeishi, eds.) pp 445-54, The Electrochemical Society, Pennington, NJ (1981).
56. R.B. Fair, *J. Electrochem. Soc.* 128: 1360-68 (1981).
57. A. Lin, D.A. Antoniadis and R.W. Dutton, *Appl. Phys. Lett.* 35: 799-801 (1979); *J. Electrochem. Soc.* 128: 1131-37 (1981).

58. S.P. Murarka, *Phys. Rev. B* 16: 2849-57 (1977).
59. A.M. Lin, R.W. Dutton, D.A. Antoniadis and W.A. Tiller, *J. Electrochem. Soc.* 128: 1121-30 (1981).
60. S.P. Murarka, *Appl. Phys. Lett.* 34: 587-88 (1979).
61. R. Pfeffer and M. Ohring, *J. Appl. Phys.* 52: 77-84 (1981).
62. S. Rigo, F. Rochet, A. Straboni and B. Agius, in: *The Physics of MOS Insulators* (G. Lucovsky, S.T. Pantelides, and F.L. Galeener, eds.) pp 167-71, Pergamon Press, New York (1980).
63. R.E. Honig, *Thin Solid Films* 31: 89-122 (1976).
64. C.A. Evans, Jr. and R.L. Blattner, *Semi. Inter.* 3(11): 109-126 (1980).
65. J.W. Coburn and E. Kay, *CRC Crit. Reviews Solid State Sciences* 4: 561-90 (1974).
66. C.R. Helms, in: *Semiconductor Silicon 1981* (H.R. Huff, R.V. Kriegler, and Y. Takeishi, eds.) pp 455-62, The Electrochemical Society, Pennington, NJ (1981).
67. D.E. Aspnes and J.B. Theeten, *J. Electrochem. Soc.* 127: 1359-65 (1980).
68. V.G. Litovchenko, et al., *Thin Solid Films* 44: 295-302 (1977).
69. E.H. Poindexter, P.J. Caplan, B.E. Deal and R.R. Razouk, *J. Appl. Phys.* 52: 879-84 (1981).
70. A.K. Ray and A. Reisman, *J. Electrochem. Soc.* 128: 2460-72 (1981).
71. K.F. Galloway, S. Mayo and P. Roitman, *J. Electrochem. Soc.* 126: 2245-48 (1979).
72. T. Sugano, F. Koshiga, K. Yamasaki, Q.V. Ho and Y. Hirayama, *Faculty Eng., Uni. Tokyo* 35: 553-627 (1980).
73. T.I. Kamins, *J. Electrochem. Soc.* 126: 838-44 (1979).
74. H. Sunami, M. Koyanagi and N. Hashimoto, *J. Electrochem. Soc.* 127: 2499-2506 (1980).
75. I. Franz and W. Langheinrich, *Solid-State Electronics* 114: 499-505 (1971).
76. T. Enomoto, R. Ando, H. Morita and H. Nakayama, *Jap. J. Appl. Phys.* 17: 1049-58 (1978).
77. R.R. Razouk, M.E. Thomas and S.L. Pressacco, *J. Appl. Phys.* 53: 5342-44 (1982).
78. J.E.E. Baglin, F.M. d'Heurle and C.S. Peterson, *J. Appl. Phys.* 54: 1849-54 (1983).
79. H. Jiang, C.S. Petersson, and M.H. Nicolet, *Thin Solid Films*, 140: 115-129 (1986).
80. B. Schwartz, *CRC Crit. Rev. Solid State Science* 5: 609-24 (1975).
81. C.W. Wilmsen, *J. Vac. Sci. Tech.* 19: 279-89 (1981).
82. J.R. Troxell, *J. Electronic Mat.* 14: 707-728 (1985).
83. H.B. Pogge, in: *Integrated Circuits: Chemical and Physical Processing* (P. Stroeve, ed) pp 241-275, American Chemical Society, Washington, DC (1985).
84. J.O. Borland and C.I. Drowley, *Solid State Tech.* 28(8):141-148 (1985).

Chemical Vapor Deposition of Silicon and Its Compounds

Kenneth E. Bean

*Texas Instruments Incorporated
Dallas, Texas*

INTRODUCTION

The process technology of chemical vapor deposition (CVD) is today and has for many years been one of the key technologies in semiconductor processing. CVD technology is used throughout the various processing steps in today's semiconductor manufacturing, starting with the synthesis of the elemental silicon through thin epitaxial and polycrystalline silicon films extending on through the oxidation process, the deposition of dielectric films of nitrides, carbides, and silicides and metallization films. Tables 1 and 2 list these various CVD technologies categorically in their order of use in the semiconductor process flow. In Table 3 we list the technology,

Table 1: Chemical Vapor Deposition of Silicon and Its Compounds

- SYNTHESIS
- THIN FILMS
 - EPITAXY
 - POLY
- OXIDES
- NITRIDES
- CARBIDES
- SILICIDES

Table 2: CVD Processes and Products

- SILICON SYNTHESIS
 - SiCl₄ + H₂
 - SiHCl₃ + H₂ } CVD/FLUID BED
 - SiH₄ + H₂
 - EPITAXIAL SILICON AND POLYSILICON, PROCESSES
 - H₂ + SiCl₄
 - H₂ + SiHCl₃
 - H₂ + SiH₂Cl₂
 - H₂ + SiH₄
 - He + SiH₄
 - SILICON NITRIDE PROCESSES
 - SiH₄ + NH₃ + H₂
 - SiH₂Cl₂ + NH₃ + H₂
 - SILICON OXIDATION PROCESSES
 - THERMAL OXIDE --Si + O₂ or H₂O (steam)
 - *R.D.O.-- SiHCl₃ + O₂ + H₂
 - SILANE OXIDE --SiH₄ + O₂
 - SILICON CARBIDE PROCESSES
 - SiCl₄ + C₇H₈ OR C₃H₈
 - BORON NITRIDE PROCESS
 - B₂H₆ + NH₃ + N₂
 - BCl₃ + NH₃ + H₂
- * R.D.O. = Reactor Deposited Oxide

Table 3: Chemical Vapor Deposition for Silicon Device Processing

TECHNOLOGY	PROCESS	USES
SILICON SYNTHESIS	C.V.D. HYDROGEN REDUCTION OF SILICON HALIDE	SEMICONDUCTOR PROCESSING, ~ 2800 METRIC TONS, 1980
SILICON EPITAXY	HCl ETCHING (CVE) EPITAXIAL DEPOSITION (FROM HALIDE OR HYDRIDE H ₂ REDUCTION)	BIPOLAR MOS - LSI STRUCTURES
POLYCRYSTAL SILICON	THIN FILMS THICK FILMS (FROM HALIDE OR HYDRIDE)	GATE ELECTRODE AND INTERCONNECTS. DIELECTRIC ISOLATION STRUCTURES.
SILICON NITRIDE	THIN FILM DEPOSITION FROM AMMONIA AND (SiH ₄ OR SiH ₂ Cl ₂)	MASK FOR DIFFUSION, IMPLANT, OR PREFERENTIAL OXIDATION. ETCH STOP, DIELECTRIC FILM, SURFACE PASSIVATION, LEADS OVERCOAT.
SILICON DIOXIDE	Si HALIDE OR HYDRIDE + OXIDANT (O ₂ , CO ₂ , N ₂ O)	DIFFUSION OR IMPLANT MASK, DIELECTRIC FILM, LEADS OVERCOAT.
SILICON CARBIDE	SILICON HALIDE AND HYDROCARBON	WEAR RESISTANT SURFACE, THIN FILM TRANSDUCER MEMBRANE, PROCESS CONTROL ETCH OR POLISHING STOP, X RAY LITHOGRAPHY MASK.
BORON NITRIDE	DIBORANE AND AMMONIA (B ₂ H ₆ & NH ₃)	DIFFUSION SOURCES X RAY MASK

the process by which the technology is formed, and the uses of this technology in silicon manufacturing. In the synthesis of ultra-high purity elemental silicon for today's semiconductor manufacturing, we may begin the process with the hydrogen reduction of an ultra-high purity silicon halide, such as (SiCl_4) or $(\text{SiHCl}_3)^{1-4}$ by CVD of the elemental silicon, on a high purity silicon rod such as that shown in (a) of Figure 1. This high temperature reduction takes place in an all quartz system under very precisely controlled high purity gas flow conditions. When this CVD reaction has reached completion we will have obtained a polycrystalline rod similar to the section shown in (b) of Figure 1. This high purity polycrystalline silicon material is then broken into small pieces, placed in a high purity quartz liner or crucible which is then heated to the melting point of silicon, 1420°C . After the thermal stability of the molten silicon pool is established, a carefully oriented seed, cut from single crystal silicon of the desired crystal orientation, is dipped into this molten silicon and then slowly rotated and withdrawn to grow the single crystal of the desired diameter.⁵ This melt is carefully doped to provide the desired conductivity type and resistivity for the slices or substrates. Figure 1c shows the top, or seed end, of a single crystal of silicon and Figure 1d shows a sawed slice from such a crystal. The standard diameter of the silicon slice used by most silicon device or integrated circuit manufacturers today is 125 mm plus or minus 25 mm while 200 mm is being developed. After the crystal is grown and sliced by the use of diamond saws the slices are ground, lapped and chemically/mechanically (chem/mech) polished to remove all surface damage introduced by the sawing, lapping and polishing operation.

EPITAXIAL DEPOSITION

Shortly after the invention of the transistor in 1947 it became evident that methods other than diffusion or the grown junction would become necessary for the production of abrupt junctions of semiconductors from germanium and then later silicon. The epitaxial process provides a method of producing these abrupt changes in concentration of doping atoms, or even in type of conductivity.

Epitaxy is the process of producing a layer of material with exactly controlled crystallographic, chemical, physical, and electrical parameters. The single crystal nucleation of this layer is controlled by the host crystal or substrate which is of a desired and carefully oriented crystallographic orientation. It also has the proper conductivity type, and carrier concentration to fit the device or circuit design. In silicon epitaxy,⁶⁻¹⁰ the epitaxial layer is usually formed by the hydrogen reduction of a highly purified silicon halide or hydride. Table 4 defines the epitaxial process and lists some of the characteristics needed for the epitaxial substrate. The epitaxial film is usually a continuous film deposited over a continuous single crystal substrate as in Figure 2. However it may in some cases be a preferentially deposited film or it may be deposited over a preferentially doped (in certain areas) substrate such as that depicted in Figure 3.

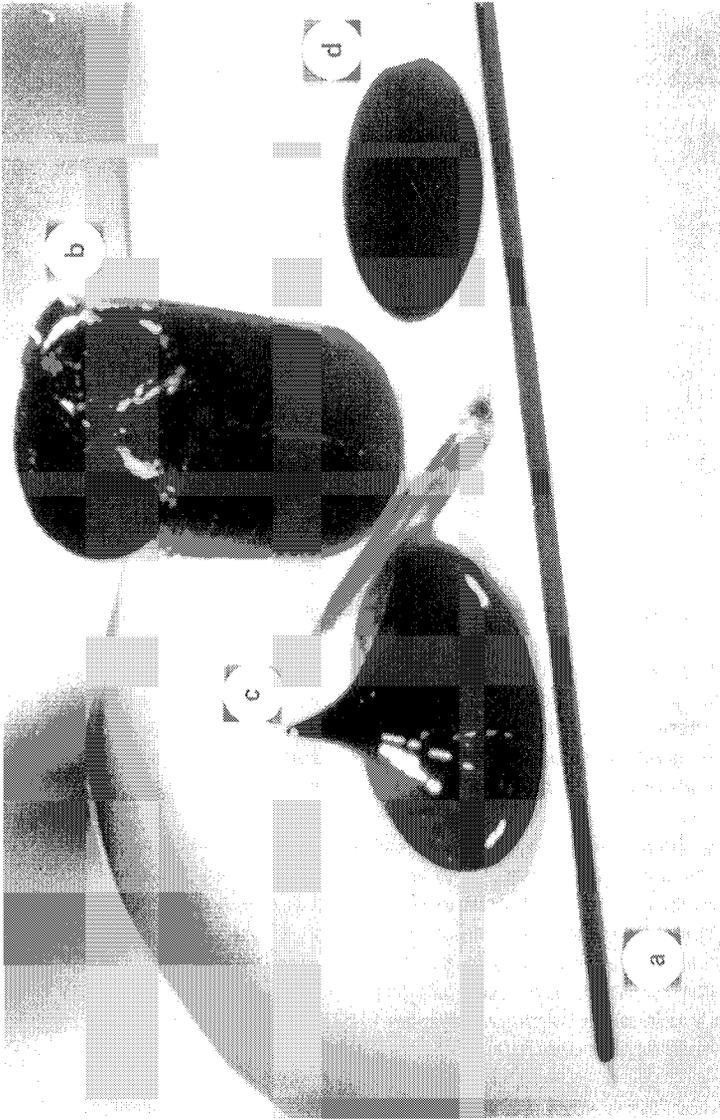
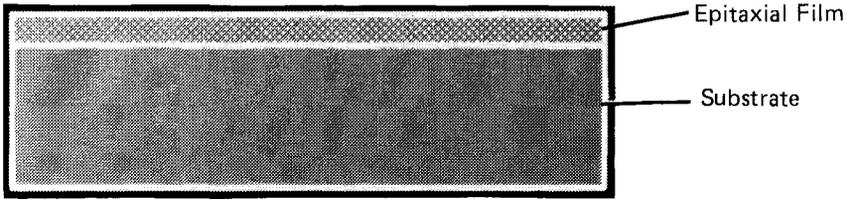


Figure 1: Silicon processing: polysynthesis, single crystal, sawed and polished slice (125 mm diameter crystal).

Table 4: Epitaxy

- Epitaxy is the process of producing a layer of material with controlled crystallographic, chemical, physical and electrical parameters.
- In silicon epitaxy, the layer is usually formed by chemical vapor-phase deposition on a carefully polished and cleaned substrate.
- This substrate also has very exactly controlled crystallographic, chemical, physical and electrical parameters.
- The silicon epitaxial layer, or layers, is usually the only active semiconductor material in the device or circuit.



The epitaxial growth process involves the deposition of a thin layer of material onto the surface of a single crystal slice in such a manner that the layer has the same crystallographic orientation as the original crystal and becomes an extension of the substrate. If the layer materials are the same as the substrate, e.g., Si on Si, it is known as *homoepitaxy*, or just plain *epitaxy*. If they are different, e.g., Si on sapphire, the combination is termed *heteroepitaxy*.

During the deposition (growth) of semiconductor layers, the conventional N and P-type impurities can be incorporated into the layer. The control of their concentration, as well as the layer thickness, to the necessary tolerances, makes the silicon epitaxial process one of the most demanding steps in the IC manufacturing process. *In fact, epitaxial processes and technologies are listed in the latest DoD military critical technology list as being of such military importance as to warrant export controls.*

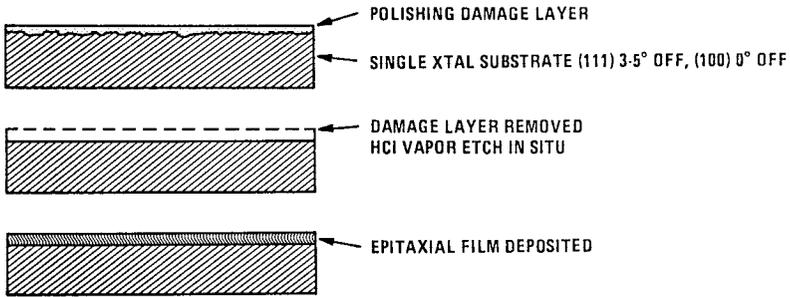


Figure 2: Epitaxy.

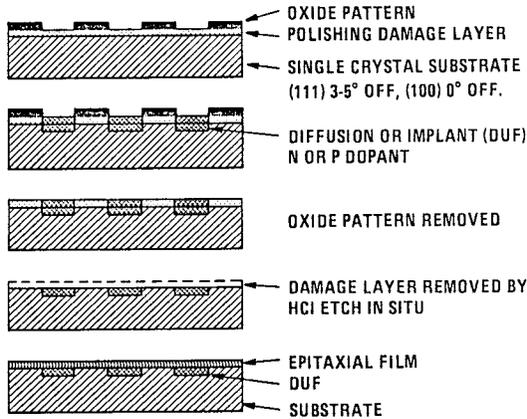


Figure 3: "DUF" epitaxy.

In the early days of silicon epitaxy most of the work or efforts were aimed towards the deposition of thin films on (111)¹¹ silicon substrates by the hydrogen reduction of silicon tetrachloride (SiCl_4), or the silicon tetrabromide (SiBr_4). In other attempts silicon tetraiodide (SiI_4) was also used. Table 5 lists the silicon bearing halides and silicon hydrides in the order of use historically and also in the order of descending energy or temperature required for the reduction. In production today most people use silicon dichlorosilane (SiH_2Cl_2). This material readily decomposes at about 1050°C. Silicon hydride (SiH_4) decomposes at an even lower temperature. However there are problems in the epitaxial deposition of thick films using (SiH_4) due to thermal decomposition in the vapor phase. When this occurs above the epitaxial substrate particles form in the gas phase which fall on the substrate resulting in the formation of spurious nucleation sites. It should also be noted that trichlorosilane (SiHCl_3), tribromosilane (SiHBr_3), Silane

Table 5: Silicon Epitaxy

- Silicon Source—Halides—Hydride
 - SiCl_4 —Early
 - SiBr_4
 - SiI_4
 - SiHCl_3
 - SiHBr_3
 - SiH_4
 - SiH_2Cl_2
- Reduction Source— H_2
- Vapor Etching—HCl
 - SiCl_4
 - SF_6
 - H_2O
 - H_2

(SiH_4) and dichlorosilane (SiH_2Cl_2) all may be thermally decomposed. This is due to the fact that there is a hydrogen atom which is liberated at high temperatures. However this is not true in the case of the silicon tetrachloride or the silicon tetrabromide. These two do not thermally decompose, and therefore must have a reducing agent, such as hydrogen, present to bring about the reaction.

HCl IN SITU ETCHING

Shortly after the beginning of attempts to do silicon epitaxy in the early 1960s it was learned that the substrate was of prime importance and must be extremely clean and free of defects at the beginning of epitaxial nucleation. All attempts to clean the substrate prior to placing the substrate in the epitaxial reactor met with high density defect levels in the epitaxial material. Due to this problem the in situ HCl chemical vapor etching (CVE) process was developed in 1963. In this process the substrates were initially cleaned and then placed in the epitaxial reactor. The final cleanup, removal of contaminants and crystallographic defects, is then done at approximately 1150-1250°C by high temperature in-situ etching, using HCl as the etchant.¹²⁻¹³ Other etches which have been experimentally used are listed in Table 5. However, the development of in-situ HCl vapor etching was a key development in the history of silicon epitaxial technology and allowed the production of large volume, low defect epitaxial material. With the development of today's better cleaning and polishing processes,

epitaxial layers may in some cases be deposited directly on the substrate without HCl vapor etching. However, in most cases in-situ HCl vapor etching is still required and in use throughout the industry.

In Figures 4 and 8 we plot the deposition rate of the halides, and the hydride as a function of mole percent halide and temperature. In Figure 8 we note that the deposition rate versus temperature is a double energy curve. In general one wants to operate in the left portion, or in the stable part, of this curve. This allows larger fluctuations in temperature with smaller affect on deposition rate. Also, in order to deposit high quality single crystal epitaxy material with a minimum number of defects one should operate in the 1 to 2 micron per minute or less deposition rate range as shown in Figure 4.

Figure 5 is a comparative listing of the processes involved in building both MOS and Bipolar devices or circuits using epitaxial material. In the Bipolar process described in Figure 5, the epitaxial film is deposited on a substrate which has been previously patterned and diffused. This process allows for the production of very high speed Bipolar devices and/or circuitry. In the MOS process shown in Figure 5, a p+ substrate is used. This p+ substrate does not interact as an active component in the MOS structure but only provides a carrier for the epitaxial film in which all of the

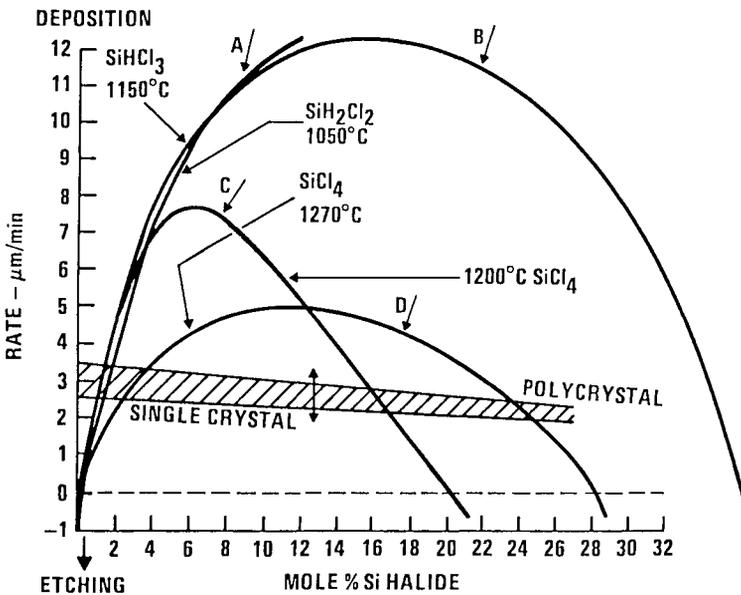


Figure 4: Deposition and etch rate vs mol % silicon halide. The deposition rate is also affected by reactor design. Curves A and B are from a vertical reactor. Curve C is from a multiple slice vertical reactor in which each slice rotates on its own susceptor. Curve D is data from Henry Therur of Bell Labs using a single slice vertical reactor.

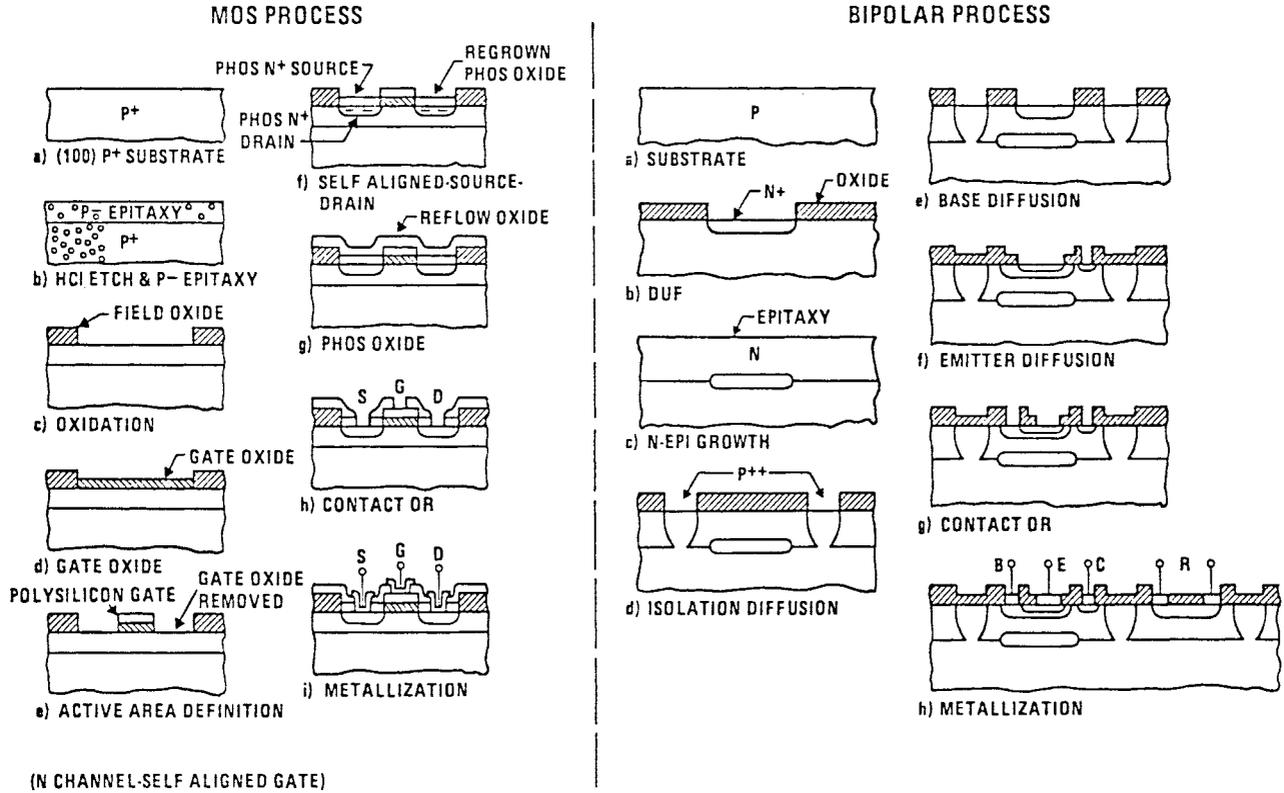


Figure 5: MOS-bipolar process comparison.

active components are fabricated. However, the p+ substrate provides a backside ground plane which eliminates effects due to substrate noise generated by the charging and discharging of p-n junction capacitance as well as process gettering action which enhances the characteristics of the MOS devices built in the epitaxial film. This structure also reduces the tendency for latchup in CMOS circuits. In today's CMOS process, many of the circuitss are built in epitaxial material, which is deposited by CVD processing. In this CVD epitaxial process, the interface at the substrate p+ material and the epitaxial p- material is of utmost importance. In order to obtain a perfect interface HCl vapor etching is normally required. This HCl step is carried out in-situ immediately prior to the epitaxial deposition at a temperature in the range of 1200°C using approximately 2% HCl in hydrogen.

As can be seen in Figure 6, the HCl etch rate increases rapidly to a temperature of approximately 1100°C. One should operate in the flat portion of the curve above 1100°C in order to have well controlled etching. If the percent HCl is too high with respect to temperature a rough etch front will develop. Figure 7 shows the ideal conditions for the useful polishing effect and the non-ideal non-useful effect. From Figure 7 one can find an

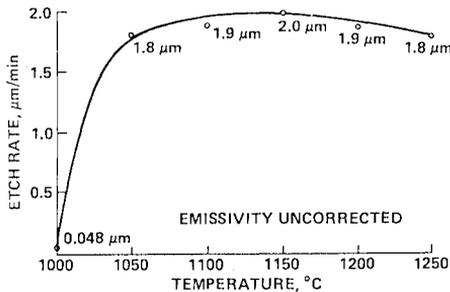


Figure 6: Etch rate (HCl-5% in H₂) vs temperature.

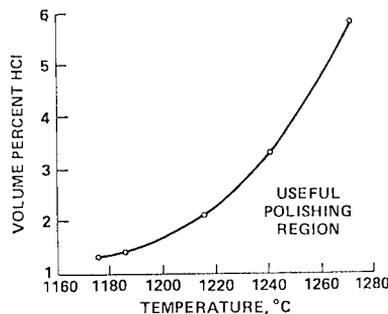


Figure 7: Useful polishing region vs % HCl temperature.

etch rate as a function of temperature for any ratio desired. For example 2% HCl at approximately 1220°C is an ideal etching condition in order to remove the impurities and damaged region of the substrate surface prior to epitaxial deposition. We are now ready to perform the epitaxial deposition. Silicon bearing halides, or the hydride, may be used in order to perform our epitaxial deposition. These materials are shown in Figure 8 with silicon tetrachloride being the silicon source requiring the highest temperature for deposition, trichlorosilane (SiHCl_3) requiring a medium deposition temperature and dichlorosilane (SiH_2Cl_2) or the silicon hydride (SiH_4) requiring the lowest temperature for deposition. In the deposition reaction, the deposition rate increases as a function of temperature. The cross-hatched bar zone shown in Figure 8 delineates the change-over point from a kinetically controlled reaction to a diffusion controlled reaction. As in the case of HCl vapor etching, one should operate in the flat portion of the curve for well controlled epitaxial deposition. The importance of hydrogen and chlorine, in the reaction cannot be overlooked. Silicon tetrachloride decomposes in the presence of hydrogen and thermal energy from the heated susceptor. This material (SiCl_4) does not thermally decompose as does the other halides (SiHCl_3) or (SiH_2Cl_2) or the hydride (SiH_4), without the presence of hydrogen as a reducing agent. The halides and the hydride all have a hydrogen atom attached to the molecule, while (SiCl_4) does not. The presence of chlorine in the silicon halide is of importance in preferential deposition.¹⁴⁻¹⁵ The by-product chlorine and/or HCl gives the silicon atom a vehicle in which it may be transported from one nucleation site to another, which may be a more desirable crystallographic site. The by-product of silicon hydride, does not contain chlorine making this material somewhat less desirable for preferential deposition unless chlorine is also added to the reaction.¹⁴⁻¹⁶ The effect of silicon atomic mobility or transport will be discussed in a later section.

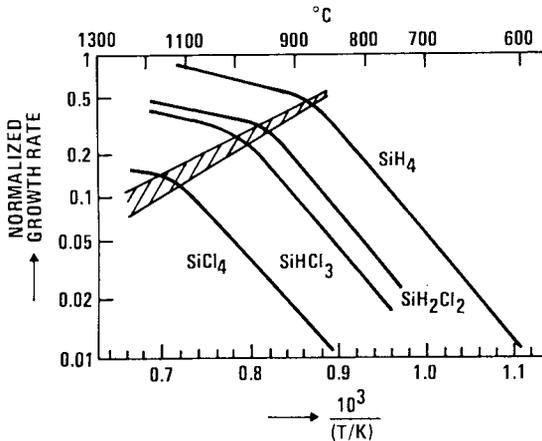


Figure 8: Silicon bearing source and deposition rate reduction vs temperature.

GETTERING

In the ever increasing packing density for today's technologies of VLSI and ULSI circuitry, silicon epitaxy plays a very important role.¹⁰ MOS circuits built in epitaxial material are in general superior in performance to those built in bulk silicon material. In the past two years considerable interest has been placed on improving the quality of epitaxial material for MOS devices. One of the recent developments which has improved the quality of epitaxial material has been the denuding of defects from the substrate surface area prior to epitaxial deposition. This denuding is accomplished by a thermal process which in most cases, consists of a high temperature cycle (approximately 1100°C), followed by a lower or intermediate temperature cycle (approximately 650°C) then followed by a high temperature (approximately 1000°C) thermal cycle. This denuding process is highly dependent upon the oxygen and/or carbon concentration in the original bulk silicon substrate. The initial high temperature cycle provides sufficient energy to dissolve the oxygen precipitates that are present and to cause out diffusion of the oxygen from the surface areas of the substrate. The intermediate thermal cycle provides energy for renucleation and growth to stability of oxygen precipitates in the center, non out-diffused, region of the slice. The following high temperature cycle provides energy for these precipitates to grow and getter oxygen, heavy metals, and other defects, during device processing, to these precipitate sites. See Figures 9 and 10.

Figure 9 shows a cross-sectional view of a denuded substrate after epitaxial deposition. The high bulk defect density is evident as is the denuded zone just below the epitaxial silicon interface. In this figure there are defects at the interface between the epitaxial film and the original substrate. This indicates that no HCl vapor etching was carried out in this process. It also shows that the original substrate surface acts as a trap to hold these defects, which may be bulk stacking faults, oxygen precipitates, heavy metals, or carbon atoms. Figure 9 is a photograph of a cleaved silicon slice. No polishing or potting was required prior to the Wright Jenkins etch in order to bring out the defects in the denuded area as well as the epitaxial film. It is also interesting to note that the defects at the substrate/epitaxy interface do not cause defects in the epitaxial film at a 1:1 ratio. It appears that only approximately one out of 10 defects at the interface actually cause stacking faults or defects in the epitaxial film.

Another method of slice/wafer processing that improves device performance is that of backside gettering prior to epitaxial deposition. There are several methods of backside gettering including sandblasting or abrading of the back surface of the substrate to produce defects or traps, and oxide or nitride films on the backside surface to produce strain fields. However the most successful appears to be that of depositing a thin CVD film, approximately 1 micron thick, of polycrystalline silicon across the back surface of the epitaxial substrate prior to final polishing of the front side. The high density of defects produced by the grain boundaries and dislocations, due to the polycrystalline film, provide a very effective gettering mechanism. During device processing heavy metals may be gettered

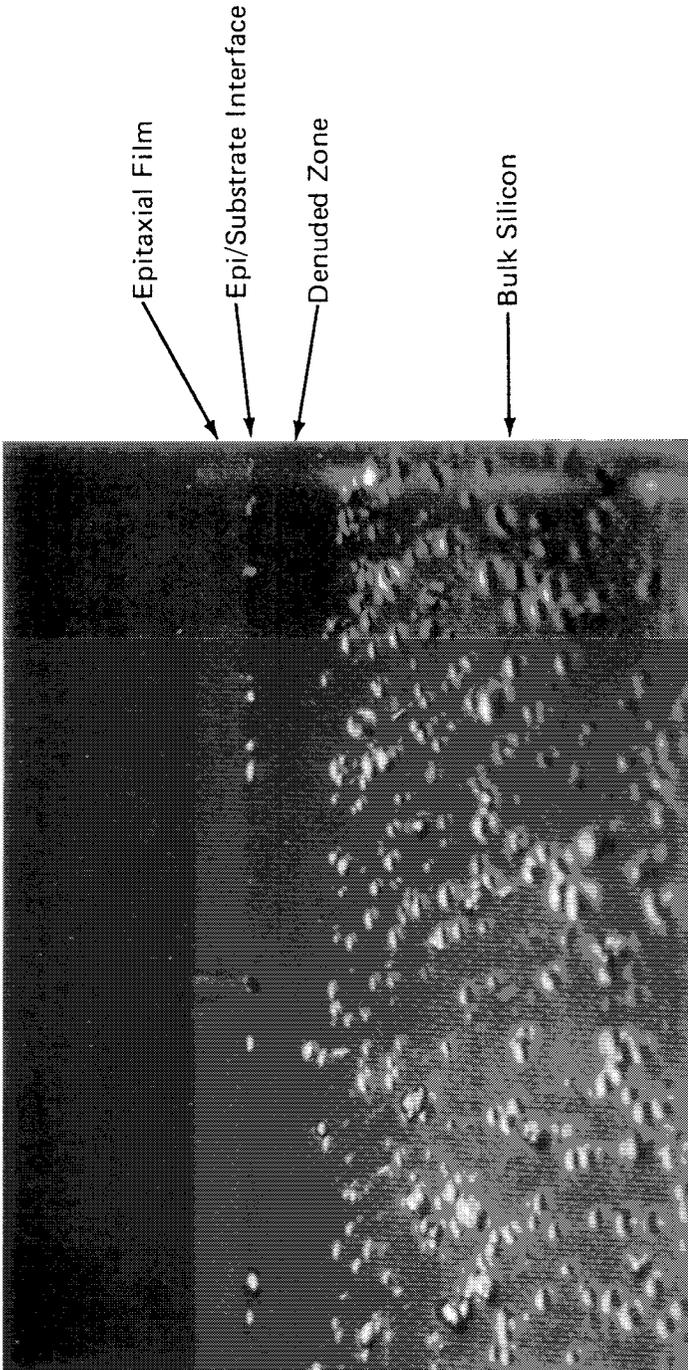


Figure 9: Cleaved cross section—epitaxy on denuded substrate. 630X P.C.

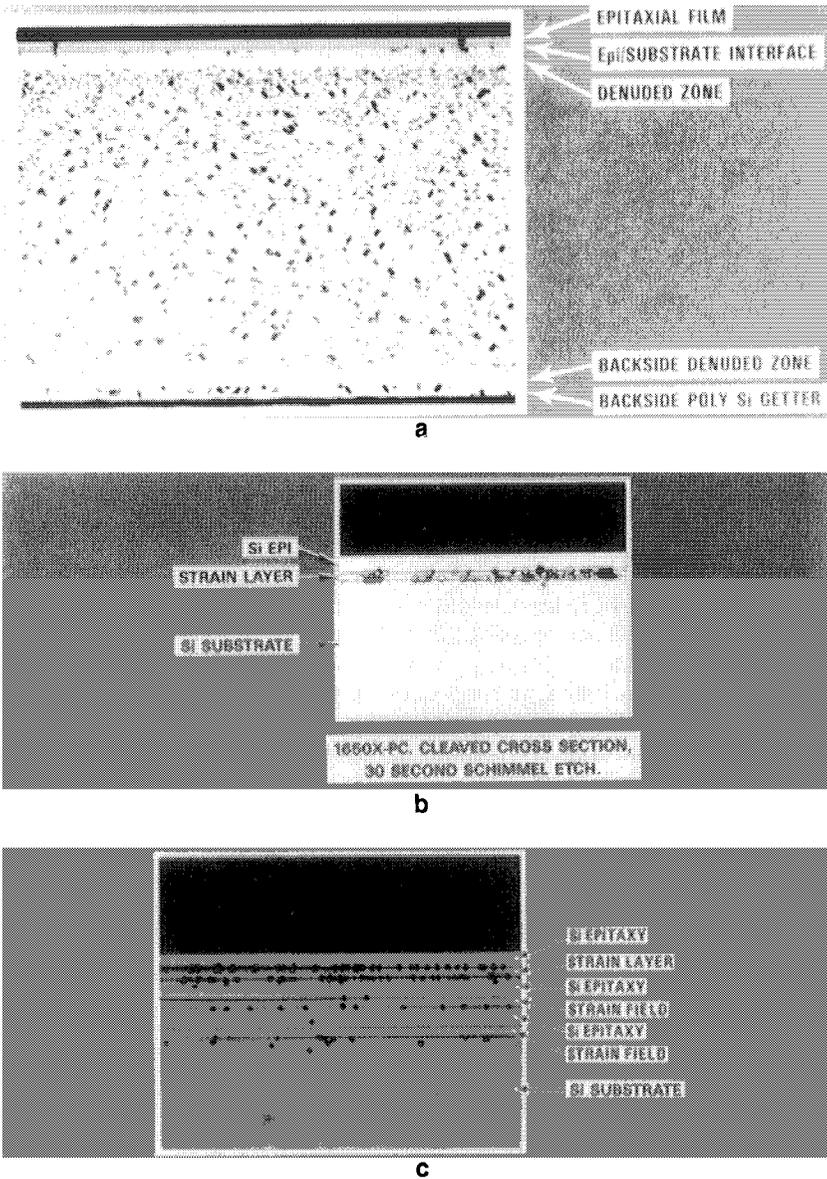


Figure 10: (a) 90° cleaved cross section, epitaxy on denuded substrate with back-side-gettering poly Si layer. 198X P.C. 5 minute W.J. etch. (b) (c) Misfit dislocation extrinsic gettering.

all the way from the front surface to the back surface where they are trapped at these defect (gettering) sites.

Figure 10a shows a cross-section of a cleaved slice which has a gettering region at the back surface, a denuded zone just above the backside gettering media, the high density bulk defect area of the slice, and then at the top surface a denuded zone just below the epitaxial-substrate interface. Again, a row of defects is noticeable at the epi-substrate interface indicating that this substrate had insufficient or no HCl vapor etching, in situ, prior to the epitaxial deposition. This also shows that one could build in intrinsic gettering at desired positions immediately below the active device region of the semiconductor circuit. This type of gettering immediately below the active surface area is very effective¹⁸ and can also be designed in discretionarily to provide gettering only at the desired circuit areas. Figure 10a is a cleaved cross-sectional view which required no polishing or potting prior to the Wright-Jenkins etch to delineate the defect region, the denuded zone, and the epitaxial layer as well as the back side gettering polycrystalline silicon film.

As mentioned above, in-situ HCl vapor etching will remove the damage sites and/or surface traps prior to epitaxy. However, one may wish to leave or form a new damage layer, for low temperature processing intrinsic gettering, in near proximity to the active device region of the structure. If so, a film can be deposited between the substrate and the epitaxial film which is intentionally doped with, for example germanium, to produce a built-in misfit dislocation strain field.^{10,17,18} Figure 10B shows a cleaved cross-sectional view of a single layer misfit dislocation, extrinsic gettering strain field, and a single layer epitaxial film. Figure 10c shows experimental multiple layers of strain field/epitaxial silicon films with increasing Ge doping in the strain fields as they were deposited. This increase in Ge doping causes a noticeable increase in the density of misfit dislocations within the strain field layers. Wright-Jenkins etch was used to reveal these damage sites.

SELECTIVE DEPOSITION

It was previously mentioned that silicon atoms, deposited from vapor phase, have surface mobility and tend to deposit in single crystal form at preferred nucleation sites. In Figure 11 we show a silicon slice which has both an oxide and an open silicon area exposed. The oxide covers most of the photographed area with the exposed single crystal area only at the top of the photograph. As a result of preferential epitaxial deposition, the five silicon octahedral growths, deposited on the oxide are all crystallographically aligned with the substrate orientation. This indicates that the nucleation was due to pinholes through the oxide and not at nucleation sites on the oxide. If silicon is allowed to deposit out in free space, with no constraints, it will always form an octahedral shape bounded by eight (111) faces. In Figure 11, we see the octahedral form of the top four of these (111) facets. Note the alignment of the (111)'s in each octahedron with each other, thus indicating that epitaxial nucleation propagates from the substrate through

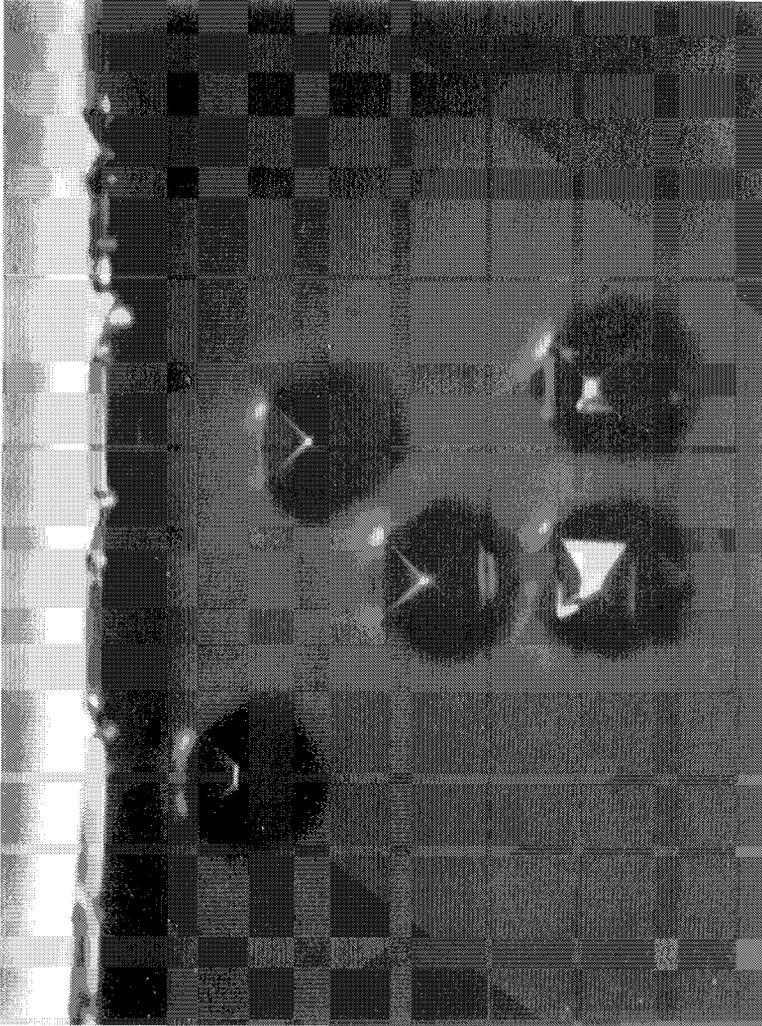


Figure 11: Silicon crystallites, octahedrons, on SiO₂. 367X.

the oxide pinhole. Note also in this figure that the surrounding area is completely free of polycrystalline silicon deposition or nucleation on the oxide. This indicates that the silicon atoms, above the oxide area, moved to a preferred site in the open silicon area. We can take advantage of this atomic mobility for preferential deposition of silicon at desired sites on an otherwise oxide or nitride covered mask substrate. Such a mask is shown in Figure 12. Shown in this figure are five micron diameter circles, on 25 micron centers, opened up through an oxide or a nitride mask on a silicon substrate. The goal in this experiment is to preferentially deposit epitaxial silicon in the five micron diameter circles but have no poly silicon nucleate on the oxide. An epitaxial diode will be formed at the interface of the original p substrate with the n epi deposit. After the epitaxial growth has proceeded up through the mask, in this case oxide, lateral spreading occurs over the oxide to form a large area "epi top" for electron beam charging. This process provides a very small area p-n junction diode with very low parasitic capacitance but with a large "epi top" beam collection area for the production of Vidicon type detectors. In this process the preferential deposition must be very complete in that single crystal silicon is nucleated in the open areas. No spurious nucleation of polycrystalline silicon can be tolerated on the oxide which would bridge across two diode "epi tops" thus causing a defect in the array. The diode density in this array is one million diodes per square inch.

Figure 13 is a top view of such an array after preferential deposition. The (100) structure is clearly evident in the epitaxial "epi tops". "Epi tops" deposited on (111) substrates show an equilateral triangle structure whereas the (100) substrate gives the perfect square "epi top" orientation. Figures 14a and 14b show two SEM photographs with 14a being a low angle SEM of the cross-sectioned substrate/epi structure and 14b being a near 90 degree cross-section after the oxide had been etched away. In 14a, the original diode area can be seen as well as the mask oxide which has been broken away with the cleavage of the slice. In the cross-section at 14b the original diode structure or size can be seen at the substrate interface. The lateral spreading, in all directions, over the oxide is approximately equal to the diameter of the original diode, thus a 3X increase in diameter and >9X increase in area.

Other examples of preferential deposition making use of the atomic mobility of the silicon atom are shown in Figures 15 through 18. Figure 15 shows the preferential epitaxial deposition of silicon in the vertical lines across the bottom portion of the slice. In this grating there are three micron wide lines of oxide with two micron wide areas of open silicon between them. Single crystal silicon is nucleated in the open silicon areas and grows up through the oxide and then spreads laterally as shown in Figure 18. Also shown in Figure 15 in the top portion of the photograph is an area with continuous oxide mask with polycrystalline silicon nucleated only on the top half of this oxide area. The lower half of the oxide area is completely clean and free of spurious nucleation of polycrystalline silicon. The silicon atoms have enough mobility, to move to preferred sites in the open silicon area or to deposit out (at super saturation) as polycrystalline silicon over the oxide. In this experiment, the deposition temperature was 1150°C and

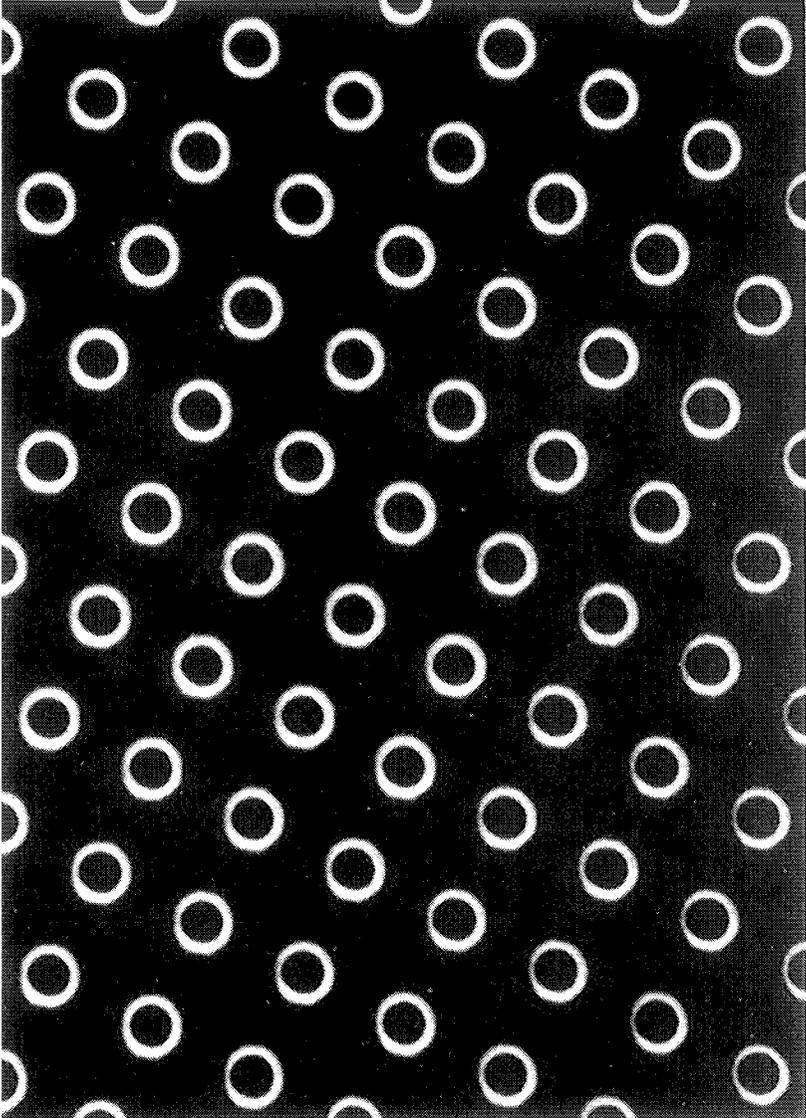


Figure 12: Tivicon mask—5 μm open circles on 25 μm centers.

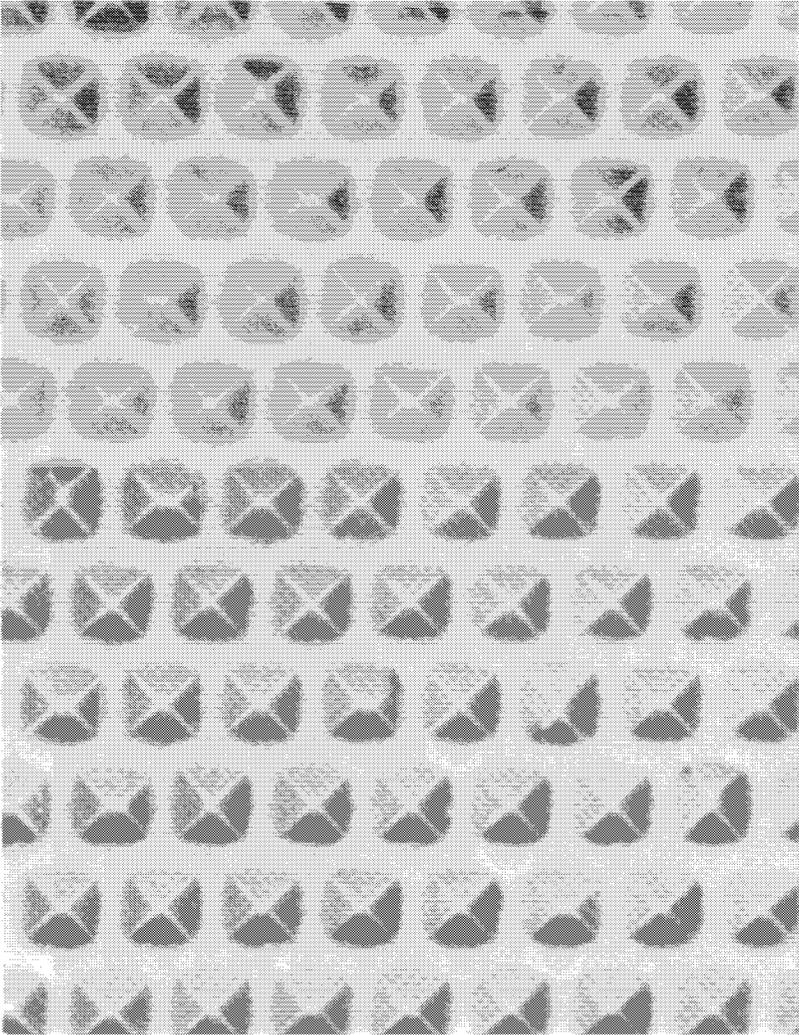


Figure 13: Approximately 20 μm (100) ODD through 5 μm oxide openings.

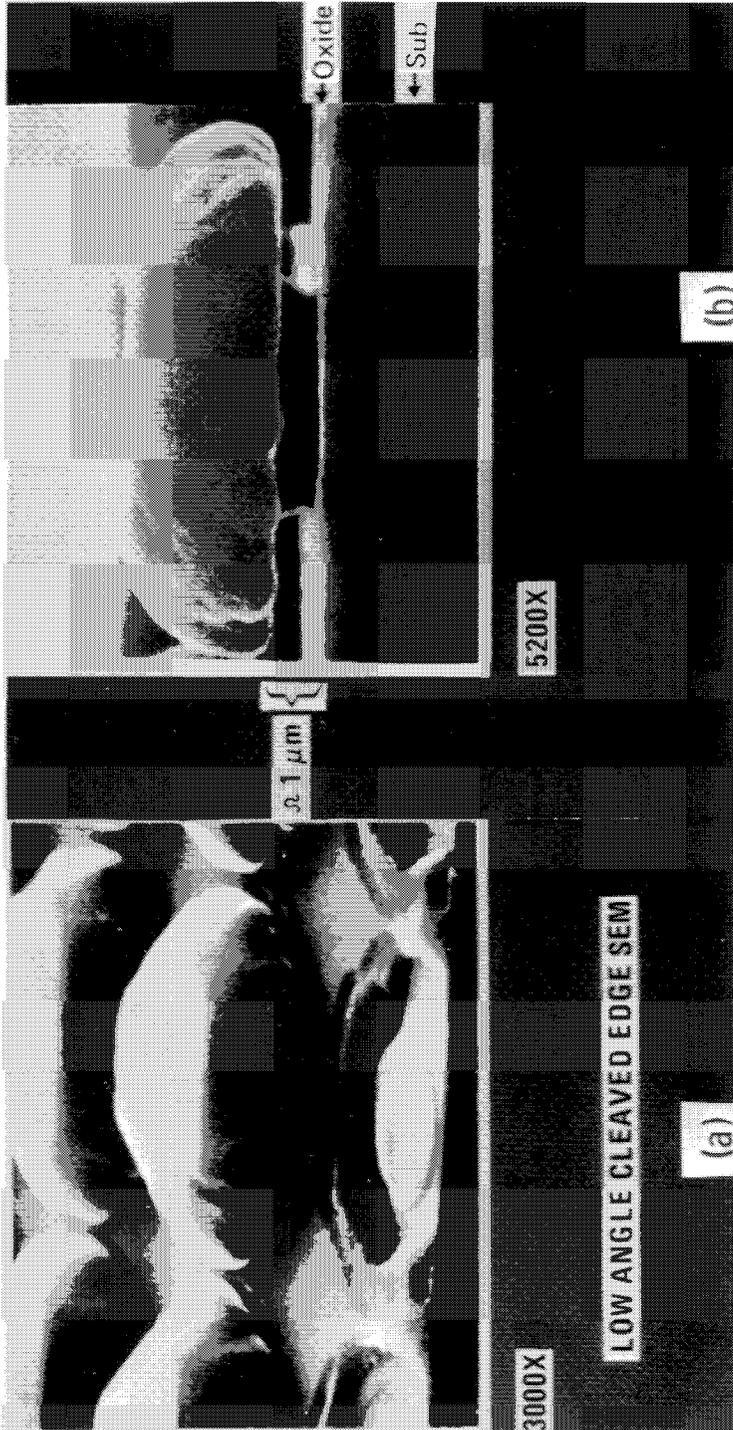


Figure 14: Preferential silicon epitaxy through oxide.

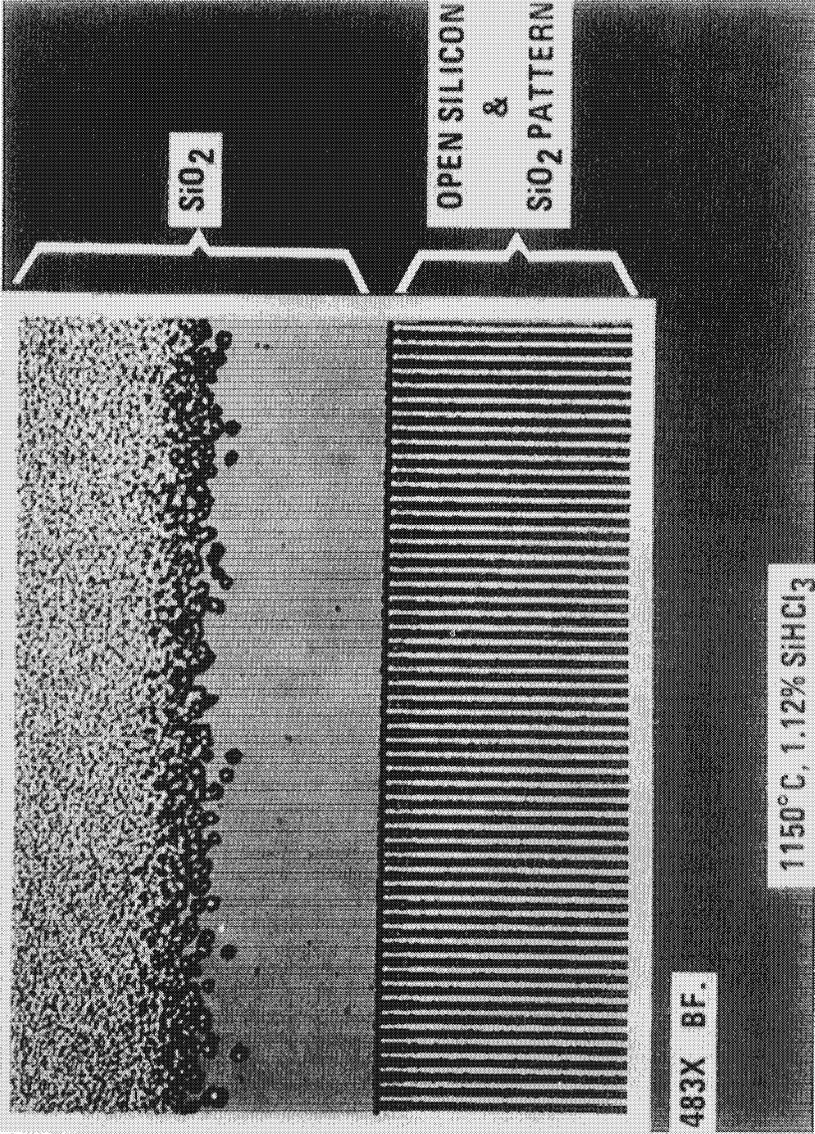


Figure 15: Silicon atomic mobility. (Magnification less than indicated.)



Figure 16: Oxide pattern inside corner effect on silicon atomic mobility. (Magnification less than indicated.)

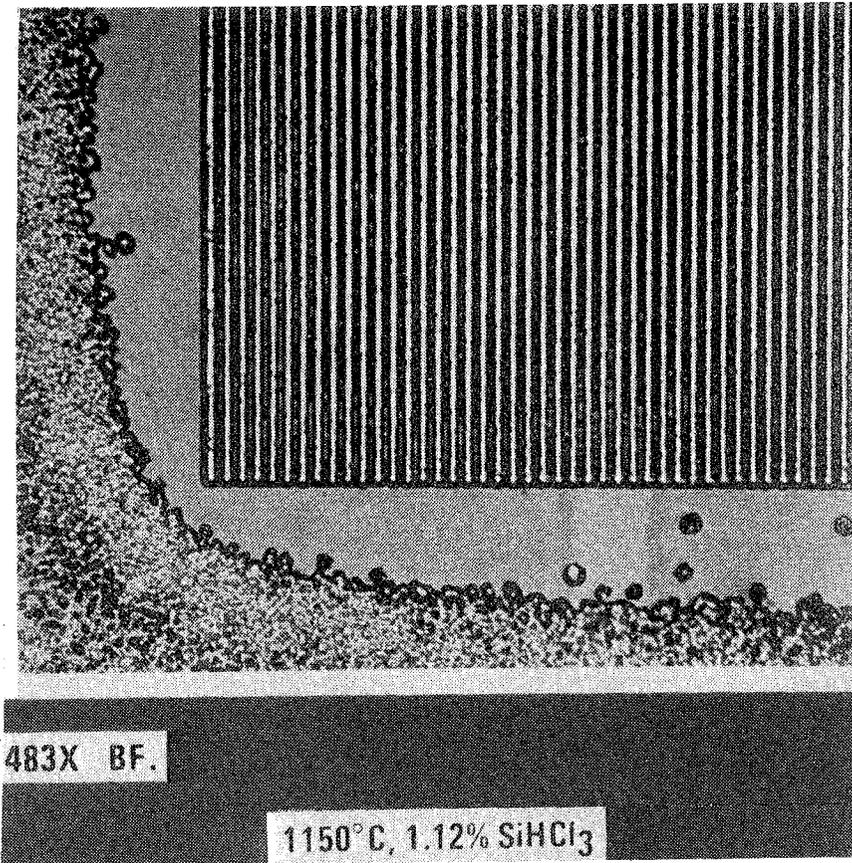


Figure 17: Oxide pattern outside corner effect on silicon atomic mobility.

the mole ratio of silicon halide (SiHCl_3) in hydrogen was 1.12%. At these conditions and mask geometry the atomic mobility of the silicon atoms was great enough for the atom to move approximately 23 microns, which is one half the distance between the open silicon pattern and the area of heavy polycrystalline nucleation over the oxide for a total of at least 47 microns. The atomic mobility is greatly affected by the halide being used, SiH_2Cl_2 , SiHCl_3 or SiCl_4 , the mask geometry (open silicon to mask area), and the operating temperature, which provides the energy to move the silicon atom. The atomic mobility increases as the temperature increases or as the mole ratio decreases, or in other words, as the energy to move the atom increases or the number of silicon atoms competing for a preferred site

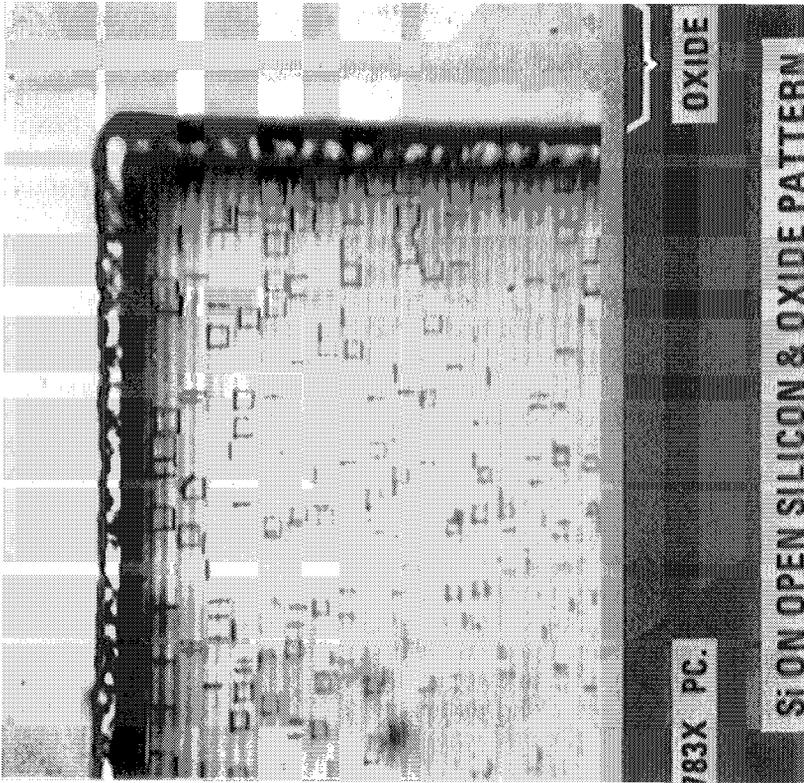


Figure 18: Effect of silicon epitaxial growth rate on oxide pattern. (Magnification less than indicated.)

decreases. It should also be noted in Figure 15 that the polysilicon crystallites along the lower portion of the poly zone over the mask oxide are much larger than those further up in the field. The larger crystallite size is due to the availability of silicon atoms migrating from the oxide region between the nucleation sites and the open silicon grating pattern. Further evidence of the influence of the silicon availability will be shown in Figure 18 for the deposition of a single crystal film over a dielectric (oxide) grating pattern.

Figures 16 and 17 also show the effect of the silicon availability on the growth of large polycrystalline crystallites over an oxide pattern. Figure 16 shows a portion of the pattern which has a large area of continuous oxide with the open silicon/oxide grating pattern across the bottom and along the right side of the photograph. Again, the epitaxial silicon nucleates in the open area of the silicon grating and grows up through the oxide and then spreads laterally over the oxide. Further note, that the zone along the bottom portion of the continuous oxide mask area is completely free of poly nucleation as is that along the right side of the continuous oxide grating pattern. Further note the increased distance from the bottom right corner of the continuous oxide up to the arc of the poly nucleation. This shows the strong influence of the available nucleation sites in the silicon at the openings along the right side and at the lower edge of the oxide pattern. In this corner area a silicon atom on the oxide can move either to the right or to the bottom and find a silicon nucleation site, thus having twice the effective mobility as those silicon atoms at the extreme left and extreme upper right of the unpatterned oxide mask area. This figure shows the affect in a convex or outside oxide pattern.

Figure 17 shows the opposite affect for a concave or inside continuous oxide pattern. In this case the available atom only has half as many available sites as in the case of the convex oxide corner. Therefore the distance from the corner of the open silicon oxide grating to the nearest polysilicon nucleation site is greatly reduced.

Figure 18 shows an enlarged view of the epitaxial silicon nucleated in the open areas of the grating, and growing up through the open area then spreading laterally over the oxide which is three microns wide. Single crystal lateral overgrowth is evident due to the (100) oriented epitaxial stacking faults. These stacking faults are all aligned with each other as well as with the edge of the mask, which is aligned with the (110) slice flat indicating epitaxial nucleation at the (100) substrate. The size of the stacking faults is considerably larger around the periphery of the grating oxide pattern. The size of the stacking fault is a method for direct measurement of the thickness of the epitaxial film.^{19,21} If one measures the length of any side of the (100) stacking fault and multiplies that length by 0.707 one will have the true thickness of the epitaxial layer at that point. This example is direct evidence that the silicon atoms moved from the oxide mask area to a preferred site in the open silicon grating. The density of available silicon atoms is greater along the periphery of the continuous oxide mask, thus giving rise to a higher rate of epitaxial deposition, and a thicker film, along the periphery of the open silicon grating area, as indicated by the epitaxial stacking fault size.

CVD OF DIELECTRIC FILMS

In today's semiconductor processing flow we quite often use CVD for the deposition of silicon nitride and in some cases silicon dioxide. These films may be used as dielectrics or as passivation films. Both atmospheric and low pressure processes are used for the deposition of silicon nitride. In either case, NH_3 , and SiH_4 or SiH_2Cl_2 are used as the nitrogen and silicon bearing sources respectively (see Table 2).

Figure 19 shows the deposition rate for silicon nitride as a function of temperature. Again the deposition of silicon nitride as in the case of the epitaxial silicon process is a diffusion limited and a kinetically controlled deposition process. One should operate in the diffusion limited portion of the curve, above approximately 850°C for the atmospheric pressure mode of silicon nitride deposition.

Figure 20 gives the deposition rate curves for the deposition of silicon nitride with a fixed amount of ammonia, 1.2%, for two amounts of silane. The refractive index and the dielectric strength of the silicon nitride film are greatly affected by the ammonia to silane ratio. In practical operation one should use a volume of ammonia greater than 10 times that of silane (see Figure 21) in order to produce good high quality dielectric films. These films should have a refractive index of 2.0 ± 0.05 . If the refractive index of the silicon nitride is less than 2.0 the silicon nitride probably contains silicon dioxide. If the refractive index is greater than 2.0 the silicon nitride is probably silicon rich²² (Figure 21).

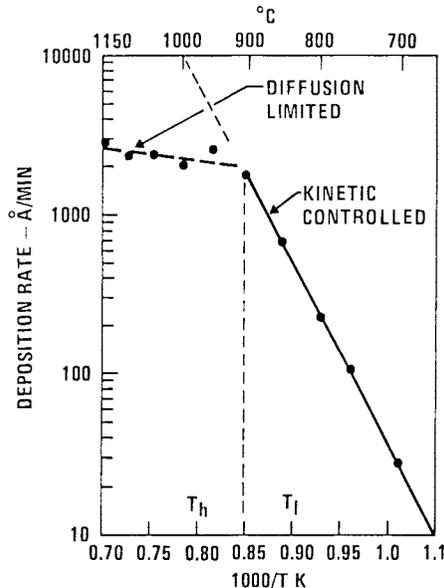


Figure 19: Si_3N_4 deposition rate vs temperature.

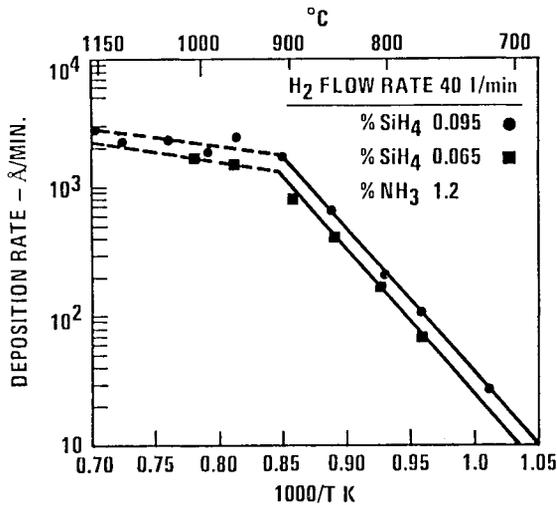
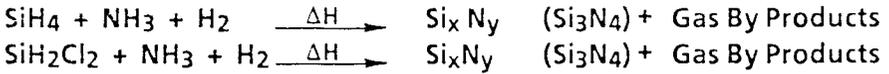


Figure 20: Si_3N_4 deposition rate vs temperature.

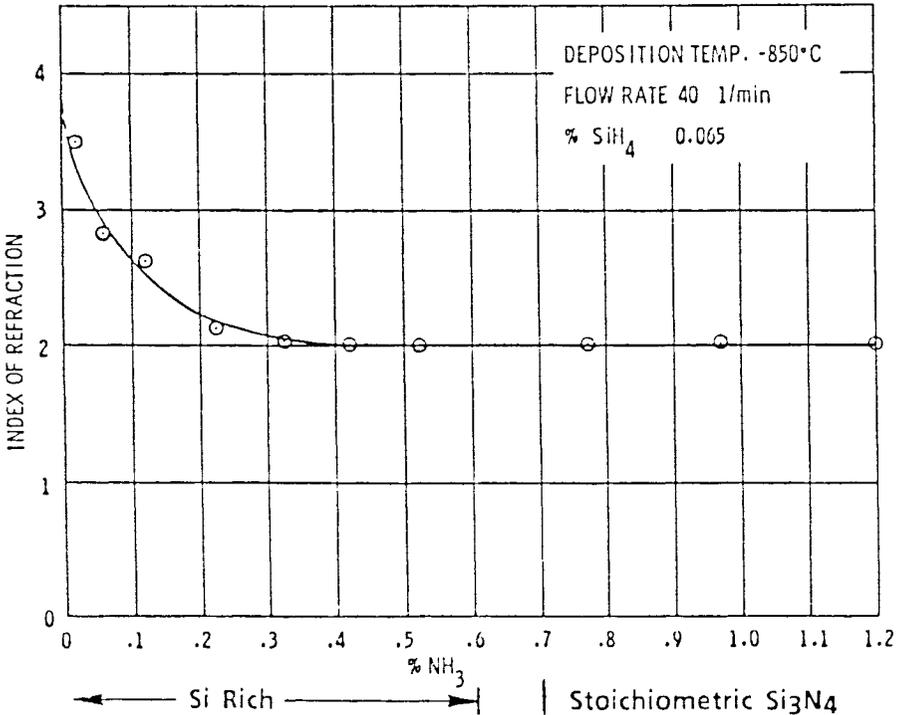
In the dielectric isolation (DI) process^{23,24} used for high-voltage, high-speed, and radiation hardened circuitry, CVD processing is used in several steps. It is used to deposit oxide, in addition to thermal oxidation, to form a pinhole-free sandwich of thermal oxide and CVD oxide. It is used in the polycrystalline silicon deposition step and may be used for the deposition of silicon nitride and/or silicon carbide films. The DI process offers advantages in several areas such as radiation hardened capability, high-voltage, high-speed, and process control capability. However, it is not widely used due to its high cost. In normal p-n junction isolation, a reverse bias junction isolates the active components within the circuit, (Figure 22). The formation of this junction, however, causes high parasitic capacitance around each device in the circuit. This parasitic capacitance decreases the speed at which the device is capable of operating. The isolation voltage between the two adjacent devices is a function of the resistivity or doping concentration of the high resistivity side of the junction. The isolation voltage in a p-n junction isolated circuit or array is normally less than 100 volts, whereas, in the case of a DI structure the operating voltage or isolation voltage is several hundred volts. In fact, in most DI structures, the isolation will not break down between adjacent components but will break down across the surface at voltages of >450-500 volts. In the dielectric isolation process (Figure 22), the p-n junction isolation is replaced in function by a dielectric film such as SiO_2 , Si_3N_4 , or combinations of SiO_2 , Si_3N_4 , and/or SiC. The replacement of this large area capacitor or p-n junction isolation around each device by a thermal oxide and/or other dielectric material effectively gives the device very high switching speed capability as well as very high voltage operating capabilities. Devices produced by the DI process also have higher radiation tolerance due to the fact that a radiation particle is

DEPOSITION SOURCE



USES

- OXIDATION MASK
- ETCH MASK
- PASSIVATION



JECS, Vol. 114, No.7, July 1967

Figure 21: CVD-silicon nitride.

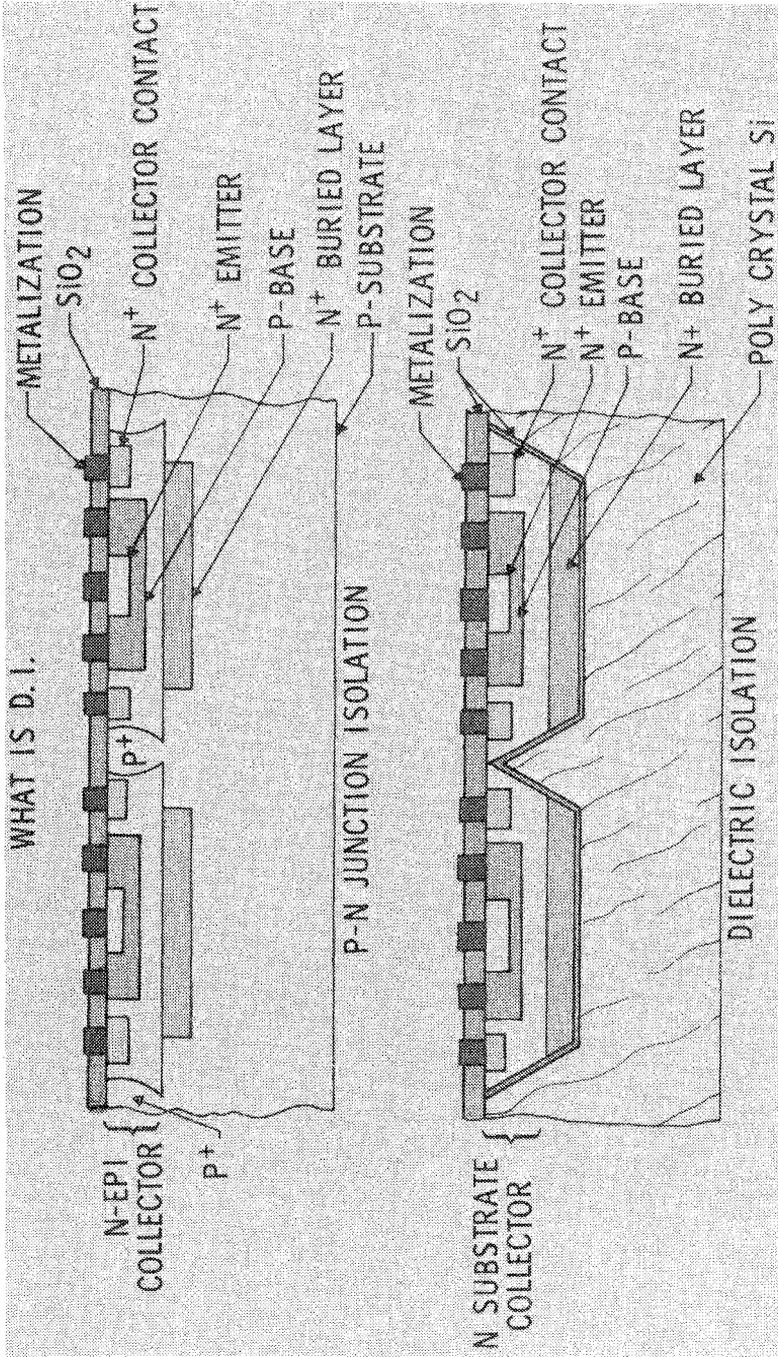


Figure 22: Dielectric isolation and P-N junction isolation.

more readily blocked by the DI media than by a p-n junction. Theoretically the DI process also offers advantages in packing density.

The effect of packing density may be seen in Figure 22 which compares the DI process with the junction isolation process. The high packing density capability of the DI over that of the p-n junction process is brought about due to the fact that the p-n junction is subject to lateral spreading of the junction isolation. Whereas the dielectric isolation is formed by a dielectric, such as SiO_2 or silicon nitride, thermally grown or CVD deposited (see Table 2) on crystallographically sharp, orientation dependent etched, crystallographic planes. In the case of the DI structure the final substrate or carrying media is a thick polycrystalline silicon film, whereas in the p-n junction isolation process the substrate or carrying structure is the original p- single crystal substrate. The formation of this thick polycrystalline structure is one of the steps which makes the DI process expensive. The process is outlined in Figure 23.

Figure 23 shows the flow process for manufacturing the DI thermal printer printhead and is in general the same process that would be required for the production of high voltage isolated, high-speed, or radiation hardened devices. The process starts with the selection of (100) orientation silicon of the desired resistivity and type for the bipolar p-n junction collector. (Not only bipolar devices but also combinations of bipolar and field effect or MOS structures can be fabricated in DI material). The second step in the DI process is to form a mask oxide by thermal oxidation across the surface of the (100) orientation slices. A photolithography mask is then used to open up the oxide in the desired patterns of the isolated devices. This step is carried out by conventional photolithography processing. After the oxide has been removed in the desired isolation pattern, an orientation dependent etch is used to etch trenches or v-shaped grooves into the (100) oriented substrate. These trenches or grooves are etched around each component that is to be isolated from its neighbor in the structure and around each circuit. These isolation grooves are crystallographically sharp and the depth of the isolation groove is controlled completely by the width of the oxide opening at the surface. Exact control of the etched depth by the oxide mask opening permits the etching of built-in thickness indicators at the same time as the orientation dependent etching of the isolation moats. These built in thickness indicators are used at the lap and polish-back process step to provide very accurate indicators of the thickness of the remaining single crystal silicon in each isolated tank. After the isolation etch step is completed the mask oxide is stripped from the surface and a thermal oxide is grown across the orientation dependent etched moats and mesas. In the case for high-voltage or radiation hardened devices, a second dielectric isolation media such as a CVD oxide is also deposited in order to ensure a pinhole-free DI structure. In the case of the electronic printer as outlined in Figure 23, a silicon nitride film (Figure 21) or silicon carbide film (see Table 6) may be deposited as the second DI. This film will also act as a wear resistant film in the final device structure. If the device is built for high-speed applications, a collector contact media such as a diffusion or thin epitaxial deposition of high carrier concentration can be deposited prior to the deposition and/or thermal growth of the DI media.

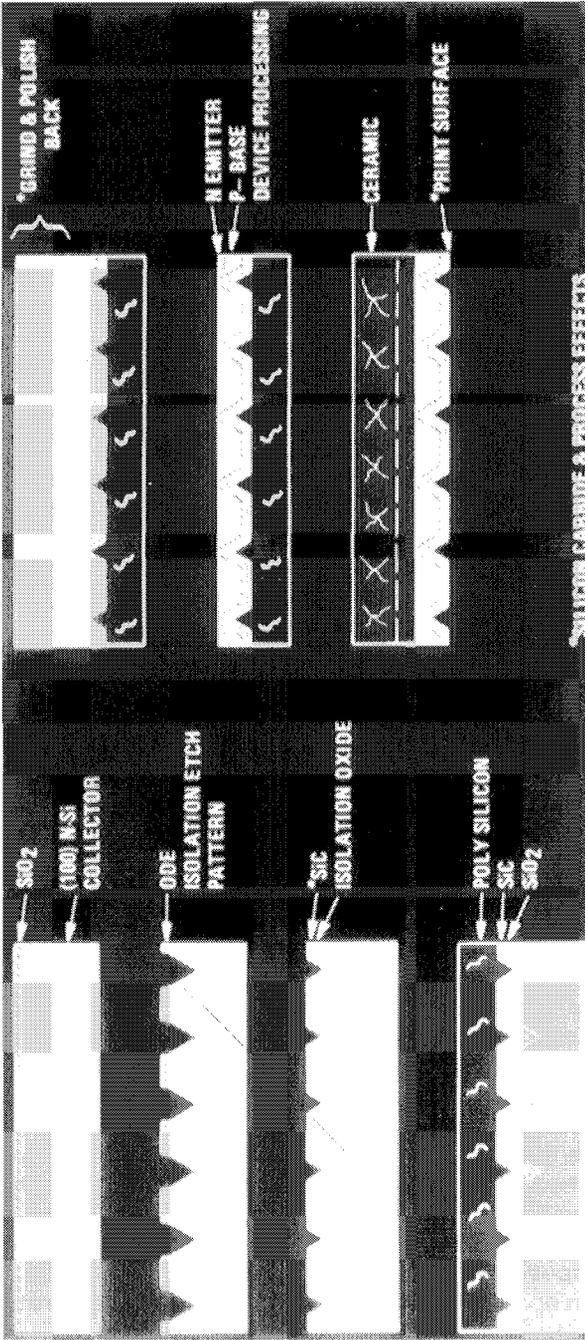
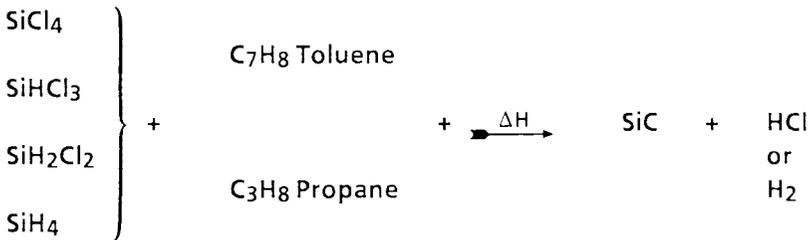


Figure 23: Thermal printer process.

Table 6: CVD of Silicon Carbide

DEPOSITION SOURCESUSES

- WIDE BAND GAP / HIGH TEMPERATURE SEMICONDUCTOR
- WEAR RESISTANT SURFACE
- CHEMICALLY INERT ETCH STOP
- X-RAY LITHOGRAPHY MASK PELLICLE

This thin film collector contact structure will be exposed at the top surface of the final device structure giving extremely low resistance to the collector contact which provides very fast collector saturation and in turn high-speed capabilities. After the DI material is deposited and/or grown, a thick polycrystalline silicon film is CVD deposited across the structure. This film is normally the same thickness as the starting substrate, in the order of 0.5 mm in thickness. After the thick polycrystalline silicon film is deposited, the structure is inverted and the original substrate is ground and polished until the isolation moats and/or thickness indicators are visible. At this time a final polish is applied and the structure is ready for final device processing. The starting structure was the collector resistivity material, however, it is possible to have a collector contact exposed at the surface if desired. In the electronic printer process the next steps are the formation of the base and then the emitter by diffusion followed by metallization as in a conventional junction isolation processing. At the grind and polish process step the extremely hard silicon carbide and/or silicon nitride will act as a polishing stop. This aids in the planarization of the entire structure. After the metallization, which may be CVD polysilicon, silicide, or metal, is applied, in the electronic printer process the structure is bonded down to a ceramic which has matching metallization leads. The thick polycrystalline silicon layer is then removed by grinding and/or etching back to the original ODE surface which is now covered by oxide, nitride, or carbide. All of these DI materials act as a good etch stop. However the nitride and/or carbide are superior in stopping the etch and in providing a wear resistant surface for the electronic printer.

Figure 24 shows a top view of the electronic printer printhead which is a small silicon chip consisting of two 5×7 arrays. Each of the 35 components in the arrays are dielectrically isolated from each other. Each array is the size of an upper case letter or figure on a typewriter. The ODE etched scribe lines are evident in Figure 23 surrounding each pair of arrays which are still in slice form. This printhead prints silently at a speed of 120 characters per second. The use of two 35 dot printer arrays in one printhead allows the high-speed. One array prints and then the other, alternating as the printhead moves across the thermally sensitive paper to minimize thermal memory effects. This printhead structure must be thermally isolated, electrically isolated, and mechanically stable. The built in control thickness indicators are visible as five small black rectangles between the two printhead arrays. The single individual isolated moat structure in the upper left hand corner of the double array is a thermistor which senses the print head temperature and controls the voltage to prevent over heating.

Figure 25a shows a low angle SEM top view and Figure 25b shows a near right angle or 90 deg cross-sectional view of the ODE isolation moats. In (a) a portion of the 5×7 printer array is shown. In this particular design of the printhead, the thermal sensor is shown at the extreme left in Figure (a) and is located immediately adjacent to the active 5×7 printer array whereas the sensor in Figure 23 is more remotely located. The very exact, crystallographically sharp, control of the ODE isolation etch is evident both in (a) and (b) of Figure 25.

Figure 26 shows a conventional high-speed or radiation hardened circuit produced by the DI process. The black lines surrounding each individual component in the circuit is the vertical edge of the DI film.

Figure 27 shows the electronic printer double printhead following the metallization step. Again, the black line surrounding each printer dot in Figure 26a is the DI media at the surface.

Figure 27 shows an artist drawing of the cross-section as the printhead would appear at this step. Figure 27c is an artist drawing of the cross-section of the electronic printer as it would appear in its final structure. The individual printheads have been bonded to a ceramic header and the thick polycrystalline silicone has been removed from the active side of the printer head.

Figure 28 is an enlarged view of the active printhead surface showing the individually isolated components. Each component heats on command due to the emitter/collector resistivity, and then cools quickly in order to by electrical command thermally address the thermal sensitive paper. Each of these DI components are approximately 0.35 mm square at the surface region and are separated only by the thin DI material which is in this case approximately 1.5 microns thick on each wall for a total of 3 micron separation between active components.

Figure 29 shows a SEM cross-section of one of the individual ODE etched DI mesas following the thermal oxidation, CVD oxide deposition, and CVD silicon carbide deposition steps. The DI films of SiO_2 and SiC are readily evident in this SEM photograph. Silicon carbide or silicon nitride would not be required in standard processing for high voltage or radiation

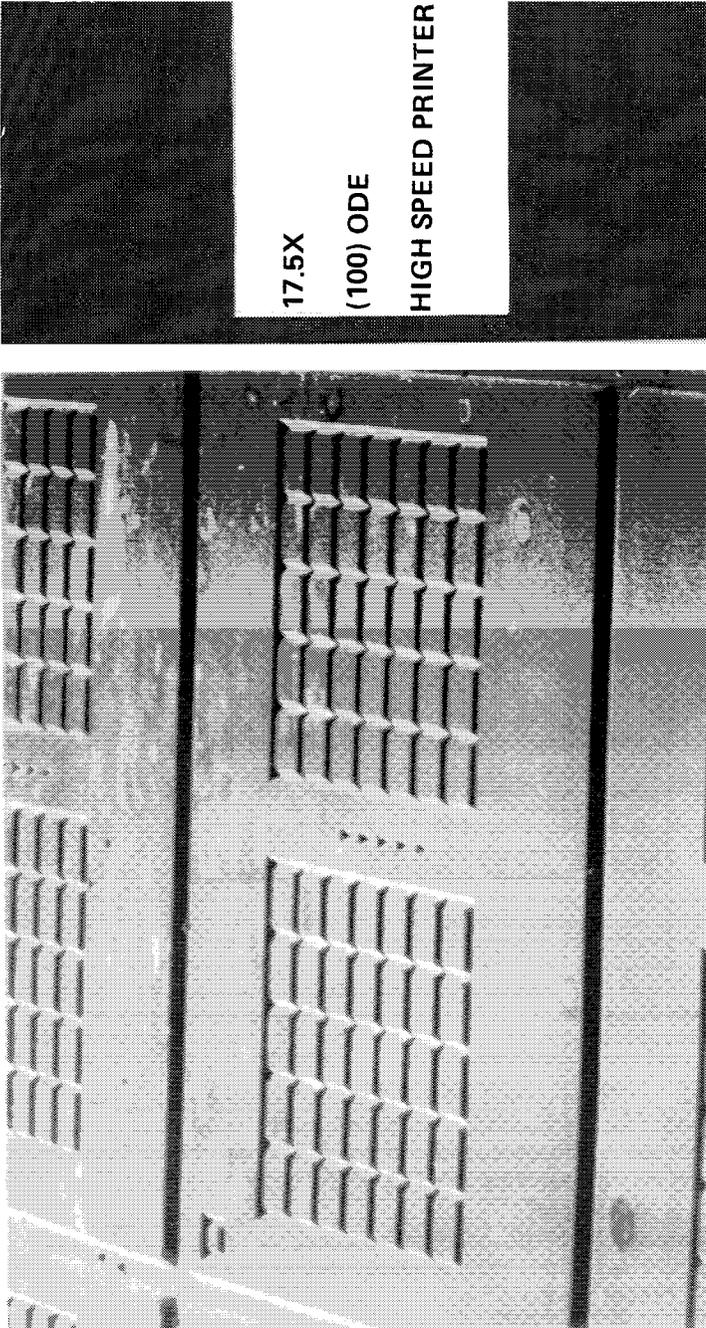


Figure 24: High speed printer silicon chip.

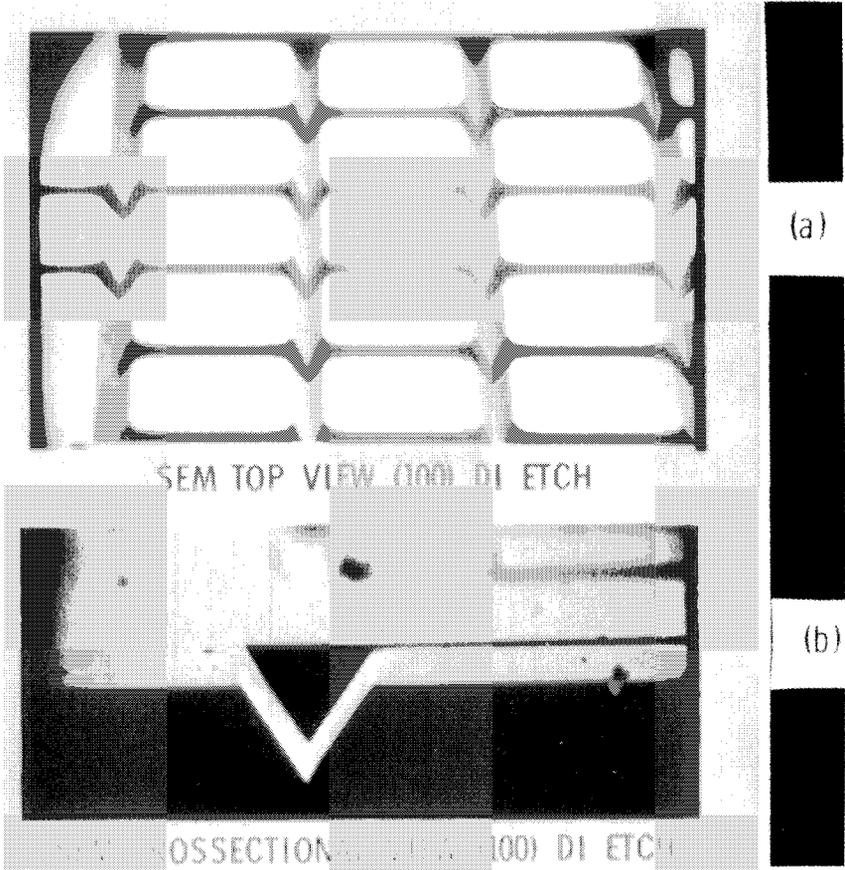


Figure 25: (100) dielectric isolation processing.

hardened circuits and is used here as a process control as well as a wear resistant surface film media on the final printhead surface.

Figure 30 shows a close up view of a portion of the 5 × 7 dot, mesa printhead array in its final form, with the ODE mesas covered with silicon carbide for wear resistance.

Figure 31 shows the relationship of the individually etched and isolated mesas to the complete printhead array as well as the relationship of the printer array to that of the ceramic chip to which it is bonded. It also shows the relationship of the ceramic chip to the electronic printer which is the size of a standard portable typewriter.

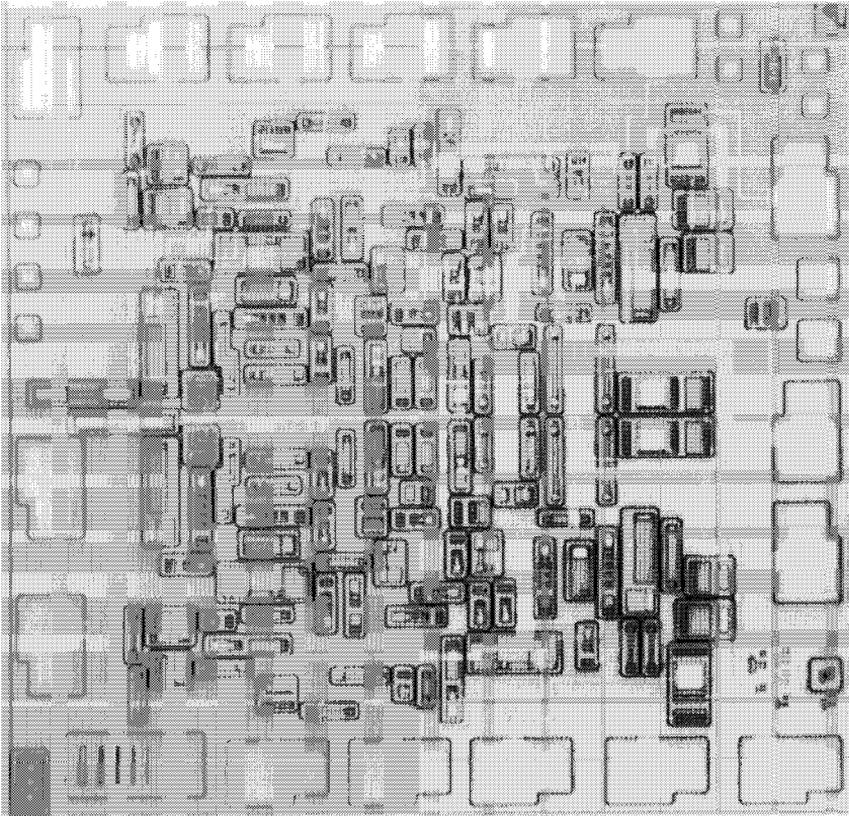


Figure 26: Completed dielectrically isolated circuit.

X-RAY LITHOGRAPHY MASK FABRICATION

CVD of silicon carbide, silicon nitride, silicon dioxide and/or boron nitride may be used in the processing of x-ray lithography mask.^{25,26} The process steps for producing such a mask are shown in Figure 32. In this case CVD deposited silicon carbide is used to illustrate the process, however, silicon nitride, silicon nitride in combination with silicon dioxide or boron nitride can also be used. The process starts by CVD deposition of a thin film of one of the above mentioned materials onto a silicon substrate of the desired diameter for the final mask. This thin film will be the final pellicle or thin membrane which supports the x-ray blocking mask material which

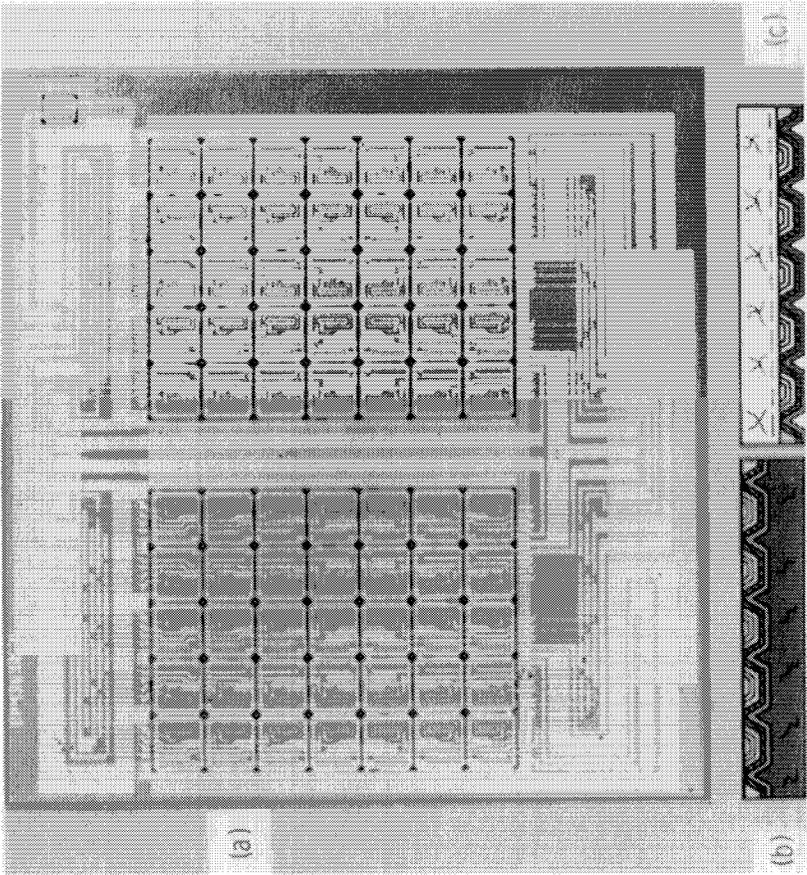


Figure 27: Electronic printer chip (a) top view photo, (b) (c) cross-sectional drawings.

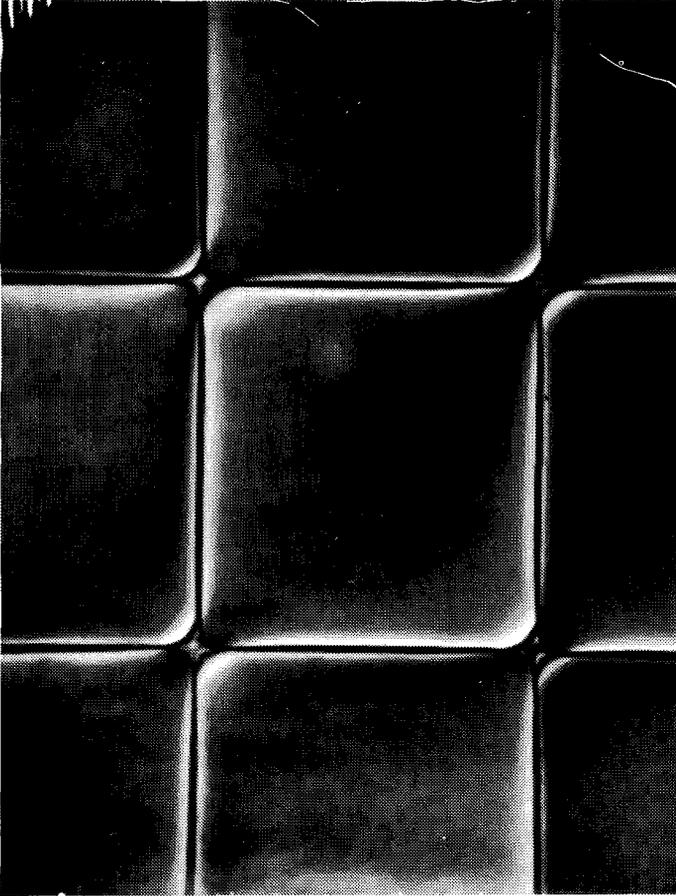


Figure 28: Dielectric isolated, electronic printer array silicon carbide film around each silicon island (electrical-thermal and physical stability).

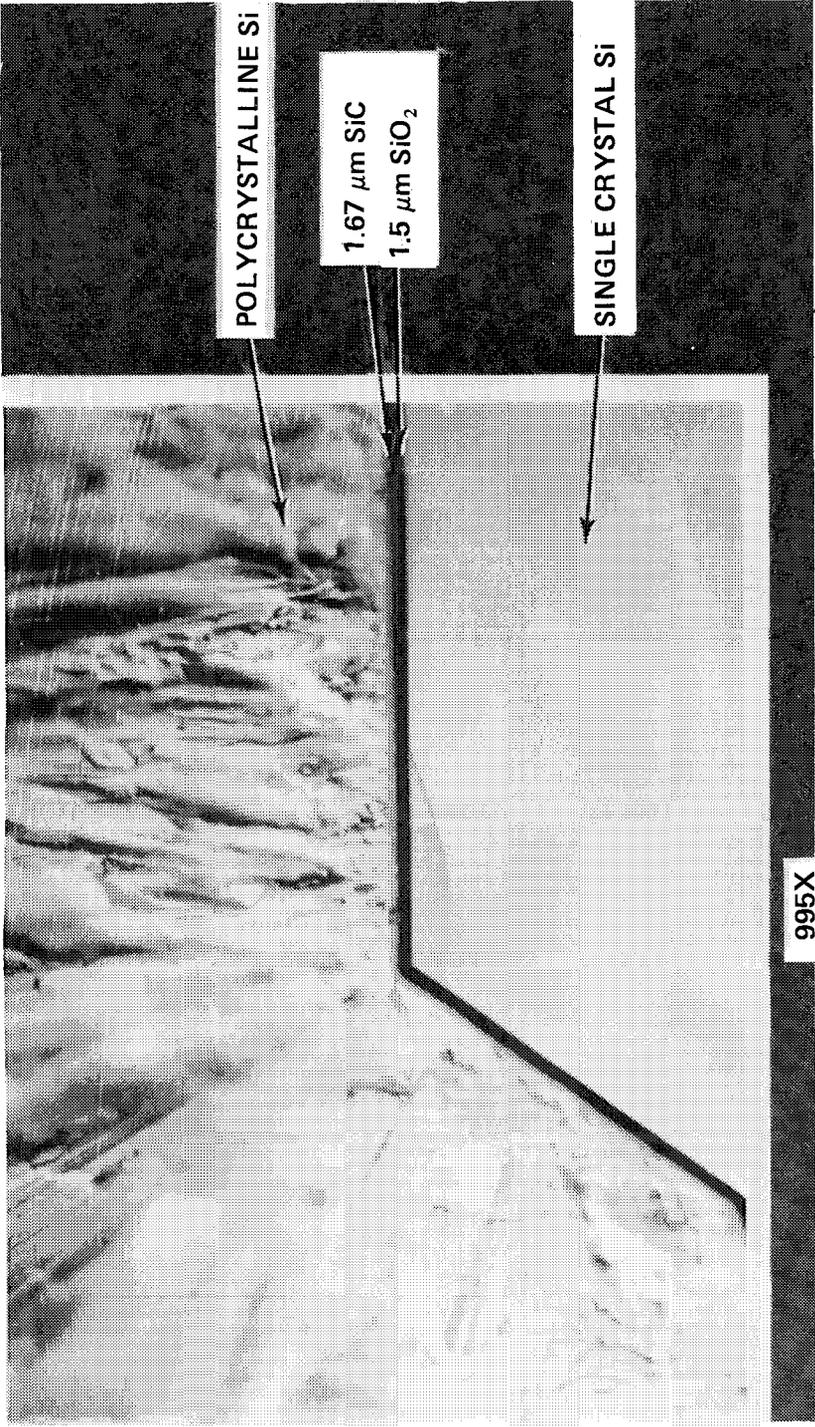


Figure 29: CVD polysilicon on SiC, on SiO₂, on single-crystal (100) Si.

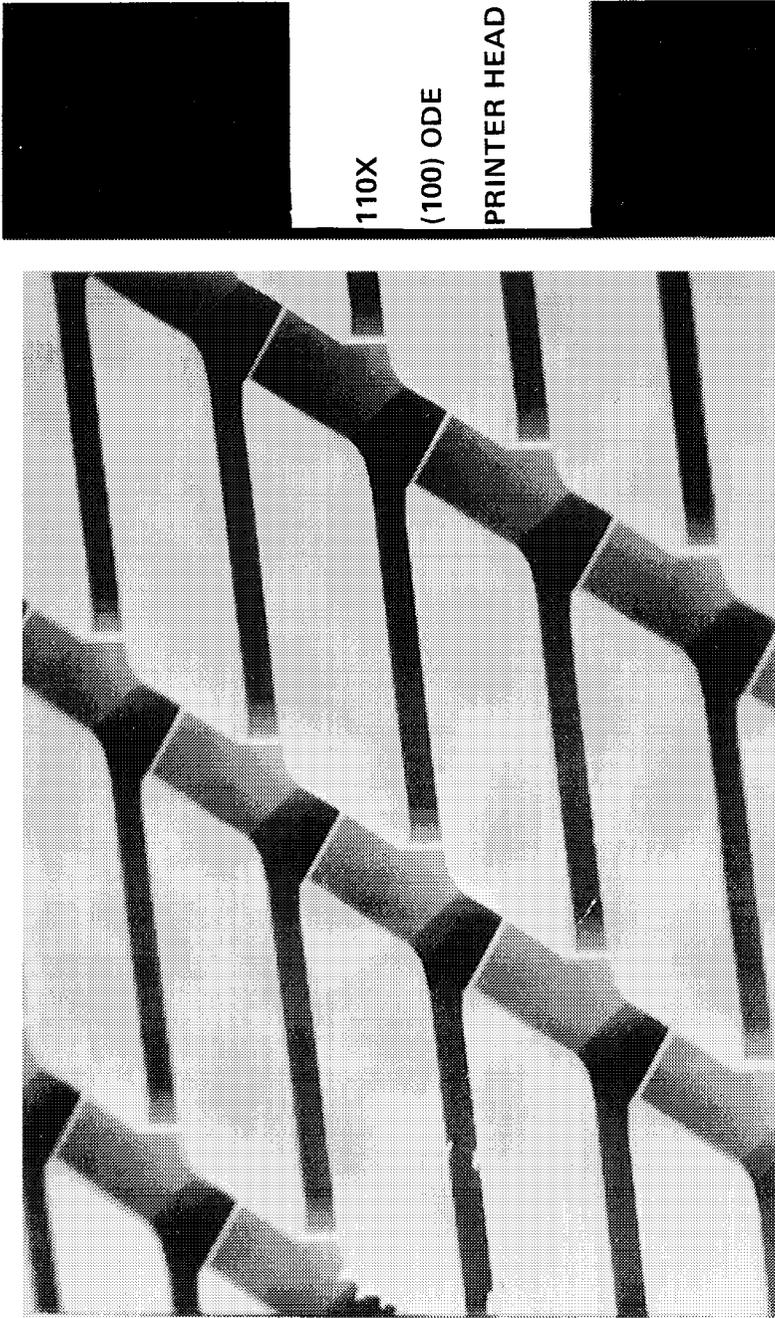


Figure 30: Mesa printhead array.

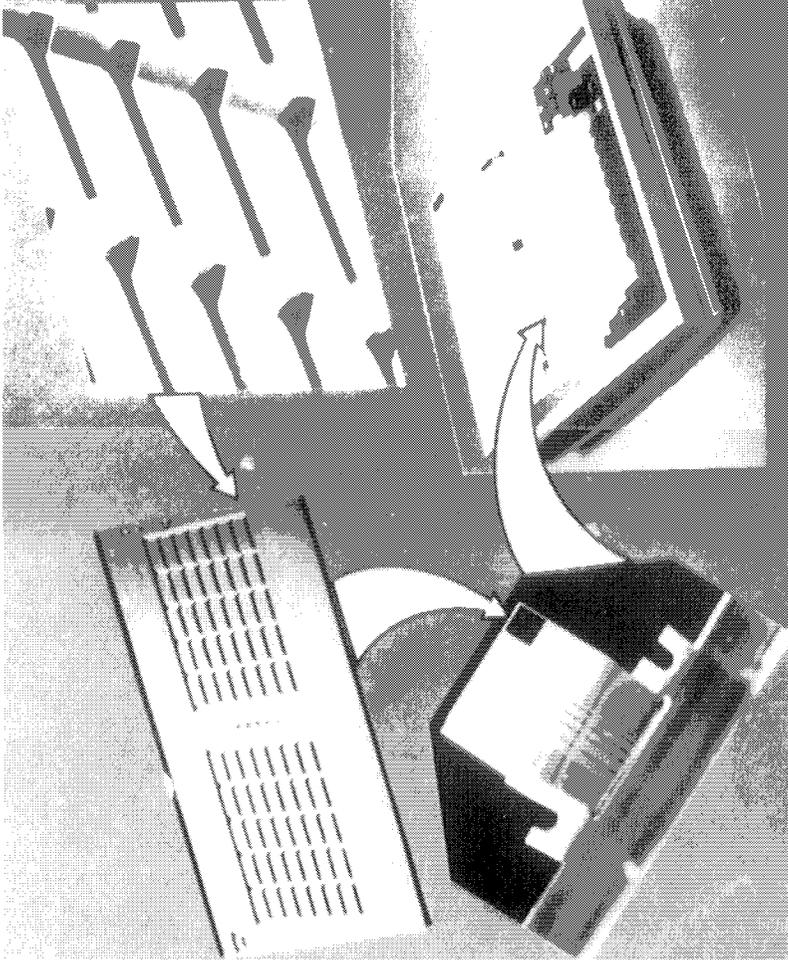


Figure 31: Composite photo of dielectric isolated printer head chip.

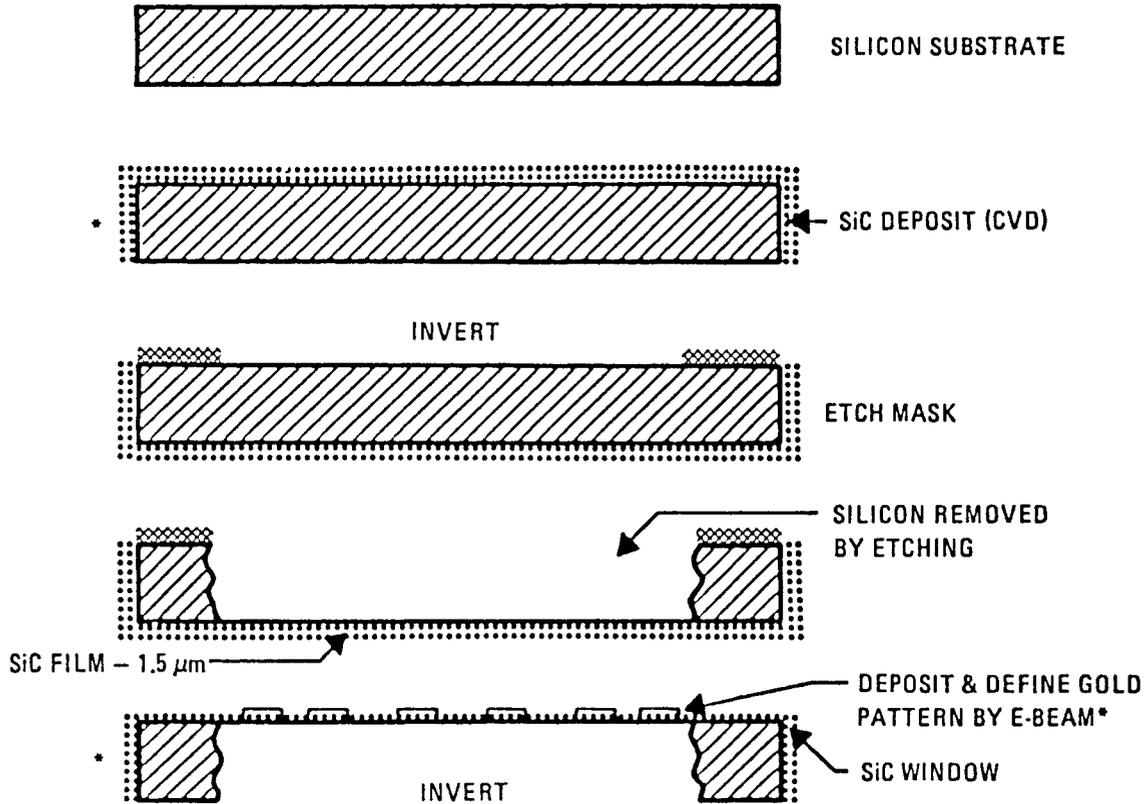


Figure 32: X-ray lithography mask fabrication.

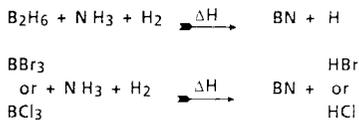
may be, e.g., gold. After the thin pellicle or window, which is transparent to x-rays and also transparent in the visible wavelength region is deposited, the structure is inverted and the silicon is etched from the central portion of the original silicon substrate leaving a supporting ring of silicon around the periphery of the structure. The gold or x-ray blocking material may be deposited and patterned prior to this etch step or it may be deposited and patterned following the silicon etching step. The silicon carbide, silicon nitride/oxide, or boron nitride film forms an exact replica of the starting substrate surface at the thin film/substrate interface. This very exact replica of the polished substrate provides a very flat, highly reflective membrane surface on which the gold is to be deposited. If silicon carbide is to be used as the pellicle one may use the hydrogen reduction of silicon tetrachloride (SiCl₄) and propane (C₃H₈), both being of electronic grade, at a deposition temperature of approximately 1200°C. The film should be deposited at as near stoichiometric conditions as can be obtained in order to produce transparent stress-free films. These conditions can be obtained by using approximately 0.89% silicon tetrachloride (SiCl₄) and 0.37% propane (C₃H₈). If one wishes to deposit boron nitride by CVD (see Table 7) one may use the boron hydride (B₂H₆), diborane, at 4.4% in hydrogen and ammonia (NH₃) 6.6% at approximately 530 degree C in an epitaxial-type reactor. In general the silicon carbide pellicle is the strongest and most stable pellicle. The silicon carbide pellicle or window may be as thin as 1.5 microns, whereas the boron nitride film must be in the order of 4-6 microns to form the transparent membrane across a 100 mm diameter mask.

Figure 33 shows the visible transmission of a thin 1.5 micron, silicon carbide pellicle across a 100 mm diameter silicon slice ring. The central portion of the silicon has been removed and the visible transmission of the silicon carbide pellicle is evident due to the transmission of the printed advertisement for the Datachron calculator photograph below it. This thin membrane or pellicle must have high x-ray transmission and high visible transmission for rough preliminary alignment of the mask by visible techniques.

Figure 34 shows a photograph of the completed silicon carbide pellicle with the gold patterned x-ray mask formed on it.

Table 7: CVD Boron Nitride

DEPOSITION SOURCE



USES

- SOLID STATE DIFFUSION SOURCE
- X-RAY LITHOGRAPHY MASK

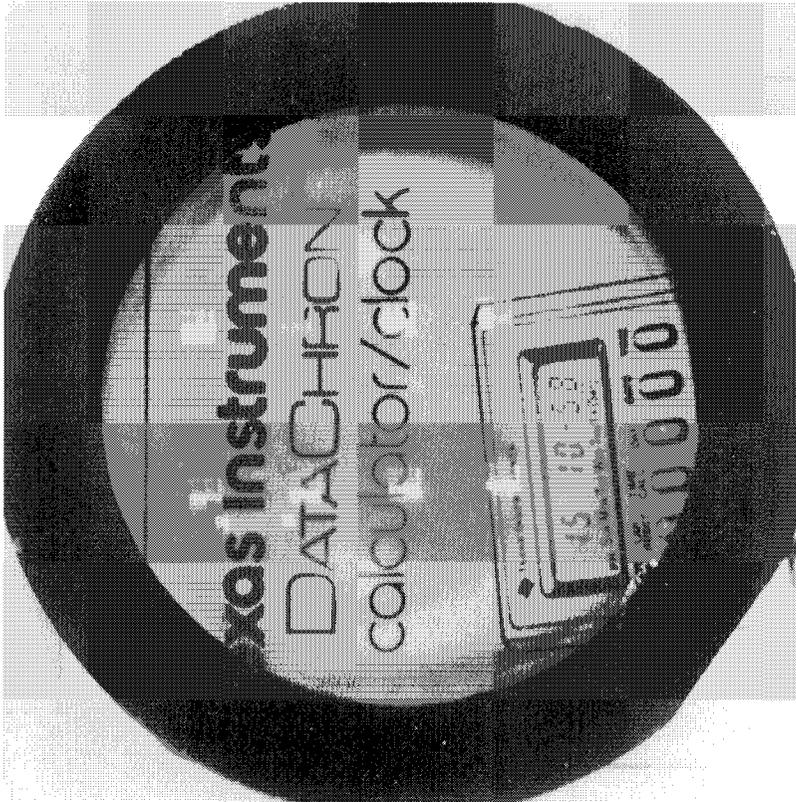


Figure 33: X-ray lithography mask, using SiC film on silicon slice (CVD).

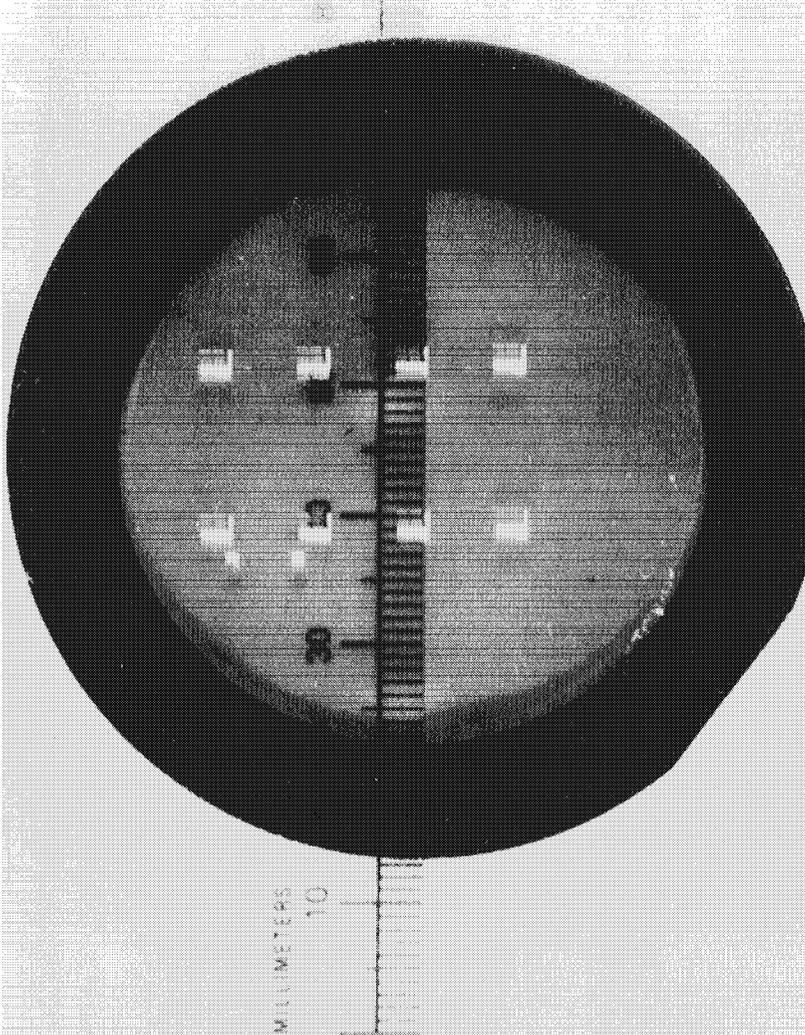


Figure 34: X-ray lithography mask, using SiC film on silicon slice (CVD).

This chapter presents a few of the many chemical vapor deposition processes used in today's manufacturing of semiconductor devices and circuits. The references direct you to only a few more of the many articles on this subject found in the open literature.

REFERENCES

1. DuPont Process: D.W. Lyon, *J. Electrochem. Soc.*, 96: 356 (1949) Process used by DuPont, Eagle Picher, and Sylvania/Siemens Process: E. Spenke, *Semiconductor Silicon*, Electrochem. Soc. 1 (1969).
2. Eagle Picher Research Laboratories, *Fourth Quarterly Report*, Contract # DA36-039-SC-66042, (July 1, 1956).
3. Eagle Picher Process, *Sixth Quarterly Report*, Contract # DA36-039-SC-64694, Aug. 1, 1956 to Nov. 1, 1956, U.S. Signal Corps. Project No. 152B. Also Final Quarterly Report, this contract Feb. 1, 1957 to May 1, 1957.
4. W.R. Runyan, *Silicon Semiconductor Technology*, vol. 12 McGraw-Hill, (1965).
5. G.K. Teal and J.B. Little, *Phys. Rev.* 78: 647 (a) (1950).
6. R.C. Sengster et al., *J. Electrochem. Soc.* 104: 317 (1957).
7. H. Mark, *J. Electrochem. Soc.* 107: 568 (1960).
8. H.C. Theurer, *J. Electrochem. Soc.* 108: 651 (1961).
9. B.A. Joyce and P.R. Bradley, *J. Electrochem. Soc.* 110: 1235 (1963).
10. K.E. Bean, W.R. Runyan and R. Massey, *Semiconductor International*, 136 (May 1985).
11. K.E. Bean, and P.S. Gleim, *IEEE*, 57: 1469 (1969).
12. G.A. Lang and T. Stavish, *RCA Rev.* 4: 488 (1963).
13. K.E. Bean and P.S. Gleim, Electrochemical Society, Fall Meeting, Oct. 1963, late newspaper, *Abs. J.E.C.S.* (Dec. 1963).
14. K.E. Bean, *Thin Solid Films* 83: 173 (1981).
15. J.O. Borland, and C.I. Browley, *Solid State Tech.* 141 (Aug. 1985).
16. H. Kurten et al, *IEEE Trans. Elec. Dev.* ED-30, 1511 (1983).
17. G.A. Rozgonyi and C.W. Pearce, *Appl. Phys. Letters* 31: 343, (1977).
18. A.S.M. Salih et al, *J. Electrochem. Soc.* 133: 475 (1986).
19. S. Mendelson, *Single Crystal Films*, Proceedings of Conference held at Philco Blue Bell, PA, (May 1963).
20. S. Mendelson, *J. App. Phys.* 35: 1570 (1964).
21. S.K. Tung, *J. Electrochem. Soc.* 112: 436 (1965).
22. K.E. Bean, P.S. Gleim, and R.L. Yeakley, *J. Electrochem. Soc.* 114: 733 (1967).
23. U.S. Davidsohn and Faith Lee, *Proc. IEEE*, 57: 1532 (1969).
24. K.E. Bean and W.R. Runyan, *J. Electrochem. Soc.* 124: 5C (1977).
25. T.L. Brewer, R.K. Watts, and K.E. Bean, Electrochemical Society Proceedings of Symposium of Electron and Ion Beam Science and Technology, 8th International Conf., *Electrochemical Society Proceedings*, 78-5, 453 (1978).
26. H. Luthje, et al, 30th International Symposium on Electron Ion and Photon Beams, N-4, (May 27-30, 1986).

4

Chemical Etching and Slice Cleanup of Silicon

Kenneth E. Bean

*Texas Instruments Incorporated
Dallas, Texas*

INTRODUCTION

Slice cleaning and wet chemical etching have been key semiconductor processing technologies since the beginning of semiconductor fabrication in the late 1940's and early 1950's. The demand for cleanliness, control of purity, and freedom from defects becomes more stringent with each advance in device and circuit complexity. This chapter discusses wet chemical etching of silicon from the standpoint of planar etching, orientation dependent etching (ODE), concentration dependent etching, and defect delineation etching. It also discusses the etch composition, the masking materials used for preferential etching, mask alignment, and applications for the above etching technologies. We will also discuss silicon slice cleanup procedures and effects thereof. Table 1 summarizes the subjects to be discussed, in order of discussion.

Planar etch is a solution that etches silicon in all crystallographic directions at the same rate. A common formulation is made up by mixing hydrofluoric acid (HF) \approx 8% by volume, nitric acid (HNO₃) \approx 75%, and acetic acid (C₂H₄O₂) \approx 17%. At 25°C this solution etches silicon slices (wafers) at approximately 5 μ m per minute, (see Table 2). Orientation dependent etches (ODE) have been developed which etch much faster in one crystallographic direction than in another. For example, a solution of potassium hydroxide and water (KOH + H₂O) in equal parts (50%-50% weight) at 80°C etches silicon in the $\langle 110 \rangle$ direction \approx 700 times faster than in the $\langle 111 \rangle$

Table 1: Chemical Etching and Slice Cleanup of Silicon

<p>*WET CHEMICAL ETCHING</p> <ul style="list-style-type: none"> • PLANAR • ORIENTATION DEPENDENT • CONCENTRATION DEPENDENT • DEFECT DELINEATION • SLICE CLEANUP "ETCHES" 	<p>*EFFECTS IN (100) SILICON</p> <ul style="list-style-type: none"> • FOUR FOLD SYMMETRY • ETCH SOLUTIONS • MASK SYMMETRY • ALIGNMENT • APPLICATIONS
<p>*EFFECTS IN (110) SILICON</p> <ul style="list-style-type: none"> • (111) TRACE-FLAT • ETCH SOLUTIONS • ALIGNMENT • APPLICATIONS 	<p>*DEFECT DELINEATION</p> <ul style="list-style-type: none"> • (100) { SECCO WRIGHT-JENKINS SCHIMMEL • (111) { SIRTL DASH LEO'S
<p>*SILICON SLICE CLEANUP</p> <ul style="list-style-type: none"> • STANDARD CLEANUPS • CHOLINE CLEANUPS 	

Table 2: Chemical Etches and Characteristics for Silicon

ETCHES	CHARACTERISTICS	COMPOSITION	RATE AND REMARKS	REF
PLANAR ETCH	ETCH UNIFORMITY	HF - HNO ₃ - HAc ≈8% ≈75% ≈17%	≈5 μm/MIN AT 25°C (IN ALL DIRECTIONS)	1
1-3-10	ETCHES P+ OR N+ SILICON "STOPS" AT P- OR N-	HF HNO ₃ HAc 1 3 10	≈3 μm/MIN (100) 25°C	2
(100) ODE	ETCHES [100] ~100 X [111] DIRECTION	KOH - NORMAL PROPANOL - H ₂ O KOH - 250 gm N PROP - 200 gm DI H ₂ O - 800 gm	≈1 μm/MIN AT 80°C, [100] "STOPS" AT P+ INTERFACE ETCHES Si ₃ N ₄ AT 14 Å/HR SiO ₂ AT 20 Å/MIN	3-5
(110) ODE	ETCHES [110] 600 X [111] DIRECTION	KOH - H ₂ O 50:50 VOL.	≈0.8 μm/MIN AT 80°C IN (110) SILICON	5-7
ETHYLENEDIAMINE	ORIENTATION DEPENDENT AND CONCENTRATION DEPENDENT	ETHYLENEDIAMINE - PYROCATECHOL - H ₂ O EDA - 255cc H ₂ O - 120cc P.C. - 45 gm	≈1.1 μm/MIN AT 100°C IN [100]. "STOPS" ETCHING AT P+ INTER- FACE. VERY SLOW ETCHING OF SiO ₂ (≈3 Å/MIN) "NO" ETCH OF Al, Au, Ag, Cu, Ni OR Ta.	5-8

direction. See Table 2 and Figure 28. If we add normal propanol to the KOH and H₂O etch, we can also etch silicon in the <100> direction approximately 100 times faster than in the <111> direction, at 80°C. See Table 2 and Figure 10.

If we mix the same mineral acids used in the planar etch solution in the ratios of one part HF, three parts HNO₃, and 10 parts C₂H₄O₂, we have an etch commonly known as Dash etch, which etches p+ silicon or n+ silicon, >7×10¹⁹ carrier concentration, much faster than p- or n- silicon. In contrast, the KOH-propanol-water etch "stops" (slows down by ≈ 20X) at a p+ interface. The ethylenediamine (EDA) etch, made up of EDA, pyrocatechol and water (see Table 2) is also both orientation dependent and

concentration dependent in etching silicon and has the advantage of etching silicon dioxide very slowly ($3 \text{ \AA}/\text{min}$) at 100°C .

In today's high density circuit technology the detection, or delineation, of material defects is of great interest and importance. Wet chemical etching is commonly used to show these defects. In general, these solvents preferentially etch the damage site due to the strained or damaged crystal lattice bonding in the defect area. Sirtl, Dash, and Secco etches are preferred for (111) crystal damage evaluation. Wright-Jenkins, Schimmel, Yang, and Secco etches are used for (100) crystal defect evaluation. (See Table 4.)

Silicon belongs to the diamond cubic crystal structure, (perhaps the most desirable crystal structure to work with). Figure 1 is a model which shows seven of the low indices planes of the diamond cubic structure we may choose to use in silicon processing. In general, the (111), (100), and (110) planes are the predominant planes used in silicon processing today. However, other planes are also predominant in etching and deposition. These planes the (331) and the (113), lie in $\langle 310 \rangle$ directions. The (221) and the (112) planes, which lie in the $\langle 210 \rangle$ direction, are also governing planes in silicon etching and deposition.

Figure 2 shows the three low indices planes commonly used in silicon processing in a more vivid display. Figure 3 shows the method of deriving the Miller indices for the crystal planes. The (111) plane, for instance, is one unit length out from the apex of lines A, B, and C in the A direction, in the B direction, and in the C direction. The plane bounded by the lines connecting A, B, and C is the (111), which is the predominant plane in silicon processing. The atomic packing in this plane is the tightest, or most closely spaced, packing density available; therefore this plane dominates the etching and deposition conditions in silicon. It is the most stable plane in the silicon crystal structure and is the most difficult on which to etch or epitaxially deposit.

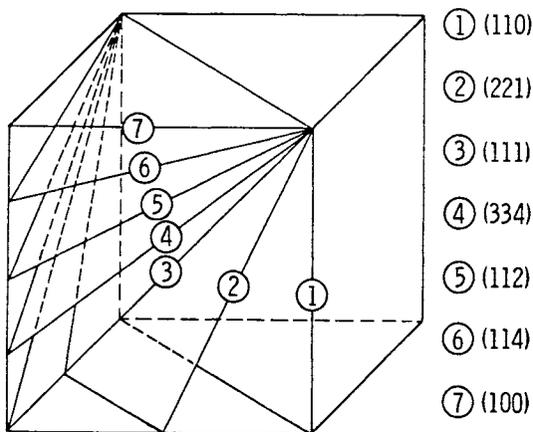


Figure 1: Low indices planes of the diamond cubic crystal structure.

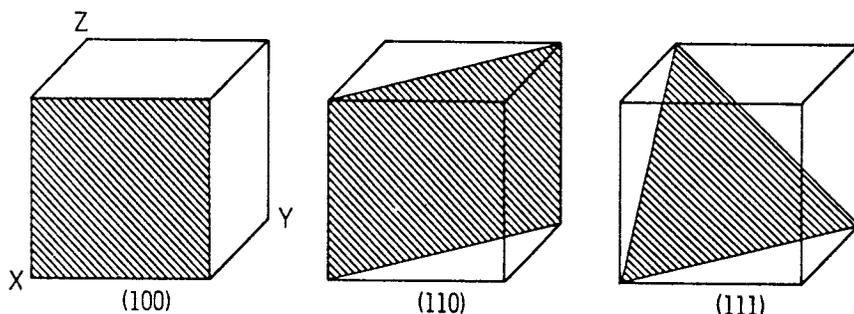


Figure 2: The (100) (110) and (111) planes.

Figure 4 is a photograph of a crystallographic model of the diamond cubic structure. In the $\langle 111 \rangle$ direction we see atoms in a very closely spaced equilateral triangular arrangement. This view shows the extremely high packing density. If we rotate this same crystallographic model 54° to the right or to the left, we are looking in a $\langle 100 \rangle$ direction and see that the atoms are arranged in a square array. The atomic packing density in this direction is slightly less, making it a more open lattice. Therefore, one would expect that etching in this direction would proceed more rapidly than into the more highly packed (111) plane. If we rotate from the (100) plane 45 or 90 degrees, using this same model then we will be looking in a $\langle 110 \rangle$ direction. The atoms are in a very open lattice structure, which exhibits the fastest etching and deposition conditions. This open lattice structure can also be used to advantage for deep ion implantation. Channeling of the ions takes place very readily in this open lattice $\langle 110 \rangle$ structure. This plane in silicon can also be used to advantage for radiation hardened circuitry. A radiation particle must travel further in this direction before colliding with a silicon atom, thus producing less radiation damage, than in the $\langle 100 \rangle$ or $\langle 111 \rangle$ direction. From this figure we can also see that the high density of the (111) plane should make it a very strong plane. The (110) plane is 90 degree to the (111) plane. The plate-like high packing density structure of the (111) planes are held or bonded together by the structure of the more open lattice (110) plane structure. Therefore, when we break or cleave a silicon slice, it will cleave along $\langle 110 \rangle$ directions between (111) planes, separating or breaking (110) bonds.

ORIENTATION DEPENDENT CLEAVING OF SILICON

Figure 5 shows a (100) and a (111) silicon slice that have been cleaved by pressing the center of each with a hard object, such as tweezers or a ballpoint pen, when the slice is lying on a pad of paper which allows it to give, therefore causing it to break. In the case of the (100) silicon slice note the fourfold symmetry of the (111) cleavage planes. Their traces intersect

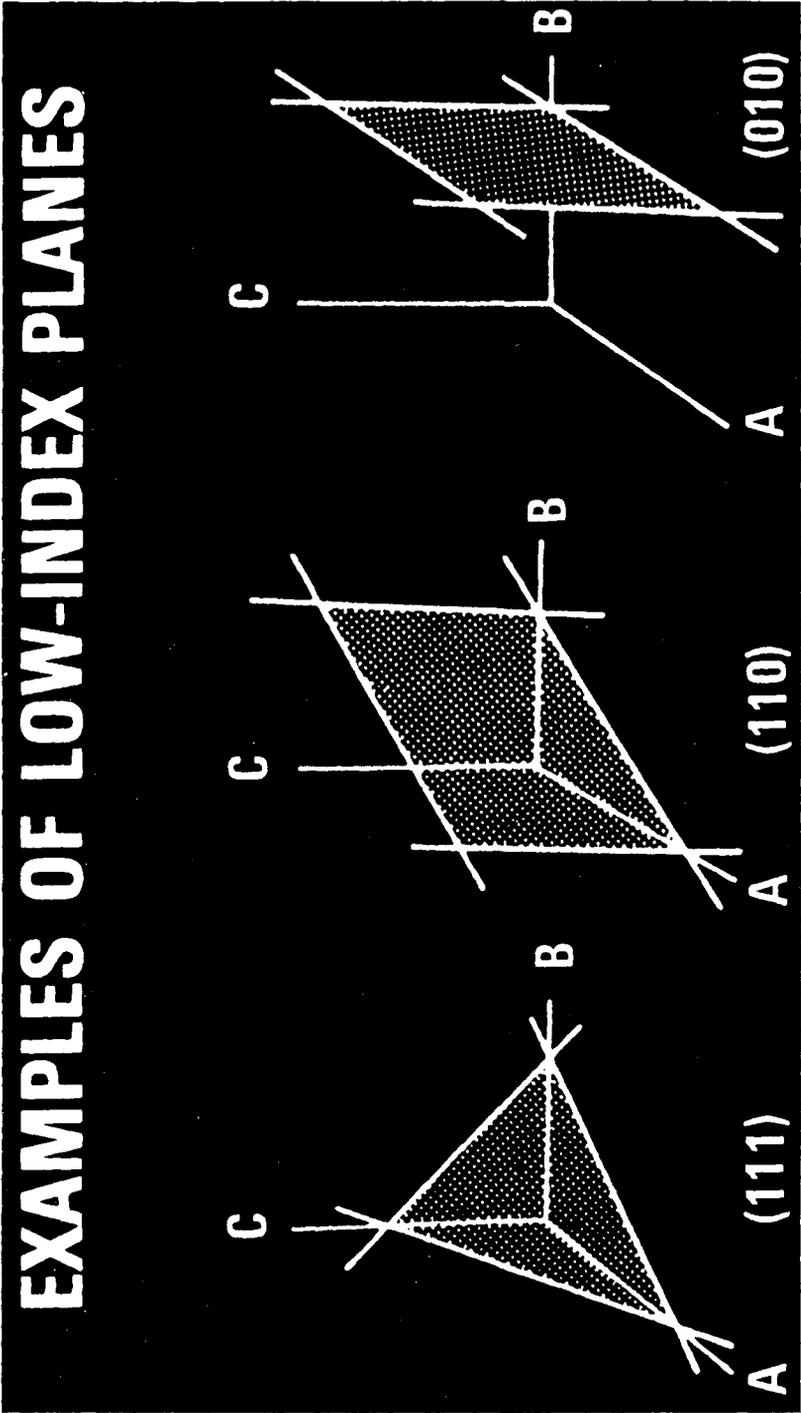


Figure 3: The derivation of the (111) (110) and (100) Miller indices.

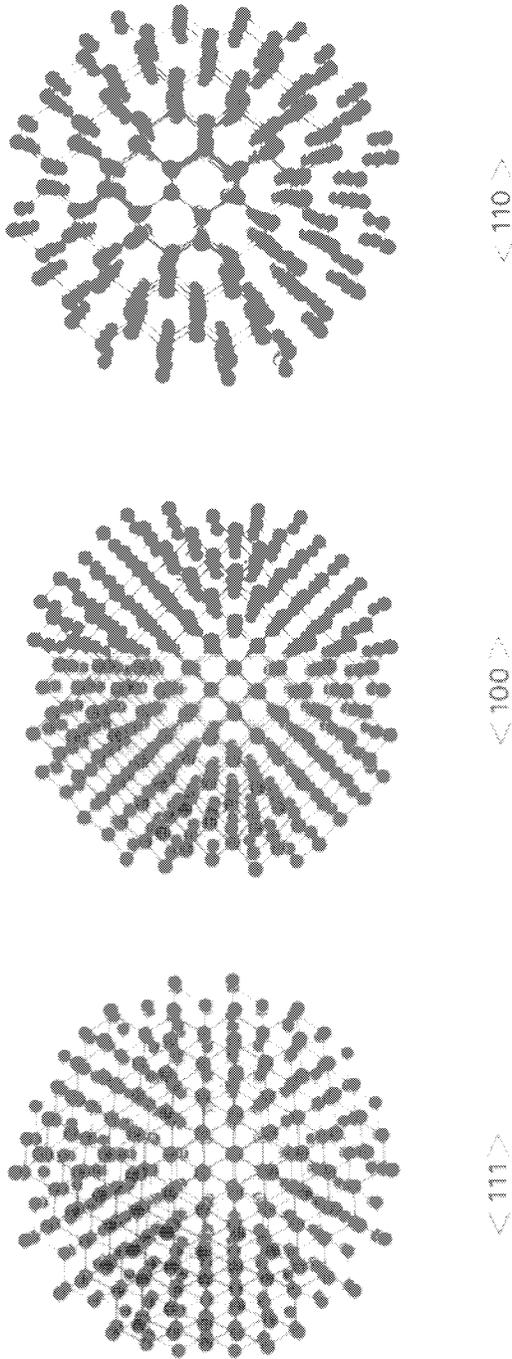


Figure 4: Model showing three low index directions in silicon (cubic, diamond structure).

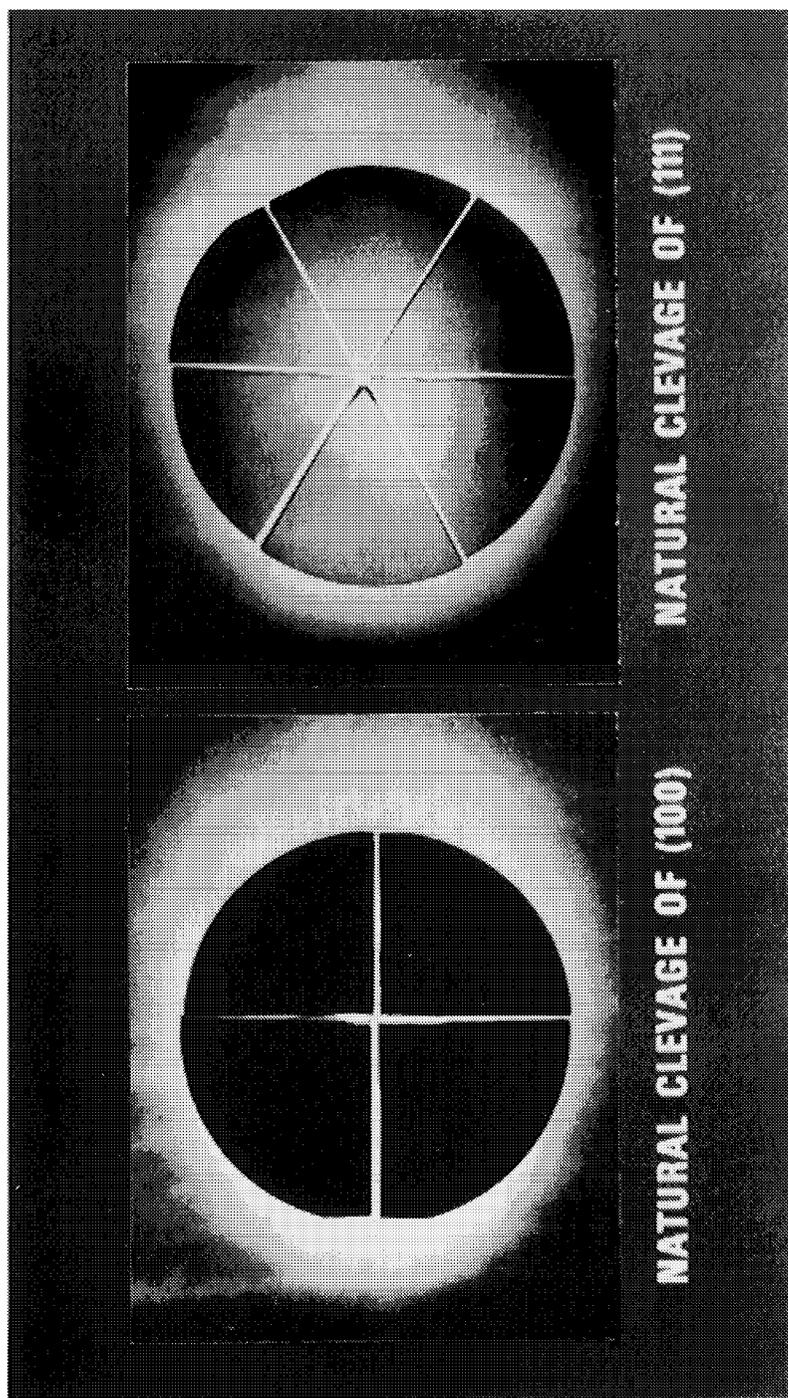


Figure 5: Cleaved (100) and (111) silicon slices.

the (100) surface in the pattern shown. These (111) planes intersect the (100) surface plane at an angle of 54.74 degrees. If we look at the cleavage plane or the edge of the cleaved border of the slice we see a crystallographic plane that is inclined to the surface at this 54.74 degree angle. In the case of the (111) slice there are three cleavage planes 120 degrees apart. When they extend all the way across the slice, they break into pie-shaped segments with 60 degree angles. The (111) cleavage planes intersect the (111) surface of the slice at 70.53 degrees. If one again cleaves the sections that have already been cleaved, they will continue to cleave, in the case of (100) into squares or rectangles, and in the case of the (111) into 60 degree triangular shapes. If we cleave a (110) silicon slice it will cleave 90 degrees to the (110) surface. If we continue to cleave these sections we will see that they form rhombic shapes as shown in Figure 6. However, if we look at the edge or the cleaved surface we will see that in all cases they are 90 degrees to the (110) surface. These are the (111) planes.

In today's silicon semiconductor processing there is great interest in MOS-type structures. For this type of structure the (100) slice orientation is usually used due to the low surface state density at the (100) silicon surface-silicon dioxide interface.

Figure 7 is a stereographic projection of the standard (001) or (100) face centered cubic crystal structure. In this projection we are looking directly at the (100) surface as we would in a (100) silicon slice. Note that the four (111) planes which intersect the (100) surface are slightly greater than half way out to the periphery of the projection (slice) in $\langle 110 \rangle$ directions. In other words, there is a (110) plane perpendicular to the (100) plane and tangent to the periphery at this point. The (111) planes are actually coming into the (100) surface at angles of 54.74 degrees. The (111) planes are also at 90 degree angles to each other. Those planes designated by the Miller indices at the periphery of the projection are known as directions. Starting at the bottom, or the periphery closest to the observer, is a (100) plane, indicating that this is a $\langle 100 \rangle$ direction. Those Miller indices indicating planes between this (100) plane and the (100) plane at the center of the projection are lying in this $\langle 100 \rangle$ direction. Moving to the right of the bottom center we have a $\langle 310 \rangle$ direction. Moving up along the periphery we find a $\langle 210 \rangle$ direction, a $\langle 320 \rangle$ direction, and then the $\langle 110 \rangle$ direction, in which the (111) predominant plane lies. Further examination of this (100) projection shows a four-fold symmetry in this (100) plane. All four quadrants are exactly alike, and the (111) planes are 90 degrees to each other, as are all other families in this projection. Note also that in this (100) projection are a $\langle 310 \rangle$ direction to the right of the $\langle 100 \rangle$ direction at the bottom of the projection and a $\langle 3\bar{1}0 \rangle$ direction to the left of the $\langle 100 \rangle$ direction at the bottom of the projection. These same two (310) planes are also to the right and to the left of the $\langle 100 \rangle$ direction at the top of the projection. Likewise, they are above and below the $\langle 100 \rangle$ direction at the right and at the left of this projection. This shows there is double four-fold symmetry of $\langle 310 \rangle$ direction planes in the (100) projection. In this double four-fold symmetry of $\langle 310 \rangle$ directions the predominant planes are the (311) and the ($\bar{3}$ 11). These planes etch and deposit rapidly in processing. The effect of these

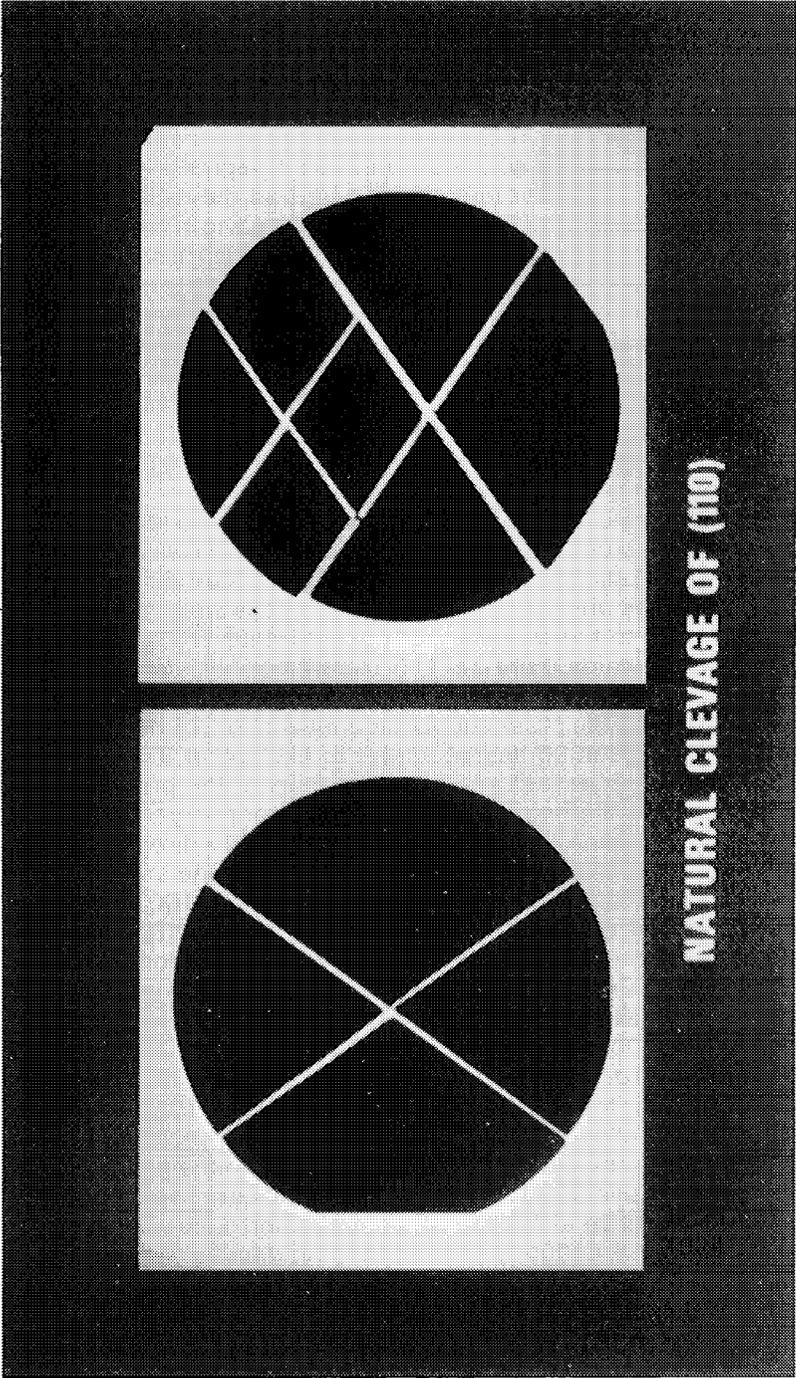


Figure 6: Cleaved (110) silicon slice.

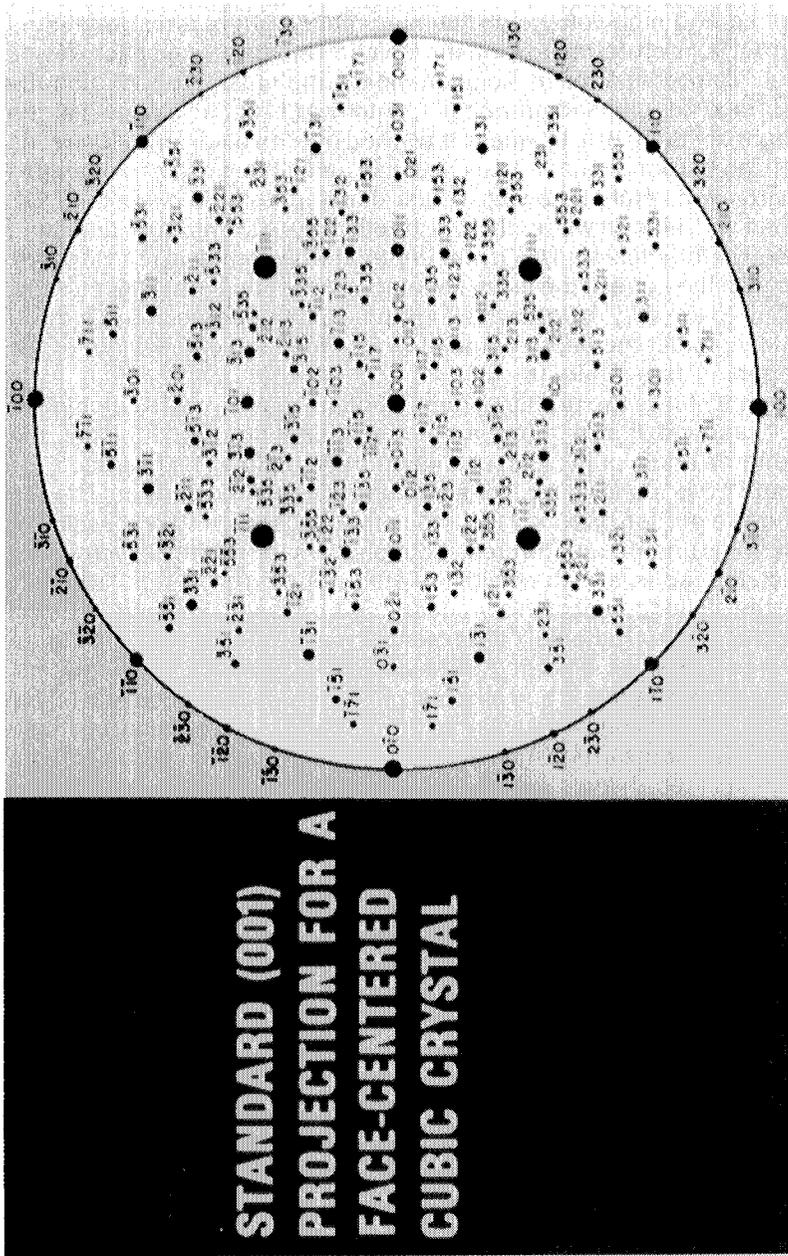


Figure 7: Stereographic projection of (100) plane.

(331) planes will be noted in future etching experiments as rapidly undercutting planes. In some silicon processing, such as the dielectric isolation process, which is used for radiation hardened devices and for extremely high-speed and high-voltage switching circuitry, orientation dependent etching (ODE) is preferred to form the isolation moats around each device or circuit. To produce these isolation moats, the mask is aligned on the (100) surface parallel and perpendicular to the (110) flat, as shown at the lower right of Figure 8. If, the mask is aligned over an oxidized silicon slice parallel and perpendicular to the (110) flat, and if the mask has openings one micron wide etched down to the silicon, through the oxide the ODE described in Table 2 will etch to a depth of 0.707 microns and then completely stop etching. This is the depth at which the two (111) traces, aligned with the edge of the mask opening, intersecting the (100) surface, meet. Due to the very exact crystallographic 54.74 degree angle, we can accurately predict the depth of the etch as a function of the width of the mask opening. See Figure 9.

Figure 9 shows the proper alignment of the mask for ODE etching of the (100) surface. It also shows a cross-sectional drawing depicting the 54.74 degree angle of the (111) plane intersecting the (100) surface.

Figure 10 is a photograph of an actual ODE etched surface. The top view shows a portion of a 5 X 7 isolation array as used in the processing of the electronic print-head. The cross-sectional view shows the exactness of the crystallographic structure with the etch stopping on the (111) planes. Early in the development of this process it was observed that undercutting occurred at convex (outside) corners of the mask area as shown in Figure 11a. Figure 11b is an enlarged SEM view of this exact crystallographic etching with etch faceting and stopping on the (331) planes, as previously mentioned. By carefully measuring the angle of intersection of these faceted planes with the (100) surface and the angle these facets make with the (110) flat, these planes were identified as (331). In Figure 12 and Figure 14, both (331) and (111) planes are shown. In order to compensate for the undercutting at the corners and to obtain square or right-angle corners, a mask corner compensation process was developed. By extending the oxide out over the area that is being undercut, we can compensate for the fast etching in the $\langle 310 \rangle$ direction.

Figure 13 shows the mask corner compensation design for (100) ODE. It should be noted that different orientation dependent etches terminate their etching on different crystallographic planes. These planes may be in $\langle 310 \rangle$ directions or in $\langle 210 \rangle$ directions, depending on whether the KOH-propanol-water solution or the ethylenediamine pyrocatechol and water solution is used. See Figure 14. Figure 15 shows a grossly over etched ODE structure. The etch time was 70 minutes, or time enough to etch 70 microns deep using a mask designed with corner compensation for a 50 micron, or 50 minute etch.

Figure 16 shows the experimentally derived information for adding corner compensation to the mask to obtain a 90 degree corner as a function of etched depth. For example, if one wishes to ODE 30 microns deep, one must extend the corner out approximately 17 microns in order to obtain a 90 degree or right angle corner. It was previously stated that when



Figure 8: Alignment of mask on (100) plane parallel with the trace of the (111) planes.

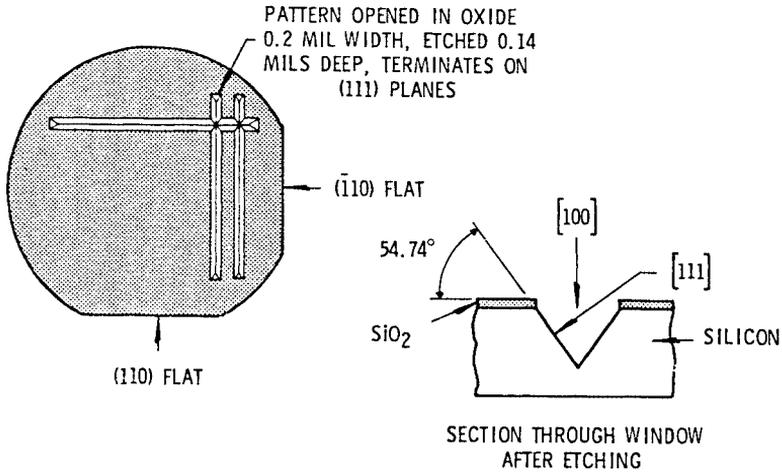
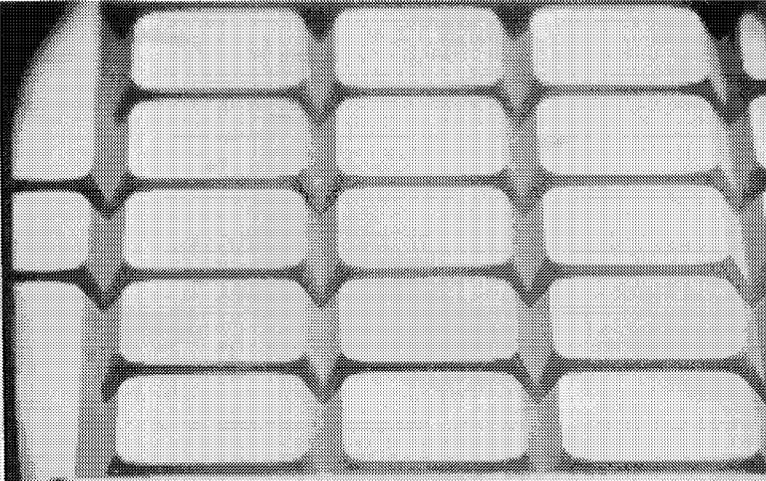
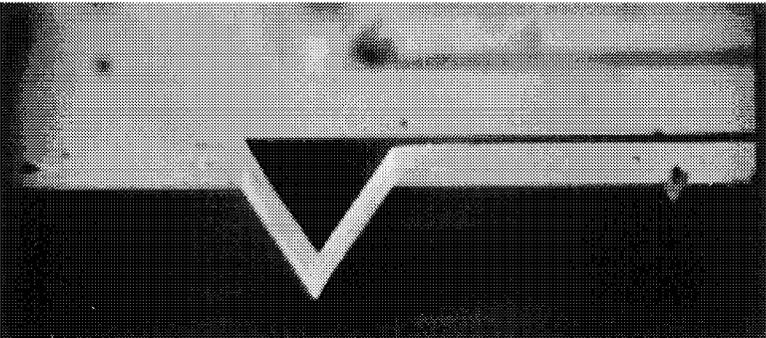


Figure 9: Alignment of mask with (110) flat on (100) plane.

the traces of the two (111) sidewall planes meet at the bottom of the etch moat the orientation dependent etch stops etching. To verify this statement, some (100) silicon slices were etched using a mask designed to etch 50 microns deep using the (100) ODE. This mask also has corner compensation designed to give square or right-angle corners at a depth of 50 microns. Figure 17a shows a top-focused view of one of these slices, which was etched for 38 minutes at 80°C. The etch depth is 38 microns. In the left-hand picture the focus is at the top of the (100) slice. In the right-side of Figure 17a the focus is at the bottom of the etch moat. Note that the bottom is flat at the etch front and is a (100) plane, since we have only etched 38 microns deep. Careful measurement of the etch width at the silicon surface/oxide mask interface shows that the etch moat is 97.5 microns wide with no measurable undercutting. Also, note that in Figure 17a, at the left top focus view there is a slight tip of silicon sticking out at the corners in all four quadrants of the etch moat. Remember the corner compensation is designed for 50 microns etch depth and we have etched only 38 microns deep. The slices were then placed back in the ODE solution and etched for an additional 30 minutes, for a total etch time of 68 minutes. We have now over-etched by 18 minutes for the corner compensation design. Careful measurement of the top of the etch moat again show that the etch moat width at the oxide is 97.5 microns, indicating that there is no undercutting or etching after the trace of the (111) planes reaches the edge of the oxide mask. The oxide corner compensation of the mask is clearly visible in both the top focus and the bottom focus views of Figure 17b. The bottom focus shows a completely v'd-out etch front with no (100) remaining. To further prove this statement, the slices were again placed in the ODE solution for an additional 30 minutes, making a total of 48 minutes over-etching for the



SEM TOP VIEW (100) DI ETCH



SEM CROSSSECTIONAL VIEW (100) DI ETCH

Figure 10: Top view and cross-sectional view of ODE etched (100) silicon.

(100) O.D.E. 50 μm

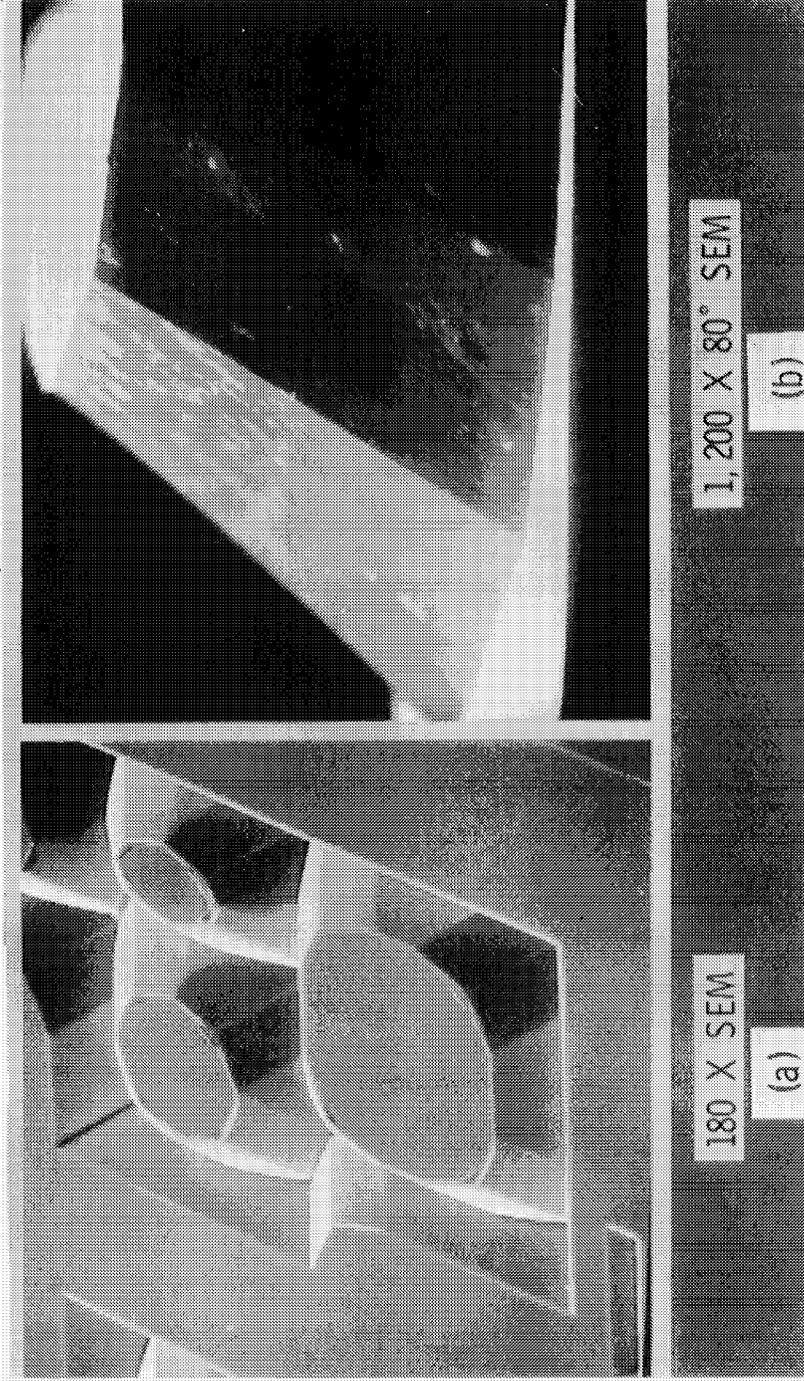


Figure 11: SEM photograph showing top view and 80° enlarged view of ODE etched (100) silicon.

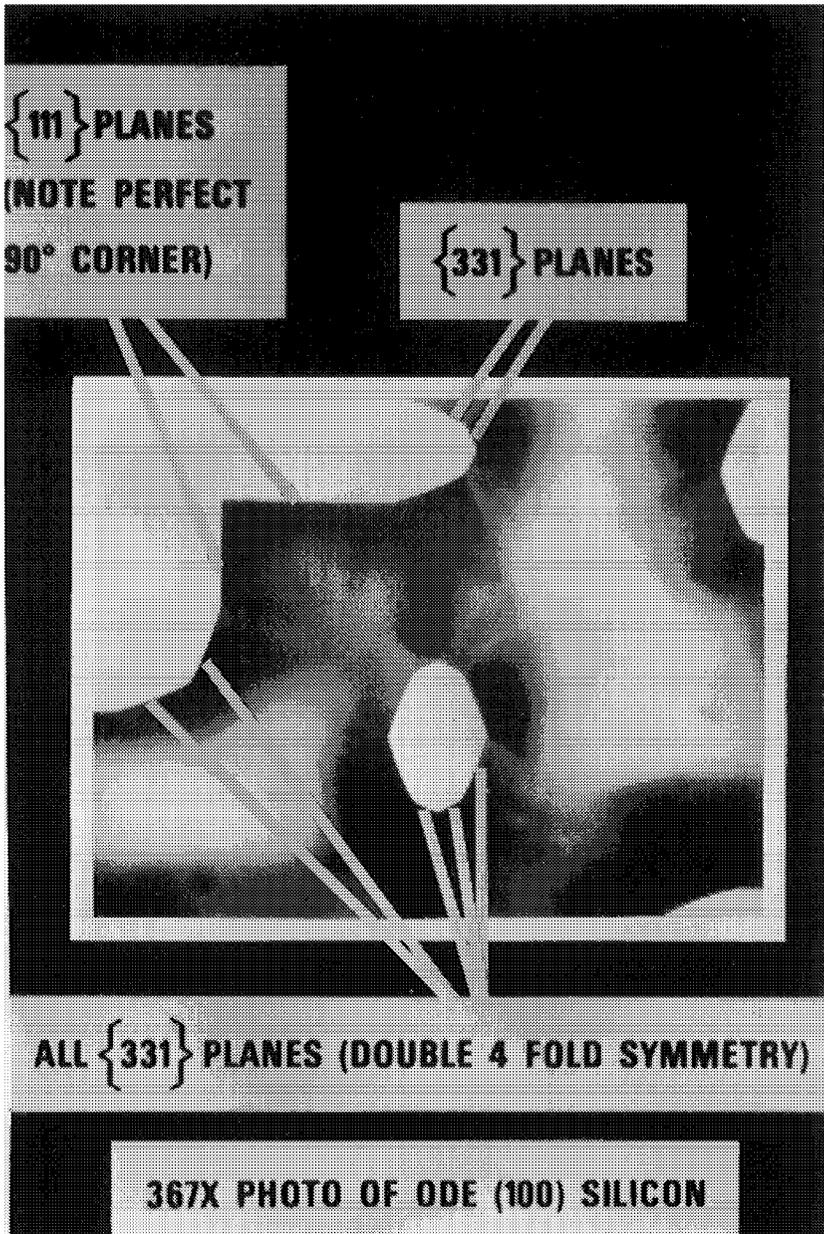


Figure 12: Photograph showing identification of (100) ODE etched planes.

CORNER COMPENSATION IN OXIDE MASK ON ODE ETCHED CIRCUIT

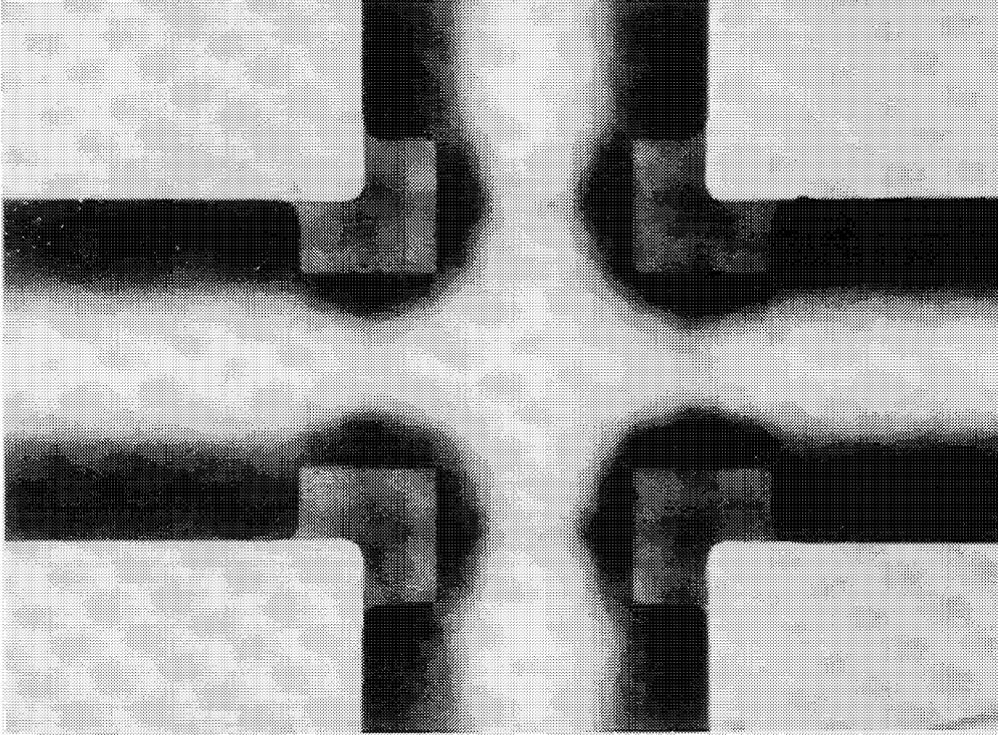
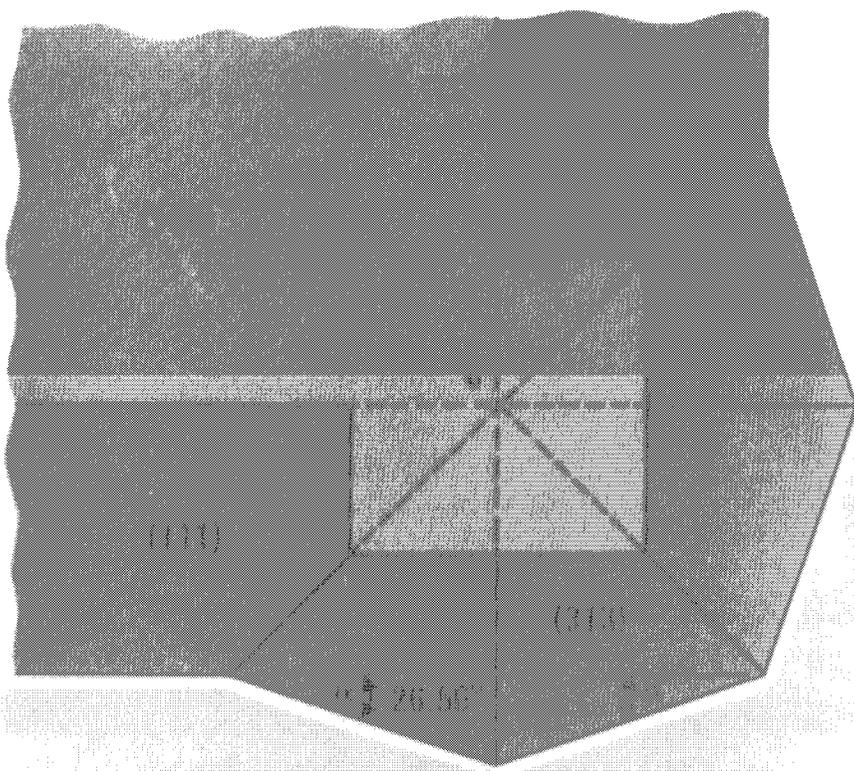


Figure 13: Corner compensated mask on (100) ODE etched silicon.

CORNER COMPENSATION



- α FOR KOH/PROPANOL/H₂O = 26.56° (311)
- α FOR ETHYLENE DIAMINE/H₂O = 18.43° (221)
- α FOR HYDRAZINE/H₂O/PROPANOL = 18.43° (112)

Figure 14: Corner faceting due to (100) ODE etching using different etches.

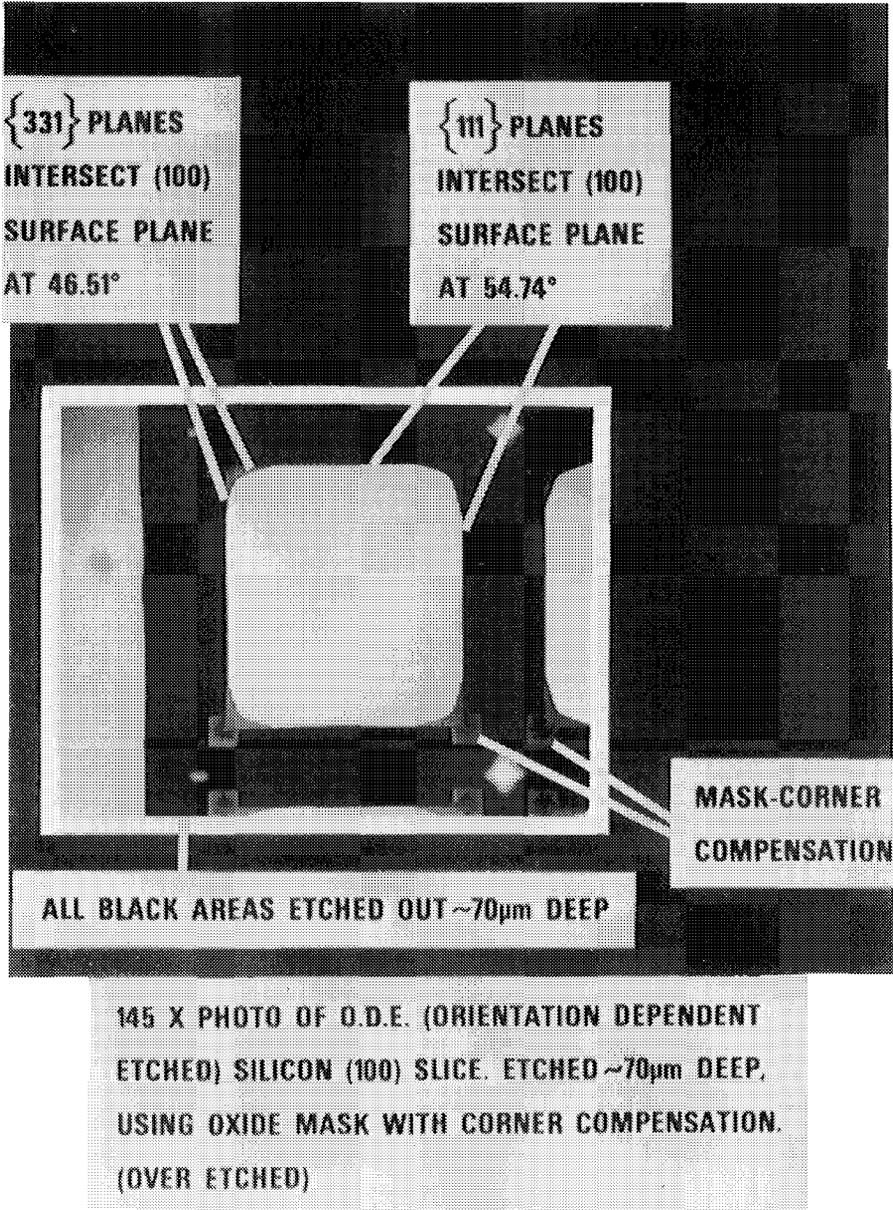


Figure 15: (100) ODE etched silicon slice.

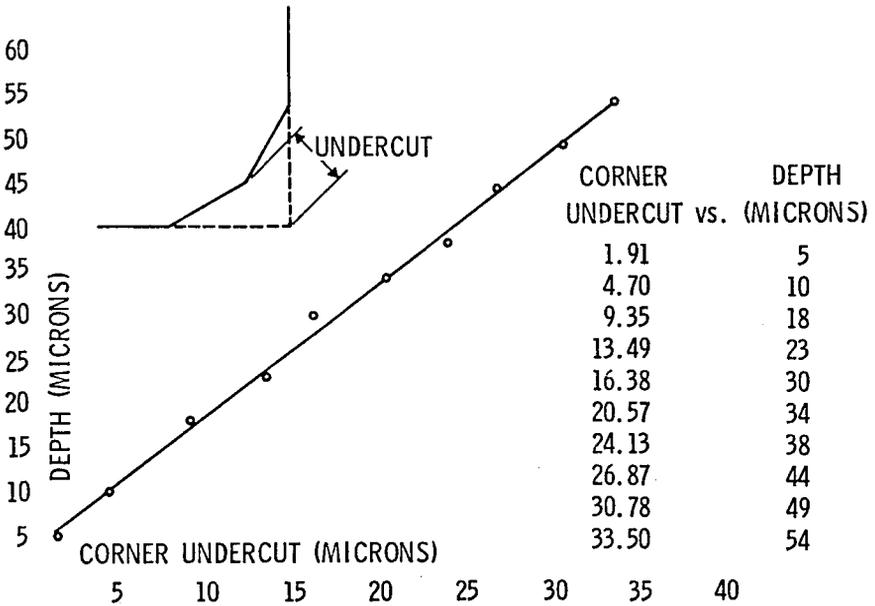


Figure 16: Graph of corner undercutting vs ODE etch depth for (100) silicon.

mask design. Again, the slice was measured at the top of the moat, at the oxide/silicon interface. The distance across the moat remained 97.5 microns, showing complete stopping of the ODE etching by the (111) plane, aligned with the mask opening. The corner compensation is clearly shown projecting out over the ODE etched moats of this greatly over etched slice both in the top focus and the bottom focus of Figure 18. The undercutting to the (331) planes is also shown.

In Figure 15 the angle of intersection of the (331) planes with the (100) surface plane was observed to be 46.51 degrees. This is considerably lower than the 54.74 degree angle of intersection of the (111) planes. We may wish to take advantage of this lower angle to more easily run metalization over these moats and/or mesas and to enhance photolithography definition. To do so, we merely align the mask parallel and perpendicular to the <310> direction, as indicated in Figure 7. The <310> direction is aligned by rotating the mask only 26.56 degrees to the right, or to the left, of the <110> direction. [This is due to the double four-fold symmetry of the <310> directions in the (100) plane]. Figure 19 shows two (100) silicon slices that were ODE etched in the same etch solution, at the same time, using the same uncompensated mask pattern. Slice (a) was aligned with the <310> direction. Note the lower etch termination plane angle is (46.51 degrees) for the (331) plane in slice (a) compared to a 54.74 degree etch termination plane angle for the (111) plane in slice (b) which was aligned with the <110> direction. Also note for slice (a) the 90 degree right angle

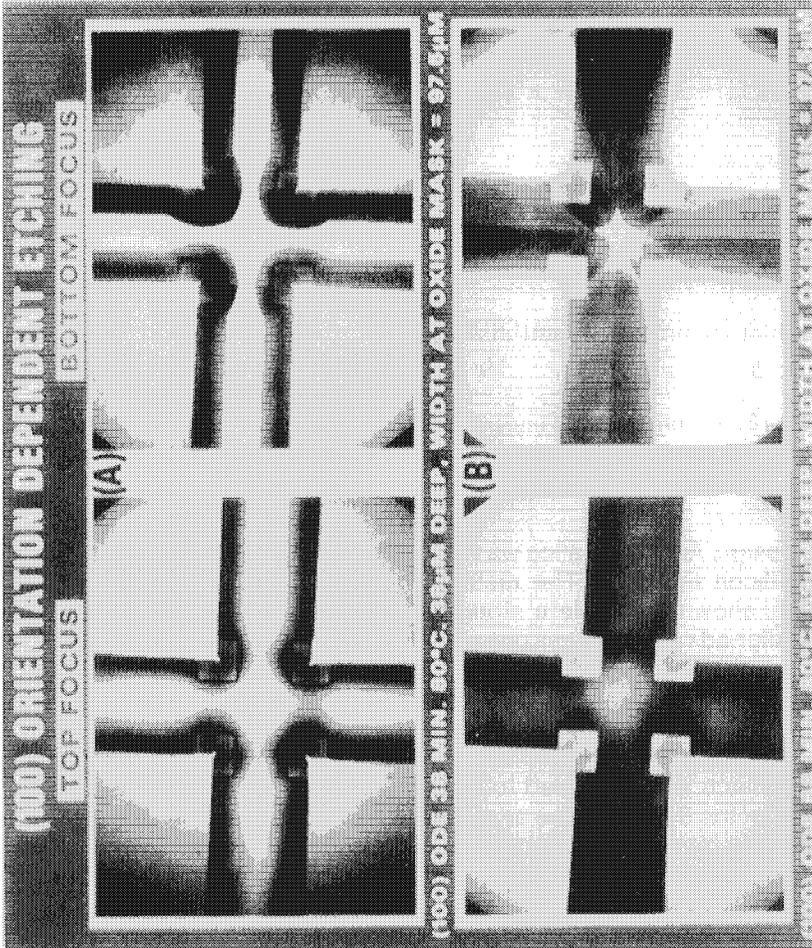
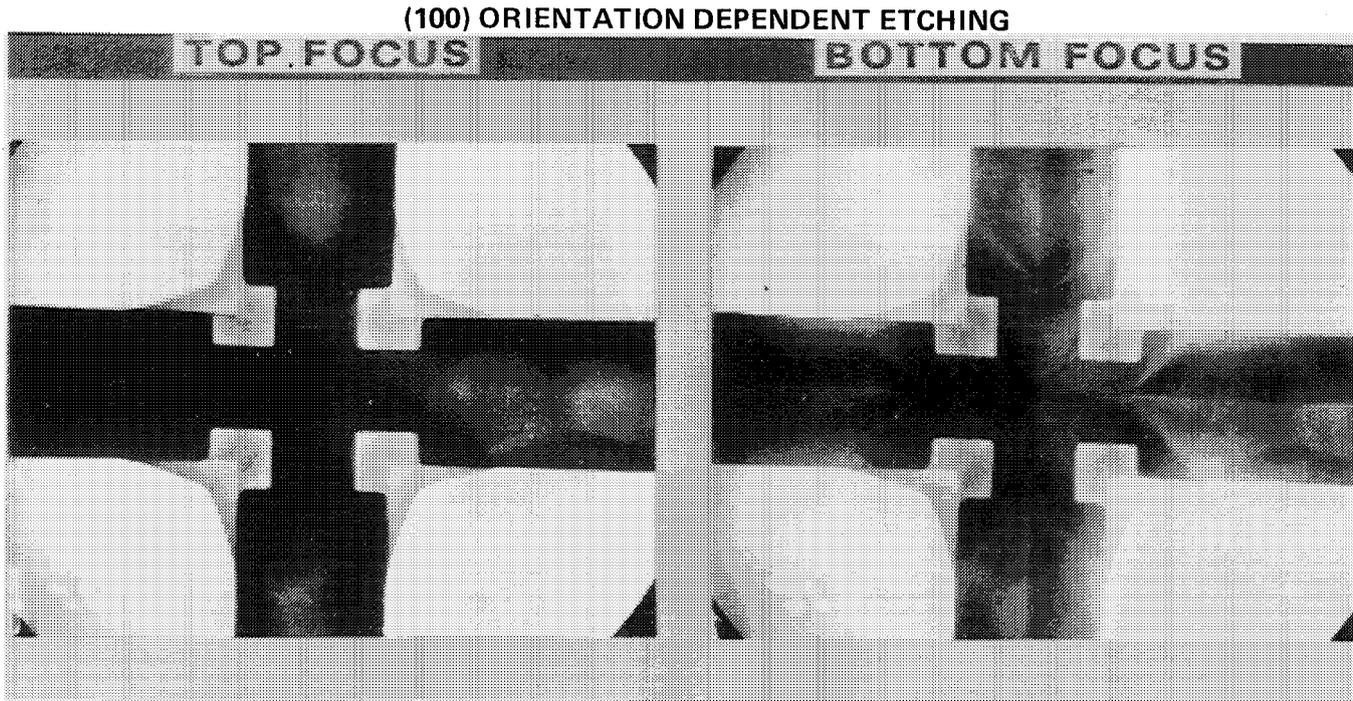


Figure 17: Photograph of (100) ODE etched surface and corner compensated mask.



(100) ODE. 98 MIN. 80°C. 69.9 μM DEEP, WIDTH AT OXIDE MASK = 97.5 μM

Figure 18: Photograph of (100) ODE etched surface and corner compensated mask.

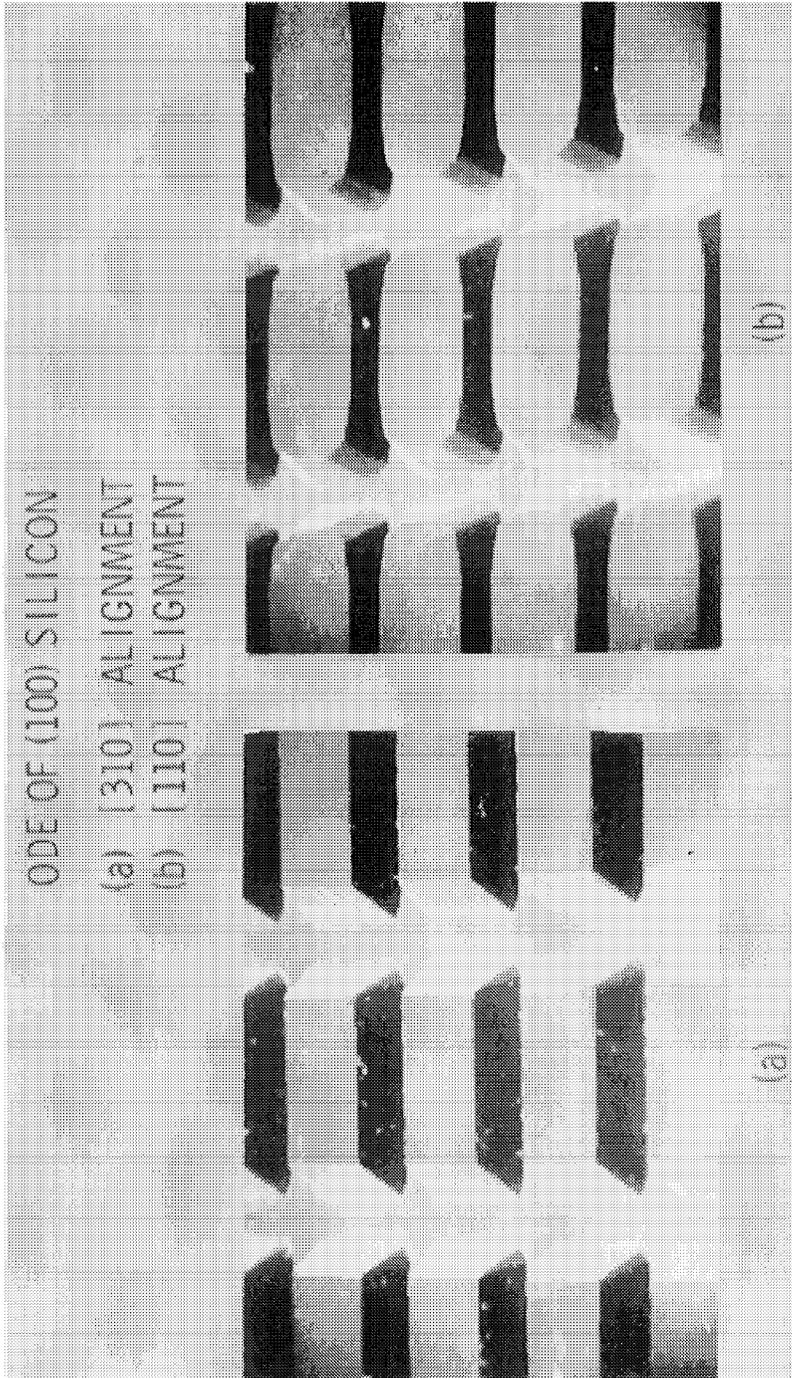


Figure 19: Photograph of (100) silicon, ODE etched simultaneously with $\langle 310 \rangle$ and $\langle 110 \rangle$ direction mask alignment.

corners, even without mask corner compensation. This is due to the moat sidewall etch termination being on the faster etching (331) planes which cause the corner undercutting in the slice aligned with the $\langle 110 \rangle$ direction. The width of the isolation moat is considerably wider, giving lower packing density in slice (a) due to undercutting of the mask all along the mesa edge. Corner undercutting, and fast etching, is observed to be due to the (331) planes at the corner of slice (b). Figure 20 shows an example of a metalization line over a silicon mesa which was ODE etched with $\langle 310 \rangle$ direction alignment. Note the low angle of the 46.51 degree (331) planes forming the mesa sides.

The ethylenediamine pyrocatechol and water etch, mentioned in Table 2, may be used as an oxide pin hole characterization tool in (100) MOS processing. If the oxide, especially thin gate oxides, has pin hole problems, these pin holes may be delineated by etching in the ethylenediamine etch solution. The delineation will result, and can be counted, due to the etch termination in the (100) silicon in the form of a perfect inverted pyramid, as in Figure 21. The etch pit will have perfect four-fold symmetry of (111) planes even though the pin hole may be irregularly shaped. The pin hole density of the oxide may be counted through the use of normal bright field, or Nomarski microscopy. Very small orifices, one micron or less, may be etched through a silicon slice using ODE and a mask designed to provide the desired orifice size. The etch depth to mask opening width is $0.707 \times W$, where W is the width of the mask opening. For example, a mask opening 10 microns wide aligned with the $\langle 110 \rangle$ direction on a (100) silicon slice will etch 7.07 microns deep and then stop etching. The reverse effect of generating very sharp points may be obtained by aligning a pattern of small squares slightly off the $\langle 110 \rangle$ direction to purposely cause undercutting. This effect is shown in Figure 22. In the photograph at the left side of Figure 22 the etch mask, has been completely undercut and the mask medium has been washed away. In the photograph at the right side of Figure 22 the etching was stopped before the mask was completely undercut, and the mask can be seen on top of the pyramid.

ORIENTATION DEPENDENT ETCHING AND ORIENTATION DEPENDENT DEPOSITION

In general, orientation dependent etching (ODE) reactions and orientation dependent deposition (ODD) reactions are complementary in silicon. Those planes that are slow etch planes, for example (111) planes, are slow depositing or growth planes. Those planes that are fast etching, for example (331) planes, are fast growth planes. The experiments depicted in Figures 23 and 24 show this effect.

Figure 23a shows an etch mask for the fabrication of an array of seven circuits or devices which are dielectrically isolated (DI) from each other. In the DI process the normal p-n junction isolation that completely surrounds each component is replaced by a dielectric material such as silicon dioxide or silicon nitride. This eliminates the very large parasitic capacitance around each component, giving rise to greatly improved switching speeds.

METALLIZATION OVER ODE MESA. [310] ALIGNMENT

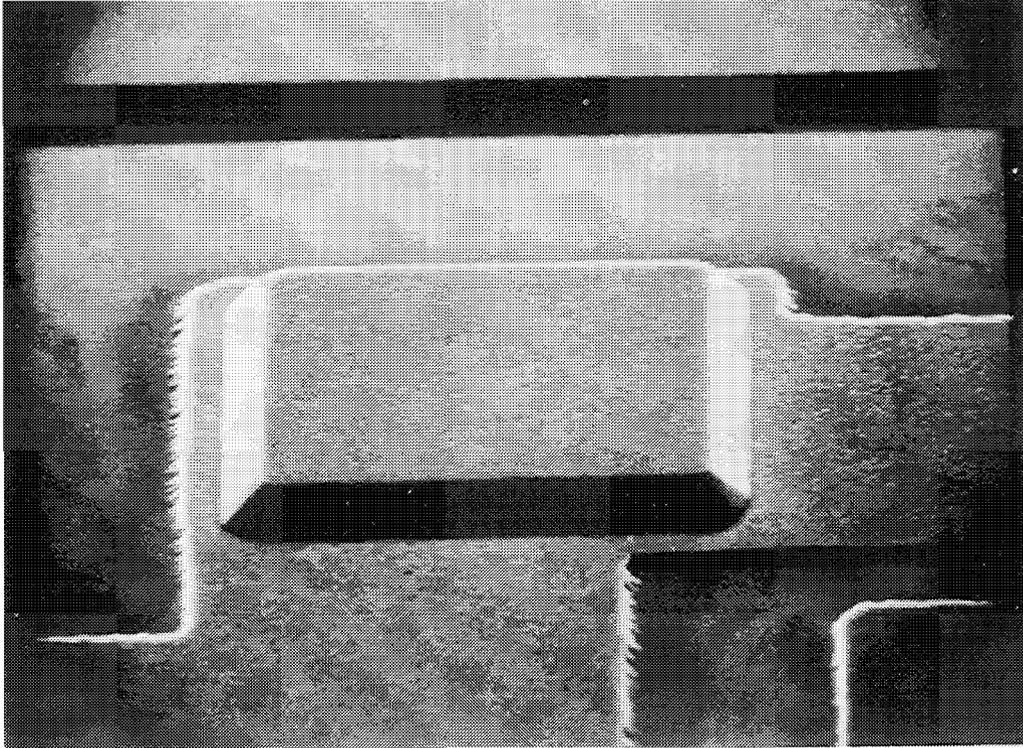
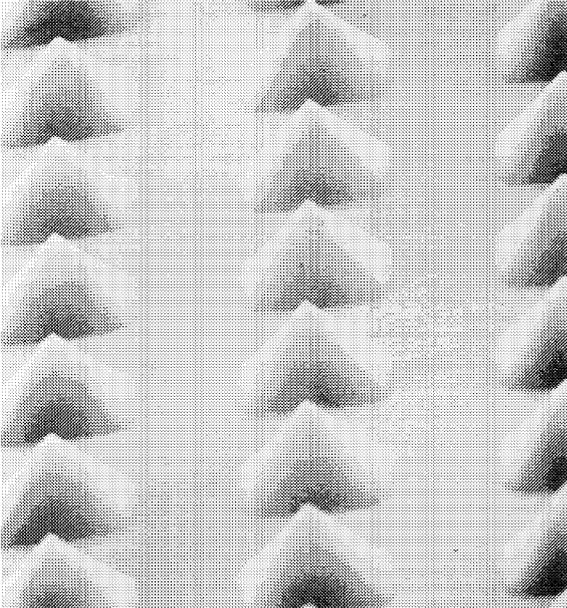


Figure 20: Photograph of metal lead contact to $\langle 310 \rangle$ aligned ODE etched silicon mesa.

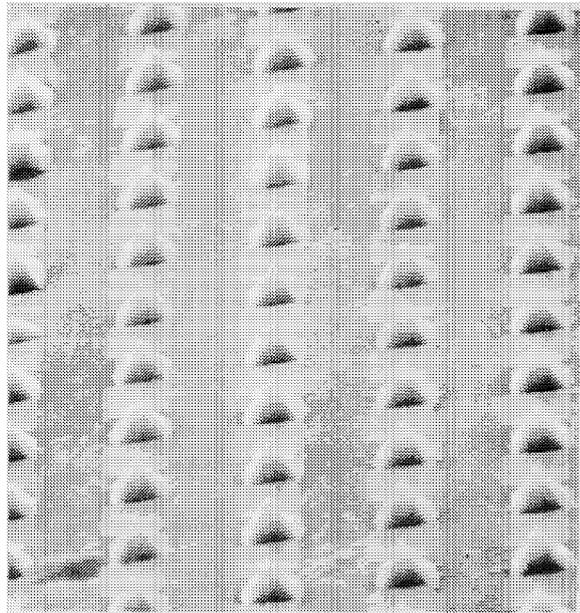


(100) ODE, SEM

Figure 21: Inverted pyramid, etch pit, in (100) silicon.



**(100) ODE, 2040 X 65° SEM
10 μm SQUARES ON 20 μm CENTERS**



**(100) ODE, 1240 X 65° SEM
NOT ETCHED TO COMPLETION**

Figure 22: Photograph of array of pyramids ODE etched in (100) silicon.

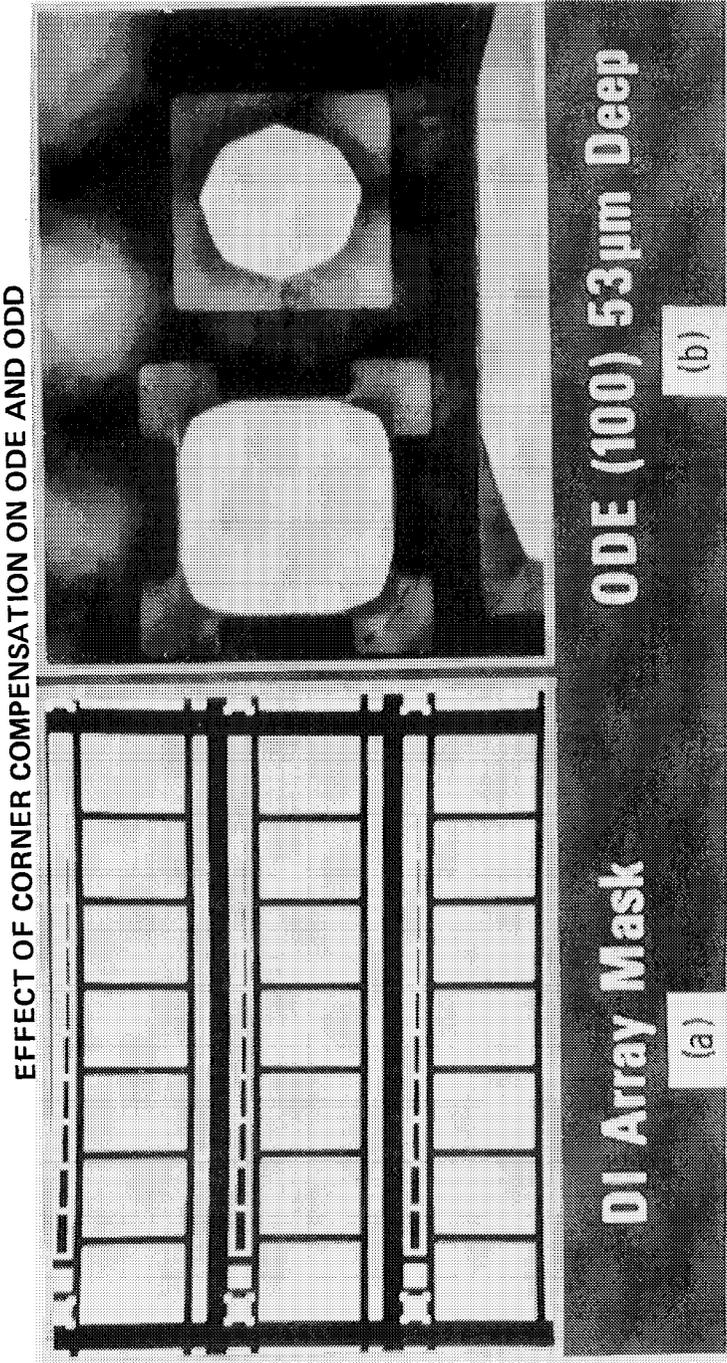


Figure 23: ODE etched, dielectric isolated, silicon array.

EFFECT OF CORNER COMPENSATION ON ODE AND ODD

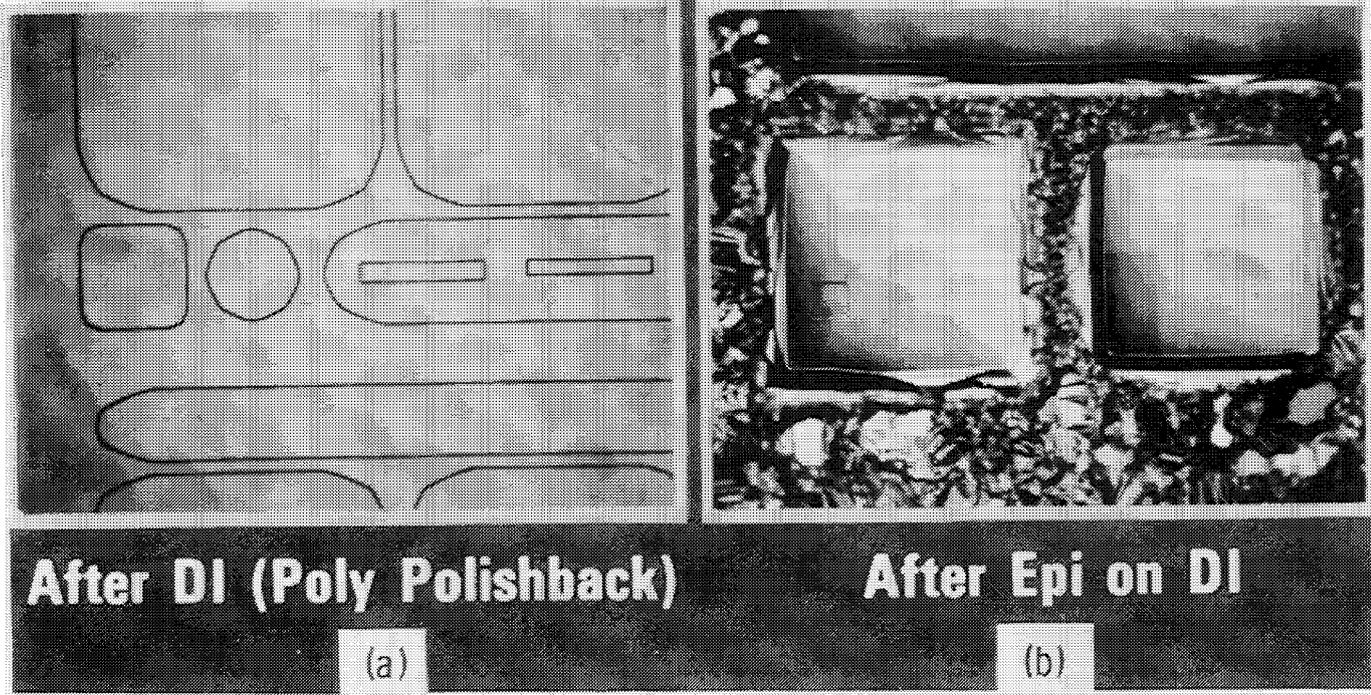


Figure 24: ODD deposition, silicon dielectric, isolated array.

It also provides much higher isolation voltage capability between components. The DI gives several advantages to the device and circuit characteristics, but has one major disadvantage, cost. However, it is used for certain applications. In the DI process ODE etching of (100) silicon is used to etch out the V-groove isolation moat. In this figure, the rows of seven large rectangles are the areas in which the DI circuitry or components will be fabricated. The row of successively narrower black rectangles are built-in thickness indicators for process control in the DI process. The two squares above and at the left of each array of seven DI islands are alignment markers. These two squares are the same size. However the one at the left has mask corner compensation. Figure 23b shows the silicon slice surface after ODE etching 53 microns deep. Note the nearly octagonal shape of the mesa on the right, which did not have mask corner compensation. It is completely bounded by the fast etching (331) planes. The mesa at the left, which has mask corner compensation, is bounded by the (111) planes except at the corners, where again the fast etching (331) planes are revealed due to over-etching. Note the large difference remaining in the surface area in the two mesas which originally were the same size. The next step in the DI process is to strip off the mask material and thermally grow silicon dioxide across the mesa structured surface. A thick polycrystalline silicon film (~ 0.5 mm) is then deposited over the oxide. The structure is then inverted and subjected to a grind and polish operation to remove the back portion of the original (100) substrate down to the tip of the V-groove isolation moats and the thickness indicators. The number of indicators showing (polished through) tells us the exact thickness of the single crystal silicon remaining in the dielectrically isolated tanks. Figure 24a is a photograph of this polished back surface. The black lines are the edges of the thermal oxide (DI) isolation. Two of the seven large isolated device tanks are visible at the top of the photo. At center left the two alignment markers are visible as a square and an octagon. At center right two of the thickness indicators are visible. Again, note the size and shape of the two single crystal alignment markers, in particular the one defined by the fast etching (331) planes, which appears as a nearly perfect octagon. Figure 24b shows these same two alignment markers after epitaxial deposition. The fast-etch (331) planes have also grown fast epitaxially, and the ODE etched octagon shape has reverted, or epitaxially grown back to a perfect square, bounded by (111) planes. This is proof that the fast etching planes are also fast deposition planes. Also, a perfect (100) epitaxial stacking fault appears at the lower left portion of the corner-compensated alignment marker.

Epitaxial stacking faults are generally undesirable, and can be completely avoided by proper substrate preparation and in situ HC1 vapor etching; however, in some instances they can be useful as process control indicators. They are readily observed, if present, by use of a Wright-Jenkins etch (see Table 4) and/or Nomarski interference contrast microscopy. Figure 25 shows a photograph of (111) and (100) epitaxial stacking faults (ESF). These ESF's can be used to readily verify the crystal orientation of the silicon substrate. Stacking faults in epitaxial films on (111) substrates will produce line faults, or equilateral triangle-shaped stacking faults with

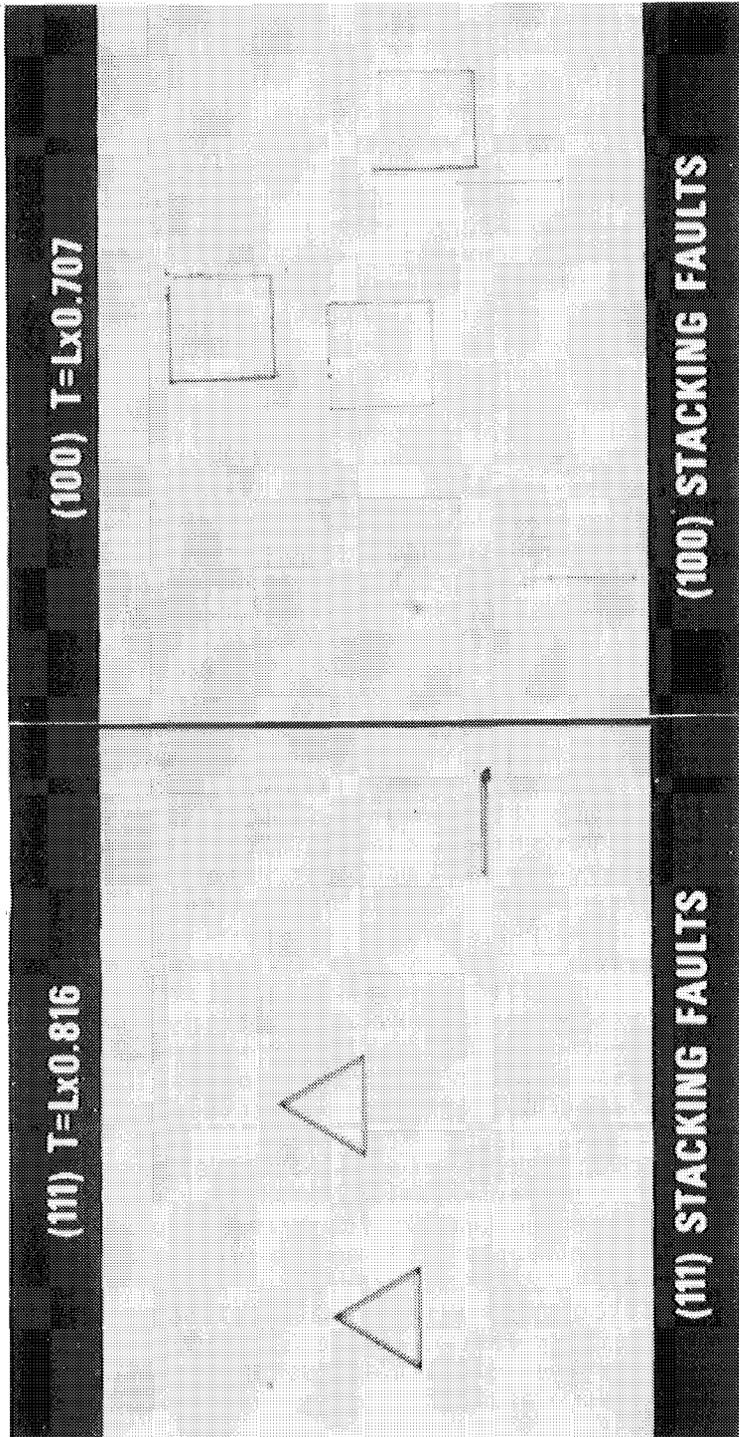


Figure 25: Photographs showing (111) and (100) silicon epitaxial stacking faults.

the lines or sides of the fault aligned with the trace of the (111) planes which intersect the (111) surface, and in $\langle 110 \rangle$ directions. The (100) epitaxial stacking faults will be line faults or squares, with the lines or sides of the square aligned with the trace of the (111) planes, in $\langle 110 \rangle$ directions intersecting the (100) surface. If the (111) triangles are not perfectly equilateral or if the (100) squares are not perfectly square, the substrate is off orientation. The ESF's usually nucleate at the substrate/epitaxial layer interface. Due to the growth control of the faults by (111) planes, and by knowing the angle of intersect of the (111) plates with the substrate plane, one can very accurately determine the thickness of the epitaxial layer. This is accomplished by merely measuring the length of the side of the ESF and multiplying it by 0.707 for (100) silicon, 0.816 for (111) silicon, and 0.5 for (110) silicon. Figure 25 shows examples of (111) and (100) epitaxial stacking faults. Further discussion of stacking faults in silicon, both epitaxial and bulk SF's, will be found in the section on defect delineation.

(110) ORIENTATION DEPENDENT EFFECTS

We now turn our attention to the orientation dependent effects of (110) silicon. From Table 2 we see that the (110) orientation dependent etch is made up of KOH and water, 50/50 by weight, and is carried out at 80°C. The etch rate in the $\langle 110 \rangle$ direction is approximately 700 times faster than it is in the $\langle 111 \rangle$ direction.

The (110) stereographic projection in Figure 26 shows that the predominant (111) orientation planes intersect the (110) surface plane of silicon at 90 degrees or at the periphery of the projection. The (111) 90 degree planes are not 90 degrees to each other as in the case of the (100) projection. The symmetry in the (110) projection is not fourfold, but mirror image symmetry. The (111) planes which intersect the (110) surface are 109 and 71 degrees to each other, thus forming a rhombus rather than a square or a rectangle. Two low angle (111) planes are observed which intersect the (110) surface plane at approximately 35 degrees. These two planes may be used for the mask alignment of thickness indicators, as was described for (100) silicon. Their effect may also be noted where we have mask alignment with the 90 degree (111) planes. These low angle (111) planes constantly tend to stop the etching, but because the mask is not aligned with the trace of these low angle (111) planes, they are not effective.

Figure 27 shows the method for aligning a mask with the trace of the vertical (111) planes at the (110) surface, thus giving the shape of a rhombic pattern. If we open lines parallel with these traces of the (111) planes through a mask, such as silicon nitride or oxide, and then etch using the (110) etch as mentioned in Table 2, we will etch moats with vertical walls which may extend all the way through the slice with practically no measurable undercutting.

Figure 28 shows SEM photographs of (110) ODE etched slices which have been etched 80 microns deep. The oxide mask pattern was 10 microns of open silicon with 10 micron spacings between openings. The

**STANDARD (110) PROJECTION FOR SILICON
(FACE-CENTERED CUBIC) CRYSTAL SHOWING RHOMBIC
PATTERN ALIGNMENT WITH THE 90° {111} TRACES**

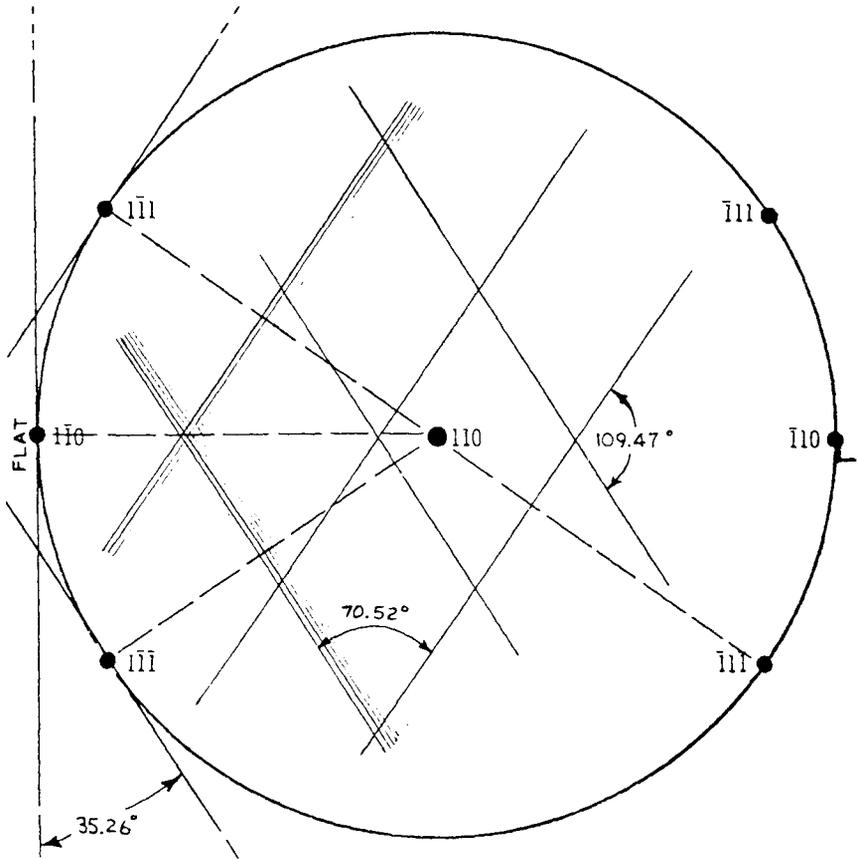


Figure 27: Rhombic shape of (111) planes for ODE alignment.

surface area can be obtained by this type of ODE etching. If a line one micron wide is opened and etched 100 microns deep, we have increased the available silicon surface for device fabrication by 200X. This vertical surface can be used advantageously for fabrication of passive semiconductor components such as capacitors, resistors, and isolation. It has also been used in the fabrication of large area, high efficiency solar cells. Methods of fabricating active components in these vertical surfaces are being explored making use of beam technologies.

Figure 29 is a 5,000X SEM photomicrograph, taken at an angle which shows the top surface as well as the cleaved edge of a (110) ODE etched

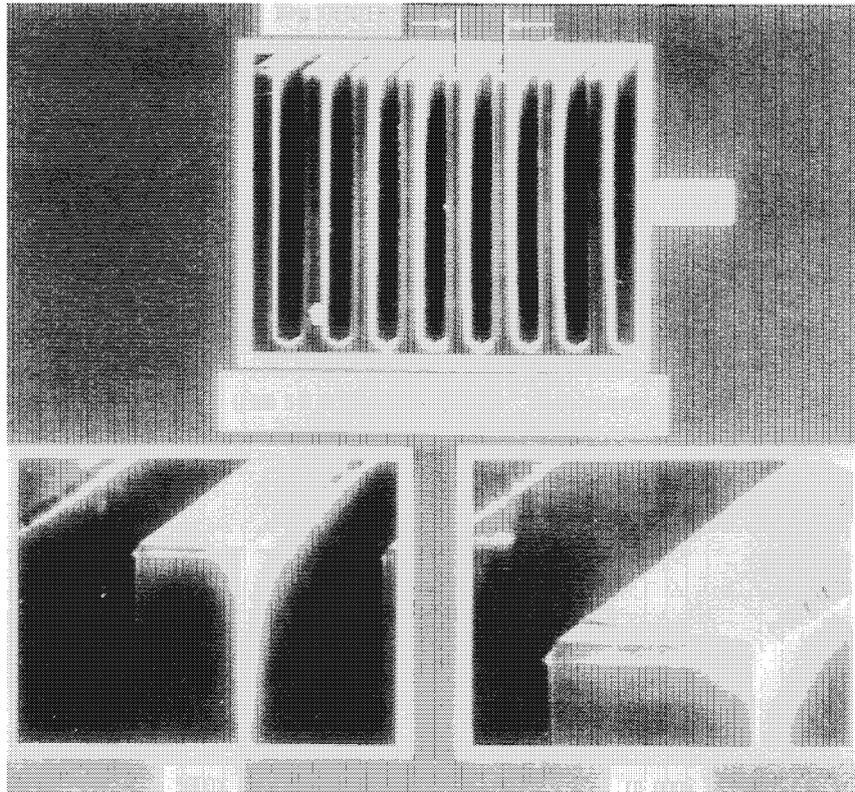


Figure 28: Photograph of cross-section of (110) ODE etched silicon. (Magnification less than indicated.)

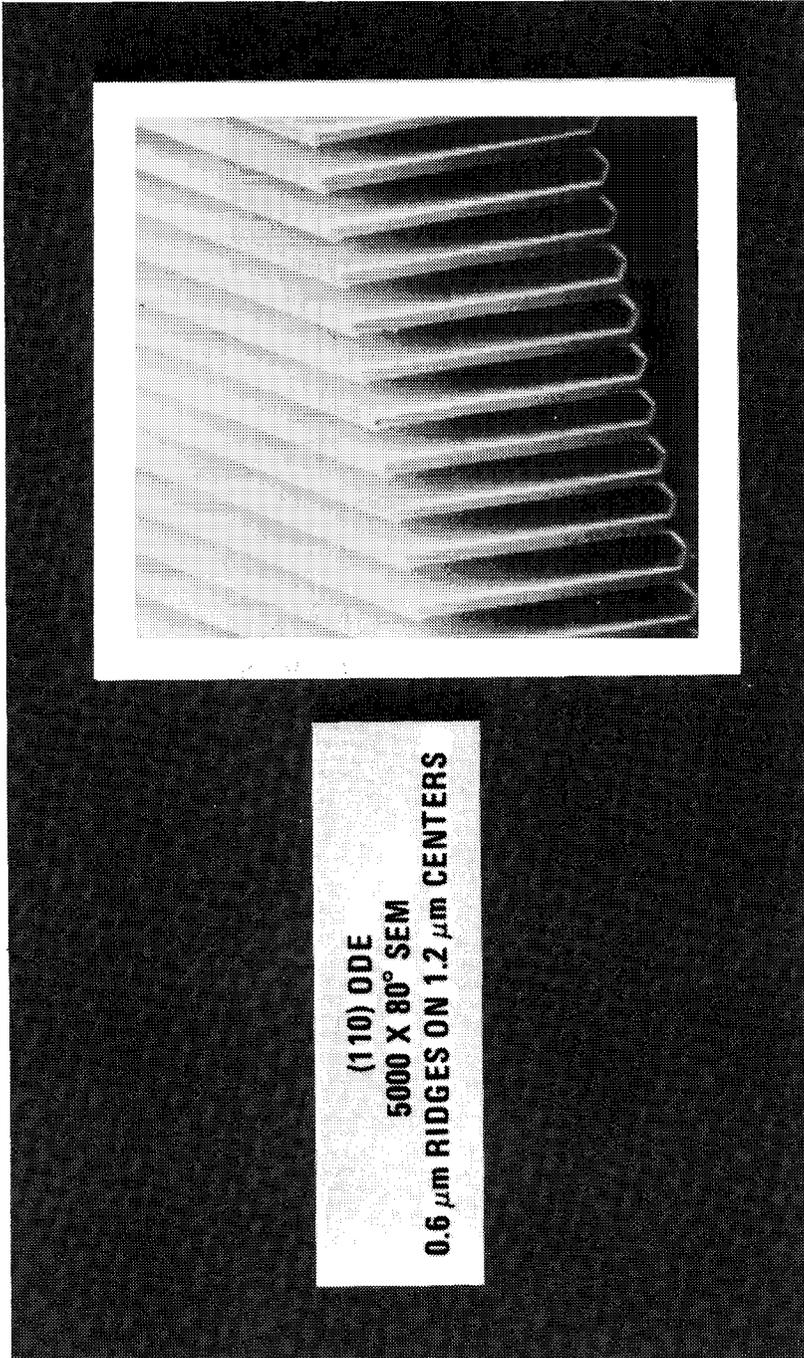


Figure 29: SEM photograph of (110) ODE etched silicon.

structure using submicron lithography. Each ridge and space in this case should be 0.6 micron with 1.2 micron centers. However, if the ridges are thinner than the spaces between the ridges, this indicates a misalignment of the mask with a trace of the (111) at the (110) surface. Again, we see the V-notches at the bottom of the etch moats due to the low angle (111) planes. The sidewalls of structures such as this may be readily doped to form junctions by standard diffusion techniques. These sidewalls are also readily oxidized by standard thermal oxidations. Therefore, by alternate etching and diffusion we can fabricate active p-n junction structures in these walls.

Figure 30 is a 570X SEM photomicrograph showing the top and the edge view of a cleaved ODE structure similar to the one shown in Figure 29. However, in this case the tops of ridges have been pointed by lightly etching in a 1-3-8 etch. These points are so exact that the surface area becomes a blackbody. Such a structure may be used as a high efficiency solar cell, which has the advantages of a large silicon surface area plus a blackbody surface area that traps all the available solar energy into the cell.

Figure 31 is a 11,500X SEM photomicrograph showing one of these ridges. Even at this magnification no radius of curvature can be detected. It should be noted that the sidewalls of these etched arrays are the (111) planes which may be tilted to any desired angle merely by cutting the slices off orientation from the (110) in the proper direction.

Figure 32 shows an array which has been etched into a silicon slice that was purposely cut 10 degrees off the (110) orientation. This effect, combined with the blackbody etching, may be used to produce optical collimation for an LED-type display with zero back-reflection to the observer. If a mask consisting of open lines in the oxide is aligned with the trace of the (111), and the slice is then turned over and the same type mask aligned with the other set of (111) traces, we can simultaneously etch these two patterns to obtain an X-Y array such as that shown in Figure 33. This figure shows an SEM photomicrograph, taken at 60 degrees, of a slice which was simultaneously etched from the top and bottom which has a section cleaved or broken out of it. In this case we have five micron openings on 20 micron centers. Where the etch from the front and the etch from the back meet, a sieve with 5 micron openings is formed. We can also electrically address these ridges in X and Y directions. These ridges are like silicon crystal whiskers, extremely strong and very flexible. Figures 34 and 35 are top view SEM photomicrographs of wafers etched simultaneously from the back and front.

Figure 36 is a close view using the SEM at 1,600X and showing the simultaneously etched top and bottom ridges. The exactness of the (111) plane is clearly evident. Also, the low angle 35 degree (111) effect is again evident in both top and bottom ridges. Figure 37 shows a low power, 45 degree angle SEM photomicrograph taken at the edge of a broken slice which had previously been ODE etched.

The previous photographs showed the effect of orientation dependent etching straight down 90 degrees to the surface. Figure 38 shows the opposite effect, orientation dependent deposition by use of CVD with

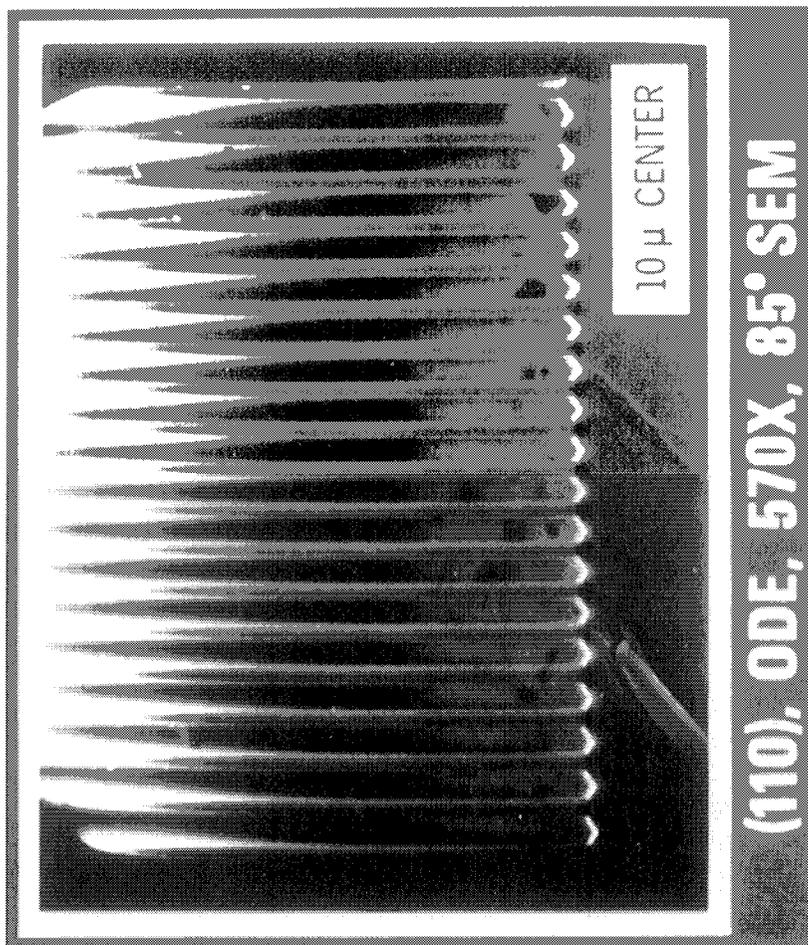


Figure 30: SEM photograph of (110) pointed ridges, ODE and 1-3-8 etched. (Magnification less than indicated.)

(110) ODE, 11.5KX, 85° SEM



Figure 31: High magnification SEM of (110) ridge point.

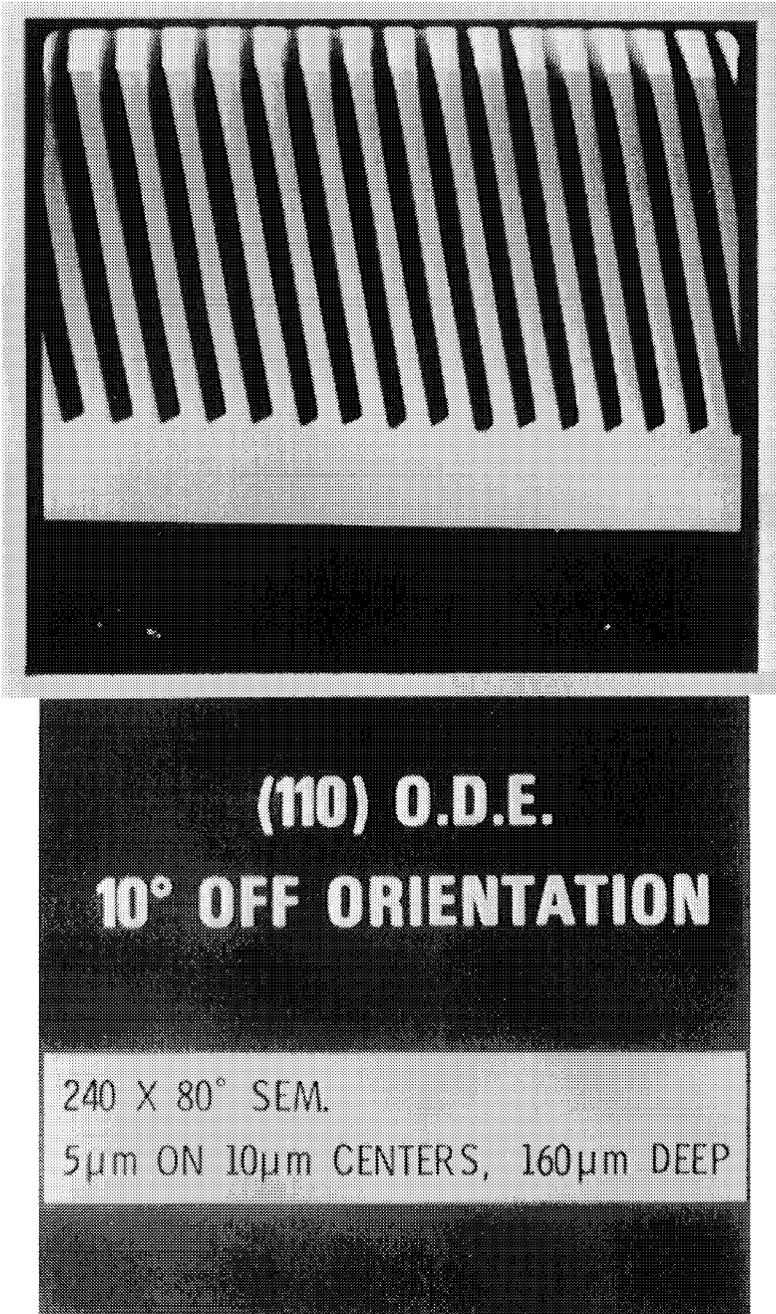
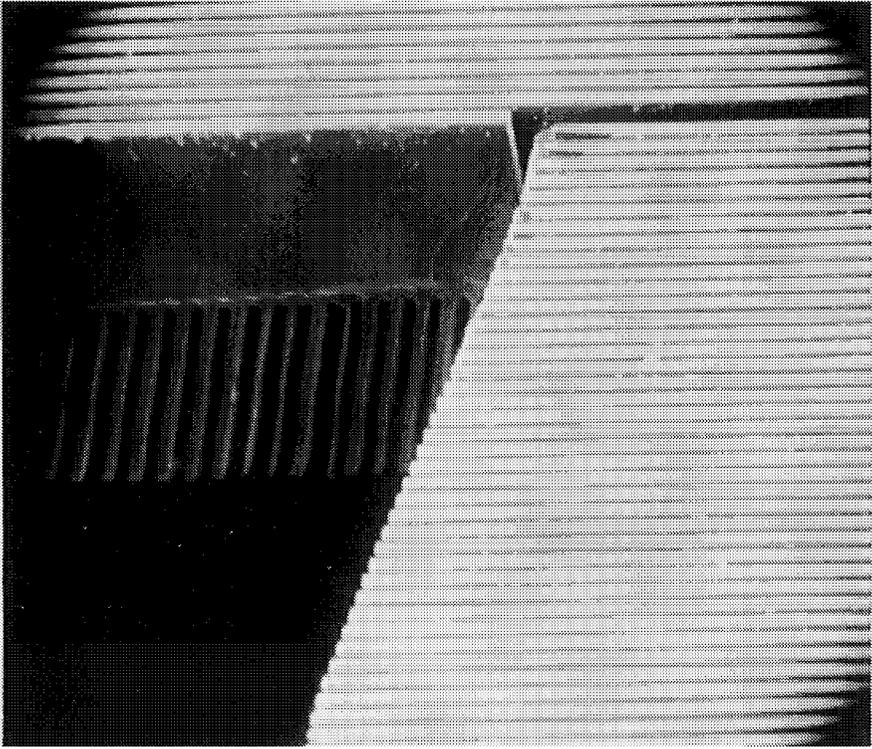


Figure 32: SEM photograph of cross-section, (110) tilted 10 (at saw) ODE etched.



(110) O.D.E. GRID

160 X 60° SEM
5 μ m ON 20 μ m CENTERS

Figure 33: SEM photograph of sectioned (110) ODE silicon, simultaneously etched from top and bottom.

ODE APERTURE GRID 500X MAG

- APERTURES - 5 μ m ON 20 μ m CENTERS
- VIEWING ANGLE - 0 DEGREES
- SLICE THICKNESS - 10.6 MILS
- PATTERN SIZE - 0.8" SQ

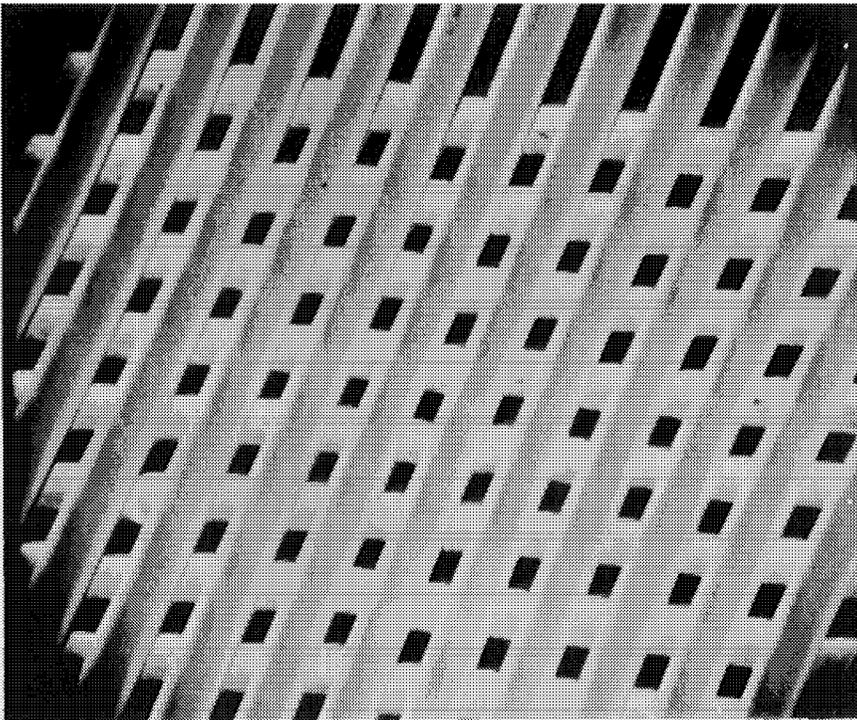


Figure 34: Top view of (110) ODE, top side and bottom side, simultaneous etched X-Y grid.

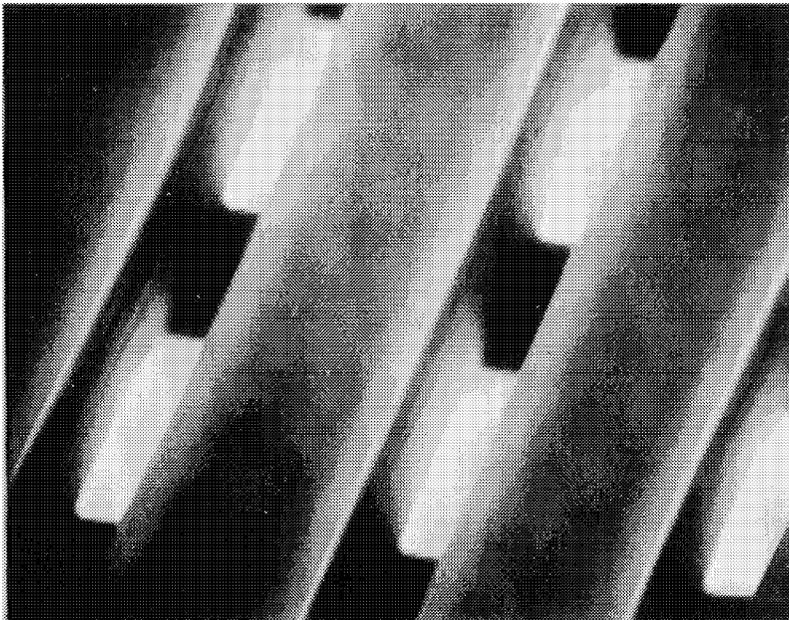
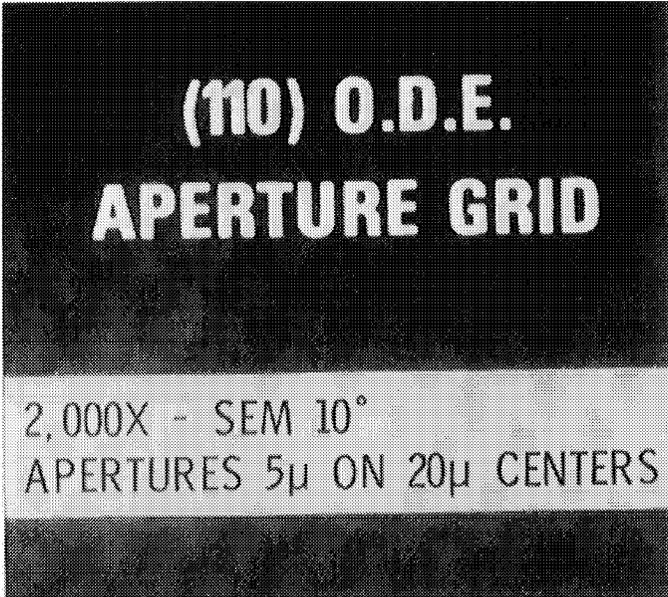


Figure 35: Magnified view of ODE etched (110) silicon as in Figure 34.

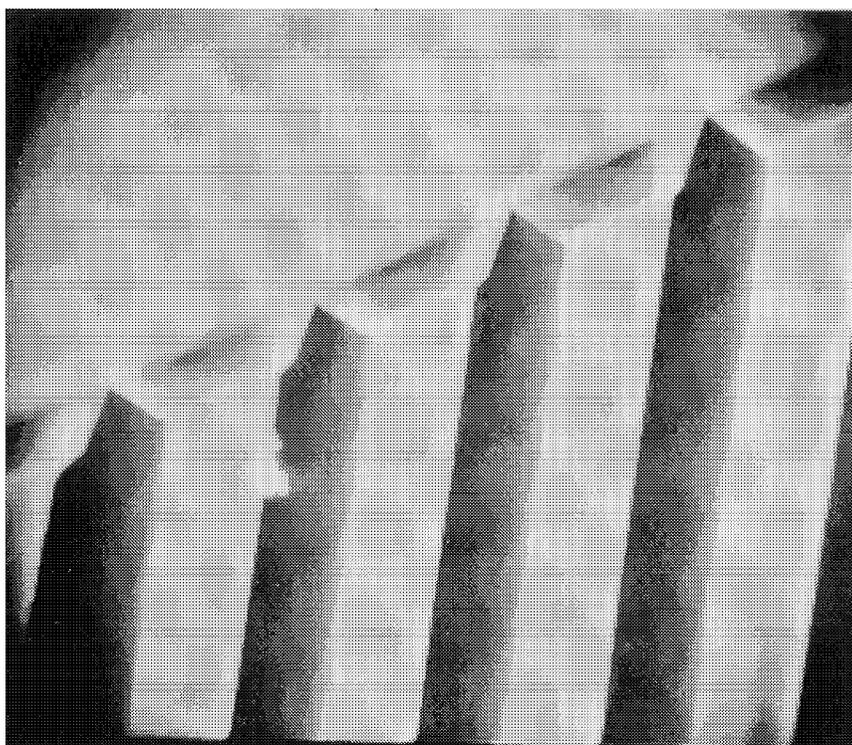
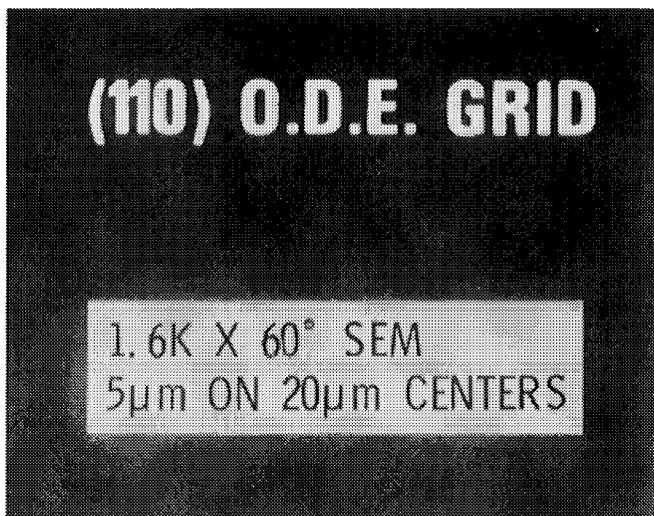


Figure 36: Magnified close up of (111) planes in X and Y directions and V etched effect of low angle (111) planes.

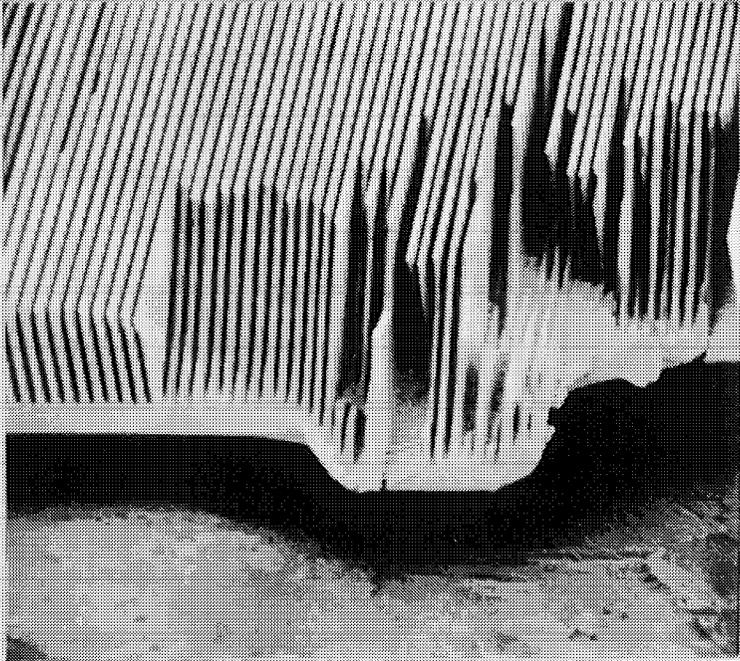
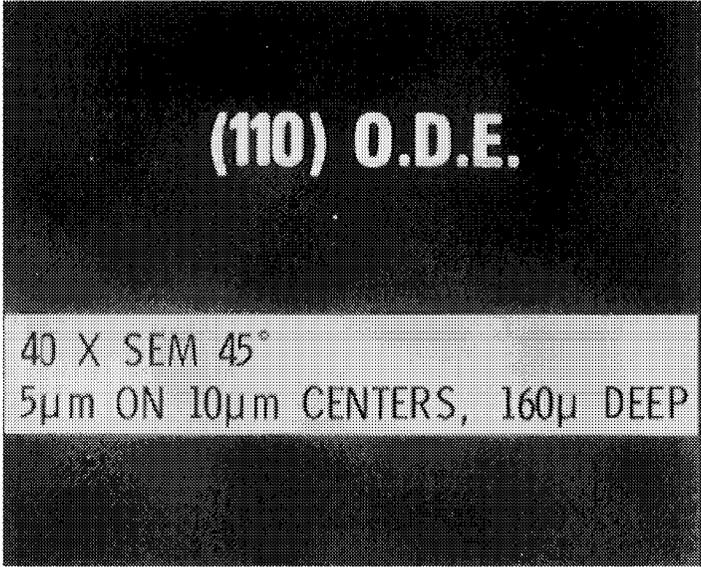


Figure 37: Cleaved or broken section of (110) ODE silicon.

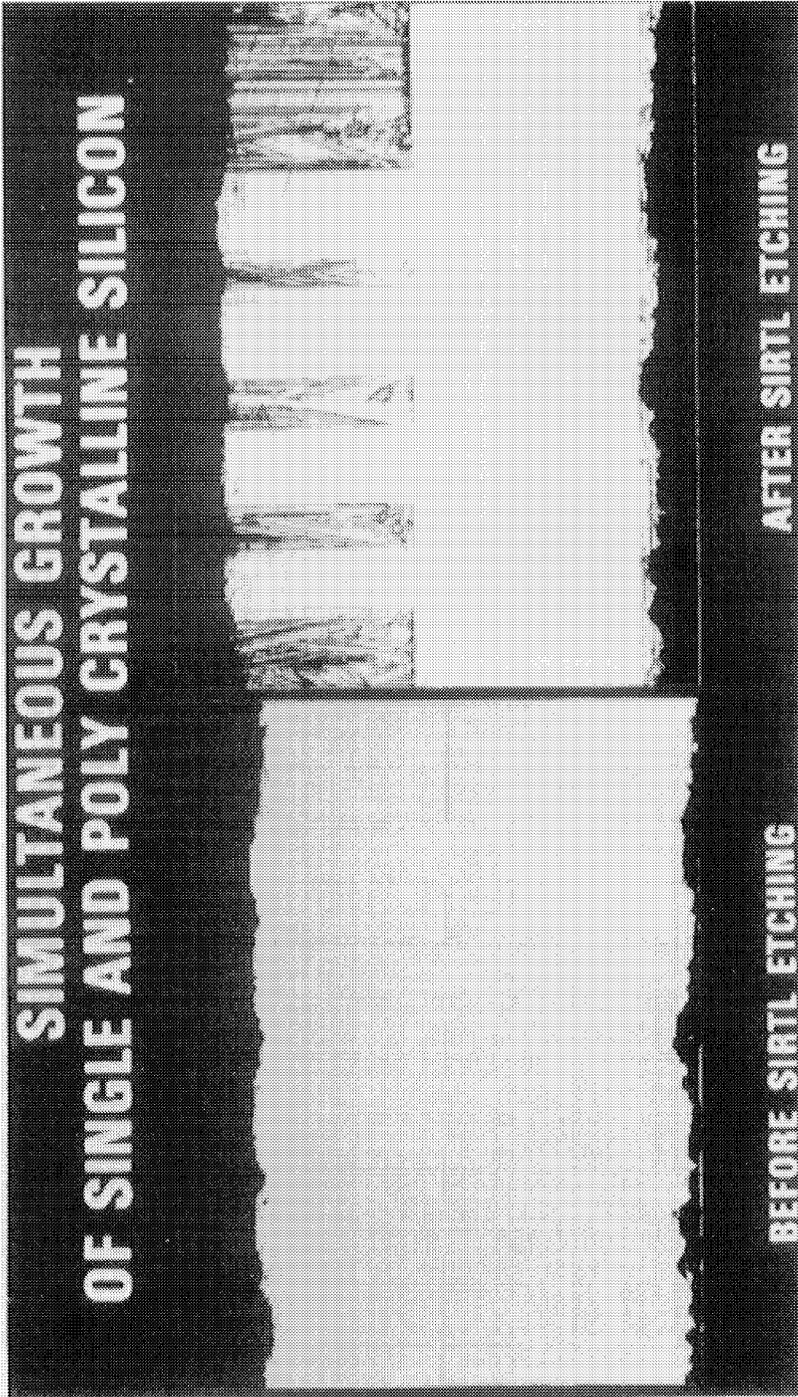


Figure 38: Photograph of cross-section of simultaneously deposited polycrystal and single crystal silicon on (110) substrate.

growth straight up in single crystal columns. In this experiment a (110) silicon slice was thermally oxidized, a mask was aligned with the trace of the (111) plane, and the oxide was opened in some areas and allowed to remain in other areas. The photograph at the left shows a cross section of the simultaneously deposited single crystal and polycrystalline silicon. The photograph at the right shows this same area after the polished cross section has been Sirtl-etched to reveal the polycrystalline and single crystal structure. It can be seen that in the areas where there was no oxide on the original substrate, the single crystal material grows perpendicular to the (110) surface. In areas where an oxide remained, we see polycrystalline silicon, highly ordered, in a (110) direction. These figures show both the effect of orientation dependent deposition with vertical growth in a (110) direction and the effect of Sirtl etching to bring out grain boundaries in polycrystalline material. Table 3 shows some of the uses of anisotropic etching in (100) and (110) silicon processing.

Table 3: Applications for (100) and (110) ODE in Silicon Device or Structures

(100)	(110)
Radiation hardened circuits	High voltage diode arrays
Electronic printer	Vertical multiple jct solar cell
Crosspoints	Wave guides
Iso planar	Sensistor
Poly planar	IR detectors
V MOS	Metallization templates
J FET arrays	High value capacitors
DI process thickness indicator	Optical collimators
Solid state pressure transducer	Black bodies
Solar cell anti-reflecting surface	

DEFECT DELINEATION ETCHING

In today's silicon VLSI/ULSI technology, defect density and, of course, defect delineation are of great importance. Wet chemical etching is used to identify and study these defects. In general, the etches used are those that etch more rapidly on strain field defect areas than on normal single crystal areas. In some cases decorating agents are also added to more clearly reveal the defect. Table 4 lists the defect delineation etches commonly used in industry today, for both (111) and (100) silicon.

Figure 39 shows a photograph of a large diameter, double twinned slice. The twin planes are revealed by a light (30 second) Secco etch. Quite often in today's processing of silicon we observe defect areas which appear to be concentric circles in the silicon slice surface. These circles may be further delineated by Secco etching. (See Figure 40.) They are normally caused by dopant segregation or oxygen segregation during crystal growth. Processing for VLSI circuitry places great emphasis on

Table 4: Preferential Etches for Defect Delineation in Silicon

ETCH SOLUTION	COMPOSITION	CHARACTERISTICS	REF
Dash	1 HF : 3 HNO ₃ : 10 HAC*	Delineates defects in (111) silicon. Requires long etch times, concentration-dependent.	1
Sirtl	1 HF : 1 (5M-CrO ₃)	Delineates defects in (111). Needs agitation. Does not reveal etch pits in (100) very well.	2
Seeco	2 HF : 1 (0.15M K ₂ Cr ₂ O ₇)	Delineates O.S.F. in (100) silicon very well. Agitation reduces etch times.	3
Wright-Jenkins	60 ml HF: 30 ml HNO ₃ : 30 ml (5M CrO ₃): 2 grams Cu (NO ₃) ₂ : 60 ml HAC • : 60 ml H ₂ O	Delineates defects in (100) & (111) silicon. Requires agitation.	4
Schimmel	2 HF : 1 (1M CrO ₃)	Delineates defects in (100) silicon without agitation. Works well on resistivities 0.6 - 15.0 ohm cm n & p types.	5
Modified Schimmel	2 HF : 1 (M CrO ₃): 1.5 H ₂ O	Works well on heavily doped (100) silicon.	5
Yang	1 HF : 1 (1.5M CrO ₃) • acetic acid	Delineates defects on (111), (100), and (110) silicon without agitation. Note: agitation - ultrasonic	6

silicon substrate purity and crystallographic perfection at the device/substrate interface. To improve this surface for subsequent device processing, oxygen out-diffusion techniques are commonly used to denude the surface areas and form oxygen precipitates to getter impurities and defects.

Figure 41 shows a cleaved (see Figures 5 and 6 as well as discussion) cross-sectional view of a silicon slice which has been subjected to an oxygen precipitation cycle. (See section on epitaxy in chapter on chemical vapor deposition of silicon and its compounds.) This slice has an epitaxial layer at the top and a backside polysilicon gettering layer on the back surface. Denuded zones can be observed in the original substrate just below the epitaxial film and just above the backside gettering polysilicon film. The central portion of the original substrate shows the large number of crystal defects remaining after crystal growth and thermal cycling. Most of these defects are in the form of bulk stacking faults and oxygen precipitates. These types of defects are readily revealed by Wright-Jenkins or Yang etch in a cross-sectioned or cleaved sample such as the example in Figure 41. It should be noted that Figure 41 is from a cleaved cross section requiring no polishing or other preparation prior to the defect delineation etching.

Figure 42 is a photograph of a grooved (100) silicon substrate showing the fairly high density of bulk stacking faults and dislocations at the bottom of the groove. The vertical lines are a result of the Philtec groover which was used to prepare the sample. In this case again the Wright-Jenkins etch was used to bring out the damage. Note that the large bulk stacking faults, or elliptical-shaped stacking faults, lie in the $\langle 110 \rangle$ direction, parallel and perpendicular to the (110) flat and to the edges of the photograph.

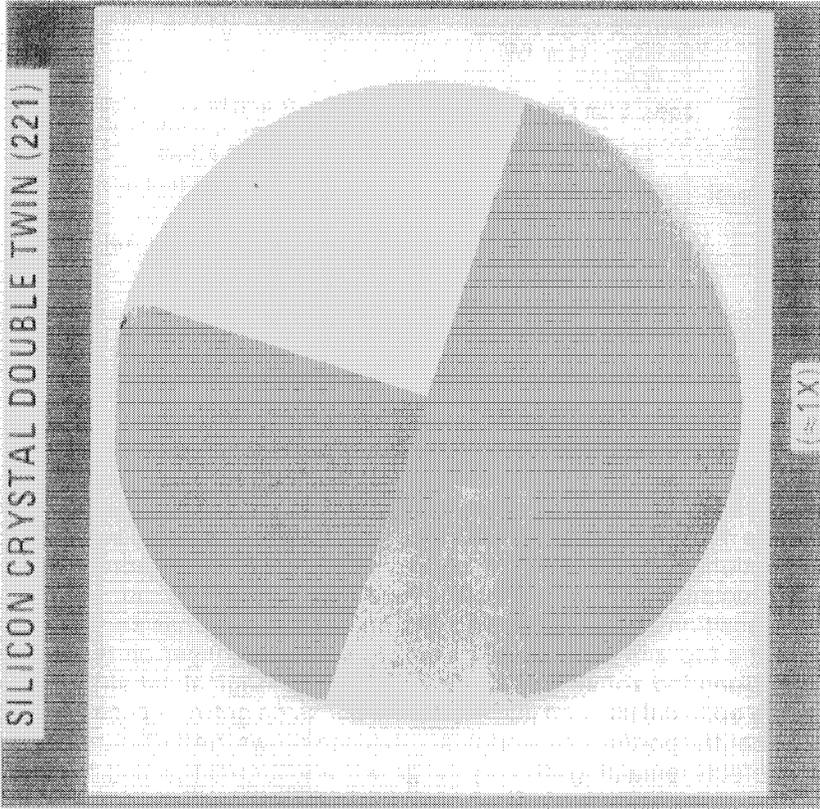


Figure 39: Use of defect delineation etch to bring out twin planes. (Magnification less than indicated.)

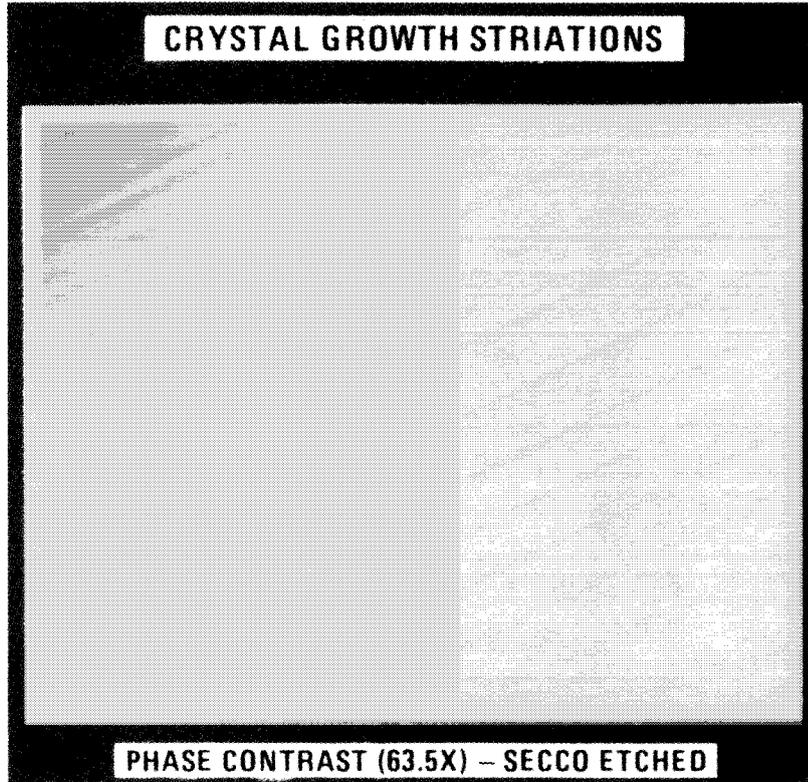


Figure 40: Photograph of slice surface containing concentric circle, growth striations. Secco etched. (Magnification less than indicated.)

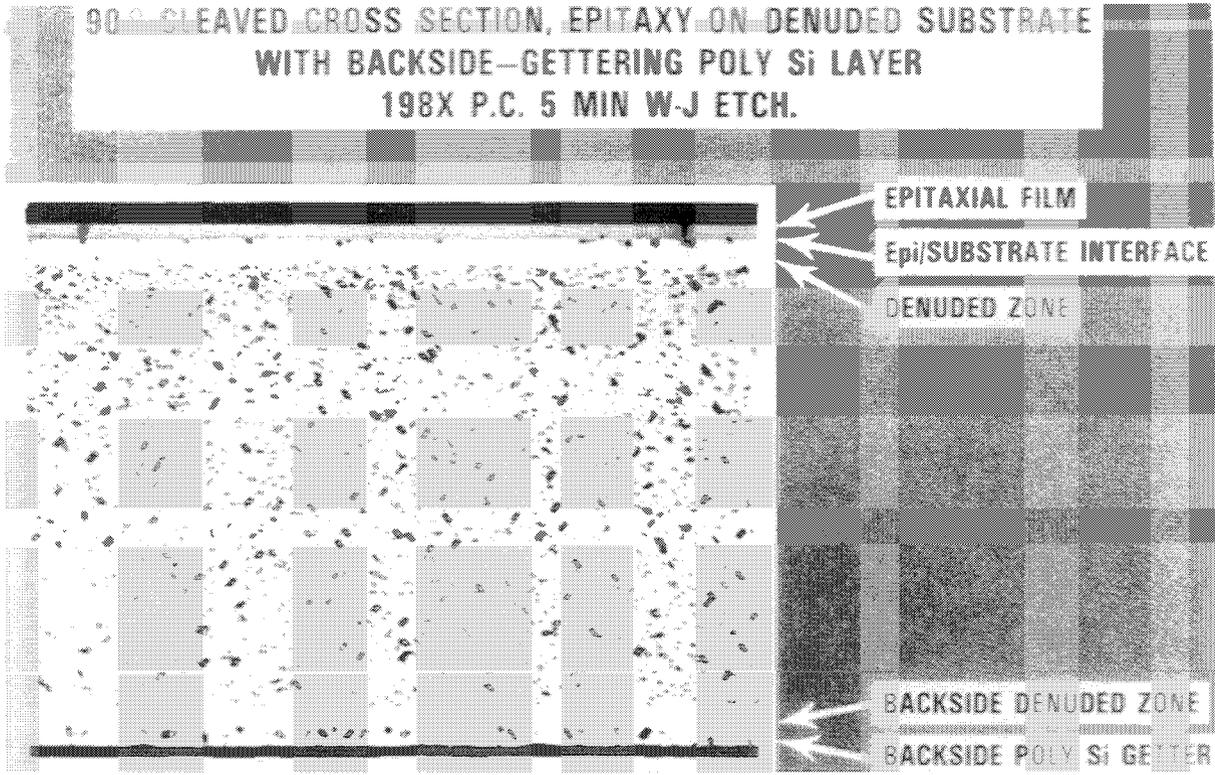


Figure 41: Photographic cross-section of cleaved surface showing bulk damage, denuded substrate, backside gettering and epitaxial layer delineated by Wright Jenkins etch. (Magnification less than indicated.)

DISLOCATIONS, LOOPS AND DAMAGE AT BOTTOM OF GROOVE

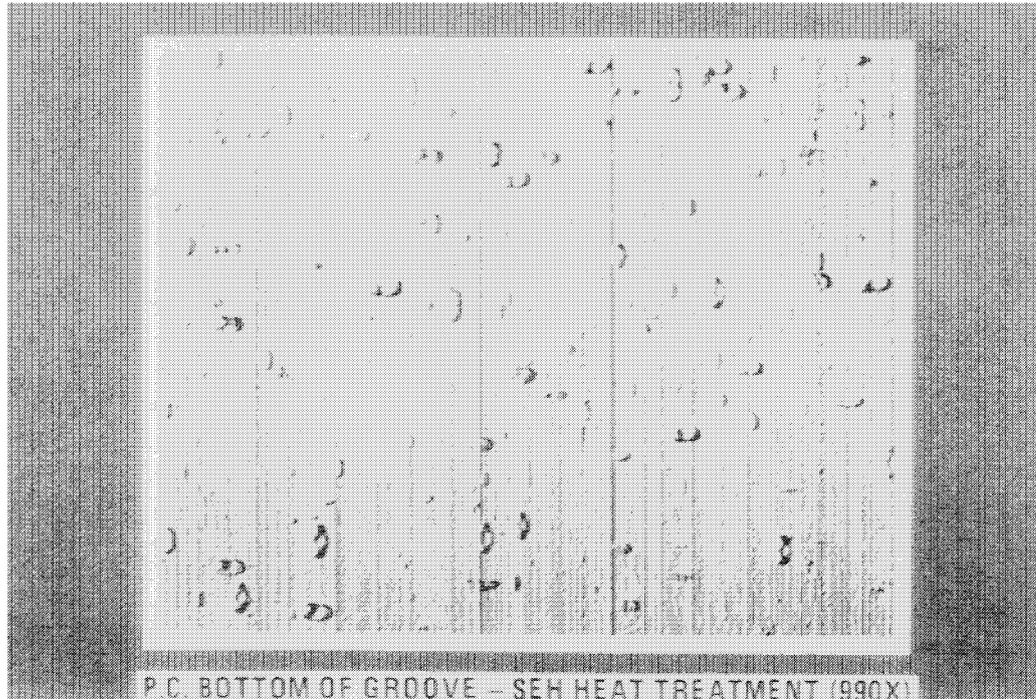


Figure 42: Photograph of grooved, and W-J etched, section of heat treated slice, showing dislocations, loops, and bulk stacking fault damage delineated by W-J etch. (Magnification less than indicated.)

Figure 43 is also a grooved section of a (100) silicon substrate. Again, note the large bulk stacking faults lying in $\langle 110 \rangle$ directions. There are also carbon related cone shaped defects lying at 45 degree angles to these bulk stacking faults. The Wright-Jenkins etch was used to bring out these defects. Auger analysis shows that the cone-shaped defect is related to excessive carbon in the silicon.

Figure 44 shows a photograph of the top of an epitaxial film which has an epitaxial stacking fault. From this stacking fault it is evident that the epitaxial material is (100) orientation. In this photograph, which was taken after Wright-Jenkins etching to bring out the dislocations or the damaged sites surrounding the stacking fault, a clear area is noted around the stacking fault. This indicates that the stacking fault acted as a getter to the defects which surrounded it. Large dislocations can be seen at the corners of the stacking fault.

Figure 45 is an enlarged view showing the large conglomeration of gettering effects at the dislocations on the corners of this same stacking fault visible after Wright-Jenkins etch. This type of gettering of small defects to a stacking fault, or to dislocations at the corners of a stacking fault, is common and is usually associated with gettering of heavy metals in silicon processing. However, in this case, Auger analysis revealed heavy carbon on the stacking fault dislocations, no detectable carbon in the denuded or gettered zone around the stacking fault, and heavy carbon in the clusters of small defects. This indicates the presence of carbon which can be gettered to dislocations or damage sites.

Figure 46 is a cleaved cross-sectional view of an epitaxial film deposited on a previously denuded substrate. The presence of a row of precipitation defects at the substrate/epitaxial interface shows the necessity for using in situ HCl vapor etching even in previously denuded substrates. A cross-sectional view of the denuded substrate shows a very clean area just below the surface. However, the surface itself acts as a trap for heavy metals and damage sites, as do the oxygen precipitations in the bulk material. It is important to note, however, that all these defects at the epitaxy/substrate interface do not introduce damage or stacking faults into the epitaxial film. Therefore a technique of designing-in a layer of defects at the substrate epitaxial interface can be used for extrinsic gettering at a zone just below the device area. (See Figures 47 and 48).

As mentioned above in situ HCl vapor etching will remove the damage sites and/or surface traps prior to epitaxy. However, one may wish to leave or even enhance this damage layer for in situ gettering in near proximity to the active device region of the structure. In this case we can deposit a built-in extrinsic misfit, dislocation strain field by intentionally doping, with germanium, an interface film between the substrate and the epitaxial film. Figure 47 shows a cleaved cross sectional view of a single layer misfit dislocation, extrinsic gettering strain field, and a single layer epitaxial film. Figure 48 shows multiple layers of strain field/epitaxial silicon films with increasing amounts of Ge doping in the strain fields as they were deposited. This increase in Ge doping causes a noticeable increase in the density of misfit dislocations within the strain field layers. Wright-Jenkins etch was used to reveal these damage sites.

DISLOCATIONS, LOOPS AND DAMAGE AT BOTTOM OF GROOVE

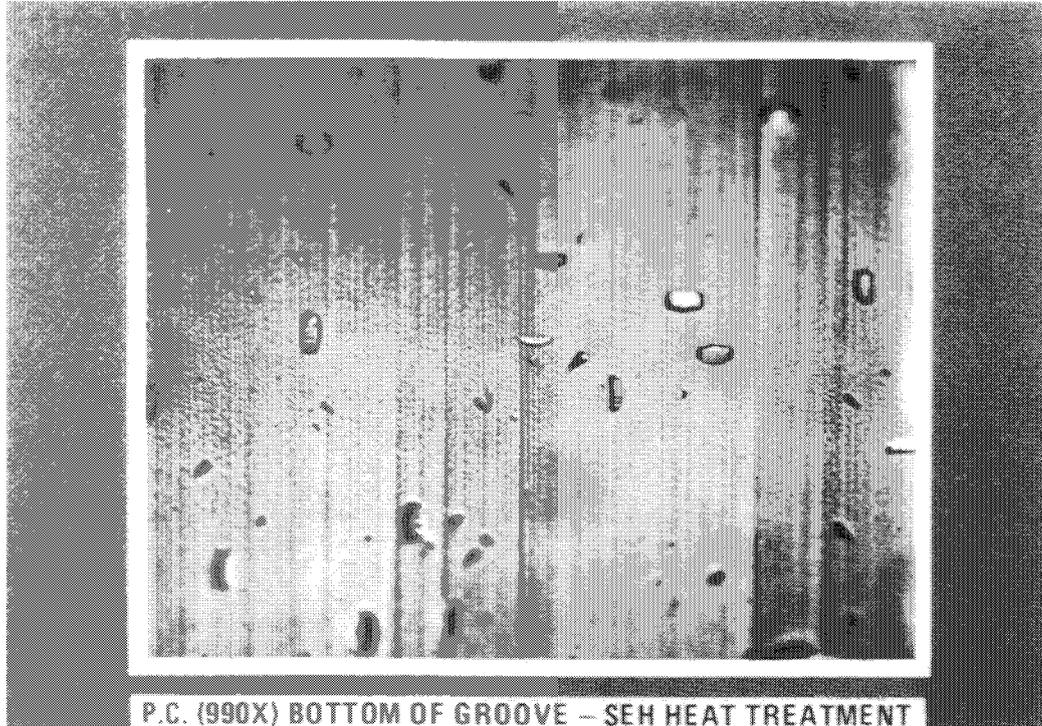


Figure 43: Photograph of heat treated, grooved and W-J etched section showing dislocations, bulk, stacking faults and "carbon" defect cones. (Magnification less than indicated.)

(100) STACKING FAULTS, DISLOCATIONS AND GETTERING EFFECT

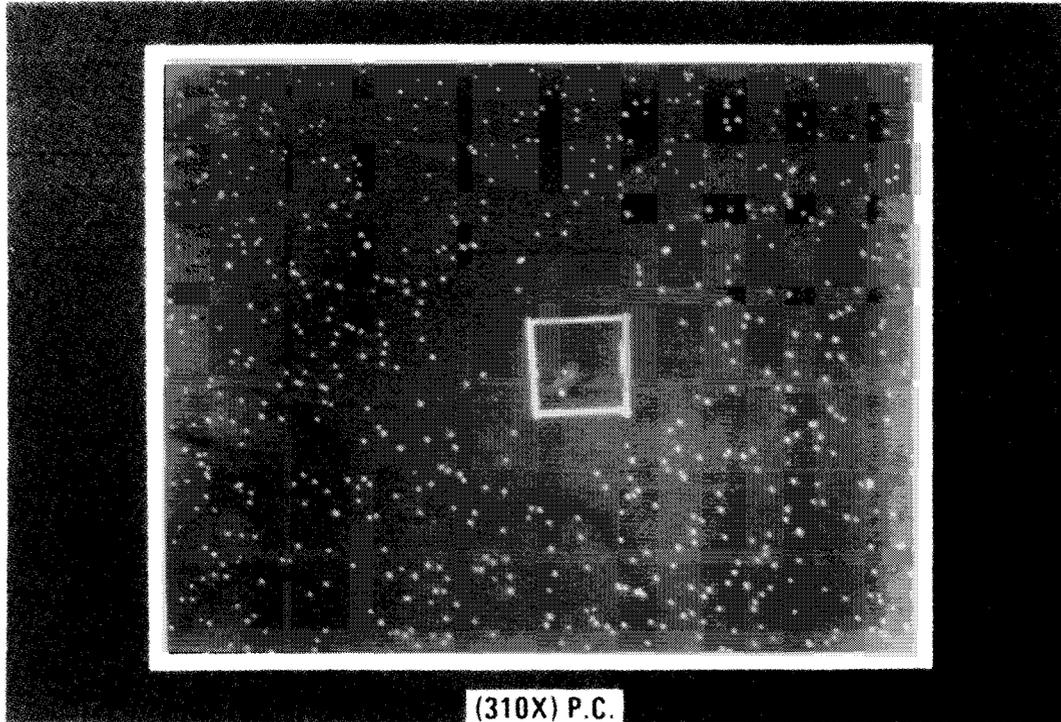


Figure 44: Photograph of (100) epitaxial stacking fault, carbon pits, dislocations and carbon gettering effect, W-J etched. (Magnification less than indicated.)

(100) STACKING FAULTS, DISLOCATIONS AND GETTERING EFFECT

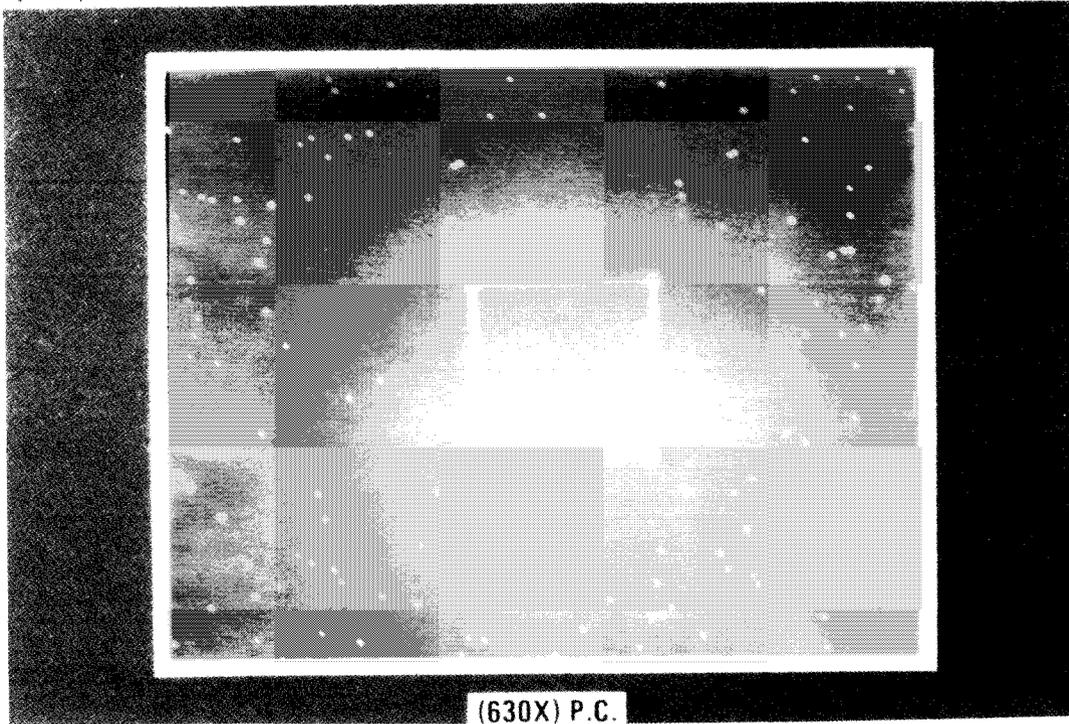


Figure 45: Enlarged view of (100) ESF and gettering. (Magnification less than indicated.)

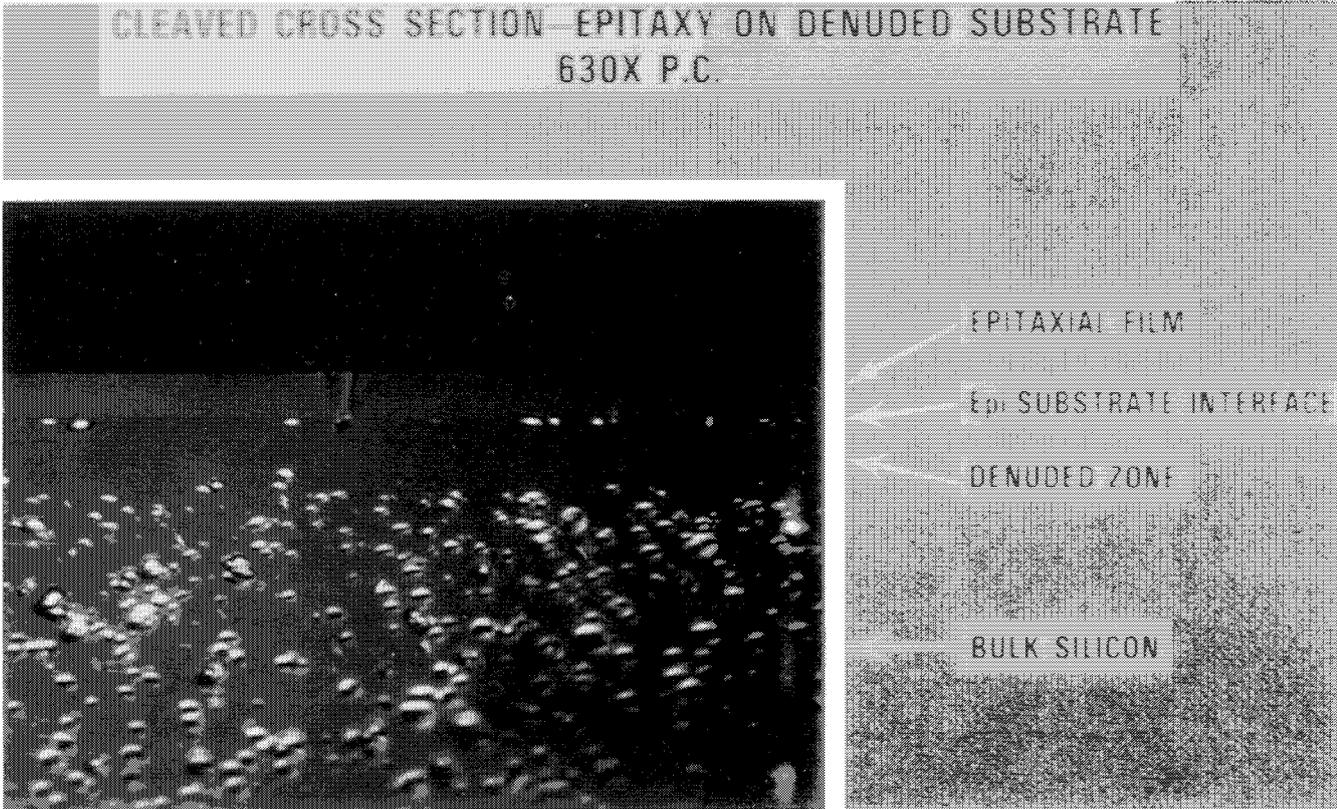


Figure 46: Cleaved cross-section, W-J etched, denuded substrate with epitaxial film. (Magnification less than indicated.)

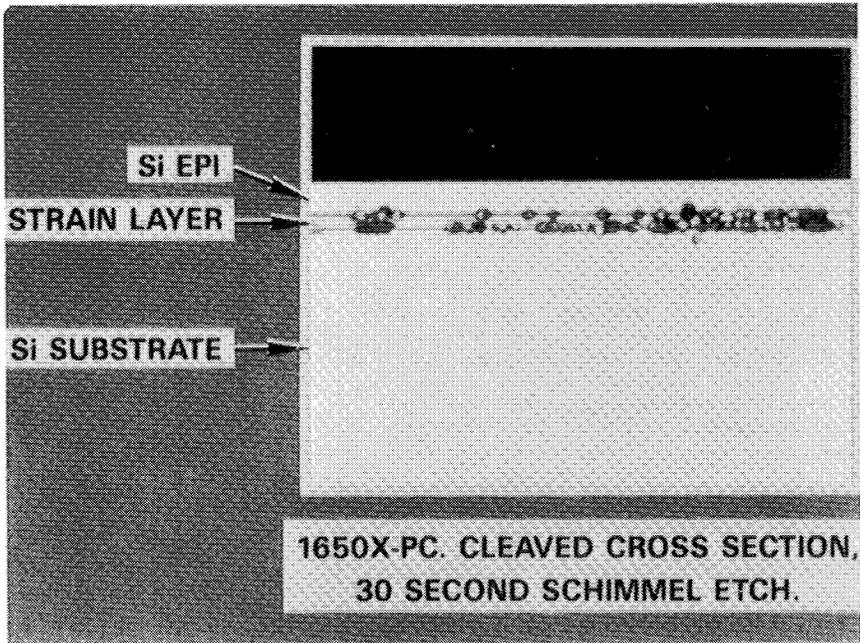


Figure 47: Misfit dislocation extrinsic gettering.

Figure 49 shows a large stacking fault in a $75\ \mu\text{m}$ thick epitaxial film. Epitaxial stacking faults of this size have a cusp around the edge of the stacking fault that extends slightly above the surrounding surface. With a microscope such as the Reichert Ultrastar with Nomarski interference contrast, in the blue phase, the upper left quadrant of the stacking fault will appear to be red, thus showing that the stacking fault cusp is slightly above the surface area. This is a method that can be used to determine if perturbations are above or below the surface. That is, if the blue phase is used in the Nomarski interference contrast, the upper left-hand quadrant will appear to be red if the perturbation is above the surface, and the lower right-hand quadrant will appear to be red if the perturbation is below the surface.

SLICE CLEANUP

During the past decade the advent of higher packing densities, in large scale integration (LSI), very large scale integration (VLSI) and now ultra large scale (ULSI), has created a demand for ultra clean processes and manufacturing processing areas. One of the first and perhaps most critical factors is the crystallographic, physical, and chemical cleanliness of the

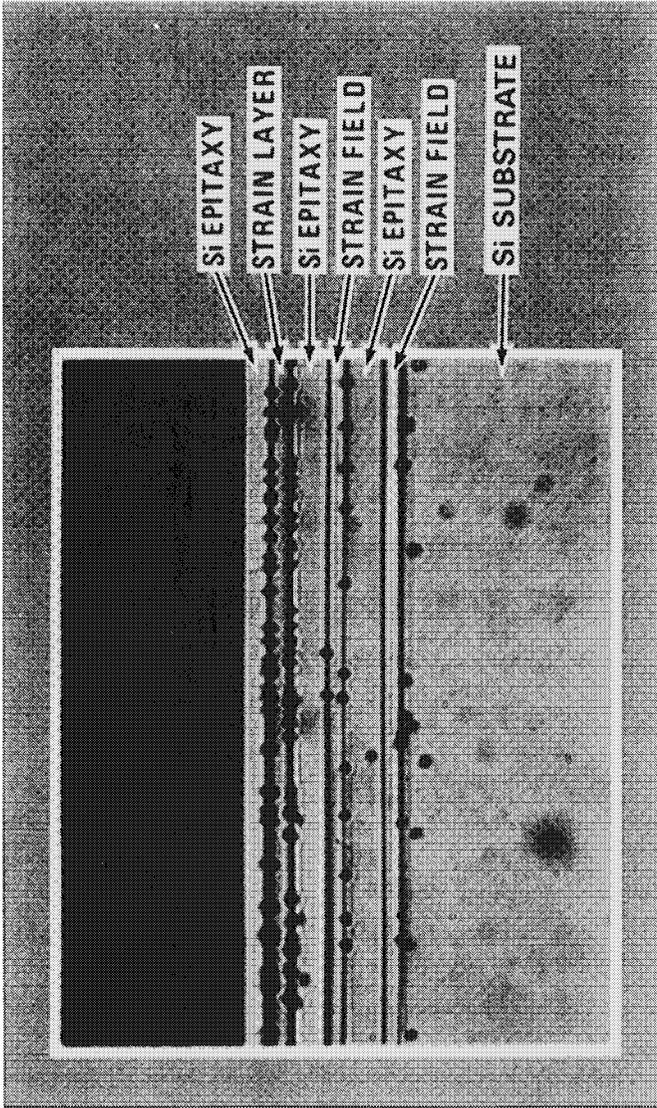


Figure 48: Misfit dislocation extrinsic gettering.

3-MIL Epi STACKING FAULT, BLUE PHASE EFFECT

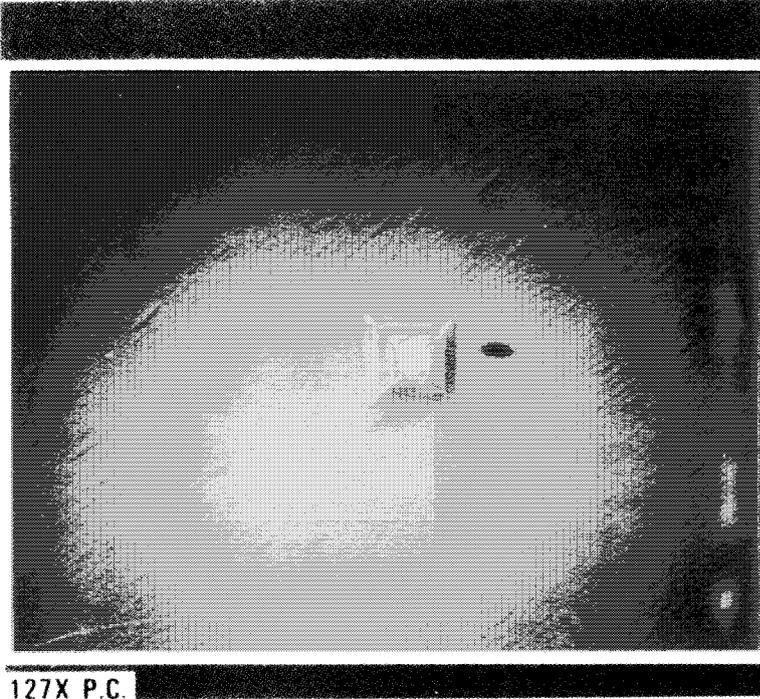


Figure 49: (100) epitaxial stacking fault, photograph taken under Nomarski interference phase contrast, no etching. (Magnification less than indicated.)

starting substrate or slice, and the subsequent maintenance of this clean condition throughout the process steps, Table 5.

Table 5: Silicon Slice Cleanup Effects on Processes and Device Performance

- Cleanliness is next to godliness.
- Affects all device processes
 - Oxidation rate
 - Oxidation induced stacking faults
 - Yields
- Effects device parameters
 - MOS - refresh time (lifetime)
 - Bipolar low current HFE (Beta)
 - Surface state density
 - Breakdown voltage
 - GOI (Gate Oxide Integrity)
 - Defect densities

Complete cleaning of semiconductor surfaces requires that particulates, organic films, and adsorbed metal ions be removed. Most cleaning procedures are based on immersion in liquid baths or liquid sprays. In addition, ultrasonic agitation or brush scrubbing may be required. In some cases high temperature vapor etching, or low pressure sputter etching may be used. A good cleanup is complicated by the fact that unless great care is taken, the materials used for cleaning may contain, and leave behind, more particulates and metals. Organic solvents are widely used, but they sometime leave residues themselves. As a result high purity water is ordinarily the last stage of a cleanup, preceded by an oxidizing acid etch to remove any remaining organics. Industry accepted standards (Table 4) for slice cleanups were reported in *RCA Review*, June, 1970 by W. Kern and D. Puotinen.¹⁴

Surface contaminants may be generally classified as inorganic or organic. Inorganic or atomic impurities such as heavy metals, e.g., gold, copper, iron and magnesium are lifetime killers in silicon processing. Sources include process equipment and chemicals, e.g., polishing equipment and media, and acid etchants. In many cases the etchants contain the contaminants, or remove the contaminant, such as gold, from the surface and then redeposit it back onto the surface. These contaminants can best be removed by cleaning solutions containing acids and complexing agents which dissolve the heavy metal and complex the ionic form to prevent plating or redeposition from the cleaning solution. The alkali ions of sodium and potassium are especially harmful in that they may form mobile charges at the silicon/oxide interface. These charges can move due to electric fields or temperature changes, thus producing inversion layers, leakage, and device instabilities. Organic contaminants, normally photoresist residues, oils (fingerprints) and waxes (from the polishing

processing) are usually chemabsorbed on the surface, and are best removed by cleaning in a hot sulfuric acid (H_2SO_4) cleaning solution or the ammonium hydroxide $NH_4OH + H_2O_2 + H_2O$ cleaning solution of Table 6.

Table 6: Cleaning Solutions

CLEANUP	SOLUTION MIXTURE				TEMPERATURE
*Choline Clean	Choline ($CH_3)_3N(CH_2CH_2OH)OH$	H_2O_2	***NCW-601A (Surfactant)	H_2O	$50^\circ C \pm 5^\circ C$ Ultrasonic 10 min, 10 min DI, H_2O , spin dry
Percent	3	1	1	95	
**RCA Clean "Basic"	NH_4OH	H_2O_2	H_2O		$80^\circ C \pm 5^\circ C$ 10 min DI water rinse
Parts by Volume	1 1	1 2	5 7		
RCA Clean "Acid"	HCl	H_2O_2	H_2O		$80^\circ C \pm 5^\circ C$ 10 min DI water rinse
Parts by Volume	1 1	1 2	6 8		
Percent	H_2SO_4 60 70	H_2O_2 40 30			

*TI — Ken Bean

**Reference from *RCA Review*, 31, #2, June 1970

***Available from: Wako USA Chem Co.

PRECLEANUP SOLVENT RINSE

Oils and organic residues remaining on the wafer from the slice polishing process may be removed by rinsing in an organic solvent such as perchloroethylene or xylene. Freon fluorocarbon solvents are also used for vapor degreasing or cleaning.

It was shown by Schwettman¹⁵ in 1978 and confirmed later by others,^{16,17} (Table 7) that the chemical cleanup process affects the oxidation rate, oxide charges, surface state density, Q_{SS} , the offset voltage, V_{FB} , and therefore the device or circuit characteristics (Table 8). In bipolar circuitry the preoxidation cleanup affects the oxide charge density, noise, junction leakage, low current gain (beta), junction breakdown voltage, minority carrier lifetime and isolation leakage between devices. In MOS circuitry, the preoxidation cleanup affects the oxide charge density, lifetime, refresh time, junction leakage, threshold voltage and breakdown voltage. (See Table 9).

In 1981 the Japanese authors¹⁸ reported the superior qualities of choline based cleanups. This cleanup, (Tables 6, 7, and 8), is based on the reactions of trialkyl (hydroxyalkyl) ammonium hydroxide (THAH) or choline, which is a strong base, and H_2O_2 . The THAH solution can be made free of sodium, potassium, and heavy metals. It is water soluble, and can be readily removed by a deionized water (DI) rinse. The choline cleanup removes

Table 7: Oxidation Rate vs Pre-Oxidation Cleanup

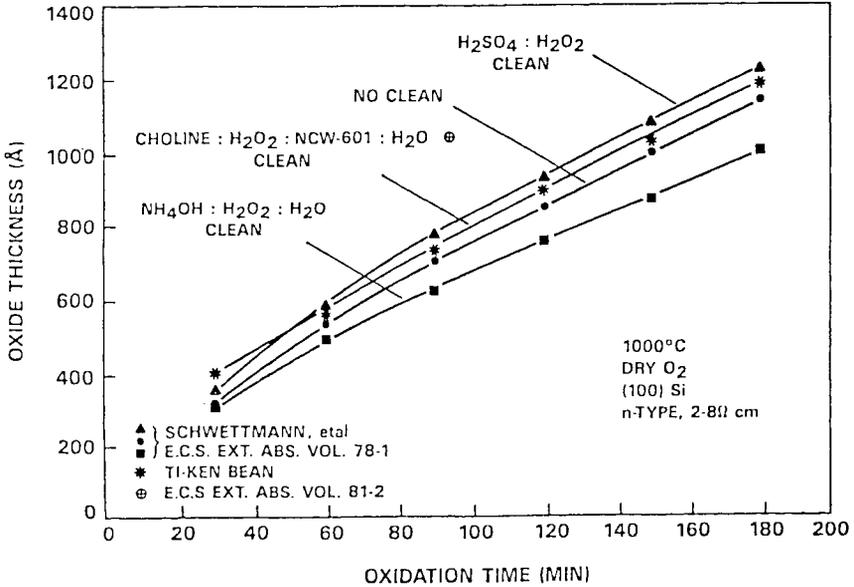
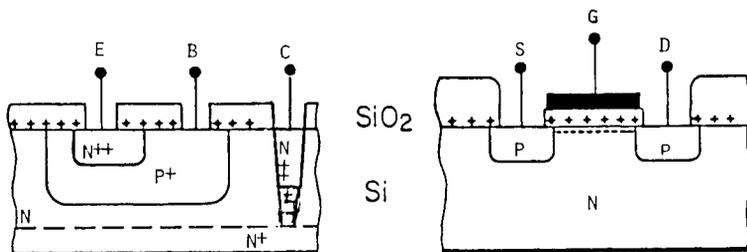


Table 8: MOS Gate Oxide Characterization vs Pre-Oxidation Cleanup

CLEANUP		W _{ox} (NS)	W _{ox} (Elec)	V _{FB}	ΔV+	ΔV-	BV Med	D Calc	E _{BR} (NS)	
STANDARD	1	409	432	-85			31.9	21	7.8	W _{ox} (NS) = oxide thickness by nanometric measure. W _{ox} (Elec) = oxide thickness by CV measurements
	2	411	429	-92	+05	-024	31.3	24	7.6	
	3	412	432	-88			32.5	4	7.9	
CHOLINE	4	410	433	-95			39.4	28	9.6	V _{FB} = flat band voltage BV (Med) = breakdown voltage (median)
	5	400	429	-89	+028	-036	39.0	44	9.8	
	6	404	427	-84			38.7	22	9.6	
STANDARD	7	269	296	-86			25.6	24	9.5	D (Calc) = defect density calculated E _{BR} (NS) = oxide electric field breakdown voltage by nanometric measure.
	8	272	298	-88	+02	-030	25.7	34	9.4	
	9	272	300	-87			26.3	20	9.7	
CHOLINE	10	276	301	-85			29.0	20	10.5	
	11	265	301	-85	+025	-024	28.7	16	10.8	
	12	262	287	-907			28.6	30	10.9	
STANDARD	13		107	-805			13.8	17	12.9	
	14		107	-813	+03	+03	13.9	25	13.0	
	15		109	-77			13.9	16	12.8	
CHOLINE	16		109	-79			13.9	13	12.8	
	17		108	-794	+03	+02	13.8	5	12.8	
	18		107	-75			13.6	9	12.7	

Table 9: Oxide Charges Can Affect the Following Device Parameters**Bipolar**

Junction leakage
 Junction breakdown
 Noise
 Low current beta
 Leakage between devices

MOS

Device threshold voltage
 Channel conductance
 Junction leakage
 Junction breakdown
 Leakage between devices

organic material since it is a strong base. It etches silicon ($3\text{\AA}/\text{min}$) and removes native oxides but does not etch thermal oxides. In an experiment comparing eight different standard cleanups, including all RCA processes, using radioactive tracers and activation analysis, we found choline to be the most effective in removing, the lifetime killer, gold (Au) from the silicon surface. Table 8 compares the MOS gate oxide electrical properties of the SiO_2/Si structure, as a function of preoxidation cleanup, for approximately 400\AA , samples 1-6, 275\AA , samples 7-12 and 100\AA , samples 13-18, gate oxides. The oxide breakdown voltage or dielectric strength of the 400\AA oxides are over 20% higher for the choline cleaned samples and over 10% higher for the 275\AA gate oxides. No difference is observed in the 100\AA oxide due to Fowler Nordheim conduction through the thin oxide.

CHOLINE CLEANUP PROCESS

The choline cleanup appears to be a good silicon slice cleanup. This cleanup removes, without swabbing, particulate matter, heavy metals, and even fingerprints that have dried overnight on the slices. Choline $(\text{CH}_3)_3\text{N}(\text{CH}_2\text{CH}_2\text{OH})\text{OH}$ is a strong base which chemically acts somewhat like KOH or NaOH but does not contain sodium or potassium ions.

The "choline clean" consists of 3% choline, 1% surfactant (NCW-601A), 1% H_2O_2 (by volume) in DI water, 40 to 45°C with ultrasonic for 10 minutes, 10 minute DI H_2O rinse, alcohol spin dry (or hot N_2 spin dry) no swab.

The choline has a rather pungent (fishy) odor and must be used in a good hood. Choline cleaning solutions have been stored overnight and

reused the following day without degradation, however, this is not recommended at this time.

For a production process a two-step cleanup using e.g., the RCA HCl (acid) cleanup followed by a choline cleanup is recommended. A multiple bath setup wherein the slices would go through at least two or three cleanup solutions, with the solution tanks being rotated from last to first as the solution is contaminated and/or used up is suggested for each process.

REFERENCES

1. B. Schwartz and H. Robbins, *J. Electrochem. Soc.* 123: 1903 (1976).
2. W.C. Dash, *J. Appl. Phys.* 27: 1193 (1956).
3. J.B. Price, *ECS Silicon Symposium* (1973).
4. E. Bassous, *IEEE Trans. Elec. Dev.* ED-25: 1178 (1978).
5. K.E. Bean, *IEEE Trans. Elec. Dev.* ED-25: 1185 (1978).
6. K.E. Petersen, *Proc. IEEE* 70: 420 (1982).
7. D.L. Kendall, *Appl. Phys. Lett* 26: 1500 (1974).
8. R.M. Finne and D.L. Klein, *J. Electrochem. Soc.* 114: 997 (1967).
9. E. Sirtl & A. Adler, *Z. Metallk* 52: 529 (1961).
10. F. Secco d'Aragona, *Phys. Status Solidi (a)*, 7:577 (1971).
11. M. Wright, *J. Electrochem. Soc.* 124: 757 (1977).
12. D.G. Schimmel, *J. Electrochem. Soc.* 126: 479 (1979).
13. K.H. Yang, *J. Electrochem. Soc.* 131: 1140 (1984).
14. W. Kern and D. Puotinen *RCA Review* June (1970).
15. Schwettman et al., *ECS Ext. Abs.* 78-1, (1978).
16. Bruce Deal, Fairchild, Personal Communication.
17. K.E. Bean, G.A. Brown, (see Tables 8 and 9).
18. H. Muraoka et al., *ECS Ext. Abs.* 81-2: 570 (1981).

5

Plasma Processing: Mechanisms and Applications

W.C. Dautremont-Smith
Richard A. Gottscho
R.J. Schutz

AT & T Bell Laboratories
Murray Hill, New Jersey

1. INTRODUCTION

In this chapter we discuss the use of plasmas in the microelectronics industry. There are basically two reasons why plasmas are used: (1) to achieve anisotropic heterogeneous chemistry; and (2) to grow materials or etch materials under conditions far from thermodynamic equilibrium. Anisotropy is required since feature widths have become comparable to feature depths. Non-equilibrium conditions are advantageous in two ways: firstly, new materials or phases can be grown which would not be thermodynamically favored in an equilibrium process; and secondly, the deleterious effects of high temperatures on devices can be avoided.

The chapter is divided into three main subsections: fundamental aspects, etching, and deposition. The origin of anisotropy and the non-equilibrium nature of the plasma will be discussed in the first section and the resultant advantages to device processing will be described in the following two sections. Since there have been several recent review articles,¹⁻¹⁶ particularly on plasma etching, we have not made an exhaustive effort to include everything that has been published in recent years on this subject.

In the first section on fundamental aspects the material is initially tutorial but rapidly progresses to a discussion of the results of recent diagnostic experiments which illustrate how power is dissipated in rf glow

discharges and how the form of power dissipation affects surface chemistry. Plasma-surface chemistry is then discussed first in terms of chemical vapor transport and then in terms of the molecular interactions between reactive adsorbates, surfaces, and products.

The second section dealing with plasma etching is oriented toward process design. The steps involved in fabricating a planar MOS silicon transistor are outlined in order to illustrate the variety of ways in which plasma etching is used in microscopic pattern transfer. Trade-offs between different reactors, and opposing design constraints, are discussed along with processes for etching specific materials.

The final section deals with plasma deposition of materials for micro-electronic applications. The plasma-enhanced chemical vapor deposition (PECVD) of silicon nitride and silicon oxide is discussed in detail, with emphasis on recent advances in techniques and property correlations. Amorphous and microcrystalline silicon deposition is discussed from the point of view of the parallels and contrasts with silicon nitride deposition. Comprehensive coverage of this widely studied field has not been attempted, but adequate reference is made to the large number of existing reviews. Emerging PECVD applications, in epitaxial semiconductor growth (Si, Ge, GaAs, GaSb), metal deposition, silicide deposition, and the deposition of non-silicon-based oxides and nitrides, are also covered. Emphasis throughout the section is on the influence of plasma parameters on the wide range of accessible film properties, and in particular in indicating those of relevance in a variety of semiconductor applications. Plasma film growth, such as oxidation and nitridation, is not discussed.

2. FUNDAMENTAL ASPECTS

The idea behind this section is to present properties of plasmas and plasma reactors which are common to all types of processes, be they etching or deposition. The emphasis is on unifying concepts. Where appropriate we have drawn upon the literature for specific examples to illustrate a concept; more often than not the examples have come from the etching literature since etching processes are generally better understood. In some instances, the subject matter has not been covered completely and some references may have been omitted for the sake of covering a wide range of unifying concepts rather than covering one or two areas in depth. To our colleagues whose work we have not included, we apologize in advance.

In Sec. 2.1, we discuss some of the fundamental aspects of plasmas and sheaths, starting with definitions of their properties as they pertain to plasma processing. Widely used equivalent circuit models of rf plasmas are discussed in light of recent diagnostic results. The relationship between the equivalent circuit parameters and processing variables will be emphasized. Next, we discuss the plasma-surface interaction. A useful and general framework for understanding both etching and deposition, chemical vapor transport (CVT) theory, is reviewed critically in Sec. 2.1. The salient features of this theory can be summarized in terms of an equivalent

circuit for the heterogeneous chemistry. Current thinking and recent experiments on the microscopic mechanisms at play in both spontaneous and ion-enhanced etching reactions are also reviewed in this section. In Sec. 2.3, we review very briefly recent efforts to model various aspects of plasma chemistry.

2.1 Plasmas and Sheaths

Plasmas mean different things to different people.¹⁷ Basically, they consist of a “soup” of ions, electrons, radicals, and stable neutrals, but, the relative numbers of these species depend on operating conditions. Typical numbers for and characteristics of the rf discharges used in microelectronic processing are given in Table 1. The primary reason for using radio frequencies is to avoid charging effects when dealing with insulating substrate material (e.g. SiO_2): dc discharges cannot be maintained with insulating electrodes because there is no direct current conduction path. In addition, we will see that rf glows have physical properties which depend on the operating frequency and which can be used to advantage in heterogeneous chemical processing. Pressures employed are such that collisions between neutral and charged species are important. Power densities employed are such that charge densities are relatively low, typically 10^8 - 10^{11} cm^{-3} , which means that collisions between charged particles or between charged particles and low density radicals can usually be neglected.

Two distinct zones are common to all plasmas: the plasma body and the sheaths (Figure 1). The body is defined as a low field region where the number densities of positive and negative species are approximately equal. In most discharges, electrons are the dominant negative charge carrier; and, because the electrons are unbound, the plasma body is a good conductor. Sheaths correspond to electron deficient, poor conducting regions which exist wherever the plasma encounters an interface (electrode, wall, etc.). Sheath formation occurs because of the difference in mobility between electrons and ions, which stems in turn from the non-equilibrium nature of partially ionized glow discharges (see Table 1): Electrons easily gain energy from the field and heat up because they do not exchange energy efficiently with the more massive neutrals with which they collide; the neutrals and ions, on the other hand, do exchange energy efficiently so that energy gained from the field is rapidly dissipated and the ion and neutral temperatures remain close to the wall temperatures.¹⁸

The voltage drop between the conducting plasma and electrodes occurs in the sheath, resulting in a large sheath electric field. This large field in turn can lead to gradients in concentration, temperature, and flux of ions and radicals to and from electrode or device surfaces. Ultimately, these gradients are responsible for the non-equilibrium, or low temperature, and anisotropic heterogeneous processing desired.

2.1.1 Response Time and Screening Distance. How does the difference in mobility between electrons and ions result in sheath formation? Consider the situation when a potential difference is applied across a

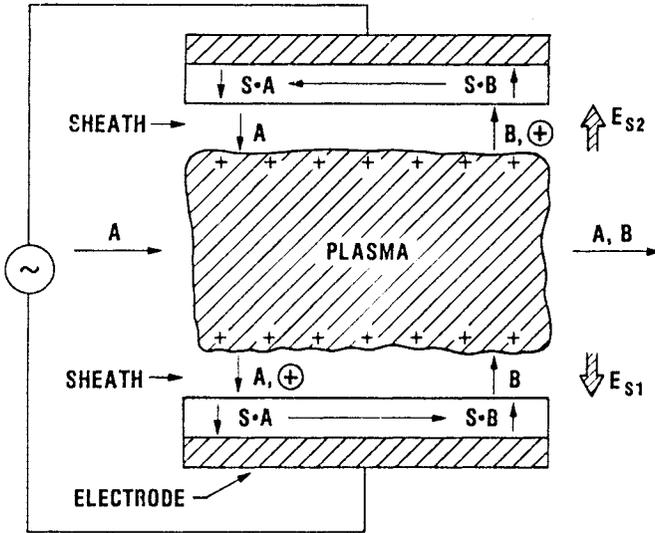


Figure 1: Schematic diagram of parallel plate plasma reactor. Reactant A enters from left and is removed along with product B at right. Positive ions, denoted by +, traverse the sheaths along the electric field lines, E_s , and impact surface S where the neutral reactions, etching on the bottom electrode and deposition on the top electrode, are enhanced (courtesy of C.B. Zarowin).

Table 1: Typical RF Plasma Properties

Pressure	10^{-3} to 10 Torr
Frequency	10 kHz to 30 MHz
Charge Density	10^8 to 10^{11} cm^{-3}
Electron Plasma Frequency	100 to 3000 MHz
Ion Plasma Frequency	1 to 5 MHz
Debye Length	0.03 to 1.0 mm
Electron Temperature	1 to 5 eV (Time dependent)
Ion and Neutral Temperature	≈ 0.03 eV
E/N (Sheaths)	10^2 to 10^6 Td ^a

^a 1 Td = 10^{-17} V cm^{-2}

plasma. The electrons rapidly drift toward the positive electrode leaving behind a net positive restoring force, which prohibits further electron depletion. A steady state is achieved when the plasma potential is sufficiently positive that electron and ion loss rates become equal. The time it takes electron plasma frequency,^{17,19}

$$\omega_e = (ne^2/\epsilon_0 m_e)^{1/2}, \tag{1}$$

where n is the total charge density, e is the electronic charge, ϵ_0 is the permittivity of vacuum, and m_e is the electron mass. This frequency is directly related to the restoring force the electron feels when extracted from the plasma (hence the dependence on the charge density, n). From table 1 we see that ω_e is much larger than typical operating frequencies for the discharges used in plasma processing. Thus, plasma potential adjustment and sheath formation will be virtually instantaneous as far as we are concerned.

The maximum distance over which charge imbalance can be maintained, in the absence of an externally applied force, is the Debye length,^{17,19}

$$\lambda_D = (\epsilon_0 k T_e / ne^2)^{1/2} = \bar{c}_e / \omega_e, \tag{2}$$

where k is Boltzmann's constant, T_e is the electron temperature, and \bar{c}_e is the mean electron speed. The screening length is inversely proportional to ω_e just like the displacement of a spring is inversely proportional to the restoring force constant. The Debye length is also proportional to the random energy, or speed, since this energy must be overcome to achieve effective shielding. Although the two are related, the Debye screening length is not to be confused with the sheath thickness, which is generally an order of magnitude larger. The relationship is not an obvious one and depends on operating parameters such as pressure, frequency, and power density.²⁰⁻²²

2.1.2 Equivalent Circuits. A simple equivalent circuit which is suitable for both dc and rf plasmas is shown in Figure 2a.^{19,23-29} Both regions of the plasma are represented by resistors in parallel with capacitors; diodes are used in the sheaths to represent the differences in ion and electron mobilities which result in rectification of the applied voltage (see below). Many of the operating characteristics of discharges used in microelectronic processing can be understood in terms of the relative impedances of each of these components (Figure 3). Consider dc or low frequency discharges, where the reactive impedance is so large that only the resistive components need to be considered. As mentioned above, the plasma body impedance will be governed primarily by electron conductivity,

$$\sigma_e = ne^2/m_e \nu_e \tag{3a}$$

$$R_P = \frac{l_P}{\sigma_e A_P}, \tag{3b}$$

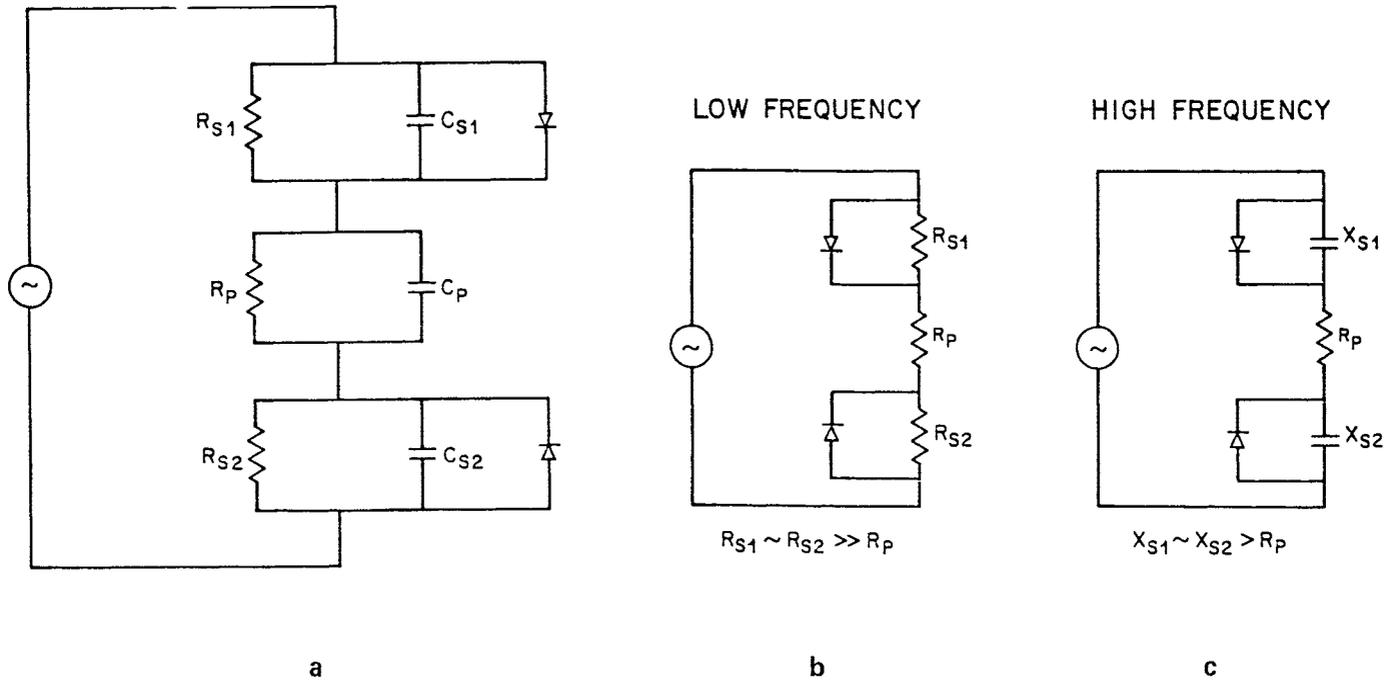


Figure 2: (a) Electrical discharge equivalent circuit. Each part of the discharge, sheath and plasma, is represented by a parallel combination of resistors and capacitors. In addition, the sheaths have diodes in parallel in order to account for rectification of the applied voltage. (b) Low frequency equivalent circuit. The resistive impedances are much less than the capacitive impedances. (c) High frequency equivalent circuit. The capacitive impedances in the sheaths are less than the resistive impedances so that a capacitive current shunt exists.

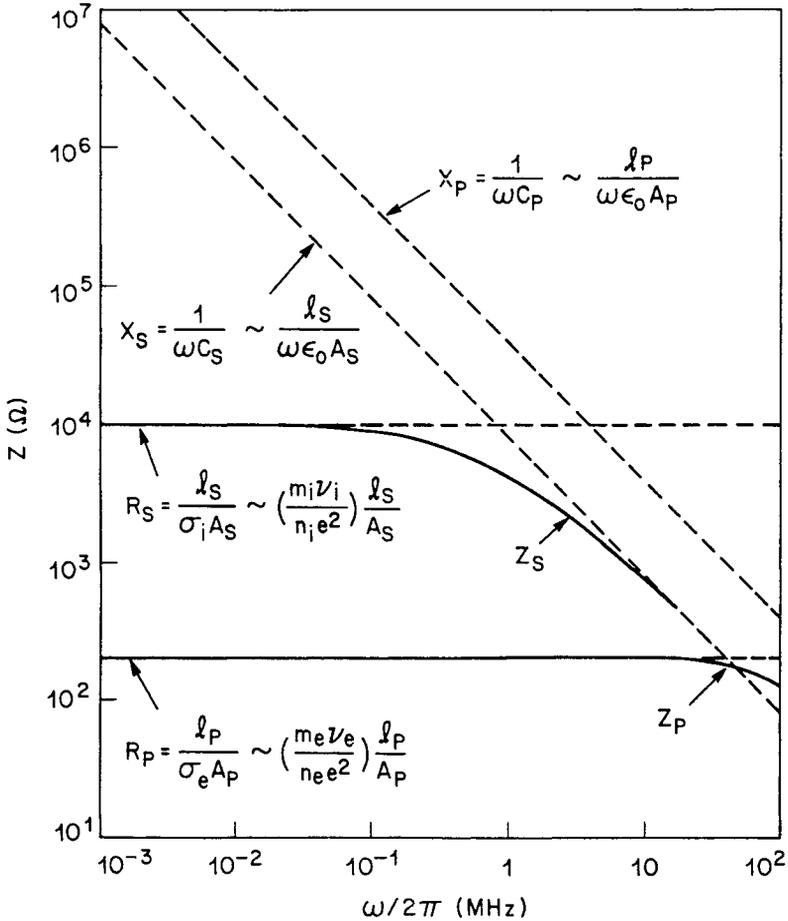


Figure 3: Plasma impedances vs. frequency. Resistances, R_S and R_P , are determined by ion and electron conductivities, $\sigma_i = 5.5 \times 10^{-7}$ mhos cm^{-1} and $\sigma_e = 10^{-4}$ mhos cm^{-1} , sheath and plasma thicknesses, $l_S = 0.25$ cm and $l_P = 1.0$ cm, respectively, and the electrode area, $A_S = A_P = 45.6$ cm^2 . The capacitive impedances, X_S and X_P , are determined by the thicknesses, areas, and frequency.

where ν_e is the electron momentum transfer collision frequency, A_p is the plasma cross sectional area, l_p is the plasma length, and R_p is the plasma body resistance. The sheath impedance will be given by the ion conductivity, defined as in Equation (3) except using the ion mass and collision frequency (see Figure 3). Primarily because of the mass difference, the ion conductivity is roughly an order of magnitude smaller than the electron conductivity; thus the sheath resistivity will always be greater than the plasma resistivity.

At higher frequencies, the sheath and plasma capacitive impedances

must be considered. The capacitances of the sheaths and plasma can be simply estimated from the area of the electrode and the sheath and plasma thicknesses, respectively (Figure 3),

$$C_{P,S} = \frac{\epsilon_0 A_{P,S}}{l_{P,S}}. \quad (4)$$

From Figure 3, we see for frequencies below 100 MHz, the plasma body impedance is predominantly resistive. However, the sheath impedance changes around the ion plasma frequency ($\omega_i \simeq 1$ MHz), from being primarily resistive at lower frequencies to being primarily reactive at higher frequencies. In other words, the sheath capacitor becomes a current shunt. Above ω_i , the ions can no longer respond to the instantaneous value of the field (see below) and so displacement instead of conduction current dominates.

2.1.2.1. Experimental Verification. Recently, experimental diagnostic techniques have been developed which allow the concentrations of free radicals and ions as well as electric field amplitudes to be measured in situ and non-intrusively.^{21,22,30-47} For a recent review, see Reference 37. These techniques allow us to see the extent to which the equivalent circuit model is appropriate.

2.1.2.1.1 Voltage distribution. How does the electric field vary across the electrode gap? According to the equivalent circuit model, we expect the field to be largest in the sheaths over the frequency range of interest to plasma processing (see Table 1 and Figure 3). In situ electric field measurements are consistent with the model. The local field is plotted as a function of position for rf discharges through BCl_3 in Figure 5. For the range of frequencies studied, 50 kHz to 14 MHz, the field is always greatest in the sheaths.²¹

2.1.2.1.2 Diode behavior. How does the local electric field vary with time in the electrode sheaths? This has been measured by spectrally resolving laser-induced fluorescence from parity mixed rotational levels of the BCl radical formed in rf discharges through BCl_3 .^{21,46,48} The different parity levels are mixed by the local electric field; the extent of mixing is dictated by the field strength as well as the excited state dipole moment and zero-field energy level splitting.^{46,48} Parity mixing is detected by recording the intensities of transitions which would be "forbidden" in the absence of an electric field and whose line intensity is a direct measure of the electric field amplitude. The technique is illustrated in Figure 4, where the field has been sampled at two different times during the rf cycle by firing the laser synchronously with the applied rf.²² Note that the change signals for the "forbidden" and "allowed" lines in Figure 4 are equal and opposite in sign because the "forbidden" component has borrowed intensity from the "allowed" component as a result of the field-induced mixing. Both measurements are made one mm from the powered electrode sheath. The upper trace is obtained at a time when the powered electrode is the momentary anode (applied voltage a maximum); the "forbidden" line in the center is weak compared to the "allowed" lines on either side. Thus, the field is small during this part of the cycle. However, when the laser is fired

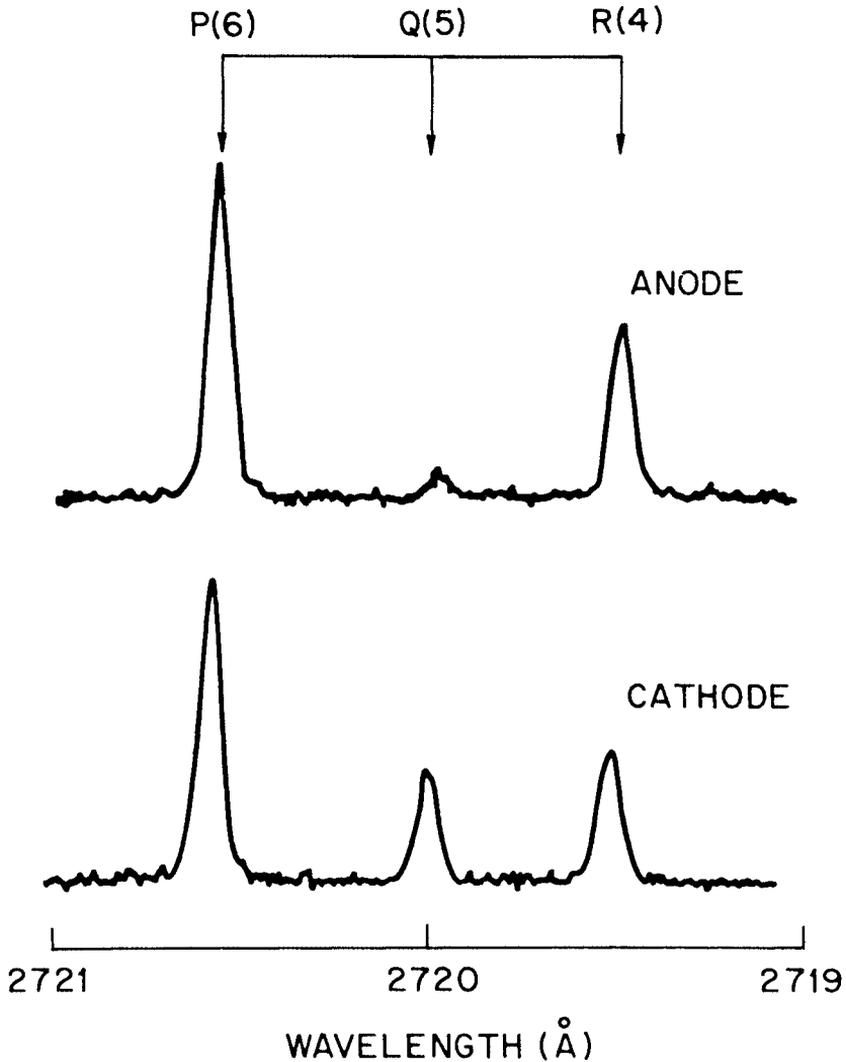


Figure 4: Spectrally resolved laser-induced fluorescence from BCl radicals formed in a 13.5 MHz discharge through BCl_3 at 0.3 Torr and 0.13 W cm^{-3} . The upper trace was obtained by exciting the P(6) transition 1 mm above the powered electrode at a point in the rf cycle when the voltage was a maximum, making the powered electrode the momentary anode. The lower trace was obtained in a similar fashion except during the cathodic part of the cycle. Note that the forbidden component, Q(5), is strongest during the cathodic cycle when the electric field is greatest (from Reference 21).

during the cathodic part of the cycle (lower trace), the “forbidden” line is comparable in amplitude to the “allowed” lines. During this part of the cycle, the electric field is strong. Thus, the applied field is rectified in the sheaths. As a result of the difference in electron and ion mobilities the plasma potential is “tied” to the anode potential. This behavior is accounted for in the equivalent circuit (Figure 2) by placing diodes in parallel with the sheath resistors and capacitors.

2.1.2.1.3 *Frequency response.* As the frequency is varied an interesting transition is seen to occur in Figure 5. The local field decreases by roughly a factor of two above 5 MHz. The peak voltage which must be applied across the plates in order to maintain constant power also decreases by this amount (Figure 6a). From the equivalent circuit model we see that this transition corresponds to a change in the sheath impedance from being predominantly resistive below ω_i to predominantly capacitive above ω_i . The total impedance decreases above 5 MHz as the sheath capacitive impedance becomes less than the ion resistive impedance (Figure 3). At constant power the ratio of voltages at high and low frequency is given by,

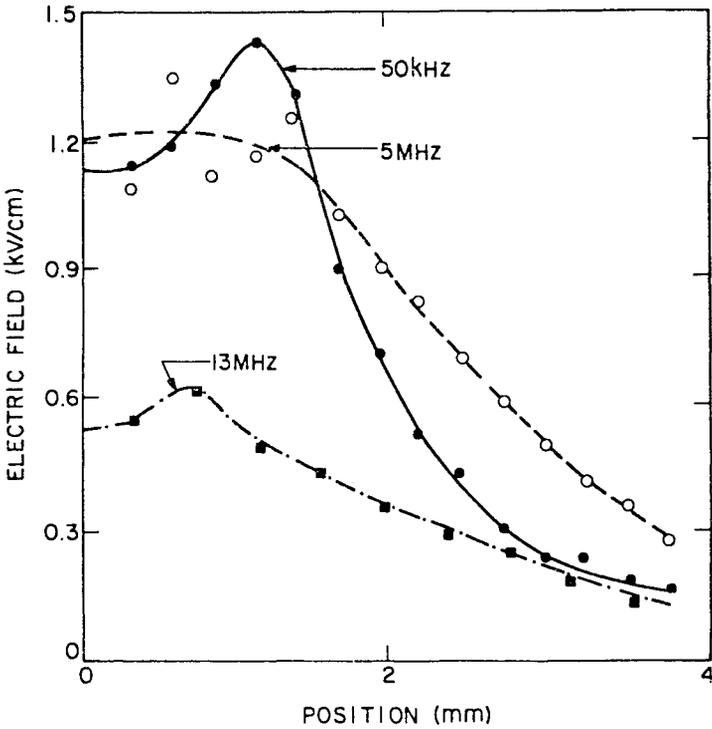


Figure 5: Preliminary measurements of electric fields in BCl_3 plasma as a function of position from the powered electrode to the plasma center for three different frequencies. The laser was fired at a time such that the powered electrode was the momentary cathode. The vertical scale is accurate to within ± 50 V/cm (from Reference 21). The counter electrode is at 16 mm.

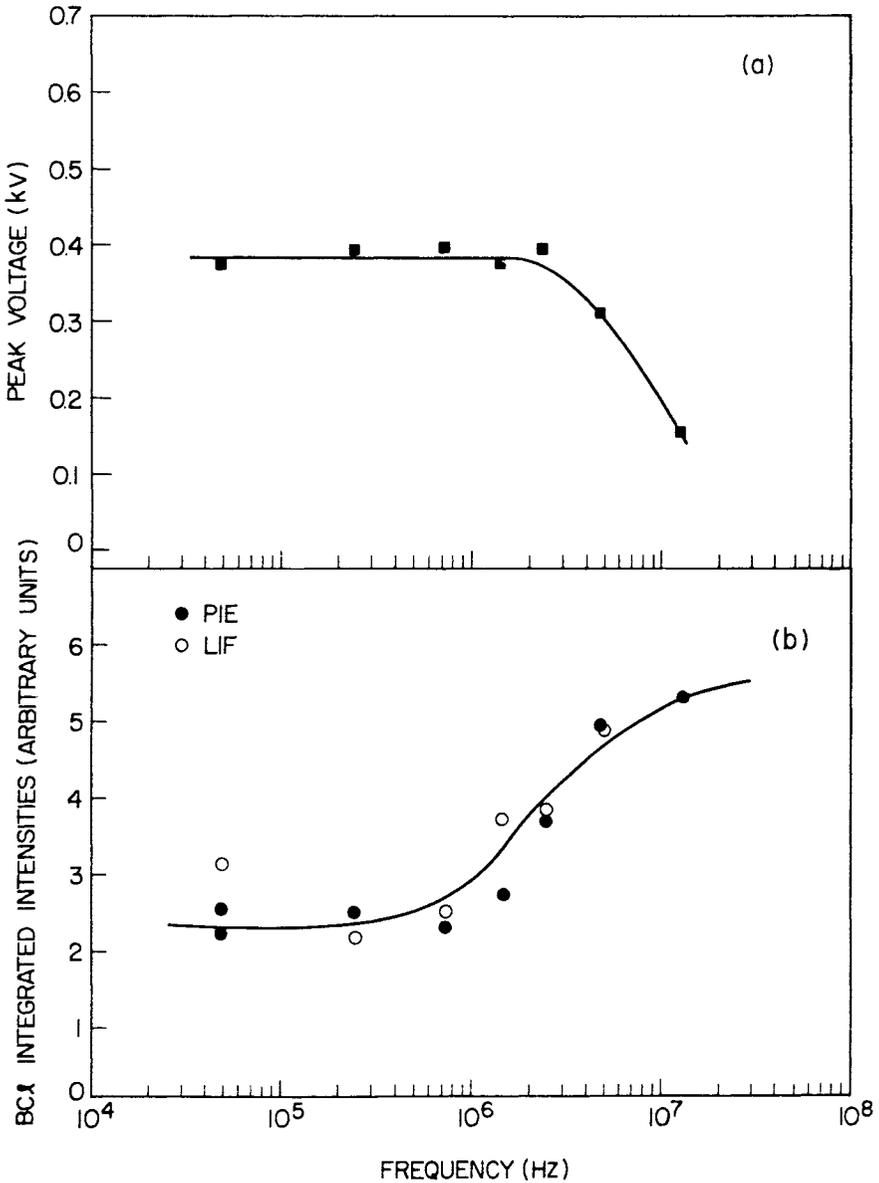


Figure 6: (a) Peak voltage for discharge through BCl_3 as a function of frequency at a pressure of 0.3 Torr and a power density of 0.13 W cm^{-3} . (b) BCl radical densities as a function of frequency. PIE refers to plasma-induced emission, i.e. excited state radicals, while LIF refers to laser-induced fluorescence, i.e. ground state radicals (from Reference 21).

$$\frac{V_h}{V_l} = \left[\frac{1 + (X_s/R_p)^2}{1 + R_s/R_p} \right]^{1/2} \approx 0.3, \quad (5)$$

which is in very good agreement with the values in Figures 5 and 6a. This transition is also apparent when one examines the voltage and current waveforms in Figure 7.²¹ Not only does the current increase and the voltage decrease above 5 MHz but the phase shift between the two increases toward 90° as the capacitive component becomes an important sheath current shunt.

2.1.2.1.4 Power dissipation. From the above discussions of voltage distribution and frequency effects we can see how and where power is dissipated in rf discharges as a function of frequency. At low frequency, the sheaths are primarily resistive and the sheath resistance is much greater than the plasma resistance. Thus, we expect power dissipation to occur primarily in the sheaths. Ions accelerated by the sheath field can dissipate their energy in basically two ways. Collisions with neutrals can result in ionization, excitation, chemical reactions and/or heating; collisions with electrodes can result in any combination of surface damage, sputtering, secondary electron emission and/or heating with either implantation or reflection of the incident ion.

The response of ions to the sheath field and the consequences of this response at low frequency can be seen very clearly in Figures 8 and 9.²¹ In Figure 8, the density of Cl₂⁺ measured in the sheath by laser-induced fluorescence is plotted as a function of time. During the positive part of the cycle when the local field is very small (Figure 4), the ion concentration builds as a result of both diffusion of ions from the plasma into the sheath and ionization by electron impact. During the negative part of the cycle, the ion concentration decreases precipitously as a result of the large cathodic fields (Figure 4) which sweep the ions out of the sheath toward the electrode. The extraction of high energy ions at low frequency causes ionization, excitation, and secondary emission of electrons. This can be seen in Figure 9 where the uv emission intensity from BCl radicals is recorded as a function of position across the electrode gap at different times during the rf cycle.²¹ When the ions are extracted during the cathodic cycle the emission is brightest because the ions and secondary electrons collide with neutrals and produce a cascade of ionization, excitation (Figure 9), and dissociation as they are accelerated across the sheath.

This periodic build-up and high-energy extraction of ions in the sheath has no dc or high frequency analog. In dc anode sheaths, ions build up to some steady-state level but are never extracted with high energy. In dc cathode sheaths, the ion concentration can never build up to a large value owing to the large, extracting sheath field. While the time-averaged flux may be similar in the dc and low frequency rf discharges, the pulsed ion bombardment in the latter may make a difference in heterogeneous reaction rates. In high frequency (i.e. above ω) discharges, the ions respond only to the average field, which is less than at lower frequencies owing to the resistive to capacitive transition discussed above. The net result at

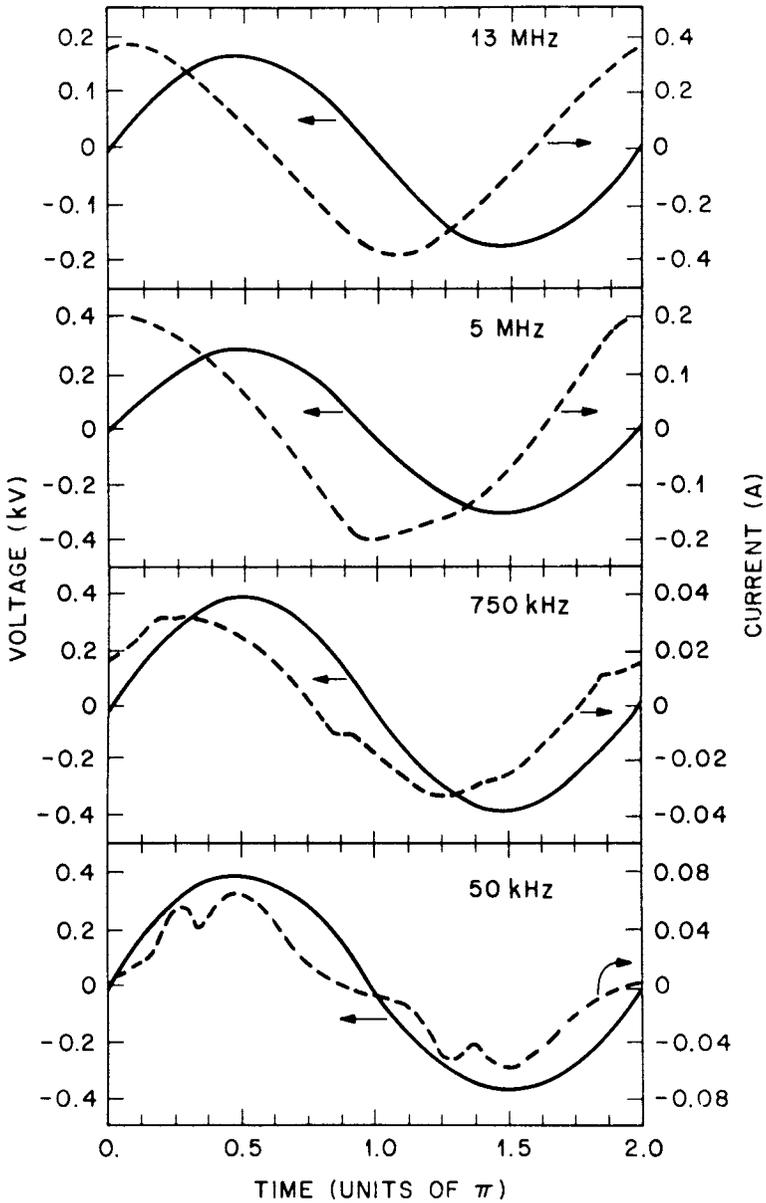


Figure 7: Current and voltage waveforms for discharges through BCl_3 at several different frequencies and a pressure of 0.3 Torr and a power density of 0.13 W cm^{-3} (from Reference 21). One unit of π corresponds to a $1/2\nu$, where ν is the rf frequency in Hz.

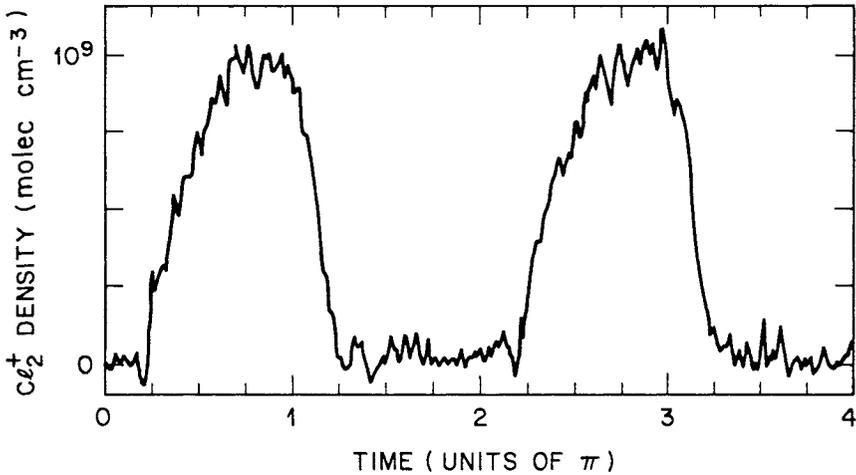


Figure 8: Cl_2^+ ion density vs. time in the sheath of a 55 kHz discharge through Cl_2 at 0.3 Torr and 0.6 W cm^{-3} . One unit of π corresponds to $18.2 \mu\text{sec}$.

high frequency is that the ions experience a smaller extraction force and again build up to some steady state concentration.

The response of ions to the instantaneous field and the change in the amplitude of this field with frequency at constant power affect the energy with which ions impact electrode or device surfaces. This is evident in the ion energy distributions measured by Bruce⁴⁹ as a function of frequency (Figure 10). At low frequency, the ions are accelerated to the full sheath potential, which is approximately the full applied potential (see Figure 5), on every half cycle as they traverse the sheath. To the extent that there is ionization and energy loss in the sheath, the ion energy distribution will be skewed toward lower energies. At 40 Pa (0.30 Torr) of Cl_2 and 100 kHz, the Cl^+ and Cl_2^+ ion energies *on average* are significantly less than the full sheath potential but the maximum ion energy is approximately equal to the full sheath potential (Figure 10).^{49,50} At 13.7 MHz, the ion energy distributions are much narrower and the maximum ion energies are much less than the peak sheath potential.

Because of the transition from resistive to capacitive sheaths above ω_p , power dissipation must shift from the sheaths to the plasma. Since the current is conducted primarily by electrons in the plasma, the dissipation mechanisms must involve electron-neutral collisions: ionization, dissociation, and excitation. The shift in power dissipation is evident in Figure 11, where the time-averaged concentrations of excited and ground state BCl radicals are plotted as a function of position across the gap. Three different frequencies are displayed; the shifts in emission intensity and radical density are indicative of the shifts in where power is dissipated. When these profiles are spatially integrated we can learn not only where but also how power dissipation changes with frequency. As frequency increases, the total, spatially-integrated densities of both excited and

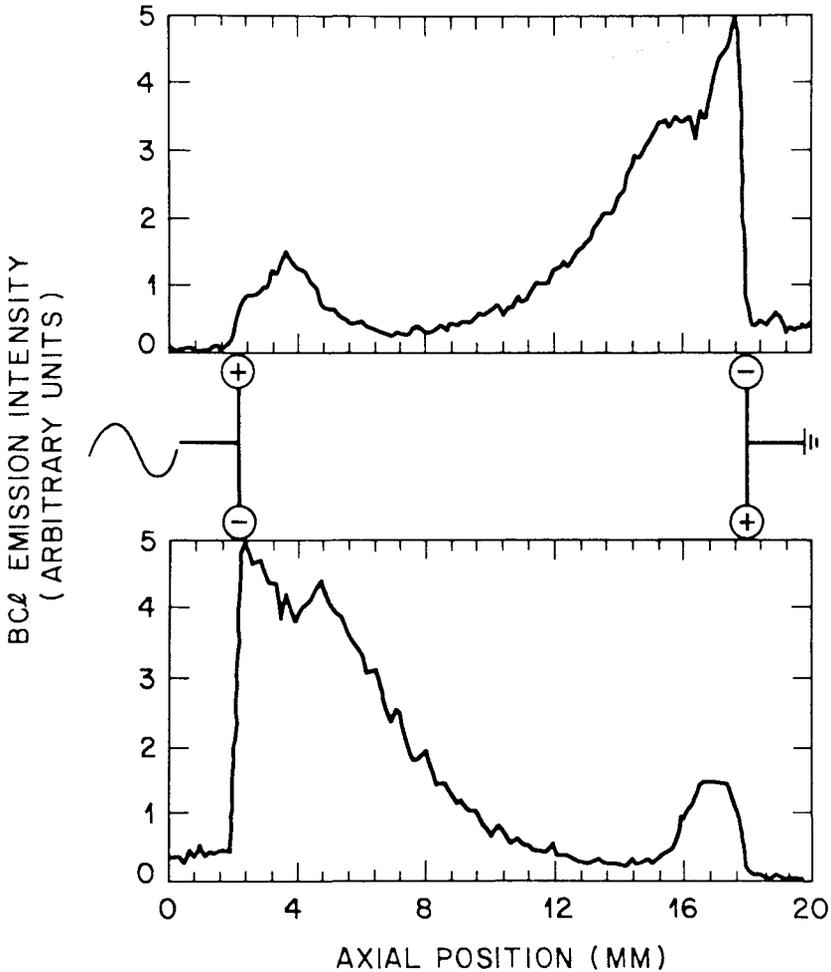


Figure 9: Time-resolved BCl emission obtained from a 50 kHz discharge through BCl_3 at 0.13 W cm^{-3} (from Reference 21).

ground state radicals also increases (Figure 6b)^{21,51} because power is no longer dissipated in the form of ion heating of surfaces and neutrals. Since total power is kept constant, the power dissipated must be converted to electron processes resulting in an overall increase in radical, ion, and excited-state production.

The net effect that operating frequency has on a particular etching or deposition process depends upon the specific surface chemistry. For example, Bruce⁵² has found that the etch rate of Si in a CCl_4 plasma increases by more than an order of magnitude as the operating frequency increases from below to above ω_i (Figure 12a). The reaction of chlorine with Si at room temperature without ion bombardment is negligible; the

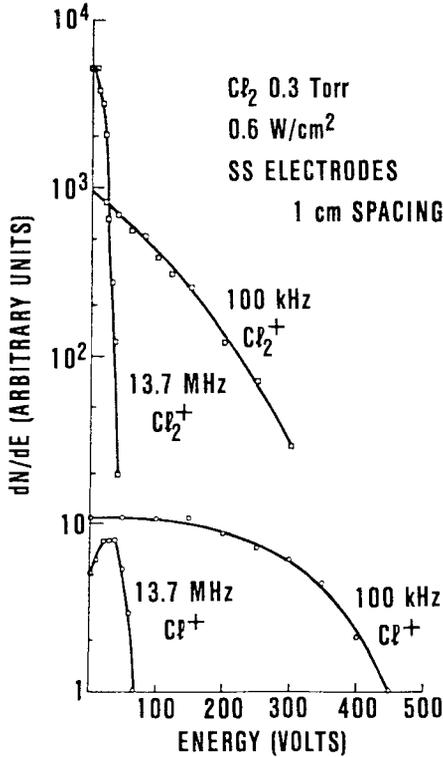


Figure 10: Ion energy distributions, dN/dE , vs. frequency for a discharge through Cl_2 (from Reference 49, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

reaction is ion induced.^{49,53-55} Thus, the plasma etching of Si is enhanced by lowering the operating frequency because of the increase in the ion energy at lower frequency (Figure 10). On the other hand, under comparable conditions the Al etch rate decreases by a factor of 4 as the operating frequency is tuned from below to above ω_i (Figure 12b). Al etches spontaneously in a chlorine environment once the native oxide has been removed.^{49,54,55} Thus, the Al etch rate increases with frequency because of the increased production of chlorine atoms and molecules as the degree of CCl_4 dissociation increases (Figures 6b and 11) despite the decrease in the bombarding ion energies.

2.1.2.2 Bias Effects. It is commonly observed that in high frequency discharges with unequal electrode areas that a dc bias voltage will build up between the two electrodes. There is a general consensus as to the cause of dc bias but not as to its scaling with discharge parameters and in particular its scaling with the electrode area ratio. First of all, unless a blocking capacitor is placed in series with the rf generator input (Figure 1), there will be a dc path to ground and no bias will be observed. The bias

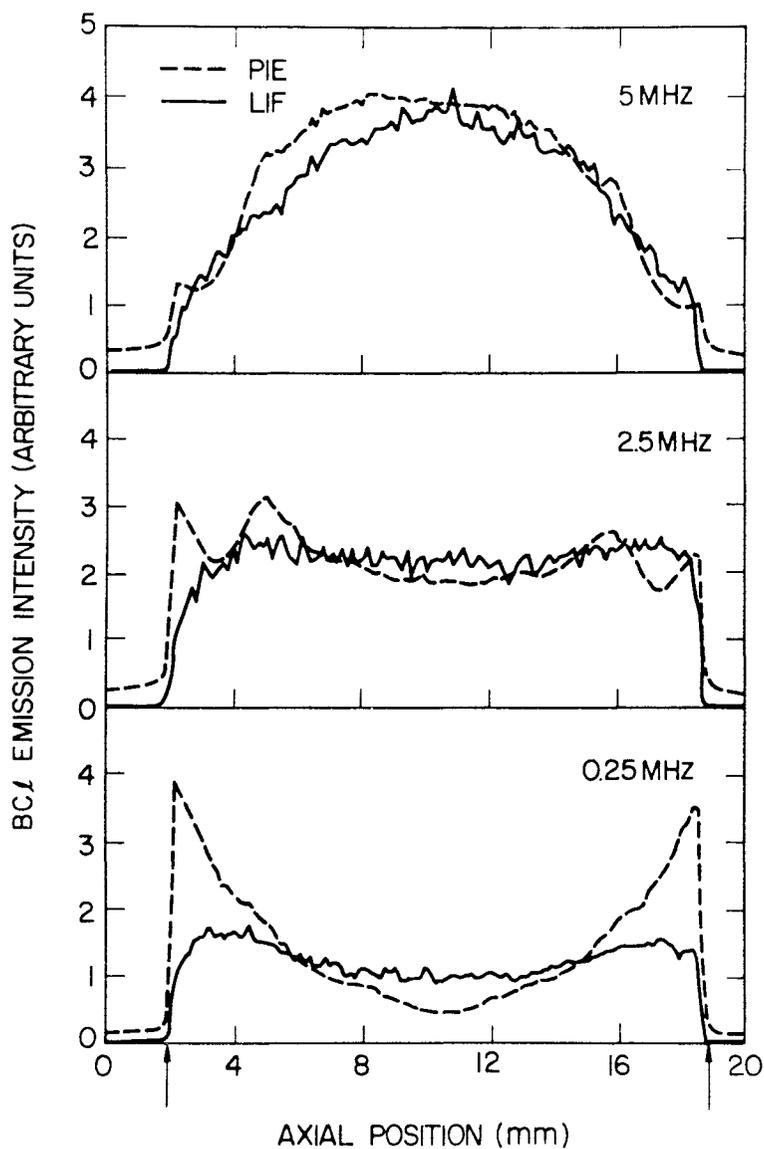


Figure 11: BCl radical spatial profiles as a function of frequency for a discharge through BCl_3 at 0.13 W cm^{-3} and 0.3 Torr . The arrows indicate the positions of the parallel plate electrodes. PIE refers to plasma-induced emission (excited states) and LIF refers to laser-induced fluorescence (ground states) (from Reference 21).

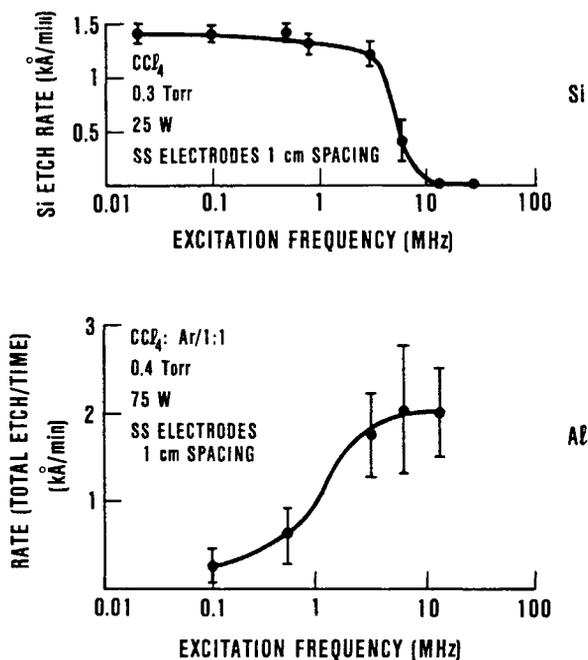


Figure 12: (a) Si etch rate vs. frequency in a discharge through CCl₄. (b) Al etch rate vs. frequency in a discharge through CCl₄ (from Reference 52, reprinted by permission of the publisher, The Electrochemical Society Inc.).

develops with a blocking capacitor when the electrodes have different areas because the ion current density to both electrodes is approximately the same.

It should be noted that an insulating substrate can serve as a blocking capacitor in the development of a dc bias voltage. However, depending on the polarizability of such an insulating layer, the sheath field may be reduced. If the insulator is polarizable it can shield the applied field and a voltage drop will occur across the substrate.

Zarwin²⁸ has shown, using the high frequency equivalent circuit model in Figure 2b, that the ratio of sheath electric fields should be equal to the ratio of areas. For equal sheath thicknesses, this implies that the voltage ratio should also be equal to the inverse area ratio:

$$\frac{V_1}{V_2} = \frac{A_2}{A_1}, \quad (6a)$$

where V_i is the potential difference between the plasma and electrode i whose area is A_i . This relationship was derived by assuming conservation of current density and sheath capacitances which are independent of current density but proportional to electrode area. One might expect this approximation to work better at higher pressures, where sheath thicknesses (and thus capacitances) are not very sensitive to voltage and current.²⁰

The “classical” work on dc bias effects to which most people refer is by Koenig and Maissel.²⁴ These authors used the Child-Langmuir law¹⁷ to relate current and voltage to sheath thickness. The sheath capacitance is assumed to be proportional to electrode area and inversely proportional to sheath thickness (see Equation 4). Furthermore, the plasma resistive impedance is neglected compared to the sheath capacitive impedance at high frequency so that the discharge acts like a capacitive voltage divider. This leads to the following equation:

$$\frac{V_1}{V_2} = \left(\frac{A_2}{A_1} \right)^4. \quad (6b)$$

This scaling relationship should be valid only at low pressures because the Child-Langmuir law applies to collisionless sheaths.¹⁷

Experimental testing of these relationships has not been extensive but the results obtained so far indicate that Equation 6a is generally more applicable. Coburn and Kay⁵⁶ measured sheath voltages as a function of the area ratio by sampling ions through a pinhole in one of the electrodes. For an Ar discharge at 0.05 Torr, they found the voltage ratio to scale more according to Equation (6a) than Equation (6b). More recently, Kohler et al.²⁹ made similar measurements on a 13.56 MHz Ar discharge operating at 20 mTorr. For a wide range of applied rf and dc voltages, the capacitive sheath model was found to be adequate in accounting for the measured sheath voltages. These experiments were done in an unbalanced (i.e. unequal electrode area) system and no attempt was made to look at the electrode area dependence.

In situ measurements of sheath fields in a 13.5 MHz discharge through BCl_3 at 0.3 Torr indicate that both the sheath field and thickness change with dc bias. The cathode field increases and the sheath *contracts* with increasing bias. No changes are observed in the anode sheath.⁵⁷ The net result is that the dc bias voltage appears across the cathode sheath only.

Control of dc bias, by changing electrode areas, rf power, or by using an external dc power supply, can be useful in controlling etching and deposition processes because the sheath voltages determine the energies of ions impacting device surfaces (for example, see References 27, 28, 58-60 and references therein). This point is discussed in further detail below (Sections 3.2 and 4.2.1).

2.1.2.3 Limitations of equivalent circuit models. From the above discussion, it should be clear that simple equivalent circuits can be useful for understanding many qualitative aspects of rf plasmas and developing intuition which can be applied in process design and trouble shooting. However, it should be equally clear that the equivalent circuit is really only a point of departure for complete understanding. The impedances plotted in Figure 3 are very crude estimates and ignore many aspects of the discharge physics and chemistry. In particular, the neglect of periodic, time-dependent variations in concentrations, energies, and fields must limit the validity of the equivalent circuit model as described above. For example, one need only look at the voltage and current waveforms for a low frequency

plasma (Figure 7) to see that the simple circuits of Figure 2 are deficient. A solution would be to consider the resistors and capacitors to have time-varying impedances. For example, the origin of the time-dependent sheath resistance can be seen in Figure 8. Once the ion density has been extracted by the field, the ion conductivity must necessarily decrease.

2.1.3 Feedstock Composition. Much has been written about the effects of feedstock composition on both gas-phase and surface plasma chemistry so we will not discuss the matter in great detail here. The reader is referred to recent review articles and references therein for more detailed information.^{4,5,10,16,61} There are basically two types of effects that feedstock composition can have on the plasma: physical and chemical. Most of the literature has dealt with chemical effects.

2.1.3.1 Chemical effects. The effects of feedstock composition on gas-phase chemistry depend upon the degree of dissociation of the feedstock constituents so that these effects are not independent of other plasma parameters such as frequency (Section 2.1.2.1.3), power (Section 2.1.5), and residence time (Section 2.1.4). Given that there is some degree of radical production, then the reactions proceed much like one would expect from fundamental chemical principles. For example, the addition of O₂ to a CF₄ discharge results in increased production of F atoms and CO and reduced production of fluorocarbon radicals and molecules.^{4,62-64,333} For example, the reaction,



proceeds rapidly and exothermically.

The net result on surface chemistry can be complicated owing to the different reactivities and sticking coefficients of various gas-phase species on various surfaces. For example, when O₂ is added to a CF₄ discharge used in the Si etch rate⁶³ (Figure 13). However, as more O₂ is added, the increase in the Si etch rate⁶³ (Figure 13). However, as more O₂ is added, the etch rate goes through a maximum and then declines because oxygen begins to compete with fluorine for Si adsorption sites.⁶³ A similar effect has been seen in the etching of GaAs and InP with CCl₃F/O₂ discharges.⁶⁵ Another effect of halocarbon oxidation is to reduce the extent of polymer deposition by oxidizing polymer precursors such as the CF₂ radical. Polymerization is a common problem resulting in slower etch rates and surface contamination.

Other feedstock recipes and the rationale behind their usage for specific etching and deposition applications are discussed in greater detail in Secs. 3.4, 3.5, 4.2.2, and 4.3.

2.1.3.2 Physical effects. The physical effects associated with feedstock composition have to do with changes in the electron energy distribution, electron density, ion energy distribution, ion density, and ion composition. These changes will, of course, affect homogeneous and heterogeneous plasma chemistry as well: radical production rates can be strongly dependent upon the electron density and energy distribution; surface reaction rates can be strongly dependent upon ion energy.

The effects of feedstock composition on the electron energy distri-

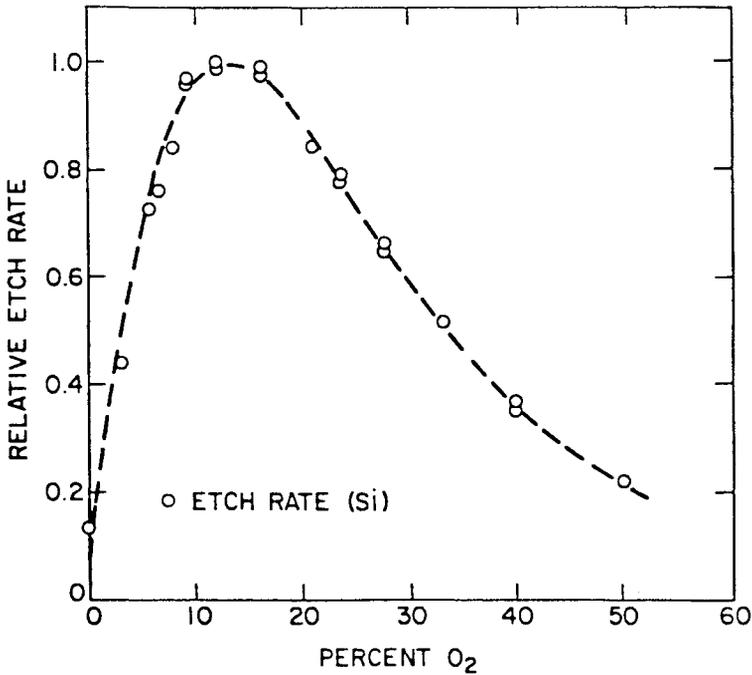


Figure 13: Si etch rate vs. O₂ concentration in a CF₄/O₂ discharge (after Reference 63).

bution can be seen in Figure 14 where the emission intensities from F and Ar are plotted as a function of the O₂ feedstock concentration in a CF₄/O₂/Ar discharge. As the oxygen concentration increases, the Ar emission intensity decreases indicating that the number of electrons with energies above the 7504 Å line threshold must also decrease. In other words, the electron energy distribution appears to cool. Note that the F atom emission lines at 7037 Å and 6856 Å, which have similar excitation thresholds, increase with O₂ concentration. The overall increase in F atom production by free radical reactions (see above and Figure 14) more than compensates for the decrease in excitation efficiency.

Two explanations for the cooling of the electron energy distribution with the addition of O₂ are plausible: (1) the enhanced production of F atoms and the introduction of oxygen, both species being electronegative compared to CF_x compounds, results in electron cooling by electron capture processes; or, (2) the lower ionization potential of O and O₂ relative to F and CF_x⁶⁶ results in electron cooling by inelastic ionization of neutral O and O₂. Of these two, the second explanation is more viable because electronegative gases do not attach high energy electrons very effectively (see Reference 67 and references therein). On the other hand, the introduction of a lower ionization potential gas necessarily reduces the average electron energy since a lower energy collision channel has been opened.

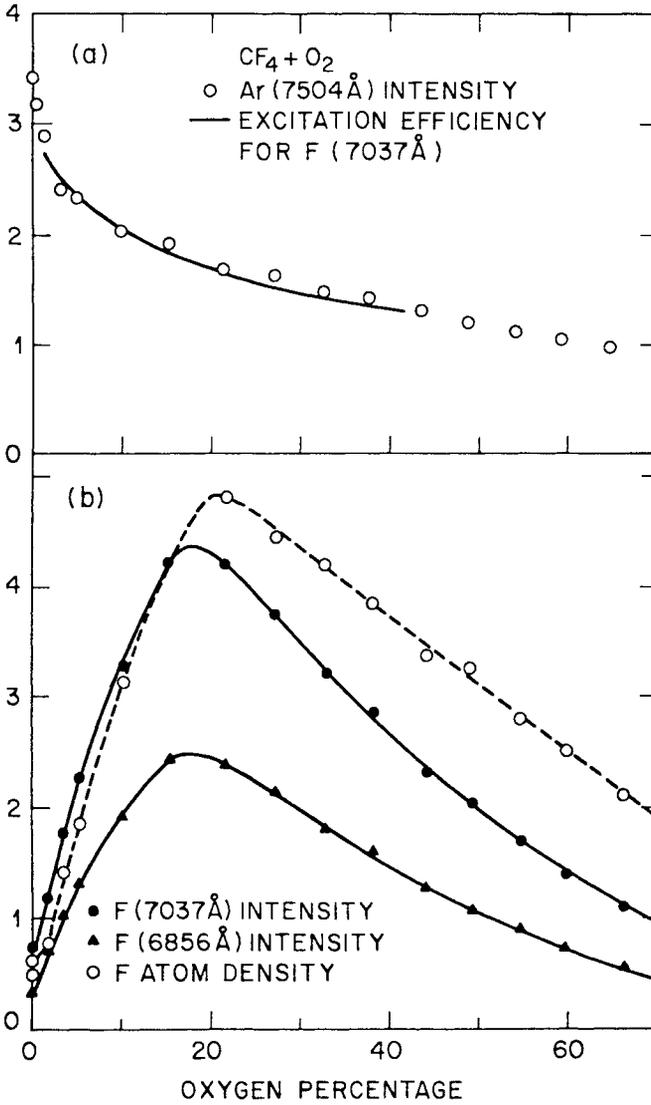
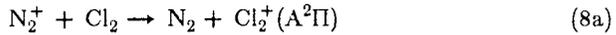
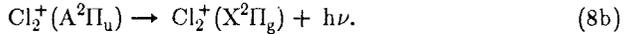


Figure 14: F and Ar emission vs. O₂ concentration in a discharge through CF₄/O₂/Ar (from References 30 and 63, reprinted with permission of the American Institute of Physics).

The importance of feedstock ionization potentials also shows up in the dynamics of ion production and loss.²² For example, small additions of O₂ to an Ar discharge cause dramatic changes in the ion composition.⁶⁸ Similarly, when Cl₂ is added to a low frequency discharge through N₂, the ion composition changes drastically. The concentration of Cl₂ in the feedstock need be only 10% for the N₂⁺ concentration to drop by an order of magnitude. The concentration of Cl₂⁺ is virtually the same in this mixed plasma as it is in a pure Cl₂ discharge.²² This effect is illustrated in Figure 15, where the time-dependent concentrations of N₂⁺ and Cl₂⁺ are plotted for various feedstock compositions. The extent of modulation in the ion concentration waveforms is a measure of the ground-state ion lifetime. For the N₂/Cl₂ mixture, the N₂⁺ lifetime is determined by the rate of charge exchange with Cl₂²² (Figure 15d). Since the ionization potentials for N₂ (15.58 eV) and Cl₂ (11.48 eV) are so different, the exchange is essentially irreversible. The charge exchange reaction must proceed rapidly for the steady-state ion concentration to be so greatly affected by small changes in feedstock composition. In this particular case, the reaction rate is probably enhanced by exchange through the excited state of Cl₂⁺,



followed by,



The excited state reaction (8a) is exothermic by 1.6 eV as opposed to the ground state reaction which is exothermic by 4.1 eV; thus, reaction (8a) is expected to lead to a resonant enhancement in the exchange rate. Another reason why the N₂⁺ density drops when Cl₂ is added to the feedstock is the electron cooling effect mentioned above. Since the electrons can interact with chlorine at a lower energy, fewer electrons will have a chance to reach the higher energies necessary for ionization of nitrogen.

An important point to keep in mind, and one which will be emphasized in Secs. 3 and 4 below, is that the effects of feedstock composition on discharge properties and chemistry may be totally out of proportion to the actual feedstock concentrations. Since ion and radical densities can be a very small fraction of the total gas density, small concentrations of feedstock additives, or impurities, may drastically alter these minority concentrations. Moreover, the role that these minority constituents play in the discharge chemistry can be substantially out of proportion to their concentration, particularly when it comes to surface processes.

2.1.4 Pressure, Flow-Rate, and Residence Time. Variation of pressure and flow-rate is commonly employed in the tailoring of a particular plasma process to a particular device application. Again, the effects of these parameters on the discharge can be roughly divided into physical and chemical effects with the caveat that the two are, strictly speaking, interrelated.

2.1.4.1 Chemical effects. The key parameter in affecting discharge chemistry is neither pressure, P, nor flow-rate, ϕ , but rather the residence

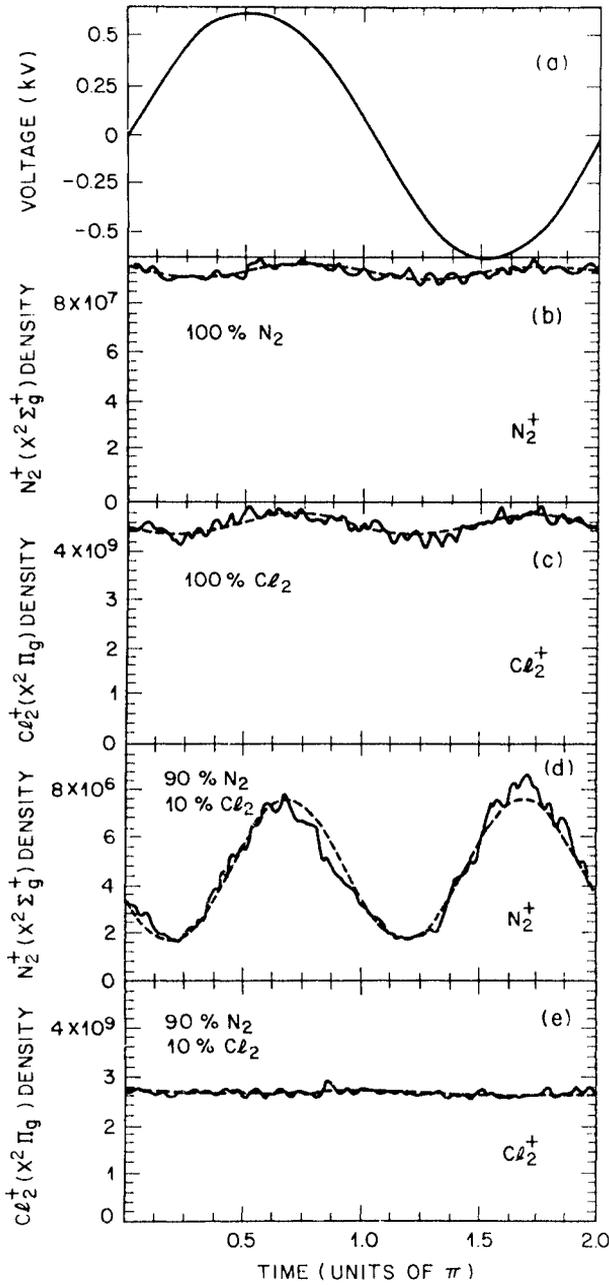


Figure 15: Time-resolved N_2^+ and Cl_2^+ concentrations measured by laser-induced fluorescence (from Reference 22).

time, which is simply proportional to P/ϕ . This can be seen very clearly in the mass spectrometric data of Truesdale et al.⁶⁹ which is reproduced here as Figure 16. They studied the plasma decomposition of C_2F_6 as a function of both pressure and flow-rate and examined the stable products downstream from the discharge with a quadrupole mass spectrometer. The final concentrations of C_2F_6 , CF_4 , and C_2F_4 vary both with flow-rate at constant pressure and with pressure at constant flow-rate. When the ratio of pressure to flow-rate is held constant, however, the final product concentrations also remain constant (Figure 16).

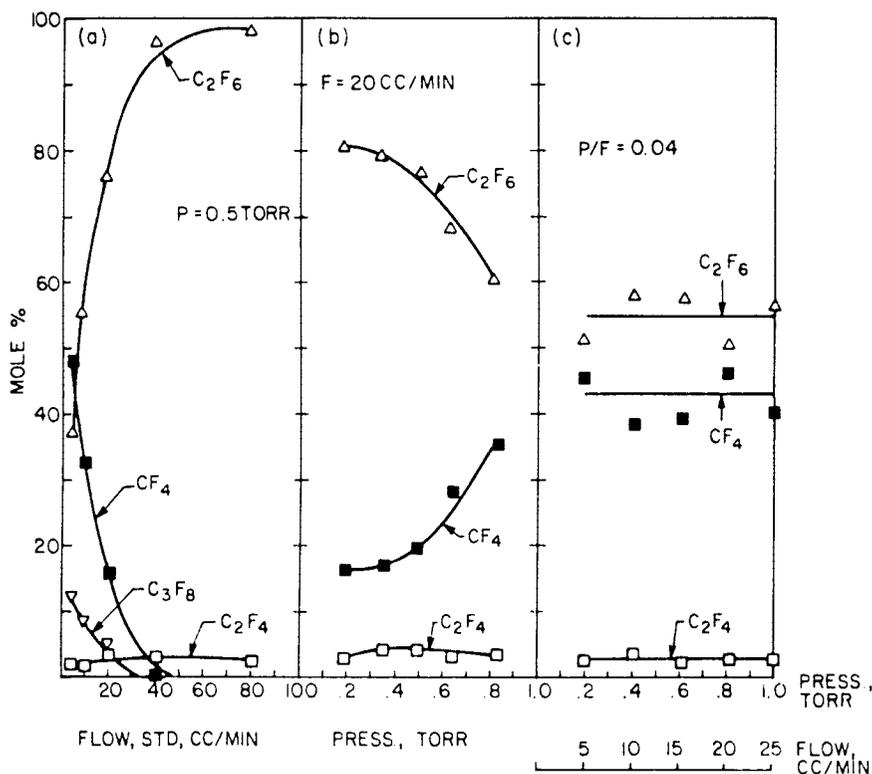


Figure 16: Composition of the effluent from a 50 W C_2F_6 discharge; (a) as a function of flow-rate at constant pressure; (b) as a function of pressure at constant flow-rate; and (c) as a function of pressure and flow-rate at constant residence time (from Reference 69, courtesy of G. Smolinsky).

The best way to vary residence time, i.e. with minimal effect on other plasma parameters, is to vary the flow-rate at constant pressure since variation of pressure can have pronounced effects on the discharge physical properties (see below). What happens to the gas-phase chemistry when the residence time is varied? If we consider the rate of decomposition

and radical production to be fixed for a given power density, then it is easy to see that the less time the feedstock gas spends in the plasma volume, the less extensive the degree of dissociation. Thus, at high flow-rates, or short residence times, the radical density will decrease with increasing flow-rate. At low flow-rates, or long residence times, the extent of dissociation may reach a limiting value and a dynamic equilibrium will be established such that the radical concentrations become flow-independent.

Let us consider the effects of flow-rate on heterogeneous chemistry when the heterogeneous rates are *not* rate-limiting: i.e. when the surface reaction probability is large. If the radical products of feedstock decomposition are the primary surface reactants, then the low flow-rate radical concentration may be negligible since the heterogeneous reaction acts as a radical sink and the overall reaction rate may be small because of the small reactant flux to the surface. At high flow rates, the feedstock may be insufficiently decomposed to provide radicals for the surface reaction and again the heterogeneous rate may be small. The overall dependence of the surface reaction rate on residence time, or flow-rate, will exhibit a maximum (Figure 17).^{19,70} The effect of flow-rate on heterogeneous chemistry when the reaction is surface rather than reactant-supply rate-limited will be discussed in Sec. 2.2.1.

2.1.4.2 *Physical effects.* The situation is very different when one considers the effects of pressure and flow-rate on the physical properties of the discharge and, in particular, the energy distributions of ions and electrons. The effects of flow-rate on ion and electron energy distributions are primarily an indirect consequence of the compositional changes discussed above. However, pressure affects these distributions directly

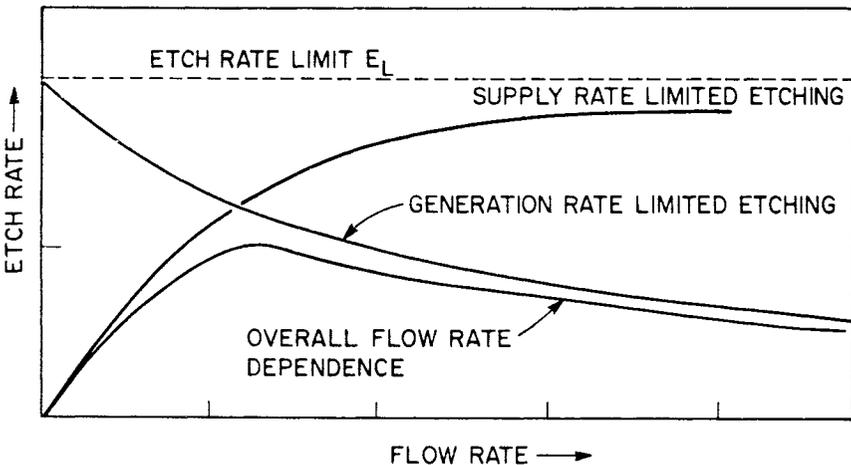


Figure 17: Generalized flow-rate dependence of etching rate for the case where the surface reaction rate is fast and the reactant is generated *in the plasma*. Two limiting cases are obvious: (1) rate is reactant supply limited and (2) rate is reactant generation limited (after Reference 70).

by affecting collision rates. For example, the energy with which an ion impacts a surface is not only dependent upon the sheath field, rf frequency, and ion mass but also the rate at which ions collide with neutrals as they traverse the sheath.^{19,56,71,72} If the collision mean free path is greater than the sheath thickness, then ions which enter the sheath from the plasma boundary will be accelerated by the full sheath potential and will impact the electrode with a narrow but highly energetic velocity distribution. Alternatively, if the collision mean free path is much smaller than the sheath thickness and if the ions lose all the energy gained from the field on each collision, then

$$E_i \approx F\lambda, \quad (9)$$

where E_i is the ion kinetic energy at the electrode and F is the sheath field, assumed to be constant with position over one mean free path, λ . The truth will generally lie somewhere between these two extremes.

Pressure also affects the electron energy distribution. Many electron-neutral collisions result in the formation of ion-electron pairs. If the electron-impact ionization mean free path is small compared to the sheath thickness, this can lead to substantial ionization in the sheath, which in turn will alter the ion energy distribution at the electrode surface and the sheath thickness.²⁰ In general, the overall charge density and plasma impedance can be expected to change with pressure in a complex fashion. Since the ion and electron collision mean free paths can be expected to scale inversely with pressure, Equation 9 suggests that the natural variable with which charged particle energy distributions and densities can be expected to scale is neither pressure nor sheath field but rather the ratio F/P . This has been long appreciated by scientists studying dc glow discharge physics. As we will see in Sec. 2.1.7, it is also a useful parameter in designing and understanding heterogeneous plasma-surface interactions.

2.1.5 Power Density. Many things can happen when the applied power density is varied. Generally, etching rates, radical densities, charge densities, and sheath fields increase initially and then saturate with increasing power. One reason why saturation occurs may be that the plasma volume often increases as the power is increased and electrons and ions acquire greater energy. This expansion may result in the discharge "finding" other grounds and discontinuities can result in measured plasma parameters. If the plasma volume can be maintained at a constant volume, e.g. by mechanical or magnetic⁷³ confinement, then an increase in power corresponds to an increase in power *density*. This in turn will result in higher electron energies, sheath potentials, and ion energies.

2.2 Plasma-Surface Chemistry

In order to understand the overall plasma-surface interaction it is necessary to understand the transport properties of the system with which we are dealing as well as the fundamental interactions between reactive adsorbate, surface, and product. In this section we first address the transport problem (Sec. 2.2.1) using recent results derived by Zarowin.^{28,58,74,75} In Sec. 2.2.2, we deal with the microscopic interactions

between reactive adsorbates and surfaces which lead to etching and deposition.

2.2.1 Chemical Vapor Transport. Plasma processing takes place at sufficiently high pressures that back reactions can be important. For example, in an etching reaction, the volatile product may redeposit either on the surface from which it desorbed, another part of the same wafer from which it desorbed (e.g. a side wall), or another surface all together (e.g. the counter electrode). For this reason it is necessary to consider the theory of chemical vapor transport in order to understand the etching and deposition of thin films in a plasma environment.

As long ago as 1926, the anomalous sputtering of certain materials was characterized in terms which we now refer to as chemical vapor transport^{58,75-77}; involatile cathode materials could be transported via the formation and subsequent decomposition of intermediate volatile compounds. For example, in the first experiments by Gunterschulze⁷⁶ As, Sb, and Bi were transported in a hydrogen discharge via the formation of the corresponding hydrides. Veprek and Marecek⁷⁷ showed that thin films of Ge and Si could be prepared by chemical vapor transport in a hydrogen plasma. Transport of the elements occurred from the cold zone where the hydrides are formed to the hot zone where they are decomposed. More recently, Zarowin has expanded these ideas to include the effects of geometry, flow-rate, and ion bombardment as well as surface temperature. We will summarize the salient features of his theory below; first, the nature of chemical vapor transport in the absence of a plasma or ion bombardment and at zero flow-rate is discussed in terms of an analogy to an electrical equivalent circuit.^{58,75} The circuit is useful only in that there are mathematical correspondences between the electrical circuit elements and elements of the heterogeneous chemistry, so that analysis of the electrical circuit provides "intuition" for the chemical vapor transport processes. In the following sections, the effects of flow-rate, loading (or geometry), and ion bombardment are considered in turn.

2.2.1.1 Equivalent circuit: Zero flow. Consider the generic reaction,



where S is the surface material to be transported, either etched, at rate α per unit area, or deposited, at rate β per unit area, A is the reactant which combines with S to form volatile compound B. The transport of these species can be neatly summarized in terms of an equivalent electrical circuit (Figure 18). Transport impedances, which can all be represented in terms of resistances, correspond to either heterogeneous reactions or diffusion. For simplicity, we consider only two surfaces, i and j, with reactive resistances,

$$R_r^i = \frac{1}{C^i \sigma^i} \quad (11)$$

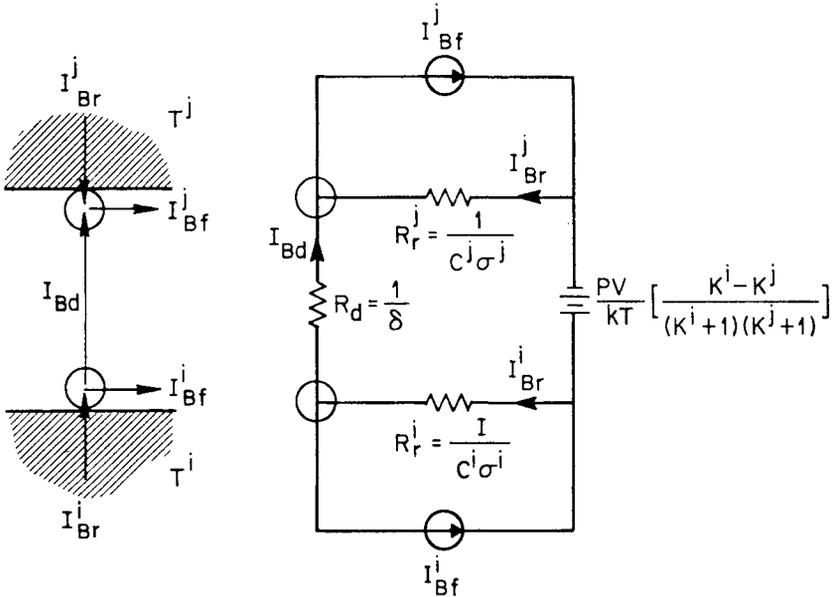


Figure 18: Equivalent circuit for chemical vapor transport. The transport of products from one surface to another is driven by an effective chemical potential difference between the surfaces which may result in turn from differences in surface temperature or ion bombardment. The effects of flow-rate are represented by current generators so that in the absence of an effective chemical potential difference, etching but not deposition may occur (after Reference 75). See text for definitions of circuit elements.

where C^i is the area of the i^{th} surface and $\sigma^i = \alpha^i + \beta^i$ is the reaction conductance. Slow reactions correspond to large resistances. The diffusive resistance between the two surfaces is given by,

$$R_d = \frac{h^2}{D}, \tag{12}$$

where $h = \Delta b / \nabla b$ is a measure of the separation between surfaces.⁷⁵ Slow diffusion corresponds to a large resistance. The *effective* chemical potential difference between surfaces i and j is given by,

$$\zeta = \frac{pV}{kT} \frac{(K^i - K^j)}{(K^i + 1)(K^j + 1)}, \tag{13}$$

where $K^{ij} = \alpha^{ij} / \beta^{ij}$ is the equilibrium constant at surface i, j . The term chemical potential difference is used somewhat loosely here. A finite value for zeta implies that the equilibrium states for the two surfaces will not be

the same (e.g. if they are maintained at different temperatures). Therefore, there will be a chemical driving force which produces transport of A and B between the surfaces. In terms of the product current,

$$I_B = \frac{\zeta}{R}, \quad (14)$$

where $R = R_r^i + R_r^j + R_d$ is the total circuit resistance. The reason for representing the surface reaction rates and diffusion rate as conductivities should now be apparent. At zero flow the product current, or etching rate, is limited by how fast etching occurs on one surface, deposition on the other, and diffusion of reactants and products in between. If the heterogeneous reaction rate happens to be limited by reactant generation from a homogeneous process, the product current will be limited instead by that process. This can be seen in a formal fashion by supposing there is a precursor to reactant A in Equation 10. In terms of the CVT equivalent circuit (Figure 19), this means that there will be an additional resistance in series with R_p . In terms of the heterogeneous rate constant, $\sigma_i^{-1} = \sigma_A^{-1} + \sigma_P^{-1}$ where σ_P is the sum of the forward and backward rates for the precursor reaction $P \leftrightarrow A$.⁷⁵

It is important to note that the above equation for the product current is valid, within the framework of the simple generic equation, for any departure from equilibrium. In equilibrium, $K^i = K^j$, there is no chemical potential difference and no net transport of A or B from one surface to another. If a chemical potential difference is maintained between surfaces i and j, for example by application of a temperature or electrical potential difference (see below and Reference 58), there will be net transport of A and B from one surface to another. Whether B is deposited on surface i and etched from surface j or vice versa depends on the sign of zeta, which in turn depends not only on any temperature differences between the surfaces but also on the reaction energetics. This can be seen most clearly by expressing the equilibrium rate constants in terms of the surface temperatures and the Gibbs reaction free energy, ΔG :

$$K^{i,j} = \exp[-\Delta G/kT^{i,j}]. \quad (15)$$

Thus, when ΔG is positive (endoergic), zeta is positive for $T^i - T^j > 0$ and the transport proceeds from the hot to the cold surface. If ΔG is negative (exoergic), zeta is negative for $T^i - T^j > 0$ and the transport proceeds from the cold to hot surface. Regardless of the reaction energetics, the direction of transport and the transport rates can be controlled by control of the differential surface temperature.

2.2.1.2 Flow effects. The effects of finite flow rate on the chemical vapor transport can be summarized by simply adding current generators to the equivalent circuit as shown in Figure 18. For small flow rates, the current generators provide an additional driving potential,⁷⁵

$$I_{B}^{i,j} = \frac{\zeta + (\phi R_d / R^{i,j})}{R} \tag{16}$$

where ϕ is the flow rate constant, i.e. the reciprocal of the reactor residence time. Now, *even in the absence of a chemical potential difference* between surfaces i and j , net transport of A and B can take place. Flow-rate can be used as a process control variable in a fashion which is beyond the simple variation of reactant concentrations by variation of residence time (Sec. 2.1.4).

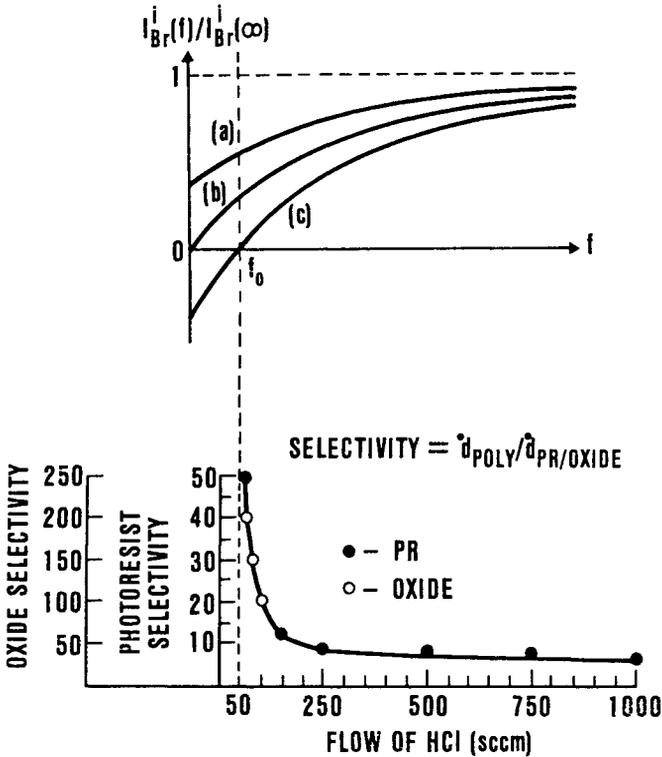


Figure 19: (a) Effect of flow-rate on transport rates showing competition between flow-rate driven currents and chemical potential driven currents (after Reference 75). The three curves label a,b,c correspond to chemical potential differences, ζ , greater than zero, equal to zero, and less than zero, respectively. (b) Example of effect of flow-rate on selectivity of polysilicon etching over photoresist and oxide (after Reference 75).

At higher flow rates, the flow current generator becomes nonlinear. By using Kirchoff's law, which says that the sum of the currents into any node must be zero (i.e. conservation of flux), Zarowin derived the complete flow dependence for the product current:

$$I_B^{i,j} = \frac{\zeta + \Delta\zeta}{R_T + \Delta R_T}, \quad (17)$$

$$\text{where } \Delta\zeta = \frac{pV}{kT} \frac{\alpha^{i,j} \phi [R_d + R^{i,j} (\phi R_d + 2)]}{\sigma^{i,j}}$$

$$\text{and } \Delta R = \phi R_d (R^i + R^j + \phi R^i R^j + 2 \frac{R^i R^j}{R_d})$$

Both the driving potential difference and the reactive impedance are modified by finite flow. At very large flow rates, the product current becomes independent of both the chemical potential difference and the flow rate and is limited only by the reactive resistance,

$$I_B(\infty) \rightarrow \frac{pV}{kT} \frac{\alpha^{i,j}}{R^{i,j} \sigma^{i,j}}. \quad (18)$$

Providing reactant production is fast, the rate is surface reaction limited.

The complete dependence of I_B on flow rate is shown in Figure 19a, where the product current normalized to its value at infinite flow rate, I_B / I_B^∞ is plotted against the reduced flow rate, $f = \phi R_d$. It is the flow-rate *relative* to the diffusive resistance which determines the overall transport rate. If the diffusion rate of products and reactants between surfaces is slow, R_d large, then the two surfaces behave somewhat independently and smaller flow-rates are required to reach the high flow-rate limiting transport rate.

As shown in Figure 19a, three situations arise because of the competition between the flow-rate current generators and the chemical potential difference between surfaces i and j (Figure 18 and Equation 17). Since $\Delta\zeta \geq 0$ in Equation 17, the flow-rate current generators can either work in concert with or in opposition to the chemical potential difference depending on the sign of ζ and which surface is considered. The three curves, a, b, c, in Figure 19a correspond to chemical potential differences greater than zero, zero, and less than zero, respectively. Note that for the case when $\zeta < 0$, the product current at surface i crosses zero at some finite flow rate, f_0 . This means that below f_0 deposition occurs while above f_0 etching occurs. At f_0 , no net transport occurs. If the chemical potential differences for different substrate materials have different signs, this competition can be used to great advantage in achieving large etching selectivities. For example, when etching polysilicon in the presence of photoresist and SiO_2 , Zarowin⁷⁵ reports selectivities of $>300:1$ (with respect to SiO_2) and $>50:1$ (with respect to photoresist) at f_0 (Figure 19). This behavior can be explained if the chemical potential difference, with respect to the counter conducting electrode, is greater than zero for polysilicon but less than zero for photoresist and SiO_2 . At flow-rates below f_0 , deposition of Si occurs on the *photoresist and oxide*. Since we have already concluded that $\zeta_{\text{Si}} - \zeta_{\text{resist,oxide}} > 0$ this means that the deposition of Si onto resist and oxide must be endoergic (see discussion after Equation 14 above). This example illustrates how we must consider chemical vapor transport between not only the two electrode surfaces but also the different parts of nominally the same surface.

2.2.1.3 *Loading.* A commonly observed phenomenon in plasma processing is that of reactor loading or the dependence of substrate transport rates on substrate area (for example see Reference 79). Under conditions where reactant generation is rate-limiting, transport between surfaces is unimportant, and flow-rate effects are negligible, Mogab⁸⁰ showed that loading can result from competition between reactant generation, homogeneous (i.e. gas-phase) reactant loss processes, and reactant loss by surface reaction. He showed that the reciprocal etching rate should be linearly related to the substrate surface area:

$$\left(\frac{dz^i}{dt}\right)^{-1} = \frac{N_0\rho C^i I_B}{M} = \frac{1}{\alpha^i \tau G} + \frac{\rho N_0 C^i}{MGV}, \quad (19a)$$

where dz^i/dt is the film thickness rate of change (i.e. the etching or deposition rate), τ is the gas-phase reactant lifetime, G is the gas-phase reactant generation rate, ρ is the substrate density, M is the substrate gram molecular weight, and N_0 is Avogadro's number. The linear relationship of Equation 19 has been observed under a wide range of conditions. This suggests that the assumptions inherent in Mogab's theory may be too restrictive. The chemical vapor transport theory described above allows us to examine loading effects when flow-rate effects cannot be ignored and when reactant generation is not rate limiting.

Under conditions where the substrate transport rate is not limited by reactant supply, the relationship between flow-rate, substrate surface area and the transport rate is contained in Equation 17. At zero flow,

$$\left(\frac{dz^i}{dt}\right)^{-1} = \frac{N_0\rho}{M} \zeta \left\{ \frac{1}{\sigma^i} + C^i (R_r^j + R_d) \right\}. \quad (19b)$$

Note that Equation 19b also exhibits a linear relationship between the reciprocal transport rate and substrate surface area. There is a direct correspondence between the parameters in Equations 19a and 19b:

$$\alpha^i \leftrightarrow \sigma^i \quad (20a)$$

$$G \leftrightarrow \frac{V\zeta}{(R_r^j + R_d)} \quad (20b)$$

$$\tau \leftrightarrow \frac{VM}{\rho} (R_r^j + R_d) \quad (20c)$$

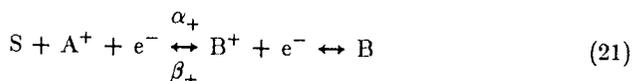
The correspondence between α and σ (Equation 20a) arises because Mogab only considers etching at surface i and assumes deposition on i or any other surface is negligible. When this is not so, α must be replaced by σ , the sum of the forward and backward reaction rate constants. In Equation 20b, we see that the volume generation rate in Mogab's theory, G , corresponds to generation by etching of "reactant" at surface j or product

transport current. In the case of zero flow, reactant A and product B are transported in opposite directions from one surface to the other when the chemical potential difference, ζ , is finite so that one surface acts as a reactant generator for the other surface. Similarly, the reactant loss time constant, τ , corresponds to the sum of the reactive resistance at surface j and the diffusive resistance (Equation 20c), both of which correspond to reactant loss mechanisms. The important point is that despite the radically different assumptions in the two theories, there is a correspondence of sources and sinks such that the overall functionality remains the same. Thus, observation of a linear loading effect is not sufficient to determine the rate-limiting process.

At higher flow-rates, the relationship between reciprocal transport rate and substrate area is no longer linear (see Equation 17). The loading effect is predicted to go through a maximum as the flow-rate is varied.⁷⁵ The sensitivity of a process to the number of wafers will depend on the flow-rate in a non-linear fashion. Although some of the aspects of this theory have been verified (e.g. that there is an effect of flow-rate on loading), more experimental data are needed to assess the range of validity of Equation 17. One difficulty in obtaining such data is the change in discharge composition which usually occurs when the residence time is changed. (see Sec. 2.1.4). This could result in a change in the rate-limiting step from heterogeneous to homogeneous. Another complication may arise when more than one reactant is important. This effect alone can give rise to a non-linear loading curve.⁸¹ If changes in flow-rate change the *relative* concentrations of these reactants further deviations from the simple theory above can be expected. To test the range of validity for the CVT equivalent circuit model, studies of spontaneous reactions, i.e. without a plasma, would be most appropriate.

2.2.2 Plasma Modified Chemical Vapor Transport. Until now we have considered chemical vapor transport without considering the effects of the plasma except to the extent that it modifies reactant concentrations and diffusion coefficients. The major influence of the plasma, however, is to modify the heterogeneous reaction rates by ion bombardment of the surfaces. Although electron enhancement of heterogeneous rates has been demonstrated^{82,83}, the sheath fields ordinarily are such as to repel electrons and negative ions from device surfaces (see Sec. 2.1.2). In this section we will see how this effect can be treated in a formal fashion within the framework of chemical vapor transport theory.^{28,58,74,75} The following section (2.2.3) will discuss the kinetic and microscopic origins of the heterogeneous chemistry and plasma modifications.

In addition to the neutral reaction, Equation 10, we consider a parallel ion-driven reaction:^{74,75}



This reaction is written in a formal fashion for simplicity. The ion need not correspond to the neutral moiety nor must the product be an ion initially. Electrons are included in Equation 21 merely to account for surface

neutralization of the incident ion as well as production of a neutral product. In fact, the following discussion is equally valid for energetic neutrals as well as ions. The key concept in Equation 21 is that *there are parallel paths for production of the final products*. Each path has a different rate constant and activation energy. In terms of the CVT equivalent circuit, the energetic ion (or atom) components correspond to resistors in *parallel* with the neutral (or chemical) reaction resistors (Figure 18). However, this analogy is of limited utility because the neutral component resistances in the presence of ion bombardment are generally different from the case where ion bombardment is absent. It is this change in the neutral component resistance which is formally responsible for synergistic effects.

From Equation 11, the total reaction rate will be given by the sum of the neutral and ionic components. If each rate is of the form⁷⁴

$$\alpha = N_A Z \exp(-E_A/kT) \quad (22a)$$

$$\beta = N_B Z \exp(-E_B/kT) \quad (22b)$$

$$\alpha_+ = N_{A^+} Z \exp(-E_{A^+}/kT) \quad (22c)$$

$$\beta_+ = N_{B^+} Z \exp(-E_{B^+}/kT) \quad (22d)$$

the total rates are given by

$$\alpha_t = \alpha[1 + f_A \exp(U_A/kT)] \quad (23a)$$

$$\beta_t = \beta[1 + f_B \exp(U_B/kT)]. \quad (23b)$$

where Z is a preexponential rate factor, $f_{A,B}$ is the degree of ionization of A, B and $U_{A,B} = E_{A,B} - E_{A^+,B^+}$ is the difference between the neutral and ionic activation energies.

Expressing the rate constants in this form allows us to explain a large number of temperature studies where Arrhenius behavior has been observed (Figure 20). For the case where the back reactions are negligible and reactant supply is constant,⁷⁴

$$\ln I_B = \frac{-E_A}{kT} + \ln [A] + \ln [1 + f_A \exp(U_A/kT)]. \quad (24)$$

For *neutral* dominated reactions, $f_A \exp(U_A/kT) \ll 1$ and the slope of I_B vs. $1/T$ gives the neutral activation energy, E_A . For *ion-dominated* reactions, $f_A \exp(U_A/kT) \gg 1$ and the slope gives $-E_{A^+}$. The fact that ion-dominated reactions give smaller slopes (see Figure 20) indicates that the ionic activation energy is smaller than the neutral activation energy. This is not surprising considering that the ions have been accelerated to relatively high energies by the sheath field. Another way to think of this activation energy difference is to think in terms of effective temperatures. Because the ion energy is superthermal, the effective temperature for the ion-surface interaction is much larger than for the neutral-surface interaction.

Thus, the same change in absolute surface temperature will correspond to a much smaller relative change for the ion-surface reaction than for the neutral surface reaction.

In general, the effect of ion bombardment is to change the surface reaction energetics.^{74,75} To see this, it is more convenient to view the

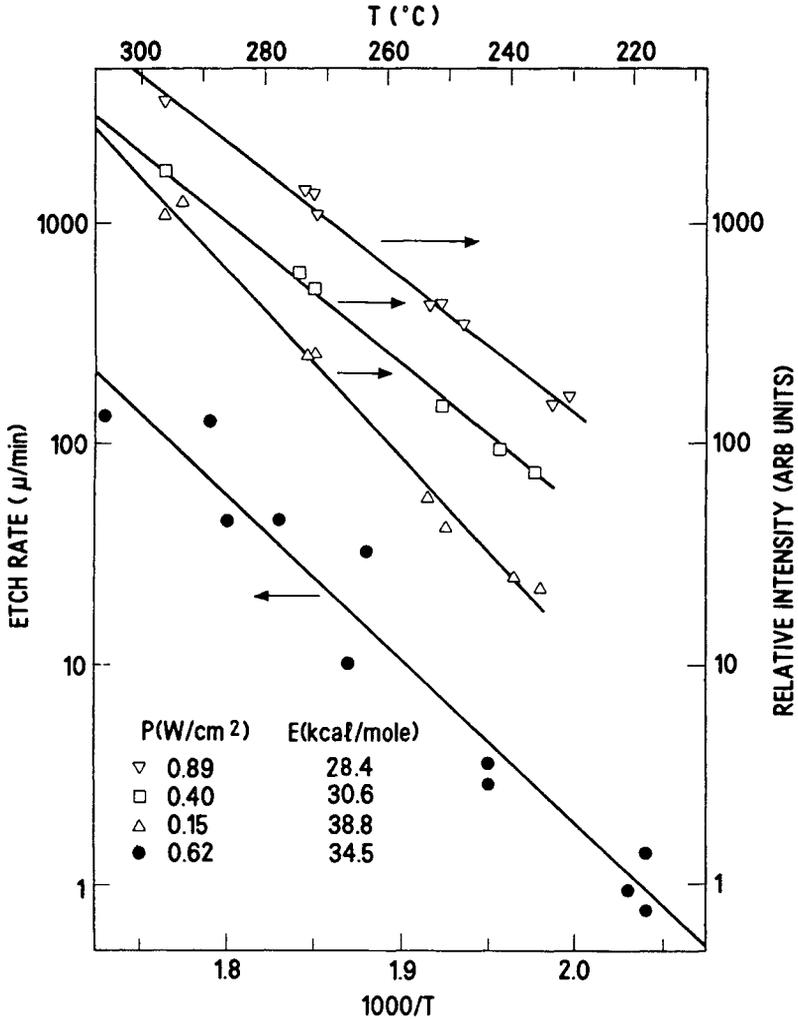


Figure 20: Arrhenius plots showing effects of temperature and power density on etch rates of InP and relative In atom emission intensity in an 0.3 Torr discharge through Cl₂. The emission intensities are proportional to etch rate. Note the smaller slopes for the higher power densities suggesting that ion-bombardment is reducing the overall activation energy (from Reference 181, courtesy of V.M. Donnelly).

parallel reaction (21) as modifying the chemical potential difference, or the degree to which the two surfaces differ from the same equilibrium state, rather than simply a parallel reactive resistance. Consider the equilibrium constants in Equation 13; the difference between K^i and K^j effectively determines the chemical potential differences which drive the transport processes. In the presence of parallel ionic reactions,

$$K_t = K \left\{ \frac{1 + f_A \exp(U_A/kT)}{1 + f_B \exp(U_B/kT)} \right\} \tag{25}$$

When the reaction is ion dominated,

$$K_t \rightarrow \exp \left\{ \frac{-[\Delta G - \Delta U - kT \ln(f_A/f_B)]}{kT} \right\}, \tag{26}$$

where K_t is the total (i.e. ionic and neutral) equilibrium rate constant, $\Delta U = U_A - U_B$. Thus, the effect of ion bombardment is a modification of the effective free energy. Alternatively, we can think of the surface temperature as being modified, $T^* = T/[1 - (kT \ln f_A/f_B + \Delta U)/\Delta G]$. If $\Delta U/kT > \ln(f_A/f_B)$, the effective temperature will be hotter than the actual surface temperature. In terms of the effective temperature, transport still goes from “hot” to “cold” for endoergic neutral reactions and from “cold” to “hot” for exoergic reactions. The difference is that “hot” and “cold” depend upon not only surface temperature but also ion energies, which can be controlled by frequency, dc bias, and pressure. Thus, we see the complementary nature of ion-enhanced chemistry and thermally enhanced chemistry. Of course, they are not the same thing. The product distributions are likely to be very different since the thermal energy deposition will be statistically distributed to the various degrees of freedom but the ionic translational energy may be disposed in very specific ways.

2.2.2.1 Anisotropy. Now that we formally understand the effects of ion bombardment on heterogeneous reaction rates we can understand how anisotropic patterning (see Sec. 3.4.2) is possible. In general we need to consider transport between not only the two electrode surfaces but among all surfaces. Specifically, when considering anisotropy, these surfaces reside on the same electrode but are mutually orthogonal. The surface which is perpendicular to the electric field lines will experience more energetic ion bombardment than the surface which is parallel to the electric field lines as long as the ion transport across the sheath is anisotropic. This will occur at higher values of F/P (see Sec. 2.1.4 and Figure 21).

The specific value of F/P which gives a particular anisotropy in the ion transport directionality will depend upon the gas composition and operating frequency (see Sec. 2.1). For example, the value of F/P , estimated from the square root of the rf power density,²⁸ required for a given degree of anisotropy is shown in Figure 21. For a pure Cl_2 discharge, where resonant charge exchange can be very effective in reducing the ion energy aniso-

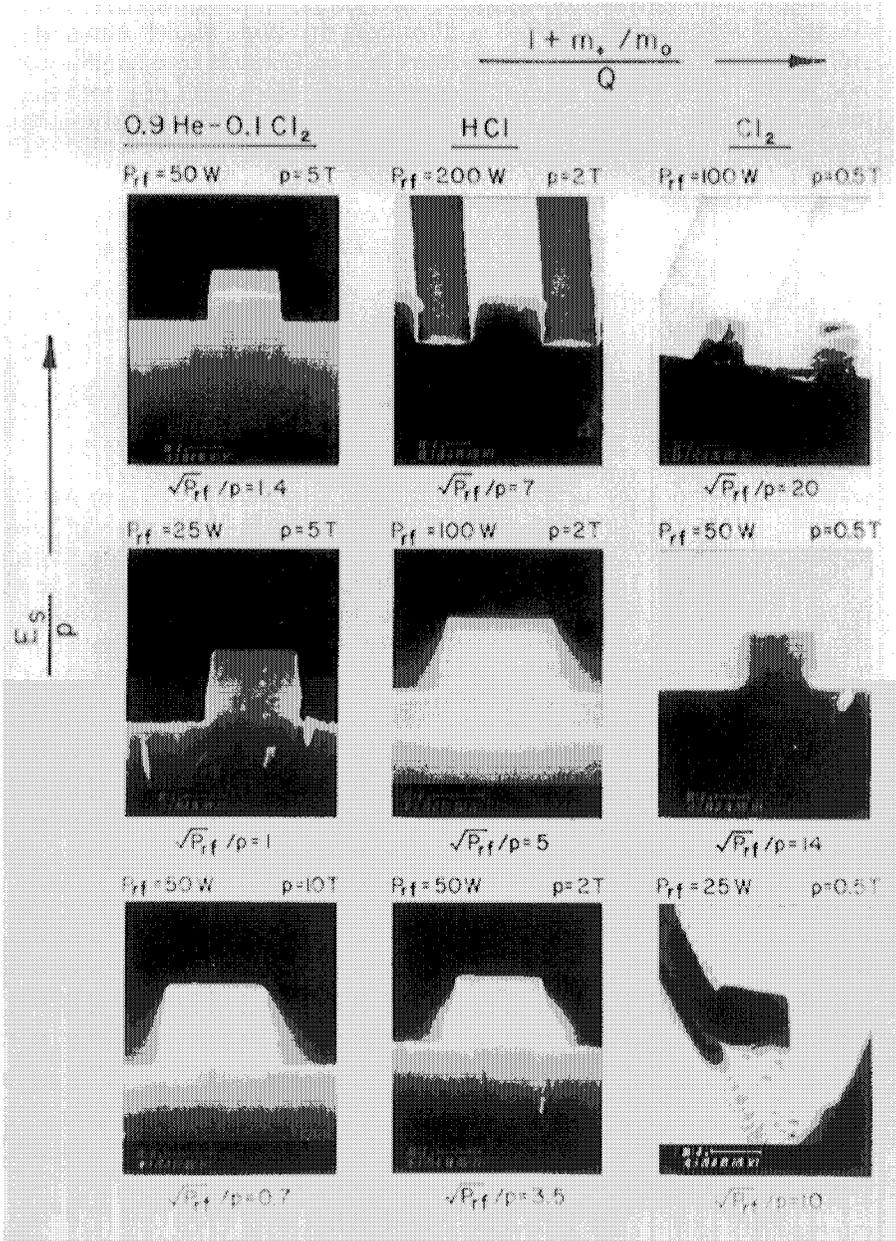


Figure 21: Anisotropy as a function of the ratio of the square root of rf power density to pressure (approximately equal to F/P) for three discharges illustrating the effect of gas composition on the degree of anisotropy in the ion energy distribution (from Reference 28, reprinted by permission of the publisher, The Electrochemical Society Inc.) and the value of F/P required for a given anisotropy.

tropy, the value of F/P required for a specific etch profile anisotropy is substantially greater than for a discharge through 10% Cl_2 in He, where the chlorine cations do not charge exchange effectively with He neutrals. The value of F/P required for the same degree of anisotropy in an HCl discharge is in between because only 50% of the neutrals, at most, will charge exchange efficiently with Cl^+ ions.

How the ion transport directionality is transferred to the surface, producing anisotropic features, will depend upon the relative neutral and ionic reaction rates, i.e. the neutral and ionic activation energies, and the ionic activation energy dependence on ion translational energy (Figures 21 and 22).^{28,74,75} For example, if the reaction is neutral dominated ($r_0 = 1$ in Figure 22), it does not matter how anisotropic the ion energy distribution is, the etched profile will be isotropic (assuming that there is not a crystalline orientation dependence to the neutral reaction rate).

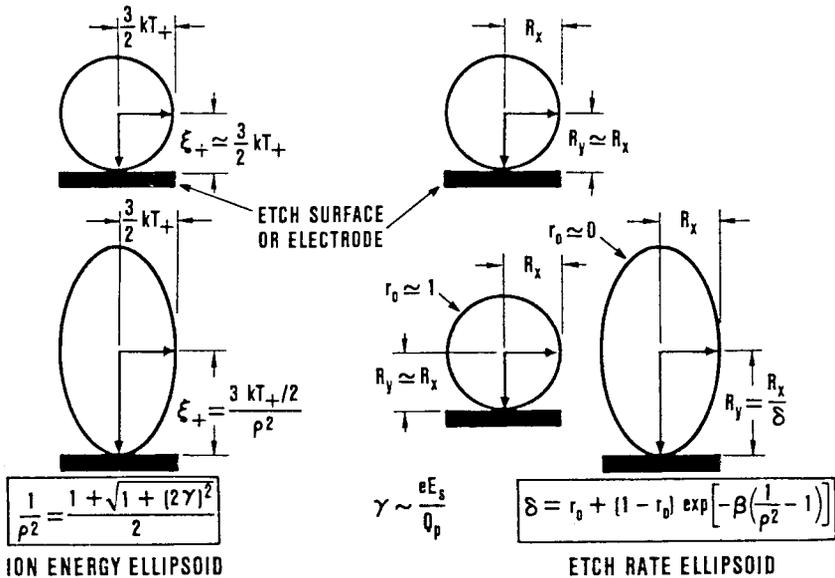


Figure 22: Schematic illustration of correspondence between ion energy distribution and etch rate profile. Anisotropy in the ion energy distribution is represented by an ellipsoid; the ratio of the minor to major ellipsoid axes is governed by the ratio of the random thermal energy, $3/2kT_+$, to the parameter ρ^2 , which is a function of F/P and the collision cross section. The degree to which anisotropy in the ion energy distribution is transferred into anisotropy in the etch rate depends upon the ratio of the neutral to the ionic reaction components. For example, two limiting cases are illustrated: when the neutral reaction is dominant, $r_0 = 1$, the etch rate ellipsoid is circular or isotropic; when the ionic chemistry is dominant, $r_0 = 0$, the etch rate ellipsoid is determined by the ion energy ellipsoid (after Reference 75).

2.2.3 Mechanisms. Until now we have considered transport phenomena in plasma processing but have said little about the microscopic mechanisms associated with heterogeneous chemistry. What are the reactants and products and how are reactants converted to products? A truly quantitative, microscopic mechanism should allow us to calculate the rate constants α and β as a function of temperature. Such detailed mechanisms are yet to be developed for surface reactions of interest in plasma processing. The most well studied system is the etching of Si by fluorinated and chlorinated compounds such as F, F₂, Cl₂, and XeF₂. First, we will consider the fluoride etching of Si in the absence of ion bombardment. When we consider the effects of ion bombardment, we will focus our attention on the etching of Si by chloride systems since chlorides do not etch Si spontaneously at room temperature. Therefore, this system will be uncomplicated by the presence of a spontaneous, neutral component and the ion component will be isolated.

2.2.3.1 Spontaneous etching of Si by fluorides. Many fluoride compounds etch Si spontaneously. The most well studied reactions to date, however, are those with F atoms, XeF₂, and F₂. Flamm et al. have studied these reactions at relatively high pressures as a function of temperature.⁸⁴⁻⁸⁷ F atoms are produced upstream from a Si surface by passing F₂ through an rf discharge at 14 MHz. The F atom concentration is measured above the Si substrate by titration with molecular chlorine.^{63,88} In this fashion, the etch rate dependence on F atom concentration can be determined. In a similar fashion, except without the discharge upstream, the Si etch rates with molecular F₂ and XeF₂ are also measured. The reaction with XeF₂ does not exhibit an Arrhenius temperature dependence over the range of temperatures and pressures studied (Figure 23) but rather shows a minimum because of competition between reaction and reactant desorption. At low temperatures, XeF₂ reactant adsorption occurs more readily but the reaction rate is slower because of the finite activation energy. Thus, as the temperature is increased, the etch rate first decreases as reactant desorbs and then increases as activation occurs (Figure 23).

What happens when fluorides come in contact with a Si substrate? The uptake of F₂ by a Si surface exposed to F₂ can be seen very clearly in the experiments performed by Winters et al.⁸ A thin film of Si was deposited onto a quartz crystal microbalance and then exposed to F₂. The change in the resonant frequency of the quartz crystal was used as a measure of the change in the mass of the film (Figure 24). At first there is a mass increase as the SiF_x surface species are formed. Continued exposure leads to etching and a subsequent mass decrease. Using Auger electron spectroscopy and x-ray photoelectron spectroscopy, Chuang⁸⁹ showed that both XeF₂ and SiF₄ dissociatively chemisorb on Si at room temperature to give an "SiF₂ like" surface. Note that the reaction of SiF₄ with the surface is equivalent to the back reaction in Equation 10 discussed above (Sec. 2.2.1). More recently, McFeely et al.^{90,91} have shown that not only SiF₂ species are present but also SiF_x, where x=1 and 3, when Si is exposed to XeF₂. Moreover, these authors conclude that SiF₂ is a relatively minor species compared to SiF₃ and SiF. However, these authors also report that the crystalline orientation affects the relative concentrations of the SiF_x

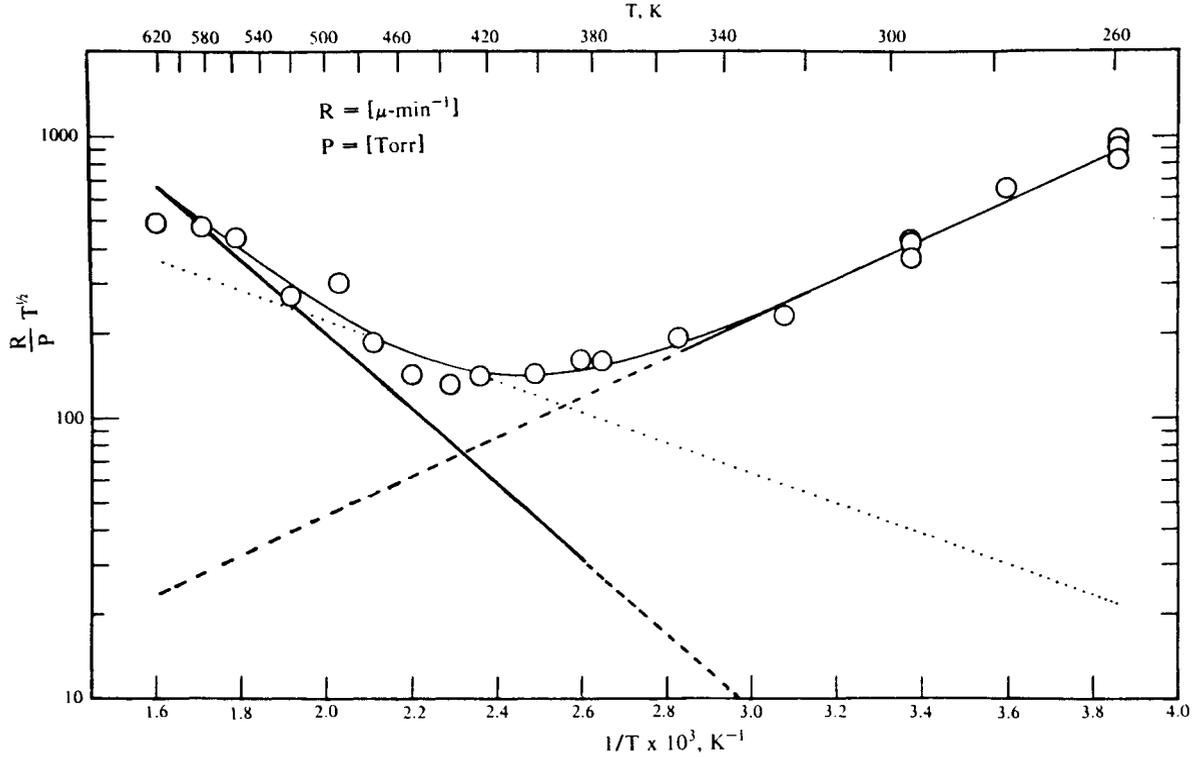


Figure 23: Etch rate of Si when exposed to XeF_2 as a function of temperature (solid curve and circles). The etch rate minimum is indicative of competition between thermal activation of the etch rate and thermal desorption of the reactant (from Reference 87, courtesy of V.M. Donnelly). The dotted line corresponds to the etch rate of Si when exposed to fluorine atoms.

species under low dosage conditions. At higher dosages when steady-state etching has been achieved, a substantial amount of fluorine appears

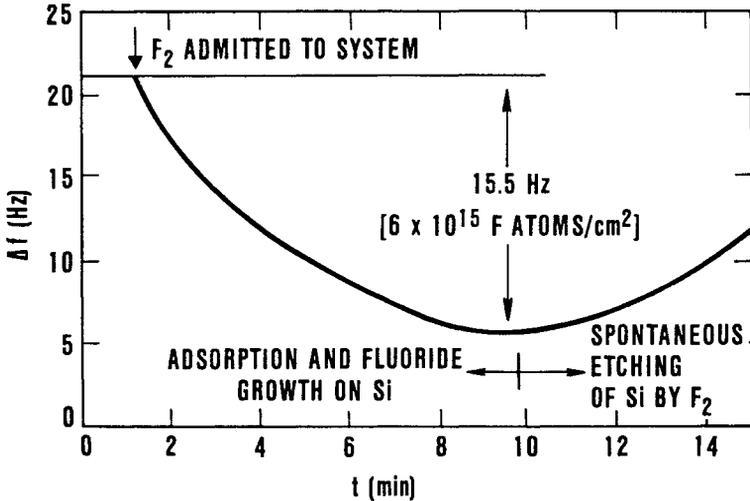


Figure 24: Quartz crystal microbalance response when Si is exposed to F_2 . The decrease in frequency is indicative of an increase in substrate mass (from Reference 8, reprinted with permission of the American Institute of Physics).

as SiF_4 which is most likely trapped beneath the surface.⁹² This observation suggests strongly that the spontaneous etching of Si by XeF_2 proceeds primarily by slow, sequential fluorination of Si followed by diffusion through the interface layer and desorption of product SiF_4 . While this mechanism would be consistent with the formation of SiF_4 , another mechanism may be active in producing other products.

What products are formed in the reaction of Si with fluorides? The only stable product which has been observed downstream has been SiF_4 but this does not mean that it is the only desorption product. Flamm et al. addressed this issue by examining the chemiluminescence which accompanies the etching of Si by fluorides.⁸⁴⁻⁸⁶ They attributed this emission to the *gas-phase* reaction sequence:



The dependence on temperature of the chemiluminescence intensity matched that of the Si etch rate suggesting that both reactions resulted from the rate-limiting reaction of F (or F_2) with a fluorinated Si surface to produce SiF_2 radicals:



It is not obvious from these studies whether or not SiF_2 is the dominant reaction product. Moreover, since Flamm et al. did not have an unambiguous means of measuring the SiF_2 radical (their conclusions rest upon their assignment of the diffuse chemiluminescent spectrum), it was not certain, initially, that SiF_2 was indeed a reaction product. However, in a molecular beam study of the reaction of Si with F atoms, Vasile and Stevie⁹³ measured directly the neutral species desorbing from the surface by mass spectrometry. By comparing the mass spectral cracking pattern of SiF_4 to that obtained from the products of the surface reaction, they concluded that 15-30% of the reaction products were radicals, presumably SiF_2 , and the rest were SiF_4 . More recently, Winters and Houle⁹⁴ measured in a similar fashion the neutral products from the reaction of XeF_2 with Si. They compared their results to not only the SiF_4 but also the SiF_2 cracking pattern and were able to determine more precisely the product branching ratios for this reaction. They concluded that the dominant product is SiF_4 (85%) with the remainder of the product distribution comprised of SiF_2 and SiF. These results are consistent with those of Vasile and Stevie. Any microscopic mechanism proposed for the etching of Si by fluorides must, at the very least, account for these various product yields.

Two reaction schemes have been proposed for the spontaneous etching of Si by fluorides. Flamm and Donnelly⁴ have proposed a concerted reaction model for the F atom etching of Si. In this model, the rate-limiting step is formation of an SiF_2 radical physisorbed on an “ SiF_2 like” surface. This radical can desorb directly to give the SiF_2 products observed or it may react rapidly to form volatile SiF_3 or SiF_4 . This model accounts for the identical temperature dependences for the Si etching rate and the SiF_3^* chemiluminescence. The reaction is said to be concerted because the arrival of reactant and formation of a new Si-F bond is coordinated with the departure of SiF_x products.

Another mechanism has been proposed recently by Winters et al.⁸ They noted that the reaction of F with Si should be similar to the reactions of O and O_2 with semiconductors and metals. The latter is believed to occur by the so-called Mott-Cabrera mechanism, where an oxide surface is formed rapidly upon exposure of the clean surface to an oxygen environment. Adsorption of O or O_2 onto the oxide layer is followed by formation of oxygen anions by tunneling of electrons from the metal substrate through the oxide overlayer. The electric field produced by the charge separation of anions and holes results in field-enhanced diffusion of the anions through the oxide layer and growth of the oxide layer at the metal-oxide interface. Winters et al. proposed that the F-Si interaction should proceed in an analogous fashion with the only difference being that the silicon fluorides are volatile so that once a new fluoride layer is formed below the initial fluoride layer, the initial layer desorbs or the underlying layer percolates out. This mechanism is consistent with the results of McFeely et al. mentioned above.⁹⁰⁻⁹²

It seems plausible that both concerted and Mott-Cabrera mechanisms may be important and that in both cases the rate limiting step giving rise to the observed activation energies might be formation of the surface fluorides. The lower, unsaturated fluorides might be formed by a concerted reaction whereas the fully saturated fluorides might result from field-assisted diffusion of F^- into the bulk.

2.2.3.2 Ion assisted etching. What happens when energetic ions are simultaneously impinging upon the surface? Coburn and Winters⁸² first demonstrated the synergistic nature of ion-neutral surface reactions by exposing Si, deposited onto a quartz crystal microbalance, to a XeF_2 neutral beam and an Ar^+ ion beam; the etch rate with both beams was greater than the sum of the individual ion sputtering and spontaneous chemical rates.

Since the pioneering experiments of Coburn and Winters, there have been many ion beam/neutral beam studies of etching reactions.^{8,53-55,94-123} Considerably more detailed data have been obtained on ion-assisted etching mechanisms from these experiments than from corresponding experiments on just the neutral reactions. A typical experimental configuration is shown in Figure 25. An ion beam of variable energy and flux is directed to a solid substrate which is simultaneously subjected to a neutral flux, either from a molecular beam or from a small background pressure. The ion sources are usually differentially pumped to prevent back diffusion of the radical into the ion source. The neutral flux is varied by either varying the background pressure or by varying the intensity of the beam source. Some workers have used substrates deposited onto quartz crystal microbalances in order to measure etching and deposition rates directly by measuring the change in the resonant crystal frequency.^{55,82} When the ion flux is measured, usually with a Faraday cup, etch rates can be converted directly into substrate removal yield per incident ion. Yields as high as 25 have been observed in the Ar^+ ion-assisted etching of Si by XeF_2 (see Figure 26).

Several studies have been done where the nascent neutral products are detected using a mass spectrometer housed in a differentially pumped chamber but with a direct line-of-sight to the reacting surface.^{53-55,94,107,114-116,121,122} Mayer and coworkers^{109,112} have used Auger electron spectroscopy to monitor in real-time the instantaneous reactant coverage in the etching of Si by Cl_2/Ar^+ . Some studies have also been done where the secondary ions emitted, e.g. $SiCl^+$ and $SiCl_2^+$, are detected and energy analyzed.^{119,120} The two systems which have been studied most extensively are the ion-assisted reactions of silicon by fluorides and chlorides, although there has been some work also on III-V^{95,99,101,106,111,122} and SiO_2 chemistry.^{96,100} We will focus primarily on the reactions between Si and Cl_2 since there have been more detailed studies of this system than any other and the reaction is ion *induced* at room temperature. We will also discuss briefly the findings in the fluoride experiments since the neutral chemistry is best characterized there (see above).

Although many measurements have been made on the $Si/Cl_2/ion$ system, the results obtained often appear to be contradictory. The reasons

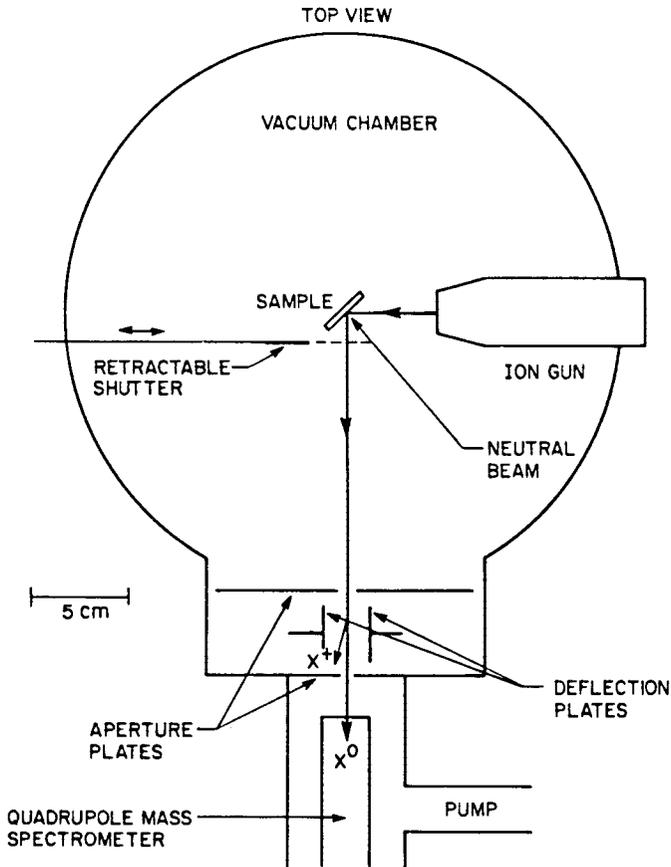


Figure 25: Typical ion-beam apparatus used in ion-enhanced etching experiments. Primary and secondary ions are deflected away from the mass spectrometer so that only neutrals are detected (from Reference 121, courtesy of S.C. McNevin).

for this are not entirely clear but one problem is certainly the wide variety of experimental conditions that have been employed (Table 2). Another possible source of difficulty is the reliability of some of the numbers summarized in Table 2. In particular, the neutral fluxes reported in the neutral *beam* experiments are in every case estimated by assuming certain characteristics of the expansion. In only one case is the neutral flux calculated in this way compared to other independent measurements.¹¹² With these caveats in mind, let us consider what these experiments tell us about the mechanisms which are important in ion-assisted etching.

First of all, it seems clear that under certain circumstances the mechanism for ion enhancement differs fundamentally from simple physical sputtering. The clearest evidence for this comes from the work by Mayer et

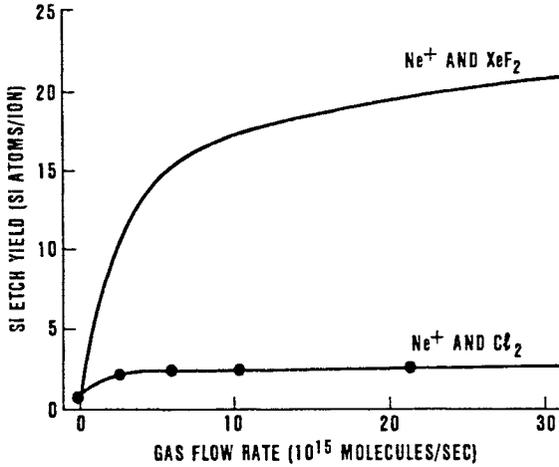


Figure 26: Si etch yields vs. XeF₂ and Cl₂ neutral flux when a 1 kV Ne⁺ ion beam is simultaneously incident upon the surface (from Reference 97).

Table 2: Ion Beam/Cl₂ Etching of Si

Reference	Cl ₂ Flux (cm ⁻² s ⁻¹)	Ion Flux (cm ⁻² s ⁻¹)	Ion Energy (keV)	Area (cm ²)	θ _{inc} (°)	Substrate
97, 98	0–6 × 10 ¹⁶	3–6 × 10 ¹³	1	0.5	0	Thermally Evaporated Si (100)
104	5 × 10 ¹⁶	1.7–6.6 × 10 ¹⁵	0–0.4	NG	0-75	Single crystals Evaporated Si (100)
54, 55	5 × 10 ¹⁶	1.5 × 10 ¹³	0.4-0.9	Broad	> 0	
109, 112	0–6 × 10 ¹⁶	0.06–3 × 10 ¹⁶	0.4-0.9	2.8	0-60	CVD undoped poly-Si NG
53, 114, 116	0.5–10 × 10 ¹⁶	2 × 10 ¹⁴	1	0.15	60	
119, 120, 123	0–2.4 × 10 ¹⁴	1 × 10 ¹³	1	0.06	0-60	Si (100)
121	0.02–4 × 10 ¹⁴	0–4.5 × 10 ¹⁴	0.3-3	0.05	45	p-type Si (100)

al.^{109,112} and Okano and Horiike¹⁰⁴ who measured the dependence of the Si etch yield on the ion beam angle of incidence, θ_i. Mayer et al.'s results are reproduced here in Figure 27. For the case where no neutral reactant is present and only a rare gas ion is incident upon the surface, the yield peaks near 60° as expected for a purely physical sputtering process (see Reference 1 and references therein). However, when Cl₂ is admitted to the system, the etch yield is observed by both groups to peak at normal incidence (0°). At this angle the ion energy tends to be dissipated into the bulk via a collision cascade or spike.¹²⁴ After correcting the angular dependence for the physical sputtering component (determined by using the Ar⁺ beam alone), Okano and Horiike¹⁰⁴ found the etch yield to follow a cos θ_i dependence, which they attributed to an ion-induced chemical

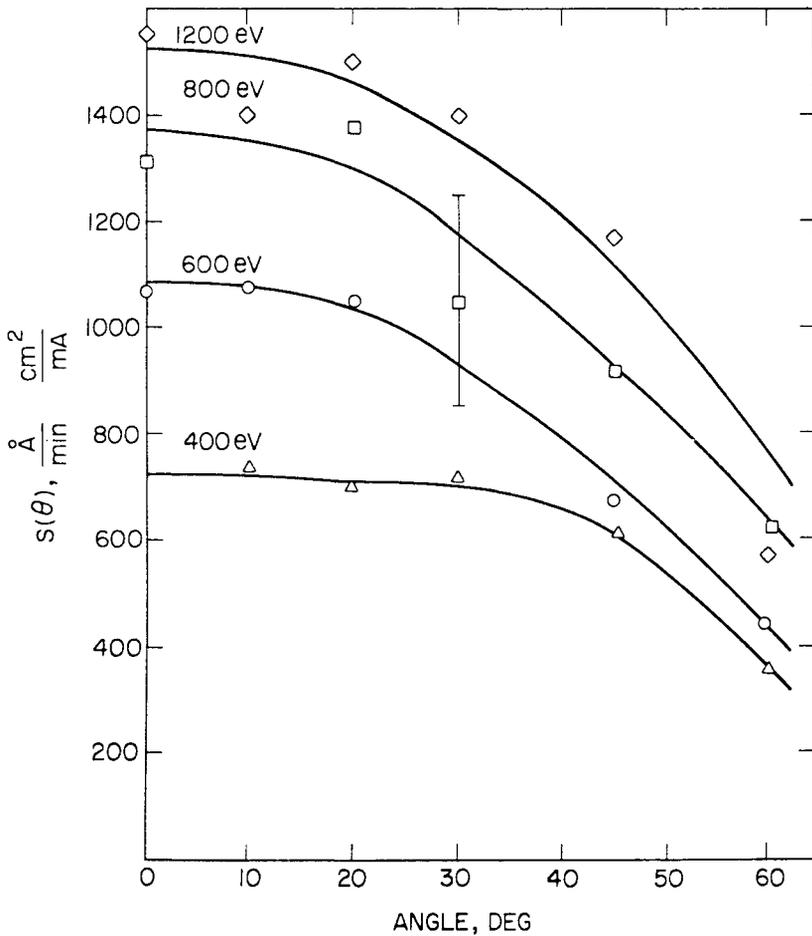


Figure 27: Si etch yield vs. Ar^+ ion incident angle and ion energy when Si is exposed simultaneously to neutral Cl_2 (from Reference 109, reprinted with permission of the American Institute of Physics).

reaction or chemical sputtering.¹⁰⁵ A distinction is made between chemical sputtering and spontaneous etching. Cl_2 does *not* etch Si at room temperature (Reference 53 and references therein) so the reaction is not ion-assisted as in the fluorine case but rather ion *induced*. The total etch yield is a superposition of the physical and chemical sputtering components.

A useful means for gaining insight into the nature of ion-induced etching is to measure the energy distributions and identities of desorbing products. Smith et al.^{54,55} measured products mass spectrometrically from the $\text{Si}/\text{Cl}_2/\text{ion}$ reaction. From the dependence of cracking patterns on ionizing electron energy, they deduced that some of the desorbing species are radical, unsaturated chlorides, probably SiCl_2 . However, they were not

able to specify how much of the product yield is comprised of these radicals. Kolfshoten et al.¹¹⁴ find that most of the products in their experiments are radicals, SiCl_2 and SiCl . By comparing the cracking patterns of SiCl_4 with those obtained in their ion beam experiments they find that only $\approx 1\%$ of the products are SiCl_4 . McNevin and Becker¹²¹, on the other hand, found virtually no evidence for radical products and by similar reasoning to that employed by Kolfshoten et al. deduced that the primary product is the tetrachloride.

Both Kolfshoten et al. and McNevin and Becker measured product time-of-flight distributions. Assuming there are no delays due to surface residence times, these distributions can be converted directly into product kinetic energy distributions. Kolfshoten et al. did precisely that and found that both the mono- and di-halide products showed a bimodal energy distribution (Figure 28). The low energy component could be explained by a Maxwellian distribution at room temperature; the high energy component could be explained approximately using the collision cascade theory of Sigmund for physical sputtering processes,¹²⁴ although a high temperature Maxwellian provides a better fit to the data. They conclude that Si is primarily *physically* sputtered by bombardment with Ar^+ despite the presence of a simultaneous flux of reactive Cl_2 . It should be noted that Kolfshoten's experiments are biased toward detection of the physical sputtering component since their angle of incidence is 60° . Although McNevin and Becker's time-of-flight distribution differs from a room-temperature Maxwellian, to account for the deviation one would have to use a sub-thermal distribution. Alternatively, a distribution of surface delay times might be important in their experiments: the surface exposed to the ion beam in these experiments was only a small fraction of the total area so that surface diffusion of adsorbed chlorine from adjacent non-bombarded regions may have been important. Since ion bombardment will desorb Cl ,¹¹² large area exposure, such as used by Kolfshoten et al. and Smith et al., might lead to a chlorine deficient surface, production of unsaturated SiCl_x radicals, and physical sputtering. Small area exposure, such as used by McNevin and Becker, might lead to the production of surface delays, saturated SiCl_4 , and thermal desorption. Since McNevin and Becker's experiments were done at $\theta_i=45^\circ$, the primary difference between their work and the work of Kolfshoten et al. and Smith et al. appears to be the beam area. An experiment where the ion-bombarded area is varied relative to the total surface area exposed to chlorine would be very useful in resolving these discrepancies.

The effect of the neutral flux on the etch yield per ion has been studied by several groups.^{97,98,109,112,121} In this case there is general agreement and the data shown in Figure 26 are typical. Gerlach-Meyer⁹⁸ interpreted these data in terms of production of an excited surface species or reactive "site" with finite lifetime. At low flux, the yield increases linearly¹²¹ because the reactive "sites" created by ion bombardment are greater in number than the available reactants and so the reactant arrival rate becomes rate limiting. At higher flux, the yield saturates because the neutral reactant is in excess relative to the rate of "site" production. Mayer et al.¹¹² showed, however, that the flux dependence observed could be accounted for

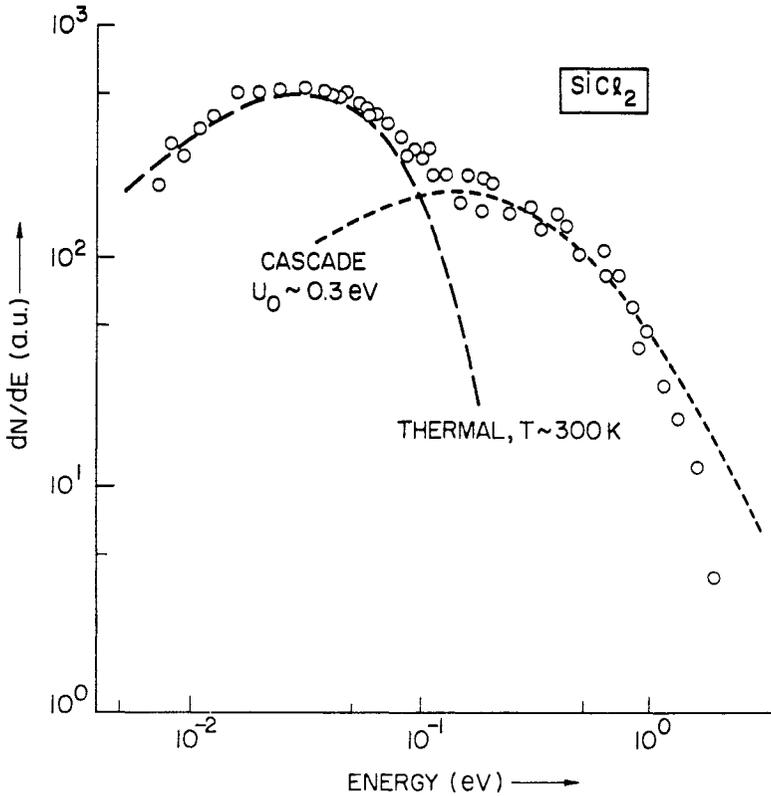


Figure 28: SiCl₂ product translational energy distribution, dN/dE, from the Ar⁺ ion induced reaction of chlorine with silicon (from Reference 114, reprinted with permission of the American Institute of Physics).

entirely by considering the competition between ion-induced desorption of surface chlorine and sticking of chlorine from the incident neutral beam. The existence of a surface excited state was not necessary to explain the neutral flux dependence. Mayer's model is consistent with etch yields measured over a wide range of ion flux as well as neutral flux. The critical parameter is the ratio of the two, I/P. The Si etch yield was found to have the following dependence upon coverage and ion current:

$$R = m\beta\theta I + \delta(1 - \theta)I, \tag{29}$$

where R is the Si etch yield, m is determined by the molecularity of the reaction (¼ for formation of SiCl₄), β is the fraction of Si leaving the surface as a chloride, δ is the fraction of unchlorinated Si sputtered directly by the ion, I is the ion flux, and θ is the chlorine surface coverage given by,

$$\theta = \frac{1}{1 + \Phi I/P}, \tag{30}$$

where Φ is the ratio of desorption to adsorption coefficients and P is the neutral flux to the surface. At high ion to neutral flux ratios, θ goes to zero according to Equation 30. This is consistent with the chlorine surface coverage as measured using Auger electron spectroscopy (Figure 29). From Equation 29, we see that with low surface coverage ($I\Phi/P \gg 1$), the etch yield reduces to the sputtering yield ($R - \delta l$) as observed (Figure 26). At low ion to neutral flux ratios, the surface concentration of chlorine saturates near unity (see Figure 29 and Equation 30), and the etch yield saturates at a value dictated by the ion flux and chemical sputtering coefficient, β (Figure 26), $R \rightarrow m\beta l$.

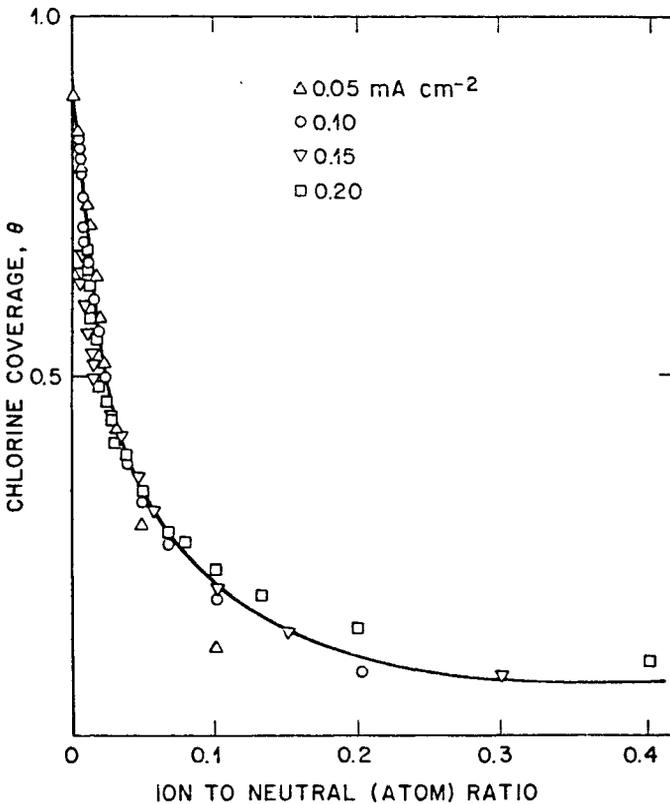


Figure 29: Surface coverage of chlorine on Si as a function of the relative fluxes of ions and neutrals to the surface. The measurements were made by Auger electron spectroscopy while the sample was simultaneously exposed to an Ar^+ ion beam and a molecular Cl_2 beam (from References 109 and 112, reprinted with permission of the American Institute of Physics).

So we see that the ion-induced reaction of Si with Cl_2 can range from physical sputtering to chemical sputtering depending on the relative fluxes of ions and neutrals, or more precisely, the surface concentration of chlorine, and the angle of incidence, energy, current density, and mass of the ion beam. The reactant surface coverage may affect not only the overall rate but also the product distributions. Chemical sputtering is favored at normal incidence where the ions are most effective in disrupting the surface structure and driving chlorine into the bulk by recoil implantation. One part of the mechanism would seem to be enhanced *production* of chlorides by ion bombardment. These volatile chlorides may then spontaneously desorb but ion bombardment could assist in this process also.

Several explanations for the enhancement of heterogeneous reaction rates by ion bombardment have been offered in the literature:^{4,8,98} (1) surface damage; (2) desorption or clearing; and (3) surface excitation. It seems likely that all of these explanations have some validity and the dominant mechanism, if there is one, will depend very much on substrate and gas as well as operating conditions such as the ratio of neutral to ion flux. For example, in the ion-induced reaction of silicon with chlorine most of the evidence above points to ion-assisted desorption and ion-enhanced product formation. The production of unsaturated chloride radicals with an energetic translational energy distribution is suggestive of sputter desorption; enhanced reaction at normal incidence suggests ion-induced damage, ion-induced surface excitation, and/or ion-assisted radical formation on the surface. In the case of silicon etching by fluorine, on the other hand, the ions probably help to desorb subsurface SiF_4 as well as sputter the partially fluorinated species residing on the surface. But ion-induced surface damage is probably not an important mechanism because the spontaneous reaction itself causes substantial surface damage. In a plasma, where deposition of etch inhibiting material may also be taking place, ion bombardment could assist in the desorption of this material as well (see Sec. 3.4.2).

2.3 Modeling

We have discussed plasmas and plasma-surface interactions but it remains to put everything together in order to describe the system as a whole. The interaction between various parameters, which will be further discussed as it pertains to particular applications below, makes the task of a global model extremely difficult. The approach taken recently by Edelson and Flamm¹²⁵ and by Kushner¹²⁶ is to select the most important homogeneous reactions and solve the corresponding set of differential equations. In many cases, the rate constants are unknown and approximations must be made. In particular, the dependence of a rate constant on temperature and/or molecular internal energy distributions is unavailable. As if this alone were not problem enough, no model to date has dealt in a self-consistent fashion with the spatial inhomogeneities in concentrations and energy distributions for both neutral and charged species let alone temporal variations in these quantities.

Another approach to modeling plasma chemistry is to perform computational “experiments.” For example, the energy distributions of ions and electrons can be calculated by the Monte Carlo method if collision and reaction rate constants are known along with the local electric field and any concentration gradients (for example, see References 127 and 128). This method has the advantage of simplicity and speed but suffers from being inherently non-self-consistent. However, it could be very useful in providing insights into how energy distributions and particle flux can be modified by various plasma parameters.

Much more theoretical and diagnostic development needs to be done before we can hope to have even a semiempirical model of some predictive value. One approach along these lines has already been reported by Mocella et al.¹²⁹ They employed the Response Surface Method. In this method, a kinetic model relating plasma operating parameters to some desired outcome, such as etch or deposition rate, is assumed. Given the form of this model, a minimum number of experiments is determined, in a statistical fashion, which will optimally determine the model parameters and, therefore, define the multi-dimensional parameter space. Once this space is so defined, the effects of any choice of operating parameters can be determined. Of course, the problem in this procedure is knowing whether or not the model which one has adapted is sufficiently accurate to provide predictive power.

3. PLASMA ETCHING

3.1 Introduction

3.1.1 Outline. This section deals with how plasmas are used in etching processes for microelectronic device fabrication. The method of patterning, using lithography and etching will be outlined. Etching process parameters and goals will be defined along with methods of affecting these parameters. Conventional commercial etcher designs are presented and examples of specific applications to Si and III-V devices are given.

3.1.2 Pattern Definition and Transfer. The formation of a layer of material which is only present in predefined areas can be accomplished by two techniques, lift off and etching. Lift off is accomplished by first defining a polymer lithographic mask which is the negative of the desired pattern (see for example References 130 and 131). Figure 30a illustrates the process. The mask is commonly referred to as photoresist, if optical lithography is used to define the polymer. Just “resist” is a more general term which encompasses polymers defined by any exposure tool including x-rays or electron beams. The layer is then deposited and the wafer is placed in a solvent which dissolves the resist, removing any part of the layer on top of the resist. Lift off can only be used when the layer deposition process does not exceed the temperature at which the resist begins to degrade, 200-300°C. This section will not deal at all with lift off techniques but will concentrate on plasma etching.

Etching requires that the layer be first deposited and a positive resist mask be defined on top. The wafer is then exposed to an environment

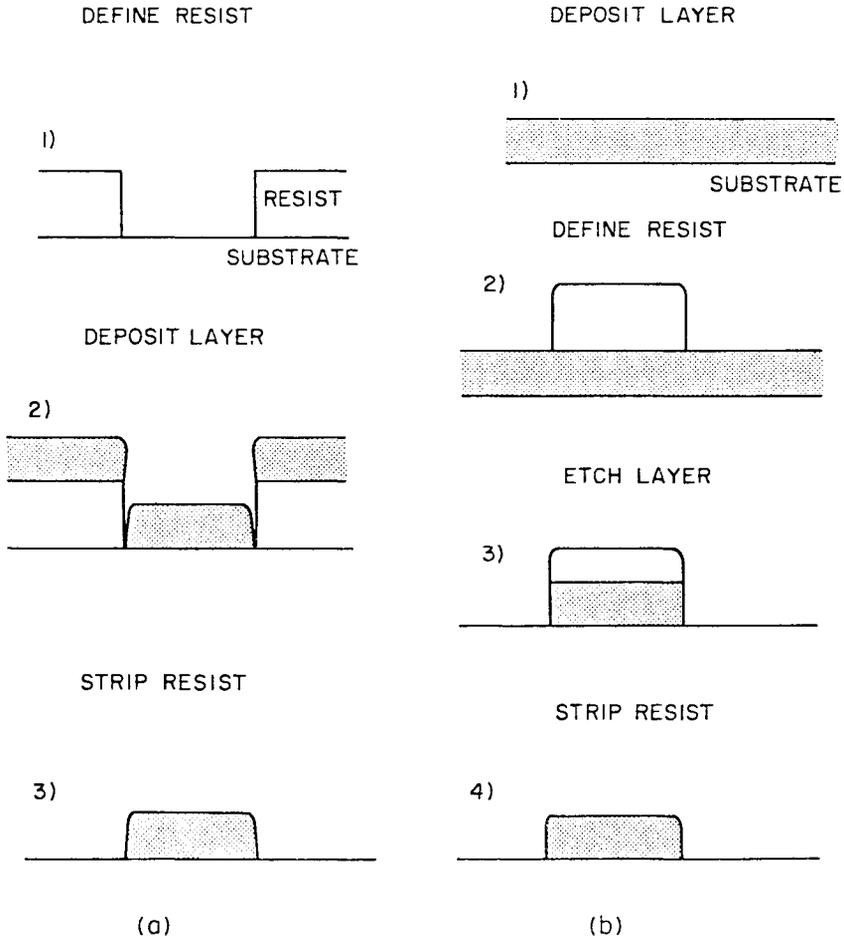


Figure 30: (a) This sequence illustrates pattern transfer using the lift off technique. (b) This sequence illustrates pattern transfer by etching.

which removes the layer where it is not covered by the masking resist. This process is illustrated in Figure 30b. This figure shows schematically that during the etching process the mask may also be partially removed.

In the following sections we will discuss chemically active plasmas, however, etching also occurs on exposure to a chemically inactive plasma where an inert feed gas such as argon is used. Here material erosion or sputtering¹³² is caused by bombardment with positive ions accelerated across the sheath. A small sputtering component is present in all plasma etching processes. The etching process known as ion beam milling¹³³ removes material by sputtering but the ionization and acceleration of the beam are performed separately, away from the etching material or target as it is called.

3.1.3 An Illustration of Plasma Etch Patterning. Exactly how plasma etching techniques are used in I.C. fabrication is illustrated in Figure 31. This cross sectional view shows part of a planar MOS silicon transistor whose formation required four etching steps which are schematically shown in Figure 32.

First, gate and source and drain (GASAD) “windows” are etched down through an isolating field oxide (FOX). Figure 31 only shows a portion of a GASAD region so no FOX is visible. A thin gate oxide (GOX) is then grown in the GASAD region and the gate material is deposited over the whole wafer. In this case the gate is low pressure chemical vapor deposited polycrystalline silicon (LPCVD poly-Si). The gate area is lithographically masked and any poly-Si outside this area is etched away. Phosphorus doped LPCVD SiO_2 , used as another dielectric isolation, is deposited and then smaller contact windows, or vias, are etched down to the transistor gate, source and drain. Only the contact to the source is shown in Figure 31. A composite conducting structure of doped LPCVD poly-Si and aluminum is then deposited, defined and etched to complete the structure.

This illustration shows how the sequence of deposition, lithographic definition and etching are used to build planar structures. In this case four etching processes are required for thermally grown SiO_2 , doped LPCVD Si, doped LPCVD SiO_2 , and doped LPCVD poly-Si/Al composite. The remainder of Sec. 3 discusses why plasma etching processes are often chosen for these and other etching applications, what equipment is typically used, what the general requirements are for such processes and some further details specific to each process.

3.2 Equipment

3.2.1 Parallel Plate Etchers. The first technologically significant plasma etching reactor is typically referred to as a parallel plate etcher. It is configured essentially like the example shown in Figure 1. A round, rf driven electrode is positioned parallel to another round, grounded electrode on which multiple wafers are placed.¹³⁴ This arrangement can also be utilized for plasma-enhanced deposition and is discussed in that regard in section 4.2.1. Similar considerations discussed there regarding pumping gases and safety also apply here.

The dc sheath potential, across which ions are accelerated, is determined by among other things the area ratio discussed in Section 2.1.2.2. Since the reactor itself is grounded the rf driven area is smaller than the total grounded area. This leads to relatively low accelerating potentials across the ground electrode sheath. For historic reasons etching in this type reactor at relatively high gas pressure (0.1-2 Torr) is referred to as plasma etching (PE).

PE is frequently characterized by high selectivity (see Sec. 3.4.4) and isotropic etching (see Sec. 3.4.2). The dominant mechanism in PE is thought to be spontaneous etching by chemical species generated in the plasma with only a small contribution from energetic ions. This is because the F/P ratio is low.

A more recent embodiment of this geometry is shown in Figure 33.

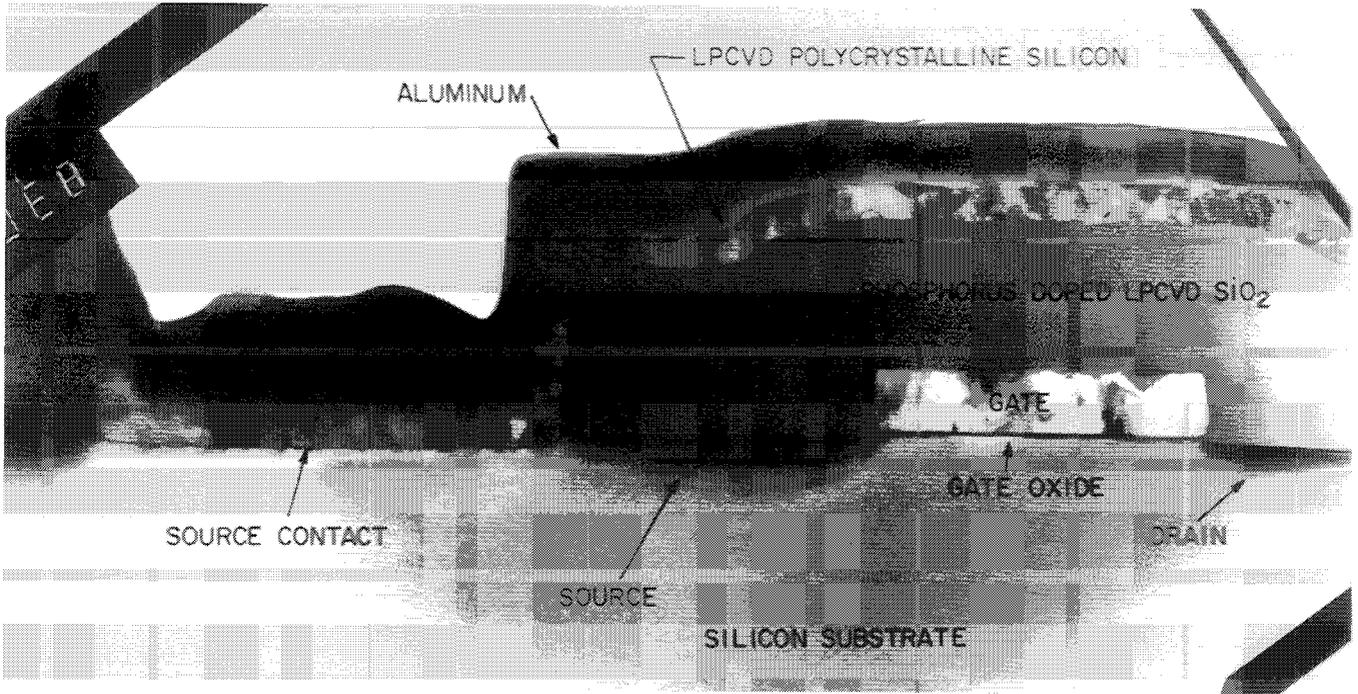


Figure 31: A cross section transmission electron micrograph of an MOS transistor (micrograph courtesy of R.V. Knoell, AT&T Bell Laboratories).

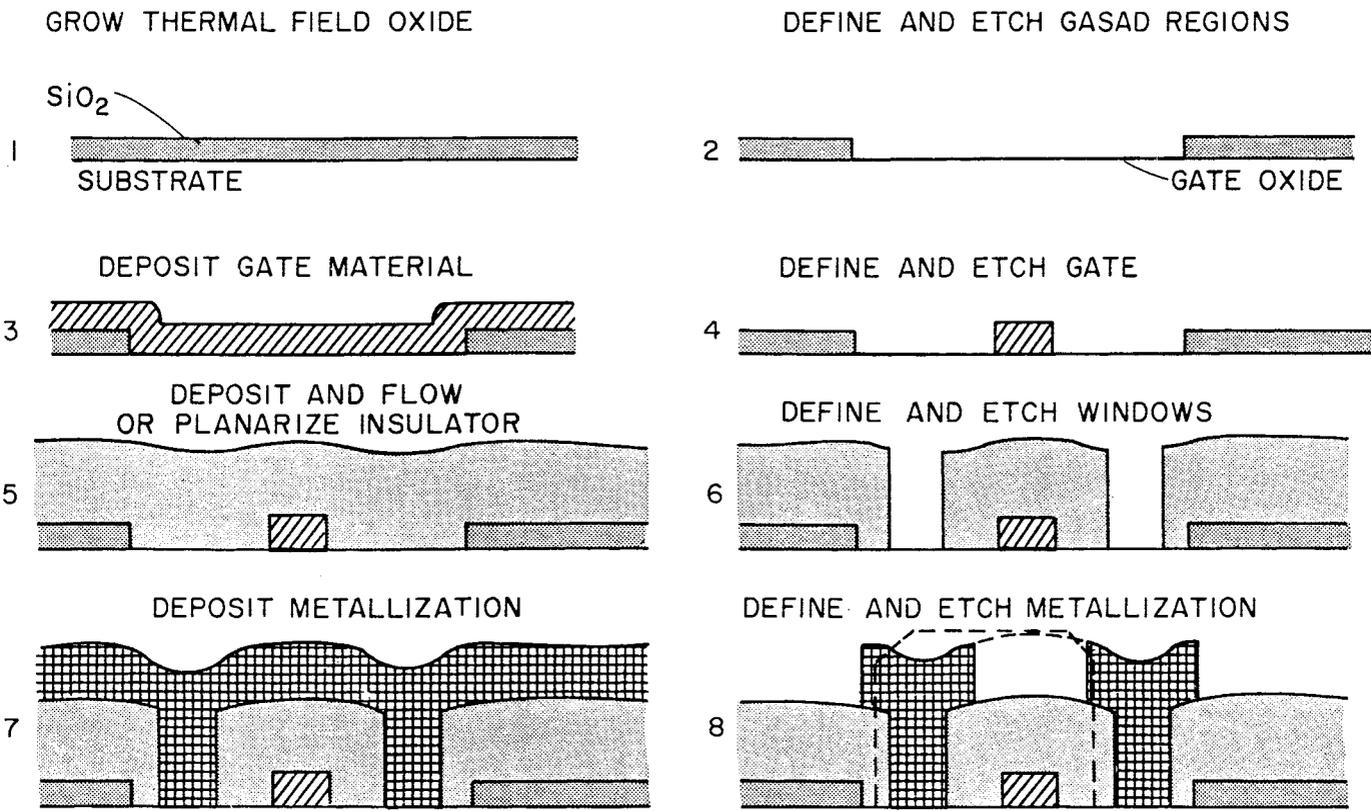


Figure 32: A sequence of patterning and etching to form an MOS transistor.

Because the rf power can be introduced into the system in a variety of ways, it is known as a “flexible” etcher.¹³⁵ It can be operated in the PE mode by connecting B in Figure 33 to an rf power supply through an impedance matching network and connecting A to ground. The opposite configuration where B is grounded and A is rf driven is commonly referred to as either reactive ion etching (RIE) or reactive sputter etching (RSE). Here the area ratio is the inverse of that in the PE configuration. This mode is characterized by relatively low plasma pressures, 5-50 mtorr, high ion accelerating potentials, lower selectivities and frequently anisotropic etching. Now ion assisted mechanisms, discussed in section 2.2.3.2, contribute more to the etching process because of the higher F/P value. The flexible etcher can also share rf power between A and B in the hope of blending etching characteristics to fill the requirement of a specific process.

3.2.2 The Hexagonal Cathode Etcher. Another RSE mode etcher of technological significance is the hexagonal cathode etcher shown in Figure 34.¹³⁶ Wafers are mounted onto the six faces of a hexagonal prism (HEX) which is the rf driven electrode or cathode. The whole metal vacuum bell jar then acts as the grounded electrode. Here an even larger difference in area between the two electrodes results in a large dc bias on the HEX which in turn produces large ion accelerating potentials. The size of the HEX allows many wafers to be etched simultaneously. Its almost cylindrical geometry causes relatively uniform rf fields and therefore uniform etch rates over all wafer positions (see Sec. 3.4.3).

Parallel plate and HEX etchers are both multiwafer, or batch, etchers. An important disadvantage of these types of etcher is that they are usually manually loaded while the chamber is vented to atmosphere. Venting the chamber allows the adsorption of contaminants, such as water vapor, onto

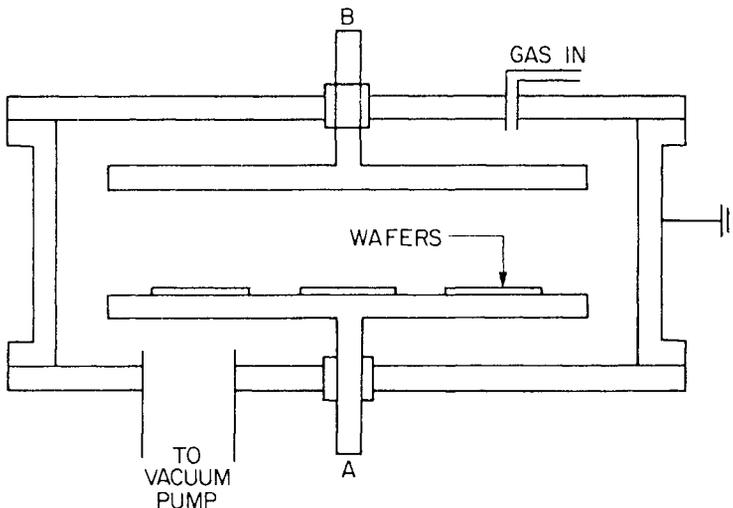


Figure 33: A schematic illustration of a parallel plate etcher which can be operated in the PE or the RSE mode (from Reference 134, © 1981 IEEE).

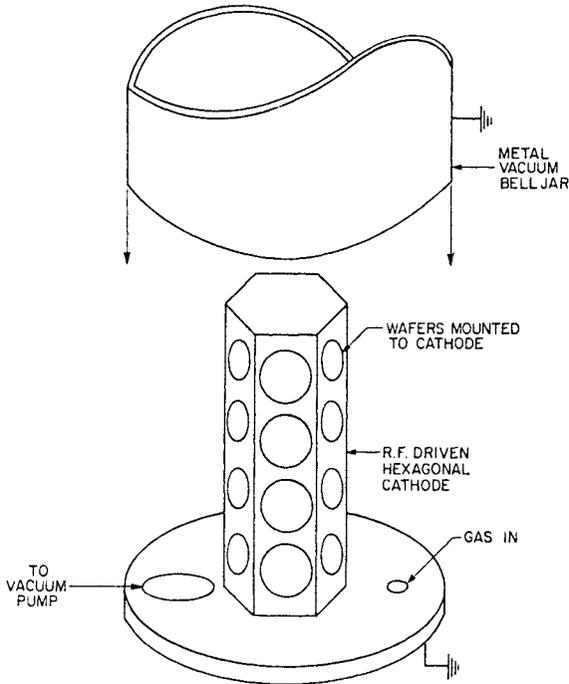


Figure 34: Illustration of a HEX etcher showing half of the twenty-four wafer positions.

the chamber walls which can change etching parameters. For example, it was shown in the case of fluorine based etching of SiO_2 , that water vapor dramatically affects the etch rate.¹³⁷ This causes differences in measured etch rates from run to run depending on the relative humidity in the room, the time the chamber is left open and the time the system is pumped before etching commences. Manual loading also increases the chance of particles falling onto wafers which can lead to defects (see Section 3.4.6).

3.2.3 Single Wafer Etcher. Ways to avoid the two problems mentioned in the last section are to automatically load the system through a load lock. A load lock is a chamber between the etching chamber and atmosphere, which accepts wafers from atmosphere, is pumped out, and then delivers wafers to the etching chamber. In this way the etching chamber is only rarely exposed directly to atmosphere. For simplicity, and added process control these systems typically etch only one wafer at a time. Single wafer etchers operate with parallel plate geometry either in the PE or RSE mode. Some additional trade-offs between batch and single wafer etchers are discussed further in Secs. 3.4.3 and 3.4.4.

3.3 Endpoint Detection

For a typical process, etching proceeds until the nominal film thickness is etched away. The process then continues until a predetermined “over-etch” time has elapsed. This overetch must be long enough to satisfy the etch rate uniformity in the system so that the layer has, in fact, been removed at every wafer position. It must also be short enough so that the underlying layer is not unacceptably eroded in areas exposed during the total over-etch time. An accurate determination of the nominal time to completion, before overetch, is desirable. This time can be obtained by visually observing the wafer through a viewing port, or determining the etch rate of the process, and calculating the nominal etch time. The latter method depends totally on the reproducibility of the process. Better ways to reliably determine the nominal etch end point are to measure some characteristic change in a plasma parameter or optical property of the wafers.

3.3.1 Voltage/Power. A simple example of endpoint determination is to hold the rf power constant and monitor the dc bias potential. Frequently, if large areas on the wafer are not masked and etching, the dc bias will measurably increase or decrease when the etch is completed and the underlying layer is exposed to the plasma. There are two possible mechanisms for this effect. The chemical nature of the plasma changes at end point because the number of product species is suddenly reduced and reactants increase. This change in the plasma chemistry affects the plasma impedance which is reflected in the dc bias—rf power relationship. Another mechanism can be important in processes where a conducting film is etching from an insulating (e.g. oxidized) substrate. As this film is removed, exposing the insulator, the effective sheath capacitance changes, leading to a change in the measured dc bias.

3.3.2 Optical Emission Spectra. The relative change in concentration of product species can be qualitatively determined by observing optical emission from excited states in the plasma. A good example of this technique exists in the case of aluminum etching where the presence of a line at 261.4 nm, as shown in Figure 35a, is a signature for the presence of AlCl .¹³⁸ Figure 35b shows the significant decrease in the intensity of this line at end point.

3.3.3 Laser Interferometry. Laser interferometry can be used to observe the etching of semitransparent films such as Si or SiO_2 .^{139,140} The relative phase shift of light reflected from the outer surface and inner surface of an etching film creates an interference pattern which changes as the film becomes thinner. This changing phase shift causes an oscillating component in the total reflected light intensity which is easily measured. The period of the oscillating intensity is related to the film thickness.

3.4 Defining Process Parameters and Goals

When developing plasma etching processes for each lithographic level, the following parameters must be considered so that the process goals are well understood.

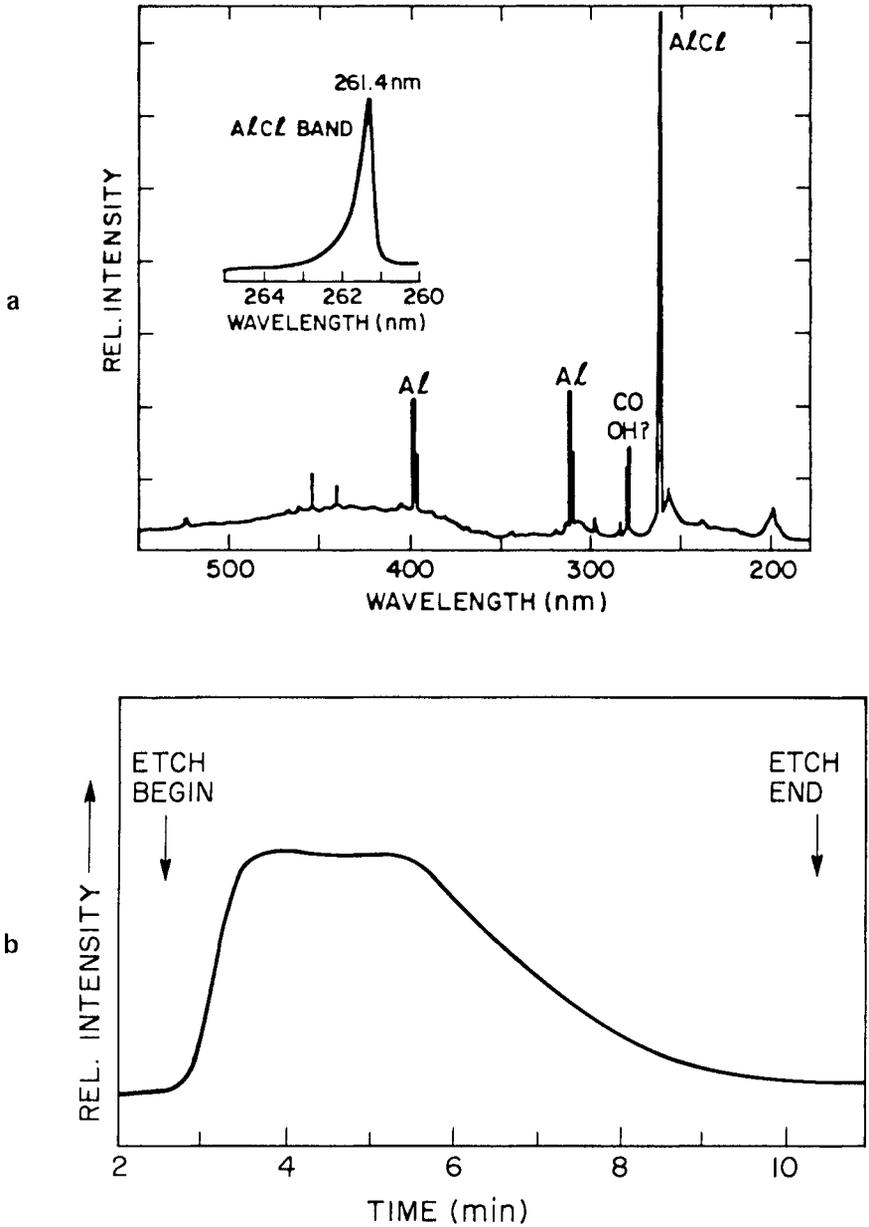


Figure 35: (a) Emission spectrum of CCl_4 plasma at 80 mTorr and 1.5 watts cm^{-2} during plasma etching of Al. (b) CCl_4 etching of Al monitored by a spectrometer set to the 261.4 nm AlCl band (from Reference 137, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

3.4.1 Material to Be Etched. The chemistry and morphology of the layer to be etched must be well understood. Details of thin film deposition processes may result in film properties which significantly affect the etching characteristics of the film. Impurity concentration, grain size, density and topography of the etching layer all play a role.

A common situation encountered when etching LPCVD poly-Si serves as a good example for the effect of film impurities on etching processes. Chlorine based plasmas typically used to etch silicon remove oxidized silicon at relatively low rates. The poly-Si etch rate in such etching processes is therefore very sensitive to the oxygen impurity concentration in the poly-Si. A low homogeneous oxygen level causes a reduction in the etch rate. On the other hand, segregated regions of high oxygen concentration can cause unetched residues commonly referred to as "grass". Figure 36 shows a scanning electron micrograph (SEM) of such a residue.

The film grain size can also affect an etching process. If the film is deposited under high temperature conditions where nucleation rates are small and grain growth rate is high then the average grain size can be as large as the film thickness. Large spherical or columnar grains can cause rough film surfaces. In addition to causing lithography problems, rough films must be significantly overetched so that total film removal is assured. Long overetch times are only possible when the etching process has adequate selectivity to the mask and underlying layer (see Section 3.4.4).

It is, therefore, extremely important to know the details of film properties before attempting to design a reproducible plasma etching process.

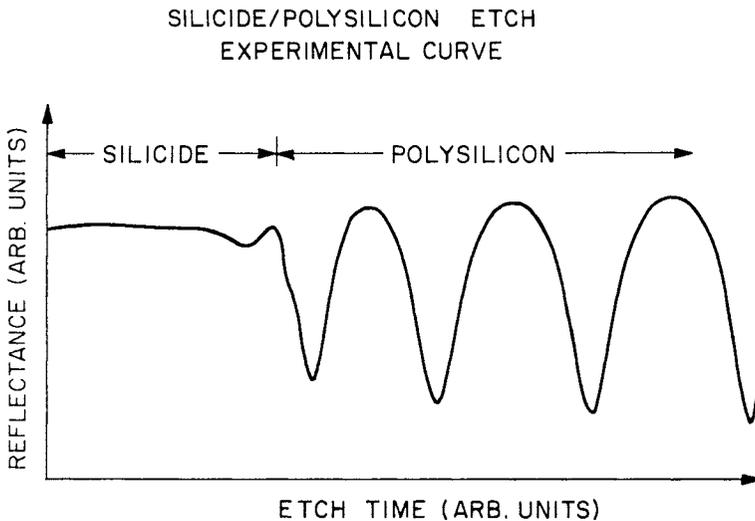


Figure 36: Experimental reflectance signal from a 6328Å laser for a silicide/poly-Si structure. The signal only shows the initial portion of the poly-Si etch (after Reference 138).

3.4.2 Feature Edge Profiles. Some plasma etches and virtually all liquid etches remove material isotropically, that is, the etch rate is the same in all directions. Figure 37a shows schematically the edge profile of an isotropically etched masked feature. As is clearly seen such an isotropic process causes an effective decrease in feature size by “undercutting” the mask. It is possible to compensate for this reduction in feature size by providing wider features in the mask, however this is not a desirable solution for VLSI devices for at least two reasons. First, the undercut, and therefore feature size, depends on etch rate, film thickness and overetch time making the feature size extremely difficult to accurately control. Second, mask compensation wastes area between features which decreases packing density and circuit speed by requiring longer, more resistant interconnections.

Anisotropic etch processes where the vertical etch rate is significantly higher than the horizontal etch rate, are therefore necessary. Figure 37b shows schematically the results of etching where the horizontal etch rate is zero. Often the anisotropic ion energy distribution is directly responsible for anisotropic etching. That is, edge profiles of etched features can be explained by a vertical etch rate which is much greater than the horizontal etch rate. In this case, if these profiles do exhibit slight undercutting it occurs at the top of the feature, as shown in Figure 37a. This undercut should be larger for thicker layers, because the region directly adjacent to the mask is exposed to the etching atmosphere for a longer time. The undercut should also be larger for longer overetch times, for the same reason.

An interesting example of how feedstock composition can affect anisotropy was reported by Mogab et al.¹⁴¹ and later discussed by Flamm et al.⁴ Poly-Si etches isotropically in the PE mode using a chlorine feedstock. It was found that adding C_2F_6 to the plasma results in features that are not undercut. This result is explained by assuming the formation of a “recom-

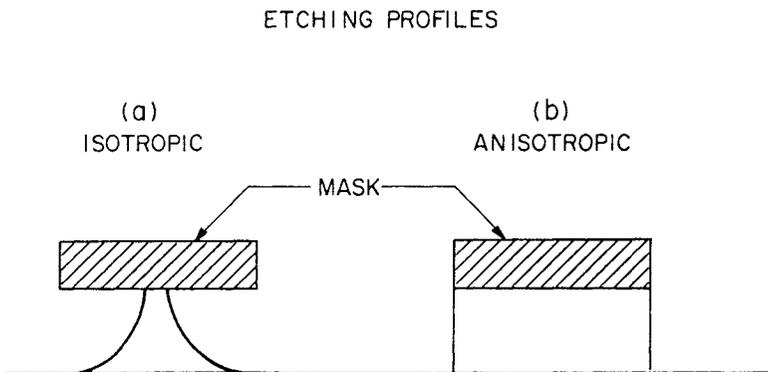


Figure 37: (a) The schematic cross section of a layer isotropically etched with a mask over a portion of the layer. (b) The cross section of an anisotropically etched layer.

binant species" or etch inhibitor on the side walls of the profile. The exact composition of this deposit on the feature walls is not known but is suspected to be a polymer. The deposit is also observed on horizontal surfaces at higher than optimum C_2F_6 flow rates. Under optimum conditions the result is etching on horizontal surfaces, deposition on vertical surfaces, and anisotropic profiles overall. This interpretation is consistent with the chemical vapor transport theory discussed in Section 2.2.1. A similar effect has also been credited for the anisotropic etching of aluminum.¹⁴²

The recombinant model assumes this passivating layer could be as thin as a monolayer, however, significantly thicker sidewall layers, easily visible in an SEM, have been reported.¹⁴³ In this case poly-Si was etching in a CCl_4 plasma. The presence of the layer was again credited with anisotropic etching. When no sidewall layer was observed the masks were severely undercut.

Undercut feature profiles are frequently observed that cannot be explained by assuming the etch to have a vertical and horizontal component. A finite horizontal component, no matter how small it is, should always cause the deepest undercut to occur directly beneath the mask (Figure 37), where the side of the feature is exposed longest to the plasma. The RSE of GaAs¹⁴⁴ in a mixture of Cl_2 and Ar results in profiles shown schematically in Figure 38. This profile cannot be explained by an etching process with vertical and horizontal components because the deepest undercut occurs well below the mask edge. Redeposition of the substrate material must also play a role in determining the final surface morphology. The loss of a protective layer on the feature sidewalls is thought to have produced the undercut profile but a detailed explanation does not exist. Such a narrowing of an etched feature at approximately half its height has also been observed in chemically assisted ion beam etching.¹⁴⁵ In this technique XeF_2 gas is ionized by electrons emitted from a hollow cathode source. Positive ions are then accelerated by grids in the direction of the etching wafer. When grid conditions cause a defocussing of the beam it diverges by as much as 20° and the features are undercut in a manner very similar to Figure 38. When the beam is collimated nearly vertical profiles are observed. This result demonstrates the relationship between the anisotropy of ion transport and feature profiles.

3.4.3 Uniformity. Ideally the etch rate should be constant, or uniform, at all etching positions. However, this goal is seldom achieved. Electric field and composition gradients can cause large variations in rate. Which of these is the dominant mechanism must be determined before any improvements can be made. Plasma composition gradients can be caused by local depletion of the active species. Such depletion can be affected by gas residence times, which in turn are controlled by source gas flow rates and loading effects. A nonuniformity which is affected by flow rates therefore points to a composition gradient mechanism. This problem can often be solved by a more uniform introduction of the source gas into the reactor.

Electric field nonuniformities are more complicated to deal with since they are the result of the interactions of virtually all of the plasma parameters. Often a change in the reactor geometry can improve etch uniformity without changing desirable properties of the etching process. With batch

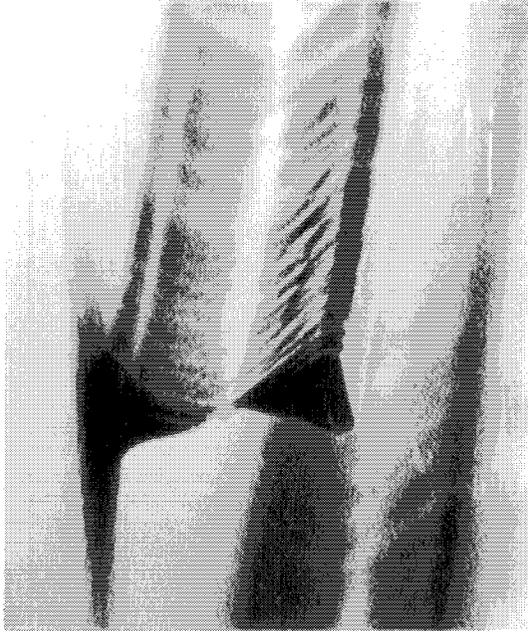


Figure 38: Profile of a GaAs substrate etched in a 4:1 mixture of Cl_2 and Ar (from Reference 143, reprinted with permission from the American Institute of Physics).

processes one must be concerned with not only intra-wafer but also inter-wafer uniformity because many wafers are etched simultaneously. Since single wafer etchers need only provide good intra-wafer uniformity, they should, in principle, provide superior uniformity with respect to batch etchers. In practice, this might not always be the case because single wafer etchers operate at higher power densities and higher etch rates in order to achieve higher throughput (Sec. 3.4.5). Uniformity is generally more difficult to achieve at higher etching (or deposition) rates.

3.4.4 Selectivity. A process is selective if it etches the desired material at a high rate and all other exposed surfaces at a low rate. Selectivity is defined as

$$S_{AB} = \frac{\text{etch rate of material A}}{\text{etch rate of material B}}$$

Two selectivities must be considered for any etching process. The first is where A is the material to be etched and B is the masking layer. The selectivity to the mask must be very high, >5 , when thin, $0.2\text{-}0.5\ \mu\text{m}$, photoresist is used. Multilayer resist schemes¹⁴⁶ which result in thick, $1.0\text{-}2.0\ \mu\text{m}$ masks relieve this constraint. The second case is where A is the material to be etched and B is the underlying layer at which the etch should stop. This selectivity is one of the most important process parameters. It is during the overetch that this selectivity becomes extremely important. The

standard etch time is either calculated from known etch rates or determined *in situ* by an end point detection technique (see Sec. 3.3). To determine an overetch time, the uniformity of the process must be known. An extremely uniform process might require a small overetch which, in turn, allows for a poorer selectivity. These two parameters are therefore interrelated.

Other factors also influence the required selectivity of a process. For example, the gate etch mentioned in Sec. 3.1.3 is shown schematically in Figure 39. Ideally, at the end of the standard etch time, the material under the mask will remain while the other material will be totally removed. However, because of the excellent step coverage of the deposition process, material on the wall of the FOX will remain. This will, of course, only be true if the horizontal etching component of the process is zero. In order to remove this “stringer” an overetch of approximately 100% must be carried out. As previously mentioned the thin GOX is used as an etch stop. If this GOX is 200Å thick and the FOX is 3000Å thick then a selectivity of at least 30 is required if removing 100Å of the GOX is permitted. Taking the uniformity into account will cause the required selectivity to be even higher.

In practice such a high selectivity along with an anisotropic etch profile is not easy to achieve. High selectivities are typically achieved in systems where chemical etch rates dominate. This is not the case when the physical sputtering contribution to the etch rate is significant. This contribution is relatively independent of etching material and impinging ion species. It does, however, increase with increasing dc bias voltage.¹³³ For this reason higher selectivities are typically achieved in the PE mode, with its inherently lower sheath voltage, than the RSE mode. Higher selectivities

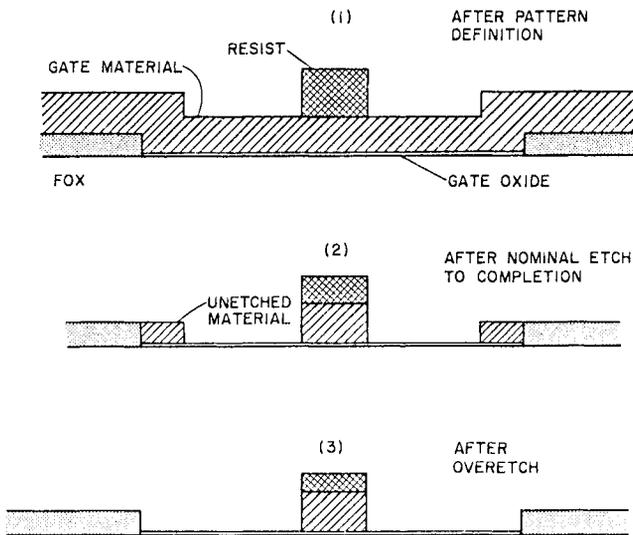


Figure 39: A sequential representation of a gate etch which illustrates the need for an overetch to remove material deposited on the field oxide wall.

can be achieved in the latter case by reducing the rf power which in turn reduces the dc bias voltage. As previously discussed operating in this regime moves in the direction of isotropic etching. A trade off therefore frequently occurs between selectivity and anisotropy.

Optimizing the selectivity can often be achieved by the proper choice of feedstock and flow rates, as pointed out in Section 2.1.3.1. A good example of affecting selectivity by blending feedstock gasses exists for Si and SiO₂ in CF₄/O₂/H₂ mixtures.¹⁴⁷ Free F atoms are the active etchant species and the F atom concentration has been shown to depend on the O₂ concentration in the feedstock. Figure 13 shows how the etch rate of Si is affected by the O₂ concentration in a CF₄ plasma.⁶³ The etch rate of SiO₂ in this mixture has a similar dependence on O₂ concentration except at higher concentrations the rate does not fall off so severely as for Si. Dilution and surface effects play a lesser role because the SiO₂ itself is a source of O₂. Adjusting the concentration of O₂ in CF₄ can be used to achieve a selectivity of at least 3 for SiO₂ over Si at 50% O₂ in CF₄.

The effect of adding H₂ to CF₄ is shown in Figure 40.¹⁴⁸ It is believed that the sharp decline in Si etch rates at high H₂ concentrations is due to a decrease in the F atom concentration caused by the scavenging of F atoms by H₂, forming HF.¹⁴⁹ In this system selectivities of 35 can be attained at 40% H₂ in CF₄. Adding H₂ to the feedstock has the effect of enhancing polymer formation in the etching chamber. Deposition of polymer onto wafer surfaces during etching is an undesirable parasitic effect which is best avoided in commercial processes.

3.4.5 Throughput. The number of wafers etched per unit time is referred to as the throughput. When the technology exists, economic

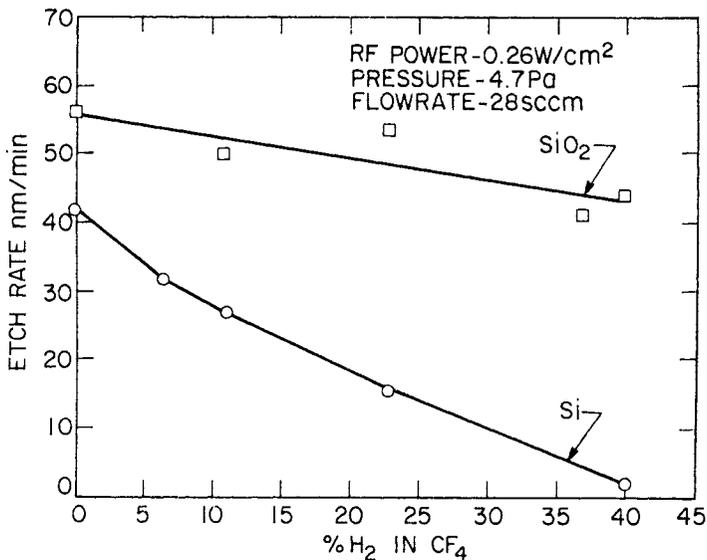


Figure 40: Dependence of SiO₂ and Si etch rates on percentage of H₂ in CF₄ feedstock (from Reference 147, reprinted with permission of the publisher, The Electrochemical Society, Inc.).

factors determine an acceptable throughput. For batch etchers, where many wafers are processed simultaneously, typically acceptable throughputs require relatively low etch rates of 100-600 Å/min. Single wafer etchers, which must achieve rates at least 10 to 20 times faster for equivalent throughputs, have to operate at high power density and high pressure. As discussed in Section 2.2.2.1 it is possible to achieve anisotropic profiles under such conditions. The high power density can, however, lead to unwanted wafer heating effects. Wafer heating can cause resist reticulation (cracking and chemical degradation) which leads to unfaithful pattern transfer. To minimize this effect wafers must be heat sunk well to a temperature controlled cathode.

3.4.6 Defect Introduction. Any local modification of the original lithographic pattern is considered a defect. A common type of defect, introduced during plasma etching is caused by particles falling onto the wafer during handling or while it is exposed to the etching environment. The former is common to all processes but the latter is related to the specific etching process.

When freons are used in the feedstock, plasma decomposition frequently results in polymer deposition onto reactor walls.¹⁵⁰ As the polymer film gets thicker it becomes unstable and flakes off onto the etching wafers. These polymer films etch readily in an O₂ plasma so that frequently etchers are cleaned by establishing O₂ plasmas at regular intervals between normal etching runs. Care must be taken when such a cleaning practice is used because different results may be obtained in a clean system than in a polymer-coated system. The dynamic equilibrium levels of active species established in the etcher is different in each case. Two components of this equilibrium are material depositing onto etcher surfaces and material being etched from these surfaces. When no polymer exists on the etcher surfaces equilibrium values of active species in the plasma might be significantly different than when polymer is coating every surface.

3.4.7 Radiation Damage Effects. In the process of plasma etching, device wafers are exposed to various types of damaging radiations: soft x-rays, uv light, ions and electrons. In the RSE mode the energy of the positive ions is considerably higher than the electrons because of the large potential drop across the cathode sheath. The total radiation dose affects the electrical properties of devices in a variety of ways.

When accelerating ions with energies of hundreds of eV, strike the surface of a solid they are implanted into the surface an average distance of 10-30Å¹³² During this implant process the crystal structure of the solid traversed by the slowing ions is disrupted. Bonds are broken and atomic positions are altered. This, in fact, may be one mechanism by which ions enhance surface reactivity (see Section 2.2.3.2).

The crystalline damage is significant when the irradiated material is single crystal and the performance of a device depends on the crystalline perfection. An example is a schottky diode formed on a Si substrate which was previously exposed to RSE. Often crystalline damage can be repaired by high temperature annealing¹⁵¹ which is already part of subsequent processing, however, the implanted species from the plasma can only be removed by chemical etching of the implanted surface layer.

Another target of radiation damage is the gate oxide and field oxide in

MOS devices. When these oxides are directly exposed to RSE, fixed positive charge is trapped in the oxide, fast interface states are created at the Si-SiO₂ interface and the oxide contains many neutral electron traps.¹⁵² Most of this damage is the result of ionizing radiation. Properties of gate oxides are typically studied by measuring changes in their capacitance-voltage (C-V) characteristics.¹⁵³ With this method the relationship between the number of "hot" electrons injected into the oxide and the flat-band voltage shift has been used to study the neutral electron trap density in gate oxides directly exposed to RSE.¹⁵⁴ Figure 41 shows how oxide damage caused by RSE can only be completely annealed out at 1000°C.

Of more practical interest is damage to the gate oxide by RSE when it is covered by the gate material. This is the case for the gate etch as demonstrated in Figure 39. It has been shown that damage then only occurs at the periphery of the oxide because of radiation absorption in the gate.¹⁵² Fortunately subsequent processing requires annealing at temperatures high enough to remove this damage so it usually presents no problems to device performance.

In the future, MOS devices will be made with extremely short (<1.0μm) gate channel lengths. These devices must have very shallow (<2500Å) source and drain junctions¹⁵⁵ which cannot be exposed to the present high temperature processing. The problem of oxide damage caused by plasma etching processes will then need to be further investigated.

3.5 Specific Etching Processes

3.5.1 Silicon and Silicides. Virtually all Si IC fabrication schemes

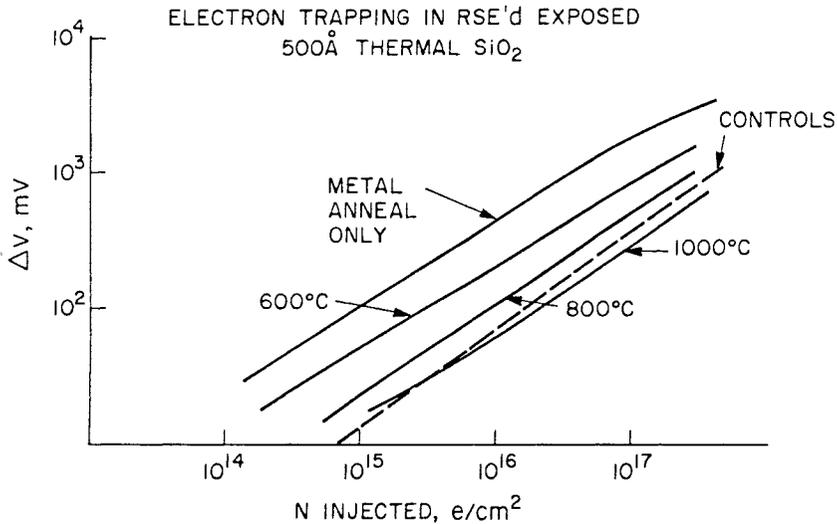


Figure 41: Electron trapping in thermal SiO₂ that was exposed during the RSE of poly-Si as a function of post-RSE thermal anneals (from Reference 152, © 1979 IEEE).

include at least one poly-Si etch process. The most common is the MOS gate etch. Recently efforts have been made to decrease the resistivity of poly-Si gates while retaining the stable gate/gate insulator (poly-Si/SiO₂) interface. This is accomplished by adding a silicide layer to the top of a normal poly-Si gate.¹⁵⁶ For this technique to be useful, it must be possible to anisotropically etch this bilayer structure, frequently called a polycide, with the high selectivity discussed in Section 3.4.4.

3.5.1.1 Silicon. A primary concern when selecting a feedstock for any etching process is that the final reaction product be volatile. Nonvolatile products lead to residues on the surface which cause either surface roughening or a total cessation of the etch. Since silicon chlorides and fluorides are volatile, etching of poly-Si can be accomplished with most Cl and F based feedstocks such as Cl₂, CCl₃F, CClF₃, CF₄, NF₃, SF₆.¹⁵⁷ Generally feedstocks containing F etch isotropically while those containing Cl can yield anisotropic profiles. The spontaneous etching of Si by free F atoms generated in the plasma causes chemical, isotropic etching.⁷ Undiluted Cl₂ can be used to etch poly-Si anisotropically in a plasma¹⁴¹ but a problem is often encountered in initiating the etch through the surface native oxide. This problem can be solved by using a Cl containing freon where the selectivity of Si:SiO₂ is not so great. The selectivity of Si:SiO₂ has been studied in the series CCl₄, CCl₃F, CCl₂F₂, CClF₃ and CF₄.¹⁵⁸ Results indicate that selectivity decreases by a factor of five as the F concentration in the source gas increases. This is consistent with Cl not spontaneously etching Si at room temperature. The characteristics of virtually all poly-Si etching processes change abruptly when the poly-Si is heavily n-type doped and activated. The etch rate then increases dramatically, and isotropic profiles are often observed. When the poly-Si is p-type doped, on the other hand, etching is indistinguishable from undoped poly-Si. If an n-type dopant is implanted and etched before it undergoes an activating anneal, it etches as undoped poly-Si. These facts lead to the speculation that a high free electron density enhances the interaction of Si with adsorbed Cl atoms leading to higher etch rates and isotropic profiles.¹⁴¹ Experimental evidence shows, however, that the etch rate does not correlate with free electron density¹⁵⁹ so the mechanism is as yet unresolved.

3.5.1.2 Silicides. Undiluted Cl₂ is ineffective in etching silicides so most polycide etching processes use Cl₂ mixtures or freons. Figure 42 shows a TaSi₂/poly-Si composite which was etched using a two-step process. The TaSi₂ was first etched using a CCl₃F feedstock and then the poly-Si was etched with a Cl₂ feedstock. For either a one step or a two step polycide etching process the selectivity of the bottom poly-Si with respect to SiO₂ is most important. There are many reports in the recent literature of one-step processes and the following few are given only as examples. BCl₃/Cl₂ mixtures are reported to etch TaSi₂/poly-Si composites anisotropically.¹⁶⁰ TiSi₂/poly-Si is reported to etch anisotropically in a CCl₄/Ar/N₂/H₂ mixture.¹⁶¹ MoSi₂/poly-Si etches in NF₃/HCl mixtures with a possible selectivity of poly-Si:SiO₂ of 30.0.¹⁶² While many mixtures of Cl and F bearing source gases can etch polycide structures, the conditions for a one-step process which results in tightly controlled profiles along with excellent selectivity (>100) to SiO₂ have not yet been reported.

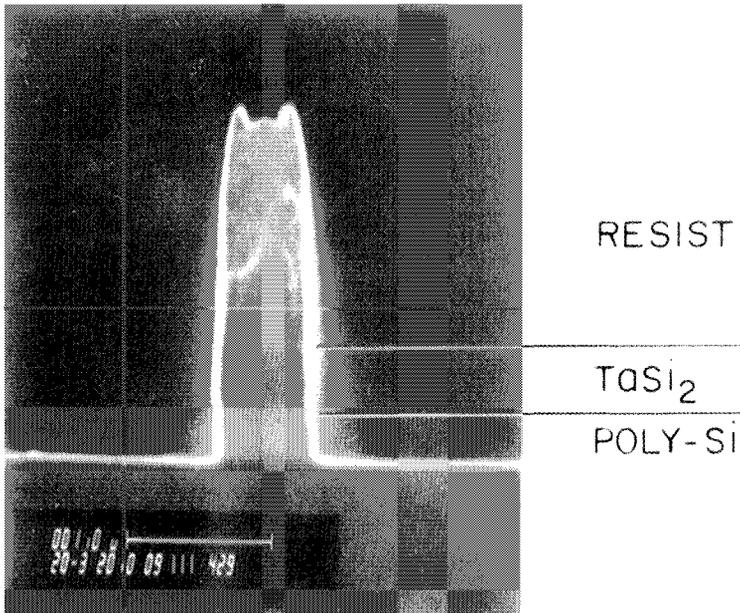
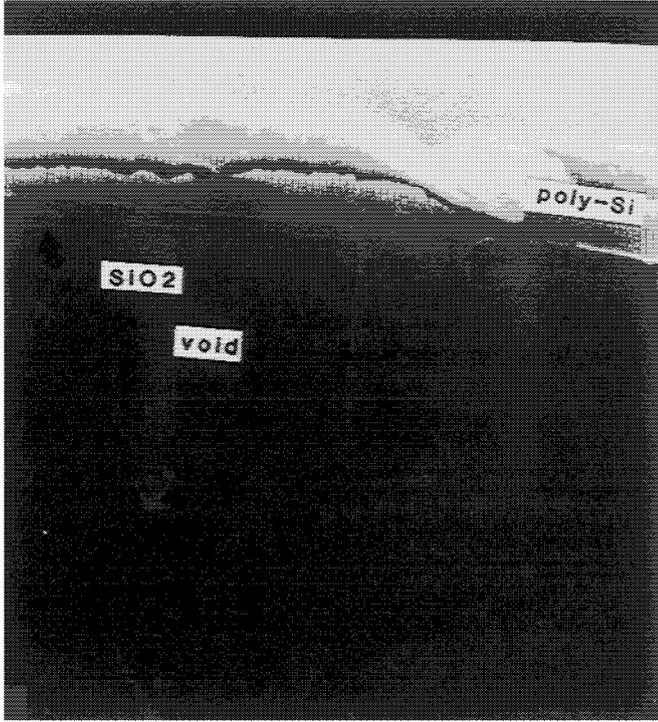


Figure 42: An SEM cross section of a TaSi₂/poly-Si structure anisotropically etched with a two step process.

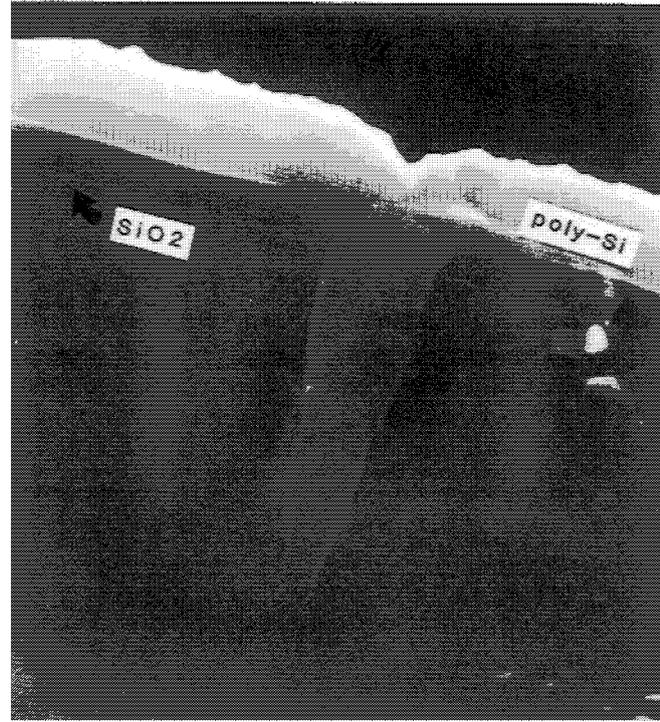
3.5.1.3 Trench etching. Another application of Si etching has recently emerged with the need for deep ($>5.0\mu\text{m}$) narrow ($<1.5\mu\text{m}$) slots or trenches in the Si substrate. These trenches can be used to isolate complementary devices in CMOS circuits.¹⁶³ Alternatively vertical storage capacitors can be fabricated inside the trenches.¹⁶⁴ A primary prerequisite for such trenches is that they not be undercut. In fact, a negative undercut is desirable so that when the trenches are refilled with an LPCVD process no void will form at the top of the trench. Figure 43a shows such a void in a refilled trench with a slight undercut. Figure 43b shows no void when the trench walls taper inward. Cl₂ based source gases are reported to etch trenches with these very high aspect ratios.¹⁶⁵

3.5.2 Etching of Thermal and LPCVD Oxide. As shown in Section 3.1.3 two major applications of oxide etching are the initial opening of GASAD regions in FOX and then the etching of contact windows through phosphorus doped LPCVD oxide (p-glass). In both these cases high selectivity with respect to underlying Si is important as is the integrity of the exposed Si surface. The GASAD region is where the MOS transistors are formed so only minimum damage can be tolerated. Contact window etching also requires only minimum residue after etching. Such residue has been shown to increase contact resistance to metals which are subsequently deposited into the windows.¹³⁵

Because of this metal deposition, profile control is very important for contact window, or via, etching. Perfectly vertical walls often create problems



a



b

Figure 43: (a) A deep, slightly undercut trench which was substantially refilled with LPCVD poly-Si. A void in the poly-Si is visible. (b) A tapered trench etched and refilled. No void is present due to the tapered trench.

because most vacuum, thin film deposition processes deposit thinner layers on the walls than on horizontal surfaces. This thinning (illustrated in Figure 44a) can lead to high resistance regions and, consequently, reliability problems. Figure 44b shows how this problem can be minimized by tapering the via walls. Tapering is accomplished by transferring an initial taper of the photoresist down into the p-glass or other insulator¹⁶⁶ as shown in Figure 45. Tapering cannot be used if vias are spaced so closely together that the tops merge together causing a thinning of the insulator. The thin insulator region at the bottom of the taper also increases the capacitance between the deposited metal and the conductor under the insulator. This causes an increased RC time constant which might limit the speed of the circuit. A tapering process demands tight control over resist profile, p-glass: resist selectivity, uniformity and overetch time.

Some gases reported to plasma etch oxide are CHF_3 , C_2F_6 , CCl_4 , CF_4 and NF_3 .^{140,143,147,167,168} All Forms of SiO_2 etch anisotropically under most

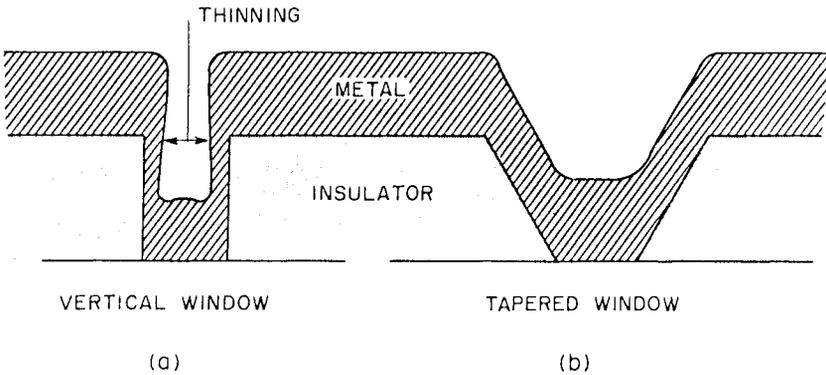


Figure 44: (a) A schematic illustration of vapor deposited Al into a contact window. Thinning of the Al is evident. (b) Only slight thinning is present with a tapered window.

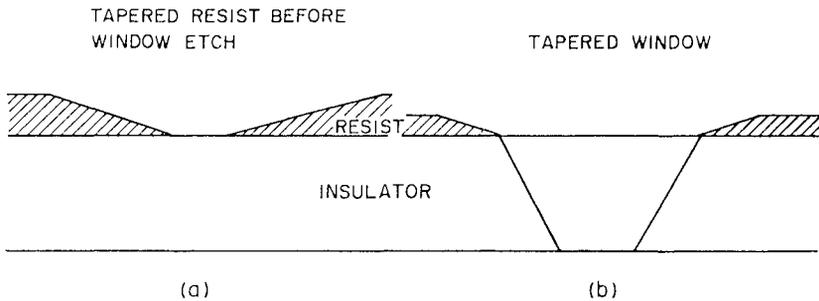


Figure 45: A sequential illustration of how a tapered resist pattern can be used to form a tapered window.

operating conditions with the first three of these feedstocks. SiO_2 can be etched anisotropically in CF_4 and NF_3 only at low frequencies and in the latter case only if it is diluted with an inert gas.¹⁴⁰ P-glass etches faster than LPCVD SiO_2 which etches faster than thermal SiO_2 ; P-glass etch rates can be as much as two times higher than for thermal oxide.

RSE of oxide in a CHF_3 plasma is an example of a system where polymer formation affects the etching process. In addition to forming polymers on the chamber walls, coating of some non-etching wafer surfaces can also occur. Polymer formation is greater at higher pressures (>40 mtorr) and higher flow rates. This deposition is thought to increase the selectivity of SiO_2 to Si by selectively depositing onto Si and not SiO_2 .

3.5.3 Etching of Aluminum Metallization. Aluminum metallization can take the form of a pure Al film, an Al/poly-Si composite structure, an Al-Cu alloy and an Al-Si alloy. While each of these metallizations schemes have slightly different etching characteristics they all share the many problems encountered in plasma etching of Al.¹⁶⁹⁻¹⁷¹

Al has a native oxide which frequently etches at a much slower rate than Al. At best this leads to an incubation period which must be accounted for when predicting clearing times. At worst, this leads to the etching surface being extremely rough because of slight nonuniformities in the native oxide thickness. A reproducible process does not have a high Al/ Al_2O_3 selectivity. Mixtures of CCl_4 , BCl_3 , SiCl_3 and Cl_2 meet this criteria and have been reported to anisotropically etch Al.¹⁷¹

Once etched, an Al pattern cannot simply be removed from the chamber into ambient atmosphere. If this is done severe corrosion of the Al takes place. Presumably a chlorinated residue on the Al reacts with water vapor in the air to form HCl which subsequently attacks the Al. Two ways of avoiding this problem are to either rinse the wafers in water immediately upon exposure to the atmosphere or to proceed, *in situ*, with an O_2 plasma etch which removes the masking resist and passivates the Al surfaces.¹⁶⁹

While it is certainly possible to anisotropically etch Al, it is frequently difficult to maintain a process which produces the same anisotropic profile from run to run over an extended period of time. The etched profile and etch rate are strong functions of the concentration of the active species, Cl, in the plasma. High Cl concentrations lead to high chemical etch rates which in turn lead to undercut Al profiles. A process which depends on a fully loaded system to deplete the Cl concentration enough to produce an anisotropically etched profile might undercut if less than a full load of wafers is etched.¹⁴² The recombinant and inhibiting mechanism has been credited with the anisotropic etching of Al⁴⁹ and the presence of a passivated sidewall layer has been corroborated.¹⁷² This passivating layer is found to etch readily in an O_2 plasma and its formation is probably threatened by residual O_2 in the chamber. Thus, the degree of anisotropy achievable in an Al etching process may be strongly dependent upon the exposure of the system to air. This problem is compounded by the formation of porous aluminum chlorides on the reactor walls. This deposit readily absorbs water vapor on exposure to air. Heating the reactor walls above room temperature or using a load reactor decreases the water vapor contamination problem.

3.5.4 Silicon Nitride. Most applications of silicon nitride in IC fabrication do not require accurate profile control during etching. This is true for LPCVD silicon nitride (Si_3N_4) or the plasma enhanced chemical vapor deposited (PECVD) silicon nitride (Si-N-H) discussed in Section 4.3.1. A typical etch for silicon nitride uses $\text{CF}_4 + \text{O}_2$ as a feedstock in the PE mode.¹⁷³ This etch suffers from poor selectivity to both silicon and silicon oxides. The selectivity of Si_3N_4 to Si is improved in plasmas which generate CF_x radicals^{148,150} and generally plasma parameters which lead to high SiO_2 : Si selectivity also increase the Si_3N_4 : Si selectivity.⁴ A recent report¹⁷⁴ shows that Si_3N_4 can be etched anisotropically in the RSE mode with a selectivity to Si and SiO_2 of 20:1. Either of two feedstocks, CH_2F_2 or CH_3F could produce these results.

3.5.5 Etching of Group III, V Compound Semiconductors. Semiconducting alloys of group III and group V elements (III-V's) are seeing increased usage because of their high mobility, large bandgap and rapid radiative recombination at wavelengths which make useful light sources. Plasma etching processes which can be used to fabricate III-V devices are being investigated and have been reviewed in the literature.^{4,175,176} This section will emphasize how these processes are different from those already discussed and what special considerations must be taken into account when etching III-V's.

As discussed in Section 3.5.1.1, a primary consideration when selecting a feedstock for a plasma etching process is that a volatile product species exist. For many etching processes previously discussed these volatile products are either chlorides or fluorides. Group III fluorides are not volatile at room temperature so fluorine is not an active etchant for III-V's. This fact can be advantageously employed by using a fluorine based plasma to etch a deposited film, such as a metal or PECVD oxide or nitride, with a high selectivity with respect to the III-V. Chlorine, bromine and hydrogen react to form either volatile or high vapor pressure compounds with most group III and V elements.^{177,178} Feedstocks containing one or combinations of these three elements are most often utilized in III-V plasma etching.^{179,180} However, even in an undiluted chlorine plasma the diverse chemical nature of the group III and group V elements often leads to non stoichiometric material removal. It has been shown¹⁸¹ when etching InP and GaAs in the PE mode with a chlorine feedstock that In and Ga enriched surfaces respectively are present after etching. The temperature dependence of the etch rates indicates activation energies close to the group III trichloride heats of sublimation.¹⁸¹ For this reason, product desorption is assumed to be the rate limiting step.

Unlike the plasma etching of single crystal silicon, the plasma etching of single crystal GaAs can exhibit crystallographic effects.⁵¹ When the etching mechanism is believed to be purely chemical, isotropic and free from ion bombardment enhancement effects, crystal facets develop under masked regions. These facets result from the dependence of the etch rate on crystallographic direction.

When the ion bombardment effects are significant, as in the RSE mode, anisotropic profiles can be obtained in III-V's¹⁴⁴ but two undesirable etching characteristics are frequently observed. The etching surface is

frequently extremely rough¹⁸⁰ and the side wall of masked features frequently slopes in such a direction as to make the feature larger at the bottom. This second effect which is illustrated in Figure 46 has been called negative undercut or overcut. Surface roughening is probably related to nonstoichiometric material removal.

Burton et al.¹⁷⁶ reviewed the application of plasma etching of the III-V semiconductors to eight III-V device fabrication processes: (1) etching via holes for through-the-chip interconnections, (2) etching wells selectively through one layer of a heterostructure, (3) etching mirror facets or windows, (4) separating chips, (5) etching integral lenses, (6) etching gratings, (7) forming mesas for restricting p-n junction area, and (8) etching channels. They conclude that plasma etching is currently the process of choice in only the first three applications. While the plasma etching of deposited layers such as metals, oxides, and nitrides is widely practiced in III-V manufacturing, the plasma etching of the semiconductor itself is clearly not as important for III-V processing as it is for Si processing. The converse is true for plasma enhanced deposition processes, so in the next section application to III-V technology will be stressed just as applications to Si technology were stressed in this section.

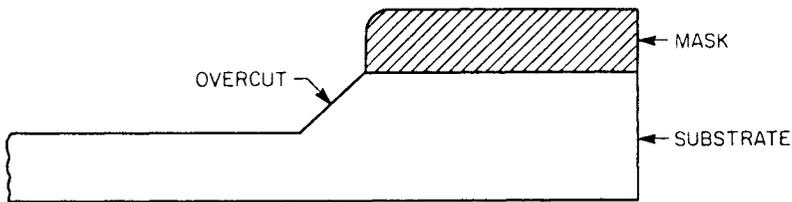


Figure 46: An illustration of a tapered, "overcut" etch profile which commonly occurs when etching III-V's (after Reference 176).

4. PLASMA DEPOSITION

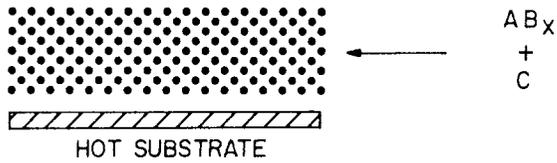
4.1 Introduction

In this section, we discuss the application of plasmas to processes in which material is deposited rather than removed. While this method of deposition is often referred to as plasma deposition (PD) or glow discharge (GD) deposition, it is more accurately described as plasma-enhanced or plasma-assisted chemical vapor deposition (PECVD or PACVD), since it is a CVD process whose rate at a low deposition temperature is enhanced by energy supplied from a plasma. In the interest of standardization, we will use the term PECVD throughout this section.

Whereas CVD is a purely chemical process, PECVD additionally involves physical processes such as ion and electron bombardment of the growing film. However, as in normal CVD, *all* the chemical constituents of the

deposited film are introduced in the gaseous phase into the reactor, whose design is basically the same as that used for the plasma etching process discussed in the preceding section. It is this property which distinguishes PECVD from reactive sputter deposition, the other common reactive plasma-employing deposition process. Reactive sputter deposition is the deposition complement of reactive ion etching (RIE) or reactive sputter etching (RSE), also discussed in the preceding section. In reactive sputter deposition, only a part of the constituents of the deposited film are supplied in the gas phase, with the other components supplied in the solid phase as the sputtering target. These deposition processes are contrasted in Figure 47. PECVD alone is the subject of this section. In recent years, there have been a number of reviews of PECVD,¹⁸²⁻¹⁹⁴ mainly from the standpoint of a thin film deposition technique. This work considers PECVD from the standpoint of its use in semiconductor processing applications. Thus in the ensuing sections, emphasis is on those materials used in semiconductor processing and on the control of film and film/semiconductor interface properties to meet the requirements of a given processing application.

PLASMA



SPUTTER

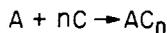
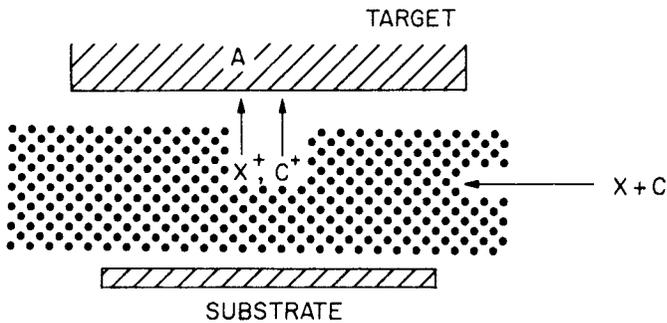


Figure 47: Comparison of PECVD and reactive sputter deposition processes, shown for deposition of a film of generic composition AC_n .

The major attraction of PECVD is its useful deposition rate at a temperature significantly lower than that for CVD. PECVD is generally performed at substrate temperatures in the range 200-350°C, compared to typical CVD temperatures of 400-900°C. As discussed in section 2, electron impact dissociation and ionization of reactants occurs in the gas phase, so that it is reactive atoms and radicals which adsorb and react on the substrate surface.^{188,189} Often thermal energy from the substrate is not necessary for reaction to occur; instead this thermal energy is usually needed to enhance surface mobility in order for desirable film properties to be attained. Although the ion temperature within the glow region of the plasma is room temperature (see Section 2.1), ions can be accelerated across the sheath to impinge on the substrate at energies ranging from a few to a few hundred eV (see Sec. 2.1.2.1.4). This ion bombardment also promotes surface mobility, as well as giving sputter-assisted desorption of volatile reaction products. *Thus PECVD processes are far from thermodynamic equilibrium and can be used to deposit materials or phases which cannot be deposited by direct CVD.* However, from the opposite consideration of radiation induced damage, another attraction of PECVD is that it is a relatively “soft” plasma process. In comparison to reactive sputter deposition, the energies of ions incident on the substrate are lower. In addition, because PECVD has a higher deposition rate for a given plasma power, the substrate is capped more rapidly by a layer thick enough to protect the underlying substrate. Thus both the damaging dose and the energy of these ions received by the substrate are considerably lower.^{195,196}

Figure 48 puts PECVD into context with other competing deposition technologies with regard to deposition temperature and process energetics. Sputter deposition and CVD are each discussed in detail in other chapters of this work. Photo-assisted CVD, also included in Figure 48, is a recently

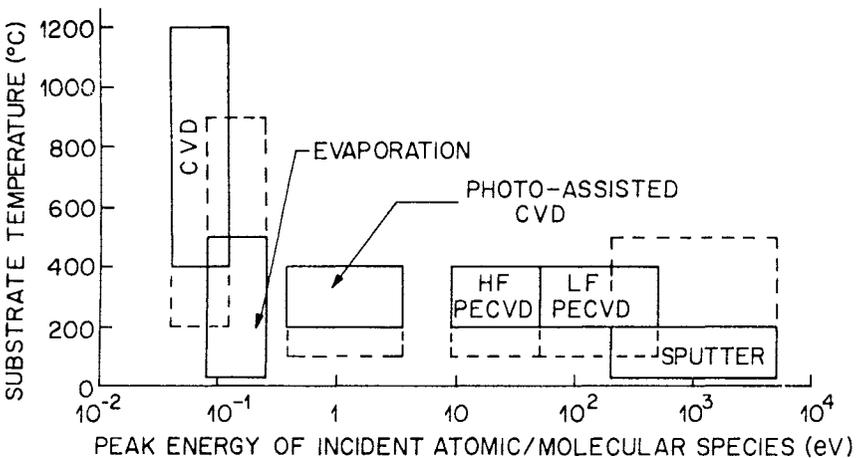


Figure 48: PECVD in substrate temperature-incident particle energy space in comparison to other deposition techniques.

developed alternative low temperature CVD technique in which the energy for gas phase dissociation of reactants is provided by ultra-violet photon absorption in place of the electron impact which occurs in a plasma. The U.V. photon flux may originate from a bank of lamps and impinge on the substrate surface,¹⁹⁷ or may be from an excimer laser whose beam is confined to a thin sheet parallel to and just above the substrate surface.¹⁹⁸ This photochemically assisted CVD process may be very important for eliminating or restricting semiconductor surface modification effects which occur due to the electron and ion bombardment associated with plasma exposure.^{188,189} A final laser assisted form of CVD does not involve gas phase photochemistry, but merely uses thermal energy from absorption of the laser beam to promote localized CVD,¹⁹⁹ and hence this deposition method is included in Figure 48 in the same parameter space as CVD, with the provision that only the deposition region of the surface is heated and not the whole substrate.

4.1.1 Applications of PECVD Materials. The range of deposition temperatures of PECVD make it particularly attractive for use on compound semiconductors susceptible to thermal decomposition. Thus many of the present applications of PECVD are in the processing of the III-V semiconductors, GaAs and InP, in the fabrication of high speed electronic and optoelectronic devices. Table 3 lists a number of these applications. Applications are also to be expected in II-VI semiconductor processing. Present Si processing is in general compatible with CVD temperatures, and since CVD is performed in simple, less-costly equipment, PECVD has not found widespread applications in Si technology.¹⁹⁰ Exceptions to this are PECVD silicon nitride as a passivation coating, and applications where the deposition needs to be performed over a temperature-sensitive organic coating,²⁰⁰ or a low melting point metallization such as Al or Au. Figure 49 shows such a Si processing application in which SiO_2 is plasma-deposited over photoresist. However, as Si device dimensions are reduced even further, lower temperature processing will be necessary in order to prevent diffusion of shallow junctions, and PECVD is expected to find further applications here.

Historically, PECVD has been used mainly to deposit amorphous, dielectric compounds of Si (SiN_x and SiO_2) and doped amorphous, hydrogenated Si itself. PECVD of these materials will be considered in detail in subsequent sections, since these are still the main commercial applications. However, the wide applicability of PECVD is beginning to be exploited. Materials which now have been deposited include single crystal, polycrystalline and amorphous elemental and compound semiconductors, metals, oxides, nitrides, silicides, carbides, and polymers. The specific materials which have been deposited are shown in Table 4, along with the source gases used and a relevant reference.

4.2 General Aspects of PECVD and PECVD Reactors

PECVD is performed by flowing the pre-mixed source gases to be reacted and their inert carrier gases (if any) into a plasma discharge in the vicinity of a heated substrate. Thus the parameters shown in Table 5, along with their typical values, are PECVD variables which must be specified in

Table 3: Applications of PECVD in Semiconductor Process Technology

	Application
Si Technology	Local thermal oxidation mask (SiN_x) Central layer of a tri-layer resist (SiO_2) Interlayer dielectric (SiO_2) Passivation (SiN_x) Contacts (silicides)
III-V Technology - High Speed Electronics - Optoelectronics	MISFET gate insulator on InP (SiO_2) H^+ , D^+ or He^+ implant mask (SiO_2) Anti-reflection coatings (on photo-detectors and LEDs) (SiN_xO_y) (SiN_xH_y) Laser facet coatings - protective and reflectivity modification. Etch Masks (wet and dry) ($\text{SiO}_2, \text{SiN}_x$) Contact area definition ($\text{SiO}_2, \text{SiN}_x$) Diffusion Mask ($\text{SiN}_x, \text{SiO}_2$) Encapsulant/passivation (SiN_x) Auxiliary lift-off layer (SiO_2) Growth Mask in epitaxial growth ($\text{SiN}_x, \text{SiO}_2$). Implant anneal cap (SiN_x).
Energy Technology	Photovoltaic material (a-Si(H)) Anti-reflection coatings on solar cells (SiN_x)
Display Technology	Thin Film Transistor (a-Si(H))

order to define PECVD conditions. It must be stressed that many of these variables are interactive, and that performance of PECVD with good spatial uniformity of deposition rate requires a dynamic balance of these parameters. Thus it is not reasonable to study the dependence of film properties on a given deposition parameter by variation of that parameter alone, since this upsets that dynamic balance (see Section 4.2.3). Instead a more appropriate approach, both from a mechanistic as well as an applications aspect, is to vary pairs of variables in tandem in such a way as to maintain that dynamic balance. For example, in order to study the dependence of deposited film properties on gas pressure during deposition, the gas pressure and the total gas flow (at constant flow ratios) can be varied together in such a way as to maintain spatial uniformity of deposition at a constant plasma power. Variation of this pair can be followed by variation of other pair combinations such as gas pressure and carrier gas flow, or gas pressure and rf power density, always in such a way as to maintain spatial uniformity. This leads to the establishment of sets of simultaneous equations from which the partial differential of film property X with deposition pressure may be derived. Ideally this should be done for more than one combination of the other variables i.e., more than one slice in parameter space should be taken. The above is quite an exhaustive and time consuming procedure.

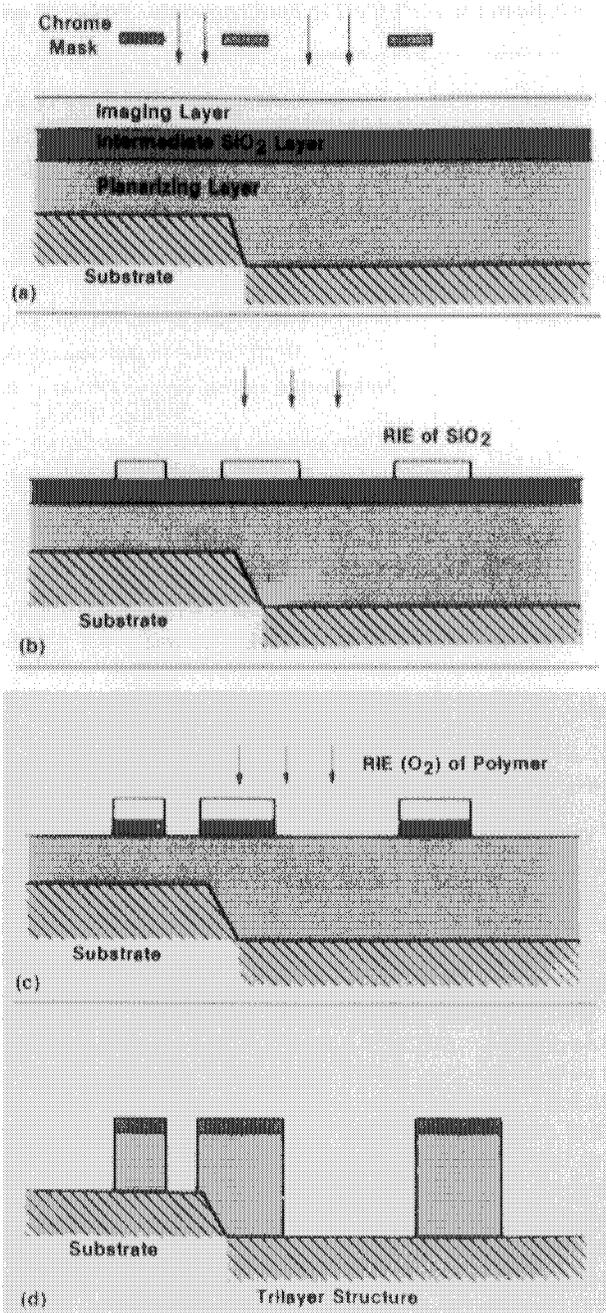


Figure 49: Schematic representation of a tri-layer resist procedure employing PECVD SiO_2 (from Reference 200, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

Table 4a: PECVD of Dielectric Oxides, Nitrides, and Carbides

Material Deposited	Reactants	Reference
Silicon Nitride	SiH ₄ , NH ₃	213, 218
	SiH ₄ , N ₂	217, 221
	SiBr ₄ , N ₂	314
	SiI ₄ , N ₂	315
Silicon Oxide	SiH ₄ , N ₂ O	218, 239
	SiH ₄ , CO ₂	240
	SiCl ₄ , O ₂	241
	Si(OC ₂ H ₅) ₄ , O ₂	242
Silicon Oxynitride	SiH ₄ , NH ₃ , N ₂ O	240
Silicon Carbide	SiH ₄ , C ₂ H ₄ or CH ₄	218
Germanium Oxide	Ge(OC ₂ H ₅) ₄ , O ₂	316
	GeH ₄ , N ₂ O	
Germanium Carbide	GeH ₄ , C ₂ H ₂	317
Tin Oxide	dibutyltin diacetate (CH ₃) ₄ Sn, N ₂ O	316
Boron Oxide	B(OC ₂ H ₅) ₃ , O ₂	316
Boron Nitride	BBr ₃ , NH ₃ (H ₂)	318
	B ₂ H ₆ , NH ₃	319, 320
Aluminum Oxide	AlCl ₃ , O ₂	300
	(CH ₃) ₃ Al, O ₂	297
Aluminum Phosphate	(CH ₃) ₃ Al, PH ₃ , O ₂	297
Aluminum Nitride	AlCl ₃ , N ₂	301
Gallium Oxide	(CH ₃) ₃ Ga, O ₂	302
Gallium Nitride	(CH ₃) ₃ Ga, NH ₃	303, 321
Phosphorus Nitride	P, N ₂	305, 306
	PH ₃ , N ₂	306
Titanium Oxide	TiCl ₄ , O ₂	318
	Ti isopropylate, O ₂	316
	TiCl ₄ , CO ₂	322
Titanium Carbide	TiCl ₄ , CH ₄	304
Iron Oxide	Fe(CO) ₅	316

Table 4b: PECVD of Semiconductors, Conductors, and Elements

Material Deposited	Reactants	Reference
amorphous silicon, a-Si(H)	SiH ₄	246, 256
	Si ₂ H ₆	261
polycrystalline silicon	SiH ₄	278
epitaxial silicon	SiH ₄	281
amorphous germanium, a-Ge(H)	GeH ₄	323
	GeH ₄	285
epitaxial GaAs	Ga, As	286
	(CH ₃) ₃ Ga, AsH ₃	287
epitaxial GaSb	Ga, Sb	288
amorphous carbon, a-C(H)	C ₄ H ₁₀	186, 324
	C ₂ H ₂	325
amorphous boron, a-B(H)	B ₂ H ₆	326
	BCl ₃ , H ₂	327
	BBr ₃ , H ₂	328
amorphous arsenic, a-As(H)	AsH ₃	329
aluminum	AlCl ₃	290
	(CH ₃) ₃ Al	290
tungsten	WF ₆ , H ₂	292
molybdenum	MoF ₆ , H ₂	292
	MoCl ₅ , H ₂	294
	Mo(CO) ₆	293
tungsten silicide	WF ₆ , SiH ₄	295
molybdenum silicide	MoCl ₅ , SiH ₄	294
titanium nitride	TiCl ₄ , N ₂ , H ₂	304
tin oxide	dibutyltin diacetate (CH ₃) ₄ Sn, N ₂ O	316

Table 5: Variable Parameters in PECVD

<i>Direct Variables:</i>	
Parameter	Typical Value
Reactant Gas Flows	1-1000 sccm
Reactant Gas Flow Ratios	1-100
Total Gas Flow (also gas flow pattern)	50-5000 sccm
Electrode Spacing	2-4 cm
Gas Pressure	200-2000 mtorr
RF Power Density	0.03 - 0.5 W cm ⁻²
RF Frequency	25 KHz - 25 MHz
Substrate Temperature	200 - 400 °C

Resultant Variables:

Deposition Rate
 Film Composition
 Uniformity of Rate and Composition
 Film Properties

Similar situations in plasma etching are being treated by Mocella et al.¹²⁹ using the statistical technique of Response Surface Methodology to generate a model parametric expression of the process. Such an approach has the potential to drastically reduce the number of experimental data points needed to optimize a multi-parameter process, and therefore its application to PECVD would be very beneficial. Thus deposition parameters may be selected to optimize a specific film property for a given processing application, and the sensitivity of that property to small variations in each parameter established in order to determine the necessary levels of parameter control.

4.2.1 Reactor Designs. All plasma deposition systems consist of the following components: gas sources, gas flow controllers, a gas manifold and distributor, a plasma chamber incorporating a heated substrate table and pressure monitoring, an rf generator, a pumping system including throttle valve, and an exhaust system. This is shown schematically in Figure 50. In commercial systems, gas flow control employs electronic mass flow controllers which can maintain absolute flows or fixed flow ratios, pressure monitoring is by species-independent capacitance manometers, and the pumping throttle valve is servo-controlled to maintain a constant chamber pressure. Many systems now employ microprocessor control.

It is the design of the plasma chamber itself, in particular the electrode and gas flow geometries, which distinguishes the various types of PECVD

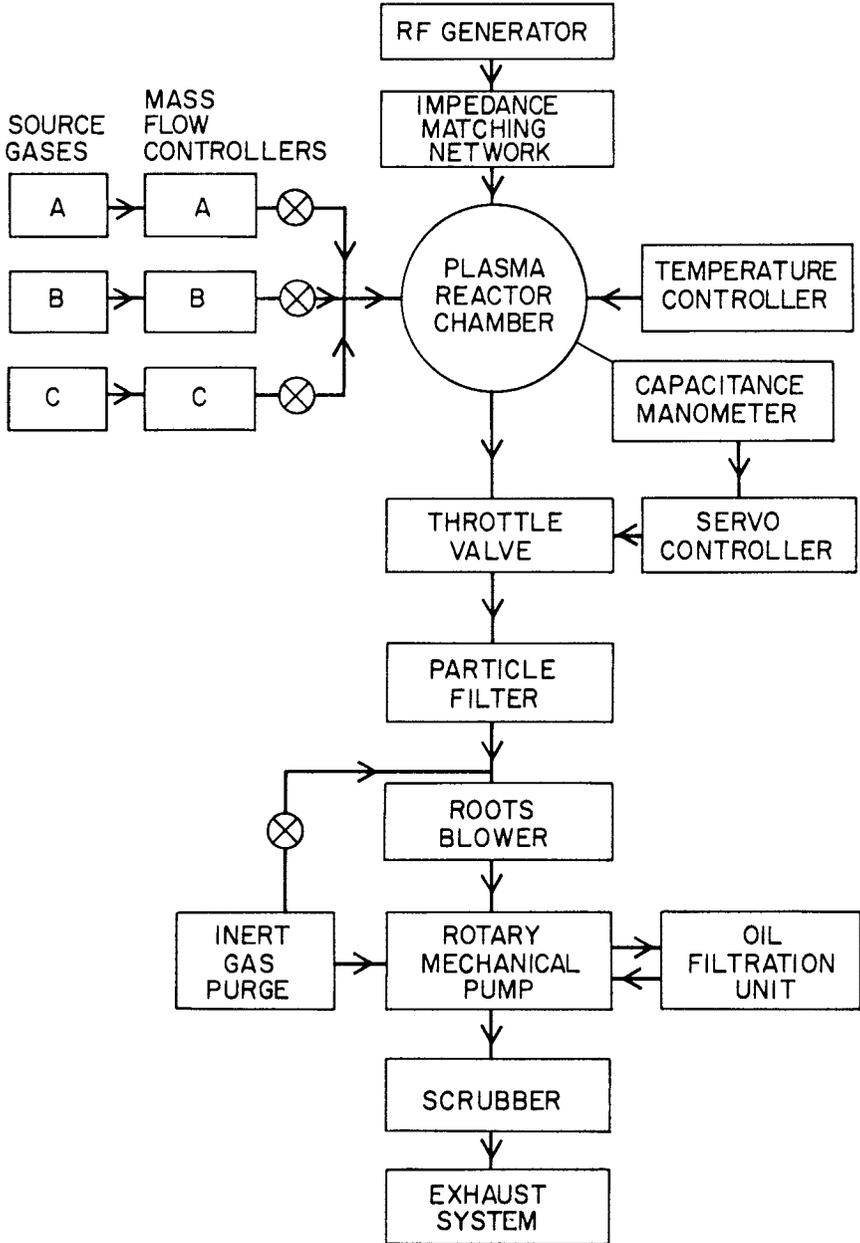
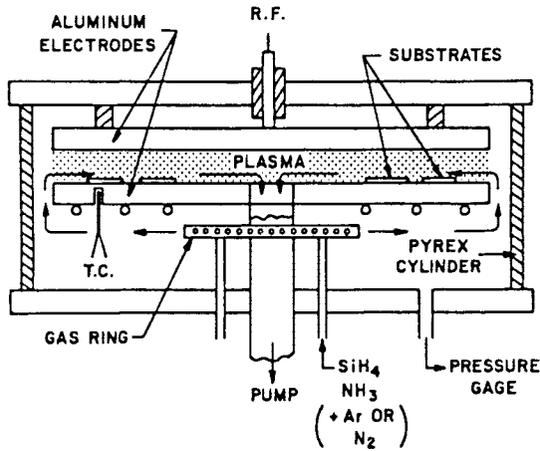


Figure 50: Schematic representation of the components of a PECVD system.

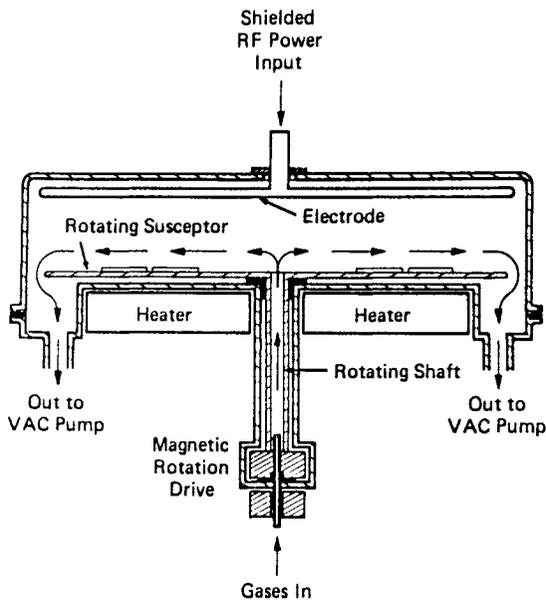
reactor. The three main categories are shown schematically in Figure 51, along with the relevant sub-categories. The class (c) shown is to some extent a sub-division of class (a) in that an individual pair of electrodes is parallel-plate, but since multiple pairs of electrodes distributed in multiple columns along a tube which is enclosed in a diffusion-style furnace are involved, it is a sufficiently different concept to merit separate description.

The parallel-plate, radial flow reactor shown in the first class was designed by A.R. Reinberg,^{134,201} for silicon nitride deposition, and is sometimes referred to as a Reinberg reactor. His original design [(a)(i)] employed inward radial flow; a later variation,²⁰² using outward flow is also shown [(a)(ii)]. The radial-flow reactor is the most commonly employed for plasma deposition. Electrode diameters are usually in the range of 25 to 55 cm, and batch processing is used. Whereas single wafer processing has certain merits for plasma etching (see Section 3.2.3), it is not a viable alternative for plasma deposition due to deposition rates (for acceptable film properties) being rather lower than etch rates that can be employed. The larger reactors are normally used for Si processing, and can accommodate about 20 four inch wafers. The smallest reactors are more than adequate for use in present III-V compound semiconductor technology. A typical process time from wafer loading to removal is about two hours, depending on at what temperature it is permissible to load wafers on to the substrate table. If native oxide growth on the surface of a III-V semiconductor wafer is to be avoided or at least limited, it is necessary that the substrate table be no more than a few tens of degrees above room temperature during wafer loading. This can significantly increase process time, particularly in the case of the larger reactors with substrate tables of large thermal mass. Use of a wafer carrier plate to give a thermal delay slightly longer than the pump-down time circumvents this problem. The final variety of parallel-plate reactor, shown in Figure 51 as (a) (iii), is the "shower head" variety, which employs a perforated upper electrode through which the reactant gases are introduced into the plasma. An advantage of this scheme is that the lower electrode (substrate table) is a continuous plate, in contrast to the annular geometry required for radial flow. A disadvantage is that cooling of the perforated, powered electrode is difficult, sometimes necessitating pulsed power plasma operation.

The second type of reactor is the tube or barrel reactor, into which the rf power usually is inductively coupled, by a coil around the tube, external to the plasma region. This type of reactor is shown as (b) in Figure 51. Capacitive coupling via external electrodes is also possible. As before, the reactor is coldwall. This type of reactor is very simple and lends itself to process research studies, but is not suitable for uniform, batch deposition needed in a production environment. However, it is particularly suitable for indirect plasma studies in which the substrate is not directly exposed to the plasma, but is mounted downstream from the glow region. In this way reactive radicals and atoms, in both excited and ground states, can arrive at the heated substrate surface if their lifetime is sufficiently long. Since the substrate is in a field-free region, energetic ion and electron bombardment is avoided. This is beneficial for avoiding or restricting substrate

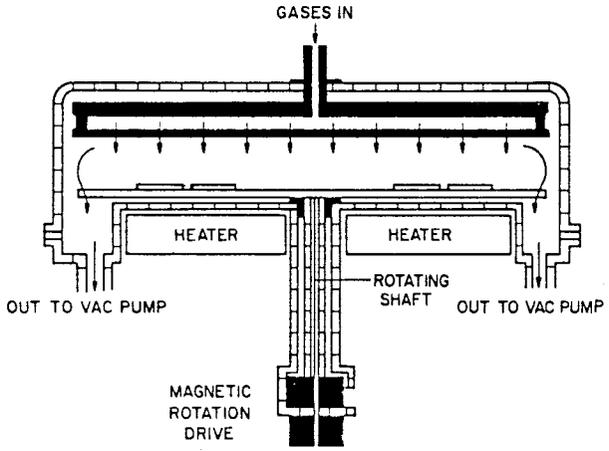


a i

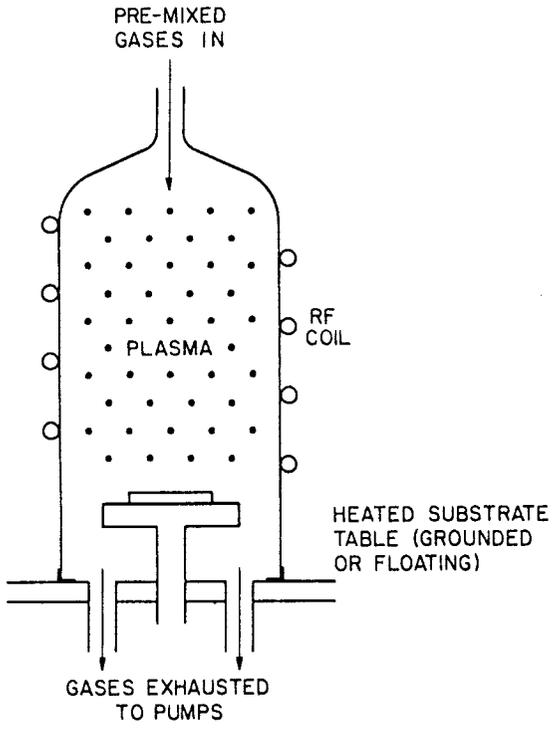


a ii

Figure 51: The main types of PECVD reactors (a) parallel-plate with (i) Reinberg-design inward radial flow (from Reference 183, reprinted with permission of the American Institute of Physics) (ii) modified Reinberg-design with outward radial flow (from Reference 202, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet), (iii) shower head gas distribution (b) inductively coupled tube and (c) hot-wall (from Reference 203, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

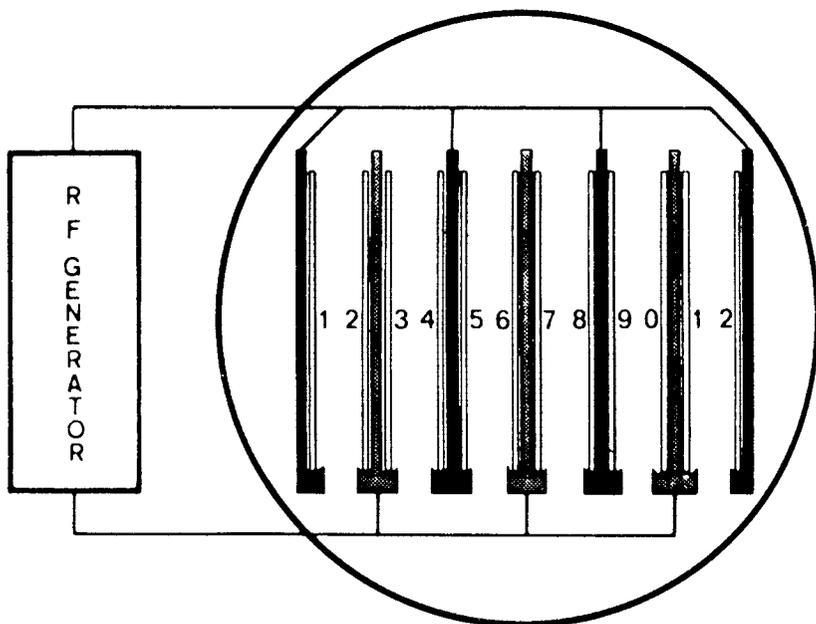
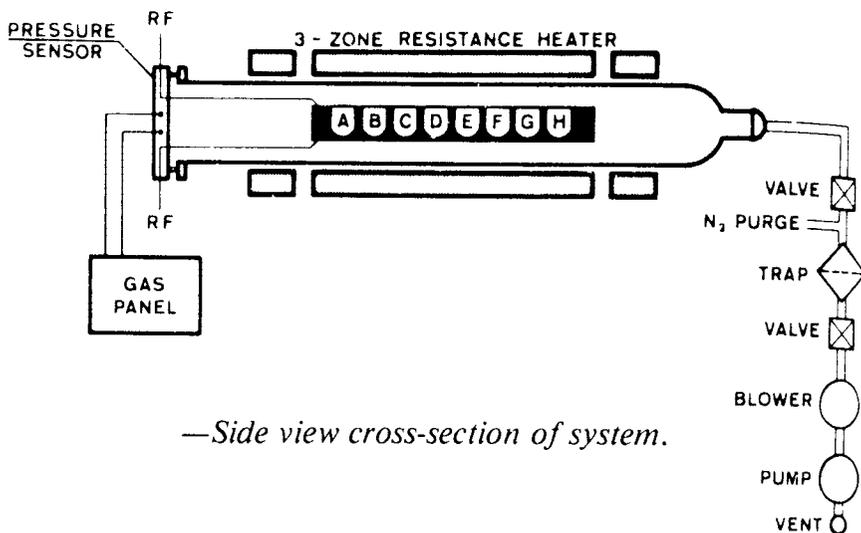


a iii



b

Figure 51: (continued)



c

Figure 51: (continued)

damage effects, but may not be beneficial to film properties (see Sec. 4.2.4).

The final type of reactor is the hot wall tube,²⁰³ shown as (c) in Figure 51. This is basically a diffusion furnace tube into which is inserted a multiple array of parallel-plate electrode pairs, usually made of carbon. Each grounded electrode can carry a single wafer in a vertical orientation. This arrangement is suitable for large, regularly shaped Si wafers, but is not suitable for the smaller and often irregularly shaped and sized III-V wafers. An advantage of this arrangement is its large wafer capacity; a commercially available system has a batch capacity of 84 four inch wafers. However in many applications its process cycle time is rather longer than that of the radial flow reactors. Since reactants are introduced at one end of the tube and both become depleted and are accelerated down a pressure gradient (thus reducing residence time) as they flow down the tube, it would appear that the only way to achieve uniform deposition is to use a large excess of reactants and hence operate at low efficiency, a possibly costly operation if very high purity SiH_4 is being used. One variation of this type of reactor²⁰⁴ pulses the applied rf power to prevent downstream depletion of reactants. Commercially available PECVD reactors have recently been reviewed.²⁰⁴

A major reactor design consideration not yet discussed is the frequency of the rf plasma. The frequency range over which reactors have been operated (≈ 30 KHz to ≈ 30 MHz) can be split into two distinct regimes, as discussed in section 2.1. In one regime, which we will refer to as low frequency rf, both ions and electrons respond to the rf field. Thus in one half-cycle of the applied rf voltage, positive ions are extracted from the glow region and accelerated across the sheath above the substrates on the grounded table. Due to the fairly high pressure employed for PECVD (≈ 1 torr), most of these ions suffer collisions during acceleration through the sheath. Nevertheless, there is a flux of energetic ions incident on the substrate with an energy distribution whose high energy tail extends as high as the amplitude of the rf voltage, which may be a few hundred volts. This is illustrated in figure 10. The width of this energy distribution depends on pressure, gas species, rf power etc., and can be as large as a few hundred eV. It is this directional ion flux which is responsible for anisotropy and enhanced etch rates in low frequency plasma etching (see, for example, Reference 140), as discussed in Section 2.2.2.1. This regime of operation extends up to a few MHz, with the exact upper limit being determined by the ion masses, pressure, etc. Above this transition frequency, we are in the high frequency rf regime, in which the inertia of the ions prevents them from responding to the rf field which is followed only by the electrons. Although there is essentially no energetic (>50 eV) ion bombardment of the substrate, there remains a high flux of low energy ions (≈ 25 eV), as also shown in Figure 10 due to the small positive dc potential of the glow region, in addition to the energetic electron bombardment. This low energy ion bombardment is also present at low frequency. The difference in extent and energy of ion bombardment fundamentally changes bulk film properties, film/substrate interface properties and in some cases deposition rates.

All the types of reactors discussed can be operated at high or low frequency, although high frequency (13.56 MHz) is generally used for tube

reactors. Low frequency operation is simpler in that impedance matching is accomplished by a variable transformer, whereas high frequency impedance matching requires an L,C network. Possibly because of the difficulty of impedance matching the complex electrode arrangement in a hot-wall reactor, low frequency is usually used in that case.²⁰³

An interesting recent development of a PECVD method has been christened electron beam assisted chemical vapor deposition (EBCVD).^{205,206} In this technique, a 2.5 cm wide, shallow electron beam of a few keV energy is injected into the reactive gases parallel to and a few millimeters above the substrate surfaces, in much the same way as an excimer laser beam was used by the same group for laser photo assisted CVD.¹⁹⁸ A localized plasma is created from which excited and reactive radicals diffuse to and react on the heated substrate surface. Source gas compositions and flows, pressure, and substrate temperature are all as for conventional PECVD. However, the advantage of the method is that the substrates do not sit on one of the electrodes responsible for maintaining the plasma. Since there is only small angle scattering of the injected electrons, the substrate surface is largely free of energetic electron bombardment. However, since the substrate carrying electrode is a boundary to the localized plasma generated above it, the glow region generally acquires a small positive bias relative to the substrate and a sheath region is formed, as described in Sec. 2.1. Thus the substrate surface is subjected to low energy, positive ion bombardment. The energy distribution of this ion bombardment is expected to be similar to that originating from a conventional, high frequency rf plasma (Figure 10), although the ion flux may be lower. Thus the technique has some of the benefits of indirect plasma deposition, but without the substrate being so remote from the plasma. Therefore, the penalty of reduced deposition rates is avoided. As with photo assisted CVD, this technique will be looked at for applications where electrical and optical modification of the semiconductor substrate surface must be minimized.

4.2.2 Source Gases. A source gas for a given element to be incorporated into a PECVD film needs to be a gaseous or volatile compound of that element. If a mixture of source gases is to be used, as is usually the case, the combination must be chosen to ensure that the component gases do not react in the absence of a plasma, otherwise gas-phase reaction will occur in the mixing manifold, giving rise to particulate incorporation and pinholing in the deposited film. The specific source gases which so far have been employed in various depositions are shown in Table 4. Note that inert gases (Ar, Ne or He) or reducing gases (e.g., H₂) are frequently used as diluents or carrier gases. Source gases may be obtained pure, or diluted, in compressed gas cylinders. Volatile liquids or solids are supplied in bubblers which are mounted in thermostatically controlled baths to control the vapor pressure. Vapor may be extracted directly or transported by carrier gas flow. Another technique which has been used occasionally is that of plasma transport, in which a plasma is used to decompose or etch a solid and the vapor-phase product is used as a source for a plasma deposition downstream (Sec. 2.2.2).

4.2.2.1 Safety Aspects. Unfortunately, a large number of the gases

commonly used in PECVD are hazardous, being either highly toxic and/or pyrophoric. Gases such as AsH_3 and PH_3 , used in III-V semiconductor growth and as dopant sources for amorphous silicon and P-doped SiO_2 , are in the first category. Gases such as SiH_4 and Si_2H_6 and the organometallics and metal carbonyls are in the second category. Thus stringent safety precautions are necessary in their use.^{207,208} This usually includes the use of exhausted gas cabinets, cross flow or body purge regulators and dedicated purge gas supplies, and stainless steel plumbing with welded joints where possible and metal gasket, gland type couplings where demounting is necessary. In some cases local safety regulations may also require the use of co-axial plumbing with inert gas purging of the outer segment. In other cases, maximum concentration limits may be imposed for the diluted gas mixtures which may be used.

The next safety consideration is the choice of pump fluid.^{206,207} Use of the inert perfluoropolyethylene pump fluids is generally the best choice, even in view of the expense. It is mandatory to avoid the risk of pump explosion if pure O_2 is one of the source gases, or if the plasma process can produce a significant amount of O_2 . Plasma deposition chambers are cleaned by plasma etching, generally using the fluorocarbon gases described in section 3. Etching is efficient due to the large quantity of F atoms produced in these freon plasmas. Since fluorine reacts with hydrocarbon oils, this is another reason to avoid their use. In PECVD processes, gas phase reaction between excited radicals continues downstream into the pumping port, with resultant fine particulate formation in the pumping lines. Many materials deposited by PECVD are refractory in nature, and in fine particle form are highly abrasive. Thus it is necessary to employ high conductance particle filters between the pumps and chamber. In addition, oil filtration units on the mechanical pumps are necessary, both for fine particulate removal and for acid neutralization to prevent pump corrosion. Other necessary pump precautions are to gas-ballast the pump with an inert gas, beginning this about 15 min prior to pumping the hazardous gases and continuing it until well after the deposition is complete. Final safety considerations are that plumbing on the exhaust side of the pump must also be of high integrity and all metal construction, since this has to carry unreacted hazardous gases as well as possibly hazardous reaction products. In the case of toxic gases, appropriate scrubbers are also necessary in the exhaust line, followed by monitoring of the effluent gas from the scrubber. Toxic gas monitors should also be available for personnel safety in the vicinity of PECVD systems.

4.2.3 Uniformity Considerations and Interaction of PECVD Variable Parameters. It is by no means simple to model deposition rate as a function of position in a PECVD reactor. Fortunately it is a relatively straightforward task to establish empirically a particular set of conditions which give spatial uniformity (assuming the reactor design itself is capable of giving uniformity). To obtain uniformity it is necessary to strike an appropriate balance between reactant supply rate and film deposition rate. This balance is a function of the other deposition parameters.

For example, the deposition rate is primarily controlled by the r.f. power supplied to the plasma, with a sub-linear increase in deposition rate

towards an asymptotic limit determined by the rate of supply of reactants. Increasing pressure at a fixed flow of reactants will increase deposition rate at a given rf power level, due to an increased residence time of reactants within the plasma. Increasing the substrate temperature may increase or decrease the deposition rate since the sticking coefficients decrease, but reaction probabilities of the reactive radicals arriving on the substrate increase. Thus if one has determined the substrate temperature, the reaction pressure and the ratio of reactant flows, based on the film properties required, uniformity is then achieved by varying the total reactant flow at a fixed power level, or by varying the power level at a fixed total reactant flow. This is illustrated in Figure 52, for an inward radial-flow reactor.

If one has a situation as shown in curve (a), one has three choices to achieve radial uniformity. If the deposition rate at the outer part of the substrate table is that desired, then it is necessary to keep the plasma power constant and increase the reactant flows, while keeping the pressure constant. Since the reactant concentration at the periphery of the table, where the gases enter the plasma, is unchanged, the deposition rate there is unaffected. Thus the rate of removal of reactants at the periphery is unchanged, but the supply rate is higher, so that the concentration of reactants arriving at mid-radius is increased, and hence so is the deposition rate there. At the appropriate flow levels, uniform deposition is achieved at the rate in existence at the outer-radius position, as shown in curve (b). Alternatively, this effect may be understood in terms of residence times. Increasing reactant flow rates at constant pressure requires increased

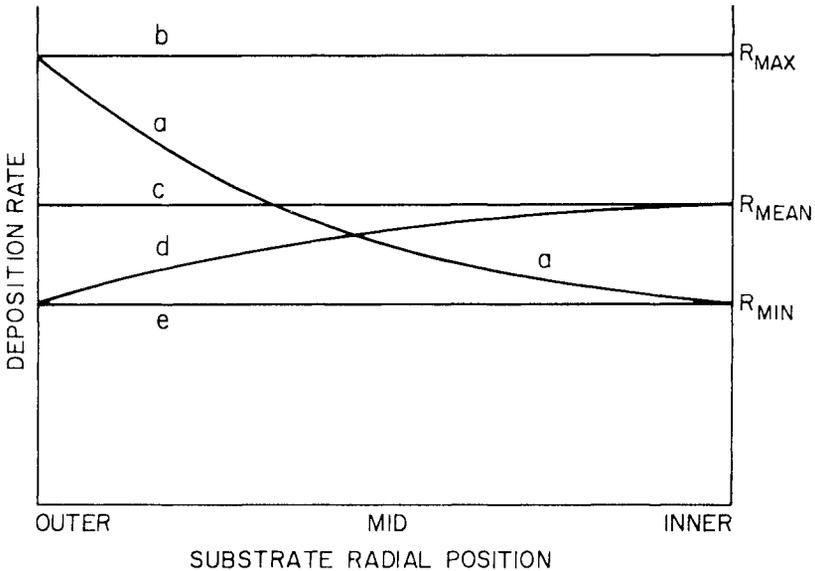


Figure 52: Deposition rate as a function of radial position, in an inward radial-flow reactor, for a variety of deposition conditions (see text).

pumping speed, and hence reduced residence times. Thus at the periphery, the higher supply rate is compensated by the reduced residence time, so that the absolute reactant depletion (and hence deposition rate) is unchanged, but the fractional depletion is reduced. As discussed in section 2.1.4, the reduced residence time may modify the homogenous and therefore also the heterogeneous chemistry, possibly influencing deposited film properties.

Conversely, if the deposition rate at the inner-radius position is that desired, then it is necessary to reduce *both* reactant flows and rf power. Reduction in rf power alone will reduce the deposition rate at the outer-radius position, but radial uniformity will be achieved at the deposition rate shown in curve (c), which is the mean deposition rate of curve (a). In other words, reduction in rf power does not reduce the overall consumption of reactants in the regime between curves (a) and (c), but continued reduction in rf power below the level producing radial uniformity as shown in curve (d) reduces mean deposition rate and consumption efficiency. In this regime, reduction of reactant supply rates can re-achieve uniformity (curve (e)) at deposition rate R_{\min} and restore consumption efficiency.

Uniformity can also be achieved by varying the pressure, and this is most easily understood in terms of residence time. At fixed reactant flows, an increased pressure corresponds to an increased residence time. Thus beginning with the situation depicted by curve (a), reduction in pressure can achieve uniformity as shown in curve (c). Conversely, an increase in pressure will produce even greater non-uniformity than that of curve (a). However, if one has non-uniformity as in curve (d), an increase in pressure can be used to produce the uniformity of curve (c). This also demonstrates that at a higher pressure, a lower plasma power is required to produce uniformity at the same deposition rate.

Experimental data exactly as depicted in Figure 52 are obtained for the PECVD of SiO_2 from dilute SiH_4 in Ar and N_2O source gas mixtures (see Sec. 4.3.2). In this reaction the heterogeneous reaction rates are rapid, so that the homogeneous chemistry is rate limiting (this assumption is implicit in the discussion of Figure 52). In a case where the heterogeneous reaction rates are slower, as for SiN_x deposition, the points made in the above discussion are relevant to the reactant supply rate rather than the resultant deposition rate. This tends to make deviations from uniformity less severe. Reinberg¹⁸⁴ has given a detailed discussion of spatial uniformity considerations, particularly from the standpoint of SiN_x deposition.

The level of radial uniformity of deposition rate achieved by the above methods in an inward radial flow reactor is shown in Figure 53 for three different deposition rates. This data is for SiO_2 deposition from a 13.56 MHz plasma fed by 3% SiH_4 in Ar and N_2O . Power density and total flow is indicated adjacent to each curve. At a deposition rate of 400 \AA min^{-1} , rate uniformity can be $\pm 0.5\%$ as shown, while $\pm 4\%$ can be achieved routinely. At deposition rates of 700 \AA min^{-1} and $1100 \text{ \AA min}^{-1}$ uniformities of $\pm 5\%$ and $\pm 7\%$ respectively can be achieved routinely. It can be seen that at the highest deposition rate shown, when deposition rates at the inner and outer radii are matched, the deposition rate in the center of the table is about 7% lower. This is thought to be a result of the gas flow pattern

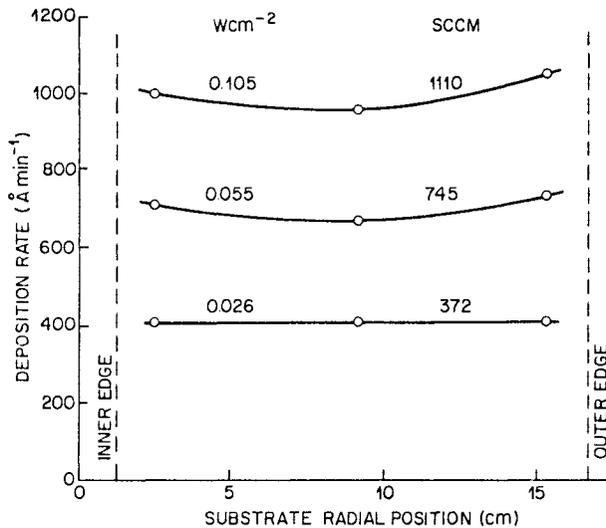


Figure 53: Actual radial uniformity achieved for PECVD of SiO_2 from 3% SiH_4 in Ar and N_2O at three different deposition rates, at plasma power densities and total gas flows as indicated. The area enclosed between the outer and inner radii is 85% of the substrate table area. (From the authors' laboratory.)

between the parallel-plate electrodes at this high flow. Gas is injected around the periphery of the lower electrode and extracted through a pumping port in the center of the lower electrode. Thus in the mid-radius substrate position, the peak of the flow distribution is closer to the upper electrode, making the mean diffusion distance for dissociated, reactive species to reach the substrate slightly larger.

4.2.4 Film Properties and Their Control. One of the attractive features of PECVD is its highly non-equilibrium nature and its large number of variable parameters. This provides wide possibilities to tailor materials properties to a given application. In addition to being performed at a lower temperature, relative to conventional CVD, another advantage is that the deposition rate is decoupled from the substrate temperature, allowing substrate temperature to become a property determining variable. A naturally attendant disadvantage is that it may be difficult to reproducibly achieve the desired properties. Even though PECVD of SiN_x has been in use for about a decade, it is still much more of an art than a science. General film properties which can be varied, and the main methods employed to measure that property, are shown in Table 6.

Film stoichiometry is controlled primarily by the ratio of source gas flows, which may be very different from the film stoichiometry. Plasma power and frequency, and substrate temperature are secondary variables for determining stoichiometry. Refractive index and wet etch rate (in an aqueous acidic etchant, such as buffered hydrofluoric acid) are both properties which are very much a function of stoichiometry, and hence are

Table 6: Film Properties and Measurement Techniques

Film Property	Measurement Technique
Stoichiometry	Rutherford Backscattering Auger Electron Spectroscopy
Hydrogen Content	Infra-red Spectroscopy Nuclear Reaction Analysis
Refractive Index	Ellipsometry Interference Effects
Etch Rate	Ellipsometry or stylus step height measurement
Stress	Induced Curvature of Substrate - Optical lever or X-ray
Adhesion	Ramped Pull to Failure, Scribing, Saw Cutting
Conformality	Scanning Electron Micrograph of step coverage
Pinhole Density	Use as a chemical or electrochemical etch mask followed by optical microscopy

determined by the same deposition parameters. Both refractive index and wet etch rate are also affected by film density. Refractive index measures polarizability per unit volume, and thus is related to density (by the Lorentz-Lorenz relationship),²⁰⁹ and so increases with increasing deposition temperature. Etch rate decreases with increasing film density, and is often related to hydrogen content. Since density increases and H content decreases with increasing deposition temperature, this produces a decrease in etch rate. While substrate temperature is the primary variable for varying H content, plasma power density, degree of ion bombardment during growth (determined by operating frequency and reactor chamber pressure), and the nature of H bonding in the source gas all influence the H content of the film.

Stress is a film property which is currently attracting much attention. Film stress is dependent on the nature of the substrate, in terms both of film nucleation effects and of mismatch of thermal expansion coefficients, as well as on deposition conditions. Thus, in general, film stress is affected by deposition rate in relation to substrate temperature, substrate temperature itself, and the extent of ion bombardment during growth. As a general empirical observation, increased ion bombardment causes the film stress to become (more) compressive. Compressive stress in a film produces convex curvature of the substrate (viewed from the coated side), so that the substrate in the region close to the film is under tensile strain. High tensile stress produces poor crack resistance and increases the film/substrate adhesion needed to prevent film peeling, whereas high compressive stress can cause blistering of the film. Low compressive stress usually indicates good crack resistance in a film, a desirable property for most

processing applications. High stress of either sign can also produce a number of other undesirable effects, such as enhanced defect migration rates, anisotropic dopant diffusion profiles, modified etch profiles, and induced interface electronic states. Thus low compressive stress is the usual goal in the design of a deposition process. In this respect, PECVD has an advantage over CVD, in which materials such as SiN_x are deposited under high tensile stress.¹⁹⁰

Increased ion bombardment of the substrate and growing film is also beneficial for improving adhesion at the film/substrate interface. The origin of this effect is both physical and chemical. Increased flux and/or energy of ion bombardment increases physical intermixing at the interface by momentum transfer processes, and also promotes additional chemical reaction or chemically driven interdiffusion. Increased substrate temperature also increases adhesion by the latter mechanism.

Ion bombardment can impart anisotropy to deposition processes in a manner similar to that for plasma etch processes (see Sections 2.2.2.1. and 3.2). Two anisotropic effects are possible in deposition, namely anisotropy of deposition rate or anisotropy of film properties. PECVD intrinsically shows a 2:1 anisotropy in deposition rate between parallel and normal to the electrode surfaces, even for isotropic arrival of depositing species. This is discussed in the following sub-section regarding step coverage. However, if ion bombardment has the effect of enhancing heterogeneous reaction rates and thus rate of film growth, the degree of anisotropy is increased. Similarly, film properties as above which are modified by the degree of ion bombardment will be expected to show anisotropy, although we are not aware of any reports in the literature. For example, the wet etch rate of a film deposited on a vertical wall could be a few times larger than that of the film on the planar, horizontal surface. If the thickness on the wall is also more than a factor two lower, this could give rise in a wet etch process to the vertical wall becoming "cleared" an order of magnitude more rapidly than the planar surface.

Conformality of coating is a general PECVD property claimed by most workers except Adams.¹⁹¹ The difference appears to be the exact definition of conformality. By conformality, most workers mean good step coverage without discontinuity over a sharp vertical step, which is either isolated or at a distance larger than the step height from an adjacent step. By this definition, PECVD does give conformal coatings, often with better step coverage than CVD¹⁸⁵ for the stress reasons discussed above. This step coverage can readily be seen from consideration of the arrival angles of the reactive species at the substrate surface. At a typical PECVD pressure of 1 torr, mean free paths are about 50 μm for the reactive atoms and radicals. Thus these species have suffered many collisions in the gas phase before arriving on the substrate surface, and therefore have a random distribution of arrival directions. This gives rise to a deposition rate proportional to the solid angle of the plasma presented to a given element of surface, so that in the absence of macroscopic surface migration film thickness on an isolated vertical wall is half that deposited on the horizontal substrate surface, as shown in part (a) of Figure 54, and part (b) of Figure 55. As the channel width is reduced, this type of coverage produces the

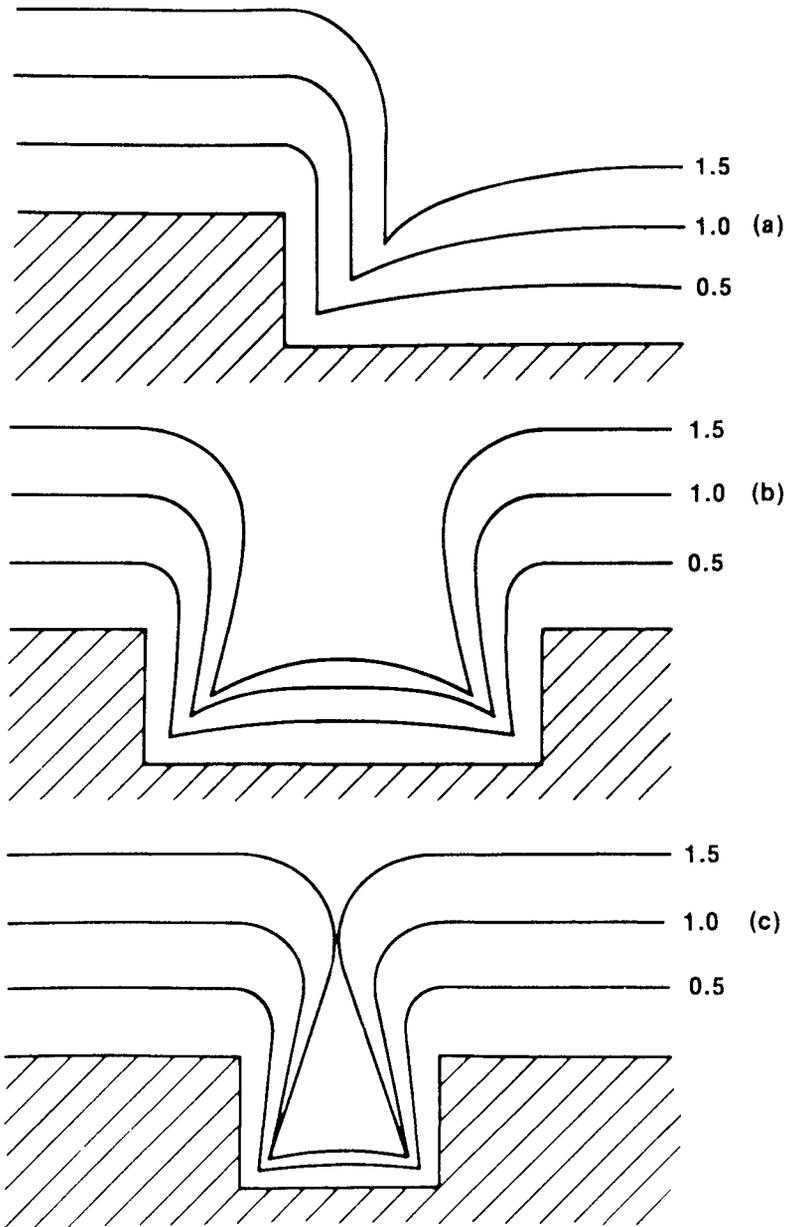


Figure 54: Calculated step coverage over an isolated vertical step (a), and over vertical-walled channels of increasing depth-to-width ratio (b and c). Film thickness is indicated as a multiple of the step height (from Reference 191, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

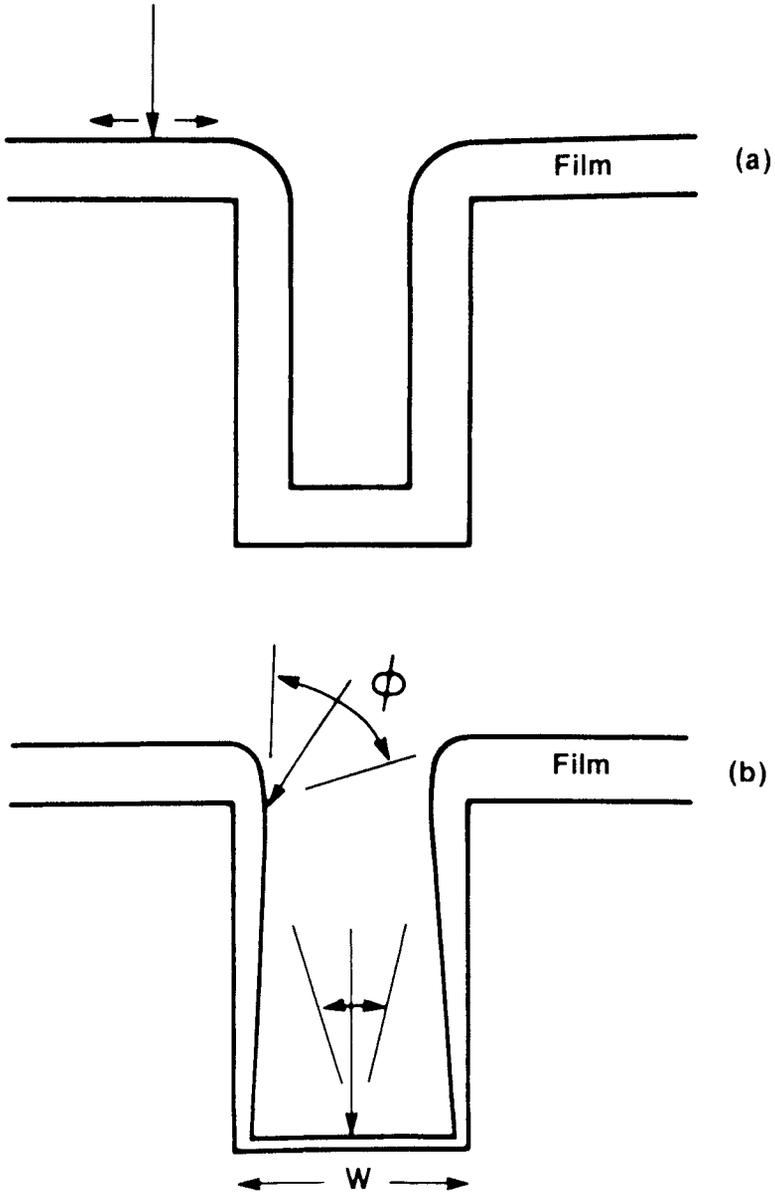


Figure 55: Schematic step coverage (a) conformal, resulting from rapid surface migration, and (b) with coating thickness determined by solid angle presented to the plasma, for isotropic reactant species arrival, with a mean free path longer than the feature size, and for surface migration small relative to the feature size (from Reference 191, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet).

profiles shown in parts (b) and (c) of Figure 54. Adams¹⁹¹ has shown that the step coverage of 13.56 MHz PECVD SiO_2 , SiN_x and a-Si follows this profile shape, implying the lack of surface migration on the scale of a few hundred nanometers. Thus coatings which are of uniform thickness regardless of surface topography, as is shown schematically in part (a) of Figure 55, are not produced by PECVD. Instead, coatings of thickness determined by the acceptance angle from which reactants are arriving with random directional distribution, with no significant subsequent surface migration, are produced, as indicated in part (b) of Figure 55. This type of step coverage is much better than that obtained from lower pressure techniques such as sputtering in which there is a significant directionality to the depositing species. Thus PECVD produces coatings with good step coverage, but these coatings are not conformal in the strictest sense.

Films with very low pinhole densities less than 1 cm^{-2} may be produced by PECVD.¹⁸⁵ Certain deposition conditions need to be met to achieve this level, dependent on the specific material and source gases involved; this will be discussed in the relevant subsequent section. There are a number of general practices necessary to obtain low pinhole density films. Substrate cleaning and wafer handling techniques are obviously of paramount importance, as is reactor cleanliness. Reactor cleaning is usually carried out by plasma etching. This should be performed before the deposited thickness on the electrodes and chamber walls is such that flaking begins to occur; for high stress films this occurs at smaller thicknesses. Coating the inside of the chamber with a thin layer of the material to be deposited directly after the plasma-etch cleaning is a worthwhile judicious practice. Non-turbulent rough pumping of the chamber down from atmospheric pressure is another beneficial practice.

4.3 Materials Deposited and Their Applications

4.3.1 Silicon Nitride. PECVD silicon nitride is not one material, but a whole family of materials. PECVD silicon nitride films, unlike those deposited by normal CVD, are not confined to the stoichiometric Si_3N_4 composition. They also contain much larger amounts of hydrogen, typically 15 to 30 atomic %. Frequently they will contain small amounts of oxygen, particularly close to the substrate interface, gettered in the initial stages of deposition. Thus a correct compositional description is $\text{SiN}_x\text{O}_y\text{H}_z$; however, for the sake of simplicity, we will write this in the abbreviated form SiN_x .²¹⁰

Properties of PECVD SiN_x which make it attractive for the semiconductor processing applications shown in Tables 3 and 7 are its hardness and good scratch resistance, its diffusion barrier properties (to H_2O , Na^+ , and most dopant atoms), its good adhesion to silicon, III-V compound semiconductors, other dielectrics and most metallizations, including Au, and its rapid and uniform plasma etching characteristics. These material specific properties are in addition to the general PECVD properties discussed in the preceding section. Schuermeyer²¹¹ has reviewed a number of these applications of SiN_x in GaAs processing, and Vanner et al.²¹² have described applications in GaAs integrated circuit fabrication.

There is now a wealth of literature concerning PECVD SiN_x , its properties and their dependence on deposition conditions. Good general coverage

of these topics has been given by Sinha et al.,^{209,213,214} Reinberg,¹⁸⁴ Mar and Samuelson,²¹⁵ Rosler and Engle,²⁰³ and van de Ven.¹⁸⁵ Maes et al.²¹⁶ have recently reported a comparative study of properties of films deposited in a wide range of commercially available PECVD systems.

4.3.1.1 Source gases and operating frequency. Source gases usually employed are SiH_4 for Si and NH_3 and/or N_2 for N. Inert carrier gases or diluents are often used also; the use of He has recently been reported²¹⁷ to be beneficial to deposition uniformity. N_2 is generally used in place of NH_3 in an attempt to reduce the H content of the film; N_2 can be obtained at higher purity than NH_3 , and also is non-toxic. A disadvantage of the use of N_2 is that the energy to produce an electron-ion pair is higher than that from NH_3 ,¹⁸⁸ and the N_2 dissociation kinetics are much slower than those of NH_3 .¹⁸⁷ This means that an increased N-source to SiH_4 flow ratio is necessary (discussed below), with either an increased rf power density or a reduced deposition rate. The latter point is demonstrated in Table 8, in which the ratio of deposition rate to plasma power density (obtained from data reported by various groups) is shown. This plasma-power-normalized deposition rate, in conjunction with the operating frequency, inversely indicates the dose of energetic incident species received by the substrate/film interface in the initial stage of film deposition. In addition, it is a simple inverse indicator of the degree of plasma enhancement required by the reaction in question. At both high and low rf frequency, this ratio is significantly smaller when N_2 is employed in place of NH_3 . This is an important consideration regarding physically-induced interfacial modification (see Sec. 4.4). It may also be observed from Table 8 that low frequency deposition, from either NH_3 or N_2 sources, produces higher deposition rates, suggestive of an ion assisted surface reaction. This is further discussed below in regard to the lower N_2/SiH_4 flow ratio required at low frequency. The first reported plasma deposition work, two decades ago, deposited SiN_x from SiH_4 and NH_3 .²¹⁸ This was followed a few years later by deposition from SiH_4 and N_2 .^{219,220} The radial flow reactor²⁰¹ was developed for SiN_x deposition.

In determining which process or process conditions to employ for SiN_x deposition for a given application, it is essential to prioritize the film (or interface) properties needed for that application. For example, for an antireflection coating application, refractive index is the primary requirement, but low stress and perhaps a low density of induced interface states may also be desirable. Table 7 indicates the necessary film properties for various processing applications in which SiN_x has been used. It is also necessary to identify which of these properties are dependent more on pre-deposition surface cleaning procedures than on the deposition conditions themselves. Having established the required film properties, the necessary deposition conditions may then be determined. Early decisions to make are whether to use NH_3 or N_2 as the N source gas, what frequency plasma is to be used, and whether pure SiH_4 or a dilute ($\leq 5\%$) SiH_4 in inert gas or N_2 is to be used. This option scheme is shown in the flow chart of Figure 56. Rarely is a choice clear-cut, in that the same result can be obtained by more than one combination of these three parameters, particularly when one allows for variation in the subsequent deposition parameters

Table 7: SiN_x Film Property Requirements for Processing Applications

Processing Application	SiN _x Property Requirements
Device Passivation	Diffusion barrier to H ₂ O, Na ⁺ etc; low stress, good adhesion; good step coverage; pinhole free; low density of induced surface states.
Diffusion Mask	Diffusion barrier to the dopant in question; low stress;* thermal stability; good adhesion; pinhole free;
Implant Mask	Low stress, good adhesion.
Epitaxial Growth Mask	Low stress,* thermal stability; good adhesion.
Etch Mask	Low stress; good adhesion; pinhole free.
Implant Anneal Cap	Low stress*; good adhesion; thermal stability; pinhole free.
Laser Facet Coating	Diffusion barrier to O ₂ , H ₂ O, Na ⁺ ; low stress; reproducible R.I.; (resistive); minimum induced optical modification; pinhole free.
Anti-reflection Coating	Adjustable, reproducible R.I.; low stress; minimum induced surface states.
Oxidation Mask	Diffusion barrier to O ₂ , OH; thermal stability.
Gate Dielectric	Low stress; minimum induced interface states; reprod. H cont. (low H content).

*at the process temperature

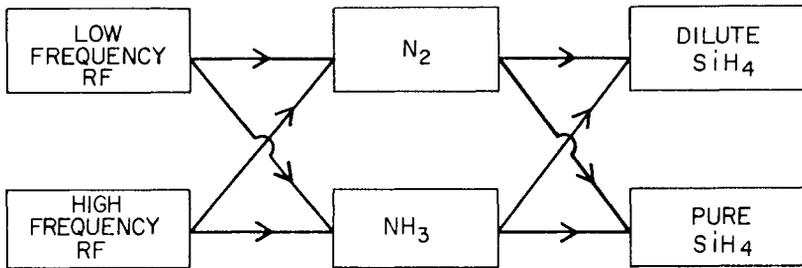


Figure 56: Decision chain for selecting a SiN_x PECVD process.

listed in Table 5. However, in some cases there is a clear choice. For example, if low hydrogen content is the highest priority, then low frequency, N₂, and dilute SiH₄ is the necessary combination of these parameters. However, this combination has been found to lead to quite high compressive stress in the film at room temperature. This may not be a drawback if the film is to be used in a high temperature application, where the compressive stress will be somewhat relaxed due to the lower linear coefficient of thermal expansion of the SiN_x relative to the semiconductor substrate. This will be discussed later in this section.

Silane dissociates very readily in a plasma, even more so than ammonia. Therefore in order to deposit films with close to bulk stoichiometry ($\text{Si}/\text{N} = 0.75$, refractive index ≈ 1.95), it is necessary to employ an excess of NH_3 . In general, a NH_3/SiH_4 and N_2 require a very large excess of N_2 . For example, in a parallel plate system at 13.56 MHz, Dun et al.²²¹ found that an N_2/SiH_4 flow ratio of 100 was needed. In an inductively coupled system, where only the N_2 was fed directly through the plasma,²²² an N_2/SiH_4 ratio of 250 was found necessary to obtain the approximately stoichiometric composition. In the early work using SiH_4/N_2 mixtures also in inductively coupled tube reactors,^{220,223} N_2/SiH_4 ratios of 200 and 670 respectively were found to be necessary. However, in a parallel plate system at low frequency (30 KHz) a N_2/SiH_4 ratio of ~ 4 (Reference 224) and 25-50 (Reference 225) was found to be sufficient although at low pressure, and high flow rates a flow ratio in excess of 120 was found to be necessary.²⁰² The origin of the apparent effect of a lower N_2/SiH_4 ratio being necessary to obtain a given stoichiometry at low (relative to high) plasma frequency is not clear. As discussed in Section 2.1.2.1.4 for a single component plasma, the degree of dissociation at a fixed power density is larger at high frequency where power dissipation is concentrated within the glow region. If this generality is still valid for a two component plasma consisting of one easily dissociated and ionized component (SiH_4) and another component with much higher dissociation and ionization energies (N_2), then one must hypothesize that the increased ion bombardment at low frequency enhances the rate of reaction between adsorbed N and SiH_x radicals, perhaps by Si-H bond breaking. This picture is supported by the higher normalized deposition rates (Table 8) and the lower H content of films deposited at low frequency, and is analogous to the etching situation depicted in Figure 12(a) and discussed in Sec. 2.1.2.1.4.

A point to be stressed regarding SiN_x deposition is the importance of a leak-tight chamber and gas input lines. Reactive Si and Si-H radicals adsorbed onto the substrate surface will preferentially react with an O containing species produced by a minor air or water leak into the plasma, even when NH_3 is used as the N source. The problem is even more severe when N_2 is used. In addition, all parts of the chamber which come into contact with the plasma should be well out-gassed before commencing a SiN_x deposition. If the substrates can be adequately protected, outgassing can be readily achieved by running an Ar or N_2 plasma, but such protection is difficult in a radial-flow reactor.

4.3.1.2 Film properties and their correlations. The range of values of properties of PECVD SiN_x films which are used in processing applications are shown in Table 9. The general methods of control of these properties have been discussed in section 4.2.4. In addition, there are a number of property dependences specific to SiN_x . The deposition parameter dependences of four important film properties are summarized in Table 10.

The refractive index increases with increasing $\text{SiH}_4/(\text{NH}_3 \text{ or } \text{N}_2)$ flow ratio due to an increasing Si/N compositional ratio in the film. This is also the cause of the increase in refractive index with increasing rf frequency, and with decreasing rf power. The latter case is due to the decrease in dissociation of the N source gas, whereas we have speculated above that

Table 8: Ratio of SiN_x (R.I. \cong 2) Deposition Rate to RF Power Density ($\text{\AA min}^{-1} \text{ W}^{-1} \text{ cm}^2$) as a Function of Nitrogen Source and Plasma Frequency($\text{\AA min}^{-1} \text{ W}^{-1} \text{ cm}^2$) as a Function of Nitrogen Source and Plasma Frequency

RF Frequency	N Source	
	NH ₃	N ₂
High (>5 MHz)	2200-600 ^a 620 ^b 770 ^c 1320 ^d	470 ^d
Low (<1 MHz)	2240 ^e 2330 ^f 1000 ^e	870 ^f 800 ^d

a. Reference 213

b. Reference 217

c. Reference 232

d. Reference 225

e. Reference 185

f. Reference 202

Table 9: The Range of Values of Important Properties of PECVD SiN_x and SiO₂ as Used in Processing Applications

Property	SiN _x	SiO ₂	Units
Stoichiometry (Si/N or Si/O)	0.7 - 1.2	0.50 - 0.55	
Refractive Index	1.85 - 2.3	1.46 - 1.55	
Optical Absorption Edge	300-450	<300	nm
H Content	15-30	0.5-8	at%
Etch Rate (in $\epsilon:1$ BHF)	30-1000	2000-4000	\AA min^{-1}
Stress	$5\text{T} - 10\text{C} \times 10^9$	$(0.5-3)\text{C} \times 10^9$	dynes cm^{-2}
Adhesion (to Si, GaAs, InP)	$>6 \times 10^8$	$>2.5 \times 10^8$	dynes cm^{-2}
Resistivity (at 10^4V cm^{-1})	$10^{11} - 10^{17}$	$10^{14} - >10^{17}$	$\Omega \text{ cm}$
Dielectric Breakdown Strength	1-6	$3-8 \times 10^5$	V cm^{-1}

the former case is due to the ion assisted nature of the heterogeneous reaction. Its increase with increasing deposition temperature is mainly due to increased density of the film. The optical absorption edge of the film also moves to longer wavelengths with increasing Si/N compositional ratio.²²⁶ The refractive index of the film has been shown to be linearly related to both its Si/N ratio²²⁷ and its (Si-H)/(N-H) ratio^{227,228} (measured by

IR absorption spectroscopy), which indicates the bonding distribution of the H content of the film. In both these references, these property correlations were for films deposited at a 50 KHz plasma frequency from NH_3 and SiH_4 source gases. However, while Claassen *et al.*²²⁷ found a linear correlation of refractive index to Si/N as measured by Rutherford backscattering (RBS), Samuelson and Mar²²⁸ did not find the same correlation with the film Si/N ratio as measured by Auger Electron Spectroscopy (AES). The RBS stoichiometry data is probably the more reliable, since that technique is directly quantitative. Maes *et al.*²²⁹ have discussed the difficulties in making quantitative AES analyses of SiN_x films. Figure 57 reproduces the refractive index correlation with (Si-H)/N-H from both References 227 and 228.

From the data in Figure 57, the refractive index,

$$n = 0.70(\text{Si}/\text{N}) + 1.39 \quad (31)$$

and from Figure 58(a)

$$n = 0.059[(\text{Si}-\text{H})/(\text{N}-\text{H})] + 1.88 \quad (32)$$

which combined yield:

$$(\text{Si}/\text{N}) = 0.084[(\text{Si}-\text{H})/(\text{N}-\text{H})] + 0.70 \quad (33)$$

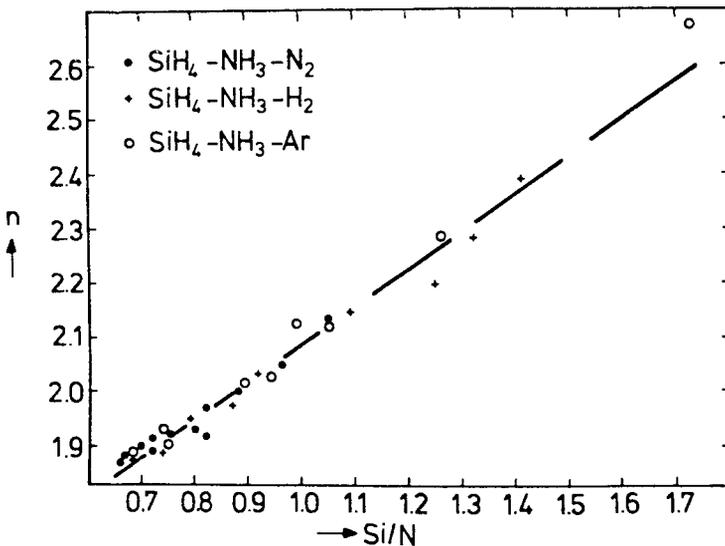


Figure 57: Correlation of SiN_x refractive index with Si/N compositional ratio of the film, as measured by RBS (from Reference 227, reprinted by permission of the publisher, The Electrochemical Society Inc.).

which indicates, as observed, that at a high (Si/N) ratio, most of the H content is bonded to Si. This is in qualitative agreement with the report²²¹ that for films deposited from SiH_4/N_2 , increasing plasma power changes the dominant H bonding from Si-H to N-H.

In addition, Claassen *et al.*²²⁷ have shown the dependence of the etch rate (in 7:1 buffered HF at R.T.) on (Si/N), or alternatively on [(Si-H)/(N-H)] by equation 33, for films deposited from SiH_4/NH_3 at 50 KHz and 300°C. This is reproduced as Figure 59. Over their fairly limited range of total H content, no correlation could be seen between etch rate and total H content. However, over a wider range of total H content for SiN_x films deposited in various commercial PECVD reactors under a wide variety of conditions, Chow *et al.*²³⁰ have shown good correlation between the etch rate and total H content (measured by the ^{15}N nuclear resonance reaction technique, described below), over three decades of variation in etch rate. These data are reproduced as Figure 60. These workers have also reported a relatively weak correlation between etch rate and film stress, with films under tensile stress etching more rapidly than those under compressive stress. As shown in Table 10, the deposition parameters which affect the etch rate are substrate temperature, rf power density and rf frequency. Increasing substrate temperature decreases the etch rate, for example by an order of magnitude from 200°C to 400°C, due to both reducing the H content and increasing the film density. Increasing rf power density decreases the etch rate, mainly by reducing the H content to the same extent as reducing the rf frequency to the low frequency regime. As also indicated for H content, use of N_2 in place of NH_3 and use of diluted SiH_4 will also reduce the H content and hence the etch rate.

Stress in PECVD SiN_x films can be varied over a very wide range, as shown in Table 9. The manner in which various deposition parameters affect stress is shown in Table 10. In general, high frequency deposition produces low to medium tensile stress, which is not good for crack resistance or step coverage. However, higher power deposition and use of dilute SiH_4 sources reduces this tensile stress, in some cases sufficiently enough to convert it to the desired low compressive stress. Use of N_2 in place of NH_3 also aids this conversion. For example, it has been reported²²¹ that at low deposition rate and an N_2/SiH_4 flow ratio of 65, the stress in SiN_x deposited on Si was changed from 1×10^9 dynes cm^{-2} tensile to 5×10^9 dynes cm^{-2} compressive by increasing the plasma power from 50 to 300W. Low frequency deposition from SiH_4 and NH_3 source gases produces low to medium compressive stress, while use of N_2 at low frequency tends to produce high compressive stress. Koyama *et al.*²³¹ have measured the change in stress in SiN_x films deposited in the same parallel-plate reactor from SiH_4/NH_3 for rf frequencies varied over the range 50 KHz to 13.56 MHz. Variation in stress was over the range shown in Table 9, with cross-over from compressive to tensile stress occurring at about 2 MHz. Martinet *et al.*²³² have also shown a similar change in stress in the SiN_x film produced by changing from 50 KHz to 13.56 MHz, although their stress values at each frequency were lower than those of Koyama *et al.*²³¹

Most of the reported stress data is for SiN_x films deposited on Si substrates. However to a first approximation similar stress values are measured for films deposited on GaAs and InP substrates. The stress

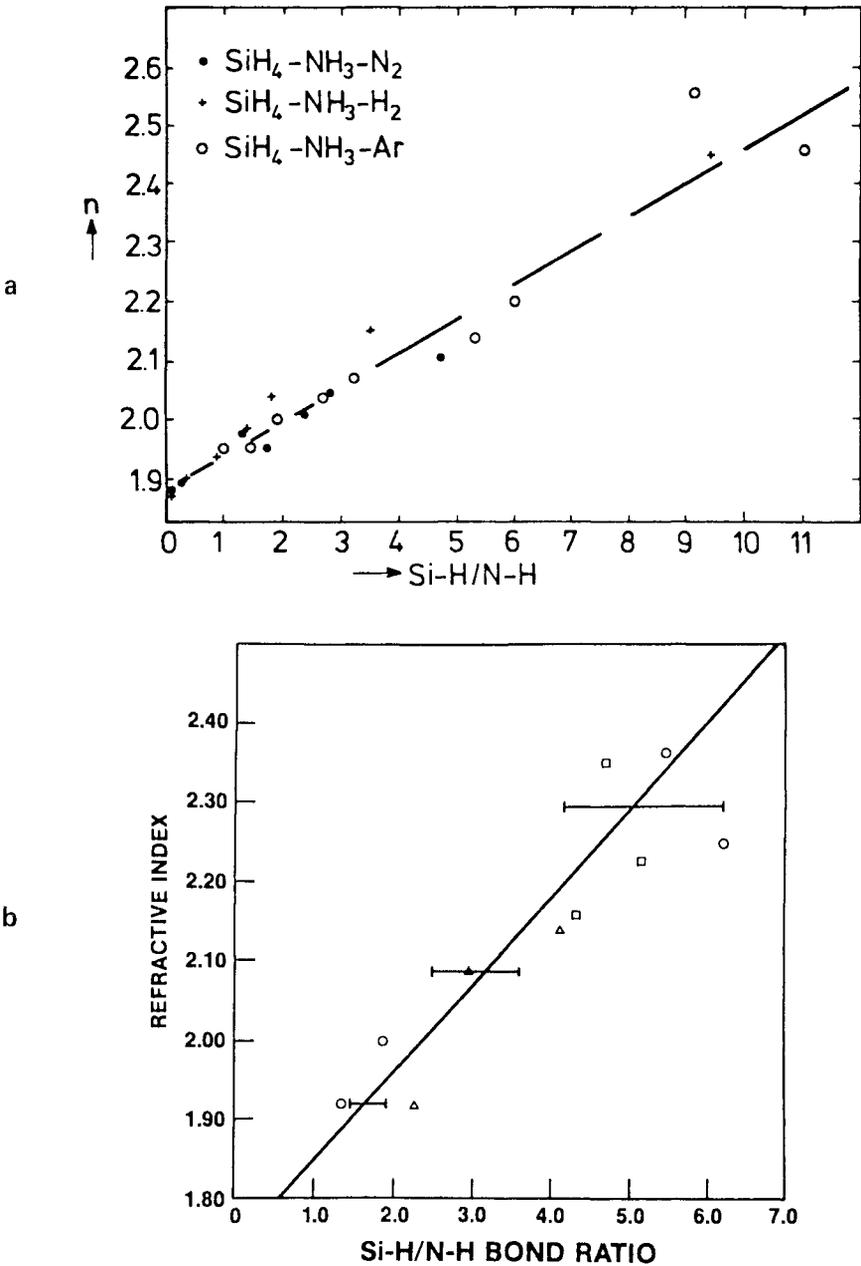


Figure 58: Correlation of SiN_x refractive index with $(\text{Si-H})/(\text{N-H})$ hydrogen bonding ratio in the film: (a) from Reference 227, deposition temperature 300°C ; (b) from Reference 228, deposition temperature 275°C (both reprinted by permission of the publisher, The Electrochemical Society Inc.)

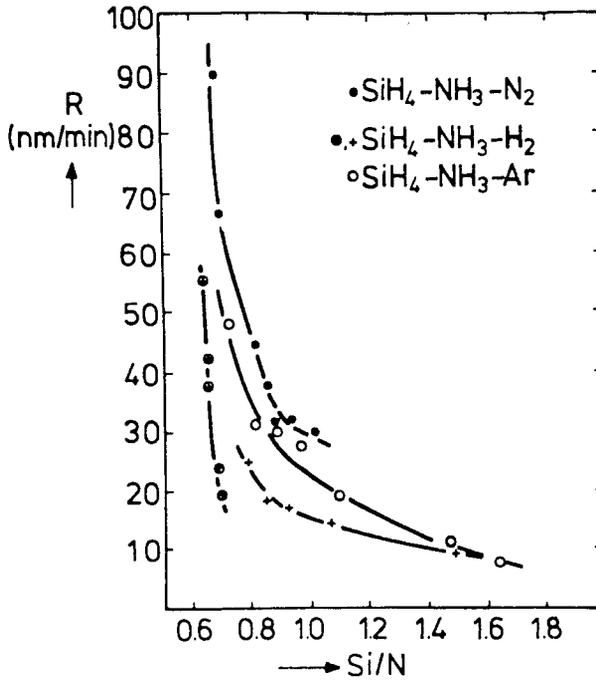


Figure 59: The etch rate of SiN_x films in 7:1 buffered HF at room temperature as a function of their (Si/N) compositional ratio. The films were deposited from SiH_4/NH_3 /carrier gas mixtures at a 50 KHz plasma frequency and 300°C substrate temperature (from Reference 227, reprinted by permission of the publisher, The Electrochemical Society Inc.).

measured in a film at room temperature can be separated into two components, one being thermally-induced during cool-down from the deposition temperature due to the difference in linear coefficients of expansion of the film and the substrate, and the other being the intrinsic stress in the film as-deposited. Retajczyk and Sinha²³³ deposited SiN_x films onto different substrates (silicon, quartz and sapphire) of known expansion coefficients, and then measured stress as a function of temperature. From these data they computed both the elastic stiffness, and the linear coefficient of expansion α of their high frequency deposited SiN_x . Elastic stiffness, $E/(1-\nu)$, where E is Young's modulus and ν is Poisson's ratio, of the SiN_x was found to be 1.1×10^{12} dynes cm^{-2} , and α determined to be $1.5 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$. This latter value is about half that for bulk Si_3N_4 . Use of these two values in conjunction with $\alpha_{\text{substrate}}$ permit the thermally induced film stress component, $\sigma_{i,T}$, to be calculated for SiN_x films deposited on other substrates, by the relationship:

$$\sigma_{i,T} = [E/(1-\nu)]_f (\alpha_f - \alpha_{\text{substrate}}) \Delta T \quad (34)$$

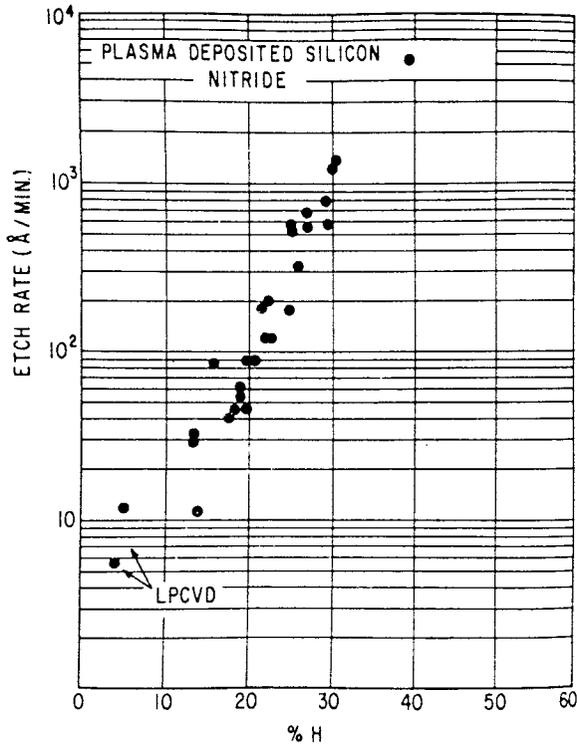


Figure 60: The etch rate of SiN_x films in buffered HF as a function of their total H content (from Reference 230, reprinted with permission of the American Institute of Physics).

Table 10: Direction of Increases in Deposition Parameters on PECVD SiN_x Properties

Increase of Effect on	Dep. Temp.	RF Power	RF Freq.	SiH ₄ /(NH ₃ or N ₂) Flow Ratio	N ₂ in place of NH ₃
Refractive Index	↑	↓	↑	↑	↑
Etch Rate	↓	↓	↑	↓	↓
H Content	↓	↓	↑	-	↓
Stress, Tensile	-	↓	↑	↓	↑

where ΔT is the difference in temperature between the deposition and stress measurement temperatures. A positive $\sigma_{i,T}$ calculated by equation 34 corresponds to a tensile stress in the film. Thus for SiN_x deposited on GaAs ($\alpha=6.4 \times 10^{-6} \text{C}^{-1}$) and InP ($\alpha=4.5 \times 10^{-6} \text{C}^{-1}$) at 300°C, the thermally-induced contributions to room temperature stress are 1.5 and 0.91×10^9

dynes cm⁻² compressive, respectively. Thus it can be seen that in most cases, the thermally-induced component is a minor contribution to the measured stress. Similar calculations are also very useful to estimate the stress in the film when used in an application at elevated temperature. This will only be an estimate, however, since loss of hydrogen and film densification will be taking place simultaneously.

Resistivity of PECVD SiN_x can be varied over many orders of magnitude by varying the Si/N compositional ratio of the film, by the methods indicated in Table 10. Sinha and Smith²³⁴ have shown that the resistivity π varies according to $1n(\pi^{-1}) \propto (\text{Si}/\text{N})$, from a value of 10¹⁶ Ωcm at (Si/N)=0.75, to 10¹⁰ Ωcm at (Si/N)=1.33. Their data are reproduced as Figure 61. Dun et al.²²¹ found basically the same relationship for SiH₄/N₂ deposited films. Note that at the highest resistivity end, resistivity is quite sensitive to trace impurities incorporated into the film. Electrical properties have recently been reported²³⁵ for low frequency PECVD SiN_x.

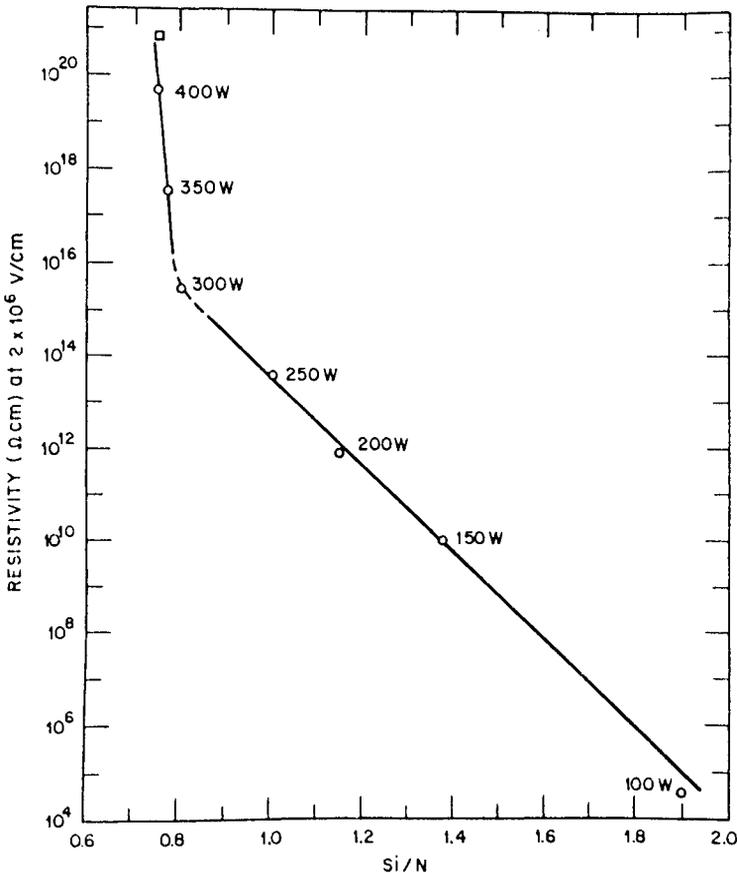
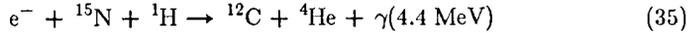


Figure 61: Resistivity vs. (Si/N) compositional ratio for PECVD SiN_x (from Reference 234, reprinted with permission of the American Institute of Physics).

4.3.1.2.1 *Measurement of hydrogen content.* Having discussed the strong influence on film properties of both total H content and distribution of that H content between Si-H and N-H bonds, it is worthwhile briefly discussing their quantitative measurement. Total H content is measured absolutely by the resonant nuclear reaction.



which occurs between an H atom and a 6.405 MeV ${}^{15}\text{N}^+$ ion. Emitted γ rays are counted. Depth profiling of the H content is achieved by scanning the incident ${}^{15}\text{N}$ ion energy upwards from 6.405 MeV. Lanford and Rand²³⁶ used H content data obtained from this nuclear reaction analysis to provide calibration of infra-red absorption band intensities of the Si-H and N-H bonds. Adams¹⁹¹ has summarized the wavenumbers and the calibration constants of the various Si-H, N-H, and O-H IR absorption bands, as well as the wavenumbers of Si-O and Si-N bands, which can be observed in the IR spectra of PECVD SiN_x and SiO_2 and a-Si(H) films. These values are shown in Table 11. The calibration coefficient K, which is related to the oscillator strength of the transition, is defined by:

$$N_H = K \int \alpha(\omega) d\omega \quad (36)$$

where N_H is the number of H bonds per cm^3 and $\alpha(\omega)$ is the absorption coefficient (cm^{-1}) at wavenumber $\omega(\text{cm}^{-1})$.

4.3.1.3 *Recent advances.* SiN_x films have very recently been deposited from SiH_4/NH_3 by electron beam created plasma-enhanced CVD,²⁰⁶ referred to as EBCVD as discussed in Section 4.2.1. In addition to the reduced energetic particle bombardment, a further attractive feature was

Table 11: The Position and Strength of the Various IR Absorption Bands Observed in PECVD SiN_x , SiO_2 , and a-Si(H)^a

Assignment	SiO_2	SiN	a-Si	K
	(wavenumber, cm^{-1})			(10^{16}cm^{-1})
OH	3620			1.4
OH	3380			0.4
NH		3340		8.2
BH			2530	
SiH	2270	2170	2000	7.1
NH		1170		
SiO	1070			
SiOH	940			
SiH, SiOH	885			
SiN		850		
SiO	805			
SiH			630	2.5
SiN		465		
SiO	450			

^aFrom Reference 191, reprinted with permission of *Solid State Technology*, published by Technical Publishing, a company of Dun and Bradstreet.

the low H content (~ 12 at %) and the undetectability of any Si-H bonded hydrogen. The origin of this effect is unclear, since in conventional PECVD, a reduction in H content is achieved by increased ion bombardment (Table 10). It will be interesting to see the level of stress in these films.

A very interesting method of improving the thermal stability of PECVD SiN_x has recently been reported by Fujita et al.²³⁷ On the basis that the H content cannot be reduced below about 15 at % (Table 9), their approach was not to minimize the H content, but instead to give it thermal stability by increasing the H bond strength. This was achieved by simultaneously incorporating fluorine into the SiN_x film, by performing its PECVD from a $\text{SiF}_4/\text{N}_2/\text{H}_2$ source gas mixture. It has previously been reported that Si-H bonds in fluorinated $\alpha\text{-Si(H)}$ have increased stability,²³⁸ and so the same effect was expected in SiN_x . Thermal stability was indeed improved, in that annealing at 640°C could be performed without loss of H. However, this was not the result of an increased Si-H bond strength, since no Si-H bonds were present in the fluorinated SiN_x film, but rather an improved N-H bond strength. Use of H_2 as a component source gas was necessary to permit deposition to occur, otherwise etching of the deposited film by the atomic F produced from the SiF_4 occurred. The function of this H_2 in suppressing etching relative to deposition was attributed to a more rapid removal of the product F atoms, due to both the reduced residence time and chemical scavenging.

4.3.2 Silicon Oxide PECVD silicon oxides are a much narrower class of materials than the nitrides, in that their composition is close to stoichiometric, with a typical range of Si/O ratio from 0.50 to 0.55, as shown in Table 9, and much lower hydrogen content than the nitrides. On this basis, we will refer to these materials as PECVD SiO_2 . Partly for these reasons, the properties of PECVD SiO_2 show much less variability than those of PECVD SiN_x as shown by their comparison in Table 9. Thus it is an easier task to deposit SiO_2 films with reproducible properties for use in semiconductor processing, where they are finding increasing application. A number of these applications are shown in Table 3. Other factors favoring the use of PECVD SiO_2 are its low compressive stress, which is almost independent of deposition conditions over a wide range, its high deposition rate at low plasma power, its lower deposition temperature and its lack of susceptibility to the effects of small amounts of air or water vapor contamination of the deposition plasma. The main applications area in which SiO_2 does not perform at least as well as SiN_x is that at elevated temperature ($>500^\circ\text{C}$) where diffusion barrier properties are required. This includes applications such as implant anneal caps on III-V semiconductors (where diffusion of the more volatile group V element is to be prevented) and dopant diffusion masks.

The usual source gases for SiO_2 deposition are SiH_4 and N_2O ,^{185,218,239,240} often with the SiH_4 diluted in an inert gas such as Ar. Other source gas combinations which have been employed are SiCl_4/O_2 ,²⁴¹ and in some early work, $\text{Si}(\text{OC}_2\text{H}_5)_4/\text{O}_2$.²⁴² The most obvious combination of SiH_4 and O_2 cannot be used, since these gases will react spontaneously on mixing (in the absence of a plasma) to form SiO_2 "smoke" in the gas phase. N_2O is chosen as the source of O, since it dissociates readily in a plasma, although

rather less readily than does SiH_4 . Because of this, $\text{N}_2\text{O}/\text{SiH}_4$ flow ratios appreciably in excess of 2 are needed to deposit SiO_2 ; the necessary $\text{N}_2\text{O}/\text{SiH}_4$ flow ratio²³⁹ is usually in excess of 15. However, N_2O is more easily dissociated than CO_2 , another possible O source, where CO_2/SiH_4 flow ratio of around 200 were needed.²⁴⁰

SiO_2 is deposited at a higher rate and lower power than SiN_x . Deposition rates employed are usually in the range $250\text{-}800\text{\AA min}^{-1}$, although rates in excess of $2,000\text{\AA min}^{-1}$ can readily be achieved, with some loss of spatial uniformity. Thus the deposition rate to plasma power density ratio in a radial-flow reactor is very high, with values ranging from $3,000\text{\AA min}^{-1}\text{W}^{-1}\text{cm}^2$ (Reference 240) to $27,000\text{\AA min}^{-1}\text{W}^{-1}\text{cm}^2$ (Reference 239). These values may be compared with those in Table 8 for SiN_x deposition. Since plasma powers are very low and spatial uniformity of deposition is sensitively dependent on plasma power (see Section 4.2.3), improved sensitivity of power monitoring and control is beneficial. For deposition from $\text{SiH}_4/\text{N}_2\text{O}/\text{Ar}$ plasmas, this may be achieved by monitoring the broadband optical emission intensity from the plasma (see Section 4.2.3), which consists almost entirely of near infra-red Ar emission.²⁴³ If Ar is not included in the plasma composition, the main emission is in the near ultraviolet from N_2 and is much weaker, and thus the technique's sensitivity for deposition rate and uniformity control is much reduced.

Effects of PECVD parameters on SiO_2 film properties were studied by Adams et al²³⁹ for high frequency deposition, and by Hollahan²⁴⁰ and van de Ven¹⁸⁵ for low frequency deposition. In the work of Adams et al., for SiO_2 deposited at a relatively low rate, many film properties were shown to have a large discontinuous change at a temperature between 175 and 200°C . Below that temperature, the films etch rapidly and non-uniformly, have reduced density and refractive index, and have higher stress. Thus films deposited below 200°C are not useful in processing applications (at least for films deposited with the other deposition parameters similar to those in Reference 239).

The refractive index of PECVD SiO_2 is mainly determined by the Si/O stoichiometry which is a function of the $\text{SiH}_4/\text{N}_2\text{O}$ flow ratio. Above a $\text{SiH}_4/\text{N}_2\text{O}$ ratio of 15 to 20 (and for a deposition temperature of 200°C or above), refractive index remains constant at 1.47. This value is very slightly higher than that of CVD SiO_2 (1.44 - 1.46) or thermally grown SiO_2 (1.464), and is probably the result of a very slight oxygen deficiency (<2%). This is supported by the film density ($\sim 2.30\text{ gm cm}^{-3}$) which is slightly higher than that of fused SiO_2 (2.20 gm cm^{-3}). This small amount of O deficiency possibly results from some Si-H and Si-OH bonding existing in the film. Lower temperature deposition, or high rate deposition at around 200°C , produces a lower refractive index due to a reduced film density. There is evidence that films deposited at low frequency have a somewhat higher refractive index, in contrast to the situation for SiN_x deposition. A value of 1.54 has been reported¹⁸⁵ for low frequency deposition in an outwards radial-flow reactor at an $\text{N}_2\text{O}/\text{SiH}_4$ flow ratio of 50. The reason for this higher refractive index appears not to be excess Si, but instead some N content in the film in place of O content. A composition of $\text{SiO}_{1.9}\text{N}_{0.15}$ was reported;¹³⁵ a few percent N content for similarly deposited films has also

been observed by XPS analysis (ESCA).²⁴⁰ Presumably this N content is a result of N_2^+ and/or N^+ ion bombardment of the growing film which occurs at both higher energies and higher fluences at low frequency relative to high frequency.

Hydrogen content of PECVD SiO_2 is much lower than that of PECVD SiN_x , even for the lower deposition temperature usually employed for SiO_2 . The reason for this is the higher reactivity of Si-H radicals towards atomic oxygen than towards N-H radicals or other reactive N containing species. Thus the heterogeneous reaction between Si-H and atomic O on the substrate surface occurs rapidly and almost to completion before the reactants are buried by further depositing material, whereas with Si-H and N-H reaction to completion does not occur.¹⁸⁵ A disadvantage of the high reactivity of Si-H and O is that under certain conditions (high rf power, high pressure, low flow rate), homogeneous gas phase reaction can occur, giving rise to particulate contamination and pinholes in the deposited film. As is the case for SiN_x deposition, increased substrate temperature, increased rf power density and reduced plasma frequency all act to reduce the H content. A typical range of values is as shown in Table 9. The H may be present in three different bonding configurations, which are Si-H, Si-OH and H_2O . Their respective IR absorption bands are shown in Table 11. H_2O can be formed in the film during deposition by reaction between H atoms produced on the surface from Si-H reaction with O, and other arriving O atoms. Its formation and incorporation into the growing film is reduced by increasing substrate temperature, but is not sensitively dependent on other deposition parameters.²³⁹ The Si-OH content is similarly dependent on substrate temperature, and at low N_2O/SiH_4 flow ratios (<20), increases with the N_2O/SiH_4 flow ratio along with the increase in Si-O bonding. As to be expected, the Si-H content decreases with increasing N_2O/SiH_4 flow ratio, with increasing rf power, and with decreasing SiH_4 content in the discharge. For films deposited at 200°C, the total H content is about 5 atomic %, distributed approximately in the ratio 9:4:1 as H_2O :Si-OH:Si-H. At a 250°C deposition temperature, this can be reduced to 2 atomic % or lower.

The etch rate of PECVD SiO_2 films in HF containing etchants does not show any strong correlation with H content, or even with overall composition. The main variation is a slight decrease with increasing deposition temperature above 200°C. For deposition temperatures below 200°C, films etch rapidly and inhomogeneously.²³⁹ Etch rates are typically about an order of magnitude faster than those of thermally grown SiO_2 , and a factor of 2 or 3 faster than those of sputtered SiO_2 . This may be due to increased porosity, although this idea is not supported by the density of PECVD SiO_2 , which is higher than that of thermal SiO_2 .²³⁹

Stress in PECVD SiO_2 also shows far less variability with deposition conditions than does that in PECVD SiN_x . In all cases for PECVD SiO_2 deposited on Si, GaAs or InP substrates, the SiO_2 is in the desirable state of being under low compressive stress. The stress magnitude is increased to values at the upper end of the range shown in Table 9 by deposition from dilute SiH_4 in Ar, or by deposition at low frequency. Stress in SiO_2 films deposited on Si are lower than in those deposited on GaAs or InP, due to

there being less of a thermal mismatch (which is quite large in all cases). The linear coefficients of thermal expansion are 0.6, 3.2, 6.4 and $4.5 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ for SiO_2 , Si, GaAs and InP respectively. Elastic stiffness coefficients have not been reported for PECVD SiO_2 , but the values for fused silica should provide a reasonable upper limit close to the correct values. Thus using a value of $[E/(1-\nu)]_{\text{SiO}_2} = 0.85 \times 10^{-2} \text{ dynes cm}^{-2}$ in equation 34, the thermally-induced components of stress in SiO_2 films deposited on Si, GaAs and InP at 250°C are found to be 0.5, 1.1 and $0.75 \times 10^9 \text{ dynes cm}^{-2}$ compressive. While these values are not large on an absolute scale, they can correspond to being the major component for the SiO_2 films of lowest stress.

The combined properties of low compressive stress and good adhesion permit very thick films of PECVD SiO_2 to be deposited without blistering or cracking occurring. Thicknesses up to $5 \mu\text{m}$ have been deposited.¹⁸⁵ Films of $3 \mu\text{m}$ thickness have been deposited on GaAs and then patterned by anisotropic, low frequency plasma etching to leave $3 \mu\text{m}$ wide by $3 \mu\text{m}$ high stripes of a few centimeters in length which have not lost adhesion to the substrate.¹⁴⁰ These stripes have then been used as proton implant masks in the fabrication of proton bombarded, gain-guided GaAs lasers.²⁴⁴

PECVD SiO_2 films have consistently high resistivity (Table 9) relative to SiN_x due to less compositional variation and a higher band gap. Additionally SiO_2 has a lower dielectric constant than SiN_x (3.8 relative to ≈ 7.0 at 1 MHz). Both these properties make PECVD SiO_2 very attractive for application as an interlayer dielectric.

The recently developed electron beam CVD (EBCVD) variation of PECVD, discussed in Sec. 4.2.1 and in the preceding section for SiN_x , was first used to deposit SiO_2 .²⁰⁵ Bulk film properties reported were similar to those of good quality PECVD films deposited at the same temperature, except for a high pinhole density. It is to be hoped that this is not intrinsic to the technique, although gas phase reaction and particulate formation could be a problem within the localized high intensity plasma, just as it is at high plasma power densities in conventional PECVD of SiO_2 .

In another recent advance, Kaganowicz et al.²⁴⁵ have employed magnetron assisted PECVD to deposit SiO_2 at room temperature. The magnetic confinement of the plasma increases its density and increases the dissociation of N_2O , so that stoichiometric films ($n \approx 1.46$) could be deposited at an $\text{N}_2\text{O}/\text{SiH}_4$ flow ratio of about three, an order of magnitude lower than that needed without the magnetic enhancement. The much reduced total gas flow, as well as the magnetic enhancement, permitted a much lower deposition pressure (45 mtorr) to be used. At higher $\text{N}_2\text{O}/\text{SiH}_4$ flow ratios, gas phase reaction was again a problem, producing particulate incorporation in the film. However, at $\text{N}_2\text{O}/\text{SiH}_4$ flow ratios of five or less, featureless films were deposited. Non-uniformity of deposition rate and film properties is a problem which will need to be overcome. Film characteristics were not reported; it will be very interesting to compare them to those of conventional PECVD SiO_2 films deposited at $200\text{--}250^\circ\text{C}$, as well as to those of films deposited at lower temperatures.

4.3.2.1 Silicon oxynitride. So far we have not mentioned intentionally-

deposited silicon oxynitrides, $\text{SiO}_x\text{N}_y\text{H}_z$. This is the same compositional formula which we quoted as being strictly correct for silicon nitride, but in this case x is rather larger. Obviously the previously discussed silicon nitride and oxide are the extremes of this compositional series, and the properties of silicon oxynitride films can be varied anywhere in between those of silicon nitride and oxide prepared under similar conditions. As discussed in Section 4.3.1, the properties of the endpoint composition, SiN_yH_z , can themselves be varied over a wide range. Thus the use of silicon oxynitride offers the opportunity of obtaining a combination of film properties, optimized for a given application, not accessible with either of the endpoint compositions.

Silicon oxynitride is deposited from $\text{SiH}_4/\text{NH}_3/\text{N}_2\text{O}$ source gas mixtures, with or without inert gas diluents. Nitrogen to oxygen stoichiometry, y/z , is controlled by the $\text{NH}_3/\text{N}_2\text{O}$ flow ratio, bearing in mind that the heterogeneous reaction between adsorbed Si-H radicals and O atoms is more rapid than that with N or N-H species. Hence most of the available O will be consumed in preference to N. The nitrogen source used must be NH_3 and not N_2 for the deposited film to have any significant N content, since the low electron temperature of the $\text{SiH}_4/\text{N}_2\text{O}$ plasma does not favor the formation of atomic N from N_2 (or from N_2O).

4.3.3 Amorphous Silicon. Amorphous silicon deposited by PECVD (glow discharge) contains a large amount of hydrogen (10 to 30 at %) present in Si-H bonds, and in recognition of this, it is generally referred to as a-Si(H). It is the H content which is responsible for the interest in this material, since it bonds to and hence neutralizes dangling bond defect states in the amorphous network. This releases the Fermi level from being pinned at the energy of defect states within the band gap. Movement of the Fermi level into either the conduction or valence band becomes possible, allowing doping to produce n and p-type material. It was this discovery a decade ago by Spear and LeComber²⁴⁶ which prompted the huge amount of attention which has been devoted to the study of a-Si(H). This was rapidly followed by reports of the application of doped a-Si(H) in photovoltaic solar cells.^{247,248} The early amorphous silicon solar cell work has been reviewed by Carlson,²⁴⁹ who has also given a more recent review.²⁵⁰ In addition to solar cells, a-Si(H) is being used in a number of other applications, including other p-n junction and Schottky barrier devices,²⁵¹ thin film field effect transistors,^{252,253} and as a xerographic photoreceptor and in Vidicon-type photoconductive image tubes.²⁵⁴ There have been a number of recent reviews of the application of a-Si(H) in these fields, including Spear et al.,²⁵⁵ various papers in Hamakawa,²⁵⁶ and Hamakawa.²⁵⁷

Hydrogenated amorphous silicon usually is deposited by PECVD at high rf frequency and low plasma power from either pure SiH_4 or SiH_4 /inert gas mixtures at a substrate temperature of 200-300°C. Variations on the above scheme are the use of a dc plasma, the use of Si_2H_6 in place of SiH_4 , and the application of a negative bias to the substrate table to increase ion bombardment of the growing film. Fritzsche²⁵⁸ has reviewed deposition conditions and resultant heterogeneities (compositional and microstruc-

tural) in a-Si(H). Doping of the a-Si(H) is readily achieved by introduction of controlled small amounts of PH_3 (for n-type) or B_2H_6 (for p-type) into the plasma. Many of the considerations discussed in the section on silicon nitride relating to the influence of PECVD parameters on the film properties are relevant here, particularly those pertaining to H content of the film, without the complicating consideration of a changing Si/N stoichiometry. As for SiN_x deposition, the absence of vacuum leaks and of outgassing is very important, since O_2 and H_2O will be gettered by the depositing film.

It should be noted that while most SiN_x deposition has been performed in radial-flow, parallel plate reactors, much of the work on a-Si(H) has been carried out in asymmetric rf diode sputtering systems (but at high enough pressure and low enough power to ensure that sputtering of the driven electrode did not occur). Thus there is not a symmetrical gas flow situation, and the bias voltage developed by the driven electrode is appreciably higher and the plasma potential lower than that in the case of the symmetrical electrodes of a radial-flow system. Hamasaki et al.²⁵⁹ have recently reported the benefits of closely surrounding the driven and substrate electrodes by a grounded mesh cage, which makes the interelectrode potential distribution look like that obtained in the symmetric radial-flow reactor, with a very low driven electrode bias. This shifted to higher energy the energy distribution of ions accelerated across the sheath into the substrates, but also eliminated the high energy tail of the incident particle energy distribution, which was due to energetic neutrals reflected from the driven electrode at high cathodic bias (as occurs in sputter deposition). This had the effect of permitting much higher rates of deposition from SiH_4 (3000\AA min^{-1} in place of $\sim 200\text{\AA min}^{-1}$ without the ground mesh) without degradation of the materials properties usually associated with high rate deposition. In fact, thermal stability of the optoelectronic properties of the high-deposition-rate a-Si(H) showed considerable improvement.

In other work, use of inert gas (He or Ne) diluted SiH_4 permitted an increased deposition rate from an increased rf power density without gas phase reaction occurring,²⁶⁰ due to the overall increased gas throughput/reduced residence time. Alternatively, disilane (Si_2H_6) has been used in place of SiH_4 , with an order of magnitude increased deposition rate, and apparently also a reduced stress, without any increase in rf power density.²⁶¹ However, this magnitude of deposition rate increase only applies for specific low power conditions. Kuboi et al.²⁶² for example, obtained only a factor two difference in deposition rate from Si_2H_6 and SiH_4 flows carrying equal Si content at low power, but no increase in deposition rate at higher powers, where presumably both SiH_4 and Si_2H_6 were being almost completely dissociated. These authors also reported an interesting effect for deposition from $\text{Si}_2\text{H}_6/\text{D}_2$ mixtures, in which it was observed that for low power deposition, the majority of the hydrogen content of the deposited a-Si was in the form of H (from unbroken Si-H bonds), whereas at high power, a disproportionately high D content was produced in the film. This is a similar effect to that observed in the deposition of SiN_x from SiH_4/N_2 mixtures,²²¹ where at low power most of the H content of the film was present as Si-H from unbroken bonds, but at high power the majority was

present as N-H. For the case of SiH_4/D_2 mixtures, this effect was not observed by Kuboi et al.²⁶²

The degree of doping also has been reported to increase the hydrogen content,²⁶³ but in addition to reduce the temperature at which the rate of loss of hydrogen is a maximum. This indicates that the average bond strength of the H has been reduced. Doping also has an effect on the microstructure of the film, changing the distribution of the H bonding sites. Hirose²⁶⁴ has studied the optical emission spectra of SiH_4/H_2 plasmas without and with magnetic confinement, and correlated the emitting species to the infra-red spectrum of the deposited film. It was found that a decrease in the SiH emission intensity relative to that of H_2 corresponded to the formation of partially microcrystalline films in which the doping efficiency was extremely high, as discussed later in this section.

Microstructure of the a-Si(H) film appears to be the key property in determining device-related properties such as doping level, carrier mobility and photoconductivity. Knights and Lujan²⁶⁵ first showed that a-Si(H) films were frequently not homogeneous random networks. They found that high power deposition at low temperature, in particular from Ar diluted SiH_4 , produced an island structure film with columns of typical a-Si(H) propagating normal to the substrate but with only low density material between these columns. Earlier work²⁶⁶ had shown films deposited under these conditions to have high densities of non-radiative recombination centers (low photoluminescence yield). At a substrate temperature of 230°C, for which the optically-active defect density was minimized for low power deposition from pure SiH_4 , there was no observable microstructure in the film down to a resolution of 10 Å. However, deposition at this temperature either at high power or from very dilute SiH_4 produced films of columnar morphology and high optically-active defect density. Deposition under these conditions, but with the addition of a negative substrate bias to give increased ion bombardment of the growing film, eliminated the columnar morphology and strongly reduced the defect density. It was postulated that the inter-column regions were not voids, but were at least partially filled by a crosslinked polysilane, $(\text{SiH}_2)_n$. The origin of this columnar morphology was identified not as a low density of nucleation centers in the initial growth, but as imperfect coalescence of islands growing from each nucleation center.

Inhomogeneous films at the opposite end of the structural order spectrum in the class of materials which still are described as hydrogenated amorphous silicon are currently of great interest. These are films in which there are microscopic quasi-crystalline ordered regions of low hydrogen content embedded in a disordered, truly amorphous network containing large amounts of hydrogen. The occurrence and growth of these quasi-crystalline regions is promoted by high doping levels, increased substrate temperature, increased plasma power density, magnetic confinement of the plasma, and high hydrogen content in the source gas mixture.^{258,264,267,268} As the quasi-crystalline content is increased, the percolation limit is exceeded and a highly conductive film results. This is the reason for the strong interest in such films, which are also referred to as microcrystalline.

The conductivity of these microcrystalline films can be varied by doping over the range from about $10^{-12}\Omega^{-1}\text{cm}^{-1}$ to about $10\Omega^{-1}\text{cm}^{-1}$ for both n- and p-type films, in contrast to an upper limit of about $10^{-2}\Omega^{-1}\text{cm}^{-1}$ for both n- and p-type amorphous films.^{246,269} For example, by employing magnetic confinement of a relatively high power plasma and a substrate temperature of 300°C , at a PH_3/SiH_4 flow ratio of 5.6×10^{-3} , Hamasaki et al.²⁷⁰ deposited n-type microcrystalline films of room temperature conductivity of $27\Omega^{-1}\text{cm}^{-1}$. Conduction activation energy was the low value of 10 meV. This same group²⁷⁰ reported p-type microcrystalline films of conductivity $7.8\Omega^{-1}\text{cm}^{-1}$, produced by a dopant flow ratio, $\text{B}_2\text{H}_6/\text{SiH}_4$, of 2.6×10^{-2} .

The increase in doping efficiency and conductivity reported by Madan and Ovshinsky²⁷¹ for doped fluorinated amorphous silicon, a-Si(F,H), deposited from SiF_4/H_2 plasmas, was originally thought to be due to improved efficiency of dangling bond termination by F relative to H, producing a reduced density of localized states. This report sparked a great deal of interest in the fluorinated material, which has been reviewed by Matsumura and Furukawa.²⁷² However, it has now been realized that the improved doping was not due to superior dangling bond termination, but was in fact due to the films being microcrystalline, with the presence of F being conducive to microcrystallization.^{272,273} As described above, non-fluorinated microcrystalline films can have equally high conductivity.

Kuwano²⁷⁴ has reviewed the relevant properties of a-Si(H) in much greater detail than the above brief discussion of some of the more pertinent points. Hirose²⁸⁴ has reviewed fabrication techniques and growth mechanisms of a-Si(H), Hamakawa²⁷⁵ has reviewed the device physics and design of a-Si(H) photovoltaic cells and Haruki and Uchida²⁷⁶ have reviewed their fabrication and performance.

4.3.4 Other Semiconductors, Including Epitaxial Growth. The semiconductor films which have been deposited by PECVD are polycrystalline silicon, and epitaxially grown single crystal films of silicon, germanium, gallium arsenide and gallium antimonide, in addition to amorphous silicon discussed in the preceding section. The application of PECVD to epitaxial growth is relatively new, and is an area where much progress is to be expected within the next few years. The closely related area in which plasmas are employed in the growth of single crystal materials, that of epitaxial crystal growth by sputter deposition, is also of much current interest, and has recently been reviewed.²⁷⁷

In the preceding section we discussed amorphous silicon films and extended the discussion to include films referred to as microcrystalline. The latter consist only partially of crystalline material, embedded in an amorphous network, with the mean size of the crystalline regions no larger than a few tens of Angstroms. The films described as polycrystalline have crystalline regions an order of magnitude larger, which are separated from each other only by grain boundaries. These polycrystalline silicon films (often, but incorrectly, referred to as polysilicon) may be deposited by PECVD under conditions which are an extension of those needed to develop microcrystallinity, in particular a further increased substrate temperature up to the $400\text{-}650^\circ\text{C}$ range. Such temperatures are still somewhat lower than those for direct CVD, the most common method of deposition of

polycrystalline silicon. The extent of choice of source gas is identical to that for deposition of a-Si(H). Those which have been employed for polycrystalline and single crystal Si PECVD are shown in Table 12.

For depositions performed in tube reactors from inductively-coupled, high frequency plasmas,^{278,279} polycrystalline Si films were obtained at the low substrate temperature of 400-450°C. In both cases the grain size was large, on the scale of a few hundred Angstroms (500Å was reported in Reference 279). In contrast, in a capacitively coupled, hot wall reactor operating at low frequency (Figure 51 c), a substrate temperature of 625°C was required for the onset of crystallization.²⁸⁰ For this substrate temperature, the grain size was only 100Å and the polycrystalline structure described as poorly defined. It is not clear whether the higher temperature needed for polycrystalline film formation was due to the use of the chlorinated silane source gas, or to the damaging effect of the energetic Ar⁺ ion bombardment present in the low frequency plasma acting in opposition to crystal growth. These low frequency plasma deposited films²⁸⁰ were in compressive stress, as is obtained for SiN_x films deposited in this type of reactor, but of even higher magnitude. Values of 6×10^9 dynes cm⁻² and

Table 12: Gases and Temperatures Employed for PECVD of Polycrystalline and Epitaxial Silicon

Source Gas Mixture	Deposition Temp. (°C)	Morphology
SiH ₂ Cl ₂ /Ar ^a	625	Polycrystalline ~ 100Å grain size
SiH ₄ /He ^b	400	Polycrystalline ~ 500Å grain size
SiH ₄ /Ar ^c	450	Polycrystalline few hundred Å grain size.
SiH ₄ /H ₂ ^d	600	Polycrystalline
SiH ₄ /H ₂ ^h	750-900	Single Crystal
SiH ₄ /H ₂ ^e	800-900	"
SiH ₄ ^f	760	"
SiH ₄ ^g	775	"
SiH ₄ /GeH ₄ ^j	600-850	"

a. Reference 280

b. Reference 279

c. Reference 278

d. Reference 330

e. Reference 281

f. Reference 331

g. Reference 283

h. Reference 282

j. Reference 332

1.2×10^{10} dynes cm^{-2} were reported for the undoped and doped films respectively. For these plasma conditions, PECVD does not offer any process temperature reduction relative to low pressure CVD from a SiH_4 source. In fact temperatures 25-35°C higher were found necessary for the onset of polycrystalline film formation.

Plasma-enhancement has been found to be very effective in reducing the substrate temperature for epitaxial Si deposition to the 600-900°C range, from the 1050-1200°C required by low pressure CVD. Pre-deposition plasma cleaning and native oxide removal from the Si substrates was found to be very important in permitting epitaxy at these low temperatures. This was performed reactively in H_2 plasmas^{281,282} or by inert gas sputter etching.²⁸³ Deposition of both epitaxial and polycrystalline Si films by PECVD is relatively new, and there is much work yet to be carried out in order to characterize these films. Reif²⁸⁴ has recently reviewed the epitaxial growth.

Germanium single crystals have been grown epitaxially on $\text{NaCl}(100)$ substrates by PECVD²⁸⁵ from GeH_4/H_2 mixtures at 450°C, about 150°C lower than required by normal CVD. A high frequency, inductively coupled plasma was employed. It was found that the plasma power had to be restricted to the minimum level possible in the initial stage of deposition until nucleation was complete. This was necessary to avoid plasma-induced damage to the NaCl surface, and subsequent damage propagation into the Ge film. After nucleation, the plasma power could be safely increased to give a usefully high deposition rate of $1700 \text{ \AA} \text{ min}^{-1}$. Electrical properties of these epitaxial Ge films look very promising.

Hariu et al.^{286,287} deposited epitaxial GaAs films onto GaAs(100) and Ge(100) substrates at temperatures above 350°C and 500°C, respectively. In contrast to conventional PECVD in which one function of the plasma is to dissociate the reactant gas molecules, the Ga and As sources were elemental, with their fluences to the substrate provided by thermal evaporation of elemental sources within the PECVD chamber. An inductively coupled plasma was maintained between the substrate and the evaporation sources, in a low pressure (~ 20 mtorr) of Ar in the earlier work²⁸⁶ and H_2 in the later work.²⁸⁷ Thus the function of the plasma was to supply energy to the growth surface to increase surface migration velocities and hence permit epitaxy at these reduced temperatures. Alternatively, one may regard the plasma as providing a locally enhanced surface temperature. A second effect of the exposure of the substrate to the plasma was the plasma etch oxide removal from the GaAs surface at the beginning of growth; this effect was also beneficial in promoting the low temperature epitaxy. Auger analysis showed the absence of interfacial oxide for the plasma-enhanced growth but its presence for a deposition without plasma-enhancement. In addition, a uniform level of oxygen was detected throughout the film but not in the plasma-enhanced grown material. Thus the plasma exposure has the added beneficial effect of reducing contaminant incorporation into the epitaxial growth. It should be noted that epitaxial growth was promoted only within a limited range of plasma power; presumably at higher plasma power, competing damaging effects of the plasma become dominant.

Using the same plasma-enhanced elemental evaporation technique, with a hydrogen plasma, Sato et al.²⁸⁸ have grown epitaxial GaSb layers on GaAs at temperatures as low as 340°C. Use of a hydrogen plasma in place of an Ar plasma was found to be beneficial to the electrical and optical properties of the GaSb layer. Electronic properties of undoped p-type layers were equal to those of layers grown by MBE or MOCVD at much higher growth temperatures.

Device quality, epitaxial GaAs layers have also been grown by Pande.²⁸⁹ The method employed was plasma-enhanced MOCVD, in which conventional Ga and As vapor phase sources (tri-methyl gallium and either arsine or tri-methyl arsenic) were used, but with the additional requirement relative to normal MOCVD that the arsenic source be passed through a plasma upstream from the substrates. The growth surface was not exposed to the plasma. Thus the approach is quite different from that of Hariu et al.²⁸⁶ discussed above. The rationale for Pande's approach is that in normal MOCVD of GaAs, the arsenic source dissociates less readily than the gallium source, and the high growth temperature and excess arsenic source flow are necessary to achieve sufficient dissociation of the arsenic source. Thus the growth temperature can be reduced by pre-dissociation of the arsenic source alone in the upstream plasma. Epitaxial layers with good morphology were obtained at 425°C, in comparison to temperatures in excess of 600°C necessary for normal MOCVD growth.

4.3.5 Metals. Many of the commonly employed metallizations in semiconductor processing may be deposited by PECVD, generally excepting the noble metals for which no vapor sources are readily available. In silicon technology, the most common metallizations at present are aluminum and highly doped polycrystalline silicon. The latter was discussed in the preceding section. For VLSI applications, the current trend is to replace these metallizations by transition metal silicides or perhaps transition metals themselves. These have improved electromigration resistance, lower resistivity (than polycrystalline Si), and can be patterned into sub-micron features. The PECVD of Al and of the transition metals Mo and W is discussed in this section, and transition metal silicides will be discussed in the following section.

Source gases for metal deposition are usually an organometallic or a volatile halide vapor, transported into the plasma by either hydrogen or by an inert gas. PECVD of Al from both types of source, AlCl_3 and $(\text{CH}_3)_3\text{Al}$, has been described by Ito.²⁹⁰ Al films deposited at room temperature had the lowest resistivity ($\sim 1 \times 10^{-5} \Omega \text{ cm}$), but this is about a factor 4 higher than that obtained by sputtering or evaporation. This was believed to be due to incorporation of trace amounts of oxygen. Films deposited at higher temperatures had higher resistivity and the oxygen content was detectable by Auger analysis. $(\text{CH}_3)_3\text{Al}$ reacts spontaneously with oxidizers, and a freshly deposited Al surface is an excellent O getter, hence very low leak and outgassing rates and high gas throughputs in conjunction with high deposition rates will be necessary to produce lower resistivity films.

PECVD of tungsten and molybdenum has been reported by Tang, Chu and Hess^{291,292} and further discussed in the review of Hess.¹⁹² These workers have employed tungsten or molybdenum hexafluoride source

gases for PECVD at high frequency in a parallel-plate reactor. In both cases the use of hydrogen as a reactive diluent was necessary to achieve deposition. This is necessary because the plasma dissociation of WF_6 or MoF_6 produces atomic fluorine, which is a very effective etchant for W or Mo. Thus the process of W or Mo film formation by reaction of WF_n or MoF_n adsorbed radicals on the heated substrate surface is in competition with the etching process. Addition of hydrogen produces atomic hydrogen in the plasma, which acts as a scavenger for atomic fluorine, allowing deposition to become the dominant process. Thus deposition rate increases with H_2 flow to a maximum value, and then begins to decrease with further increase in H_2 flow due to a reduction in residence time becoming the dominant effect. A H_2/WF_6 flow ratio of 3 was employed for W deposition, and a H_2/MoF_6 flow ratio of 7 for Mo deposition.²⁹² Metallic, smooth W films were deposited for substrate temperatures in the range 300-350°C at a power density of 0.06 W cm⁻² and 200 mtorr pressure. Auger analysis showed the W films to be oxygen free, and also did not reveal any F content. However, resistivity of the as-deposited films was about an order of magnitude higher than that of bulk W, and a factor 4 higher than that of high temperature CVD W films. Higher hydrogen flow during deposition produced lower as-deposited resistivity. In all cases a short high temperature anneal could reduce the resistivity to close to the bulk value, apparently without significant grain growth. Hence, trace amounts of F incorporation, below the detection limit of Auger analysis, which could be out-diffused at elevated temperature were suspected to be responsible for the high as-deposited resistivity.

4.3.6 Silicides. Silicides of tungsten and molybdenum have been deposited by PECVD from W and Mo halide source gases in conjunction with silane. Akimoto and Watanabe²⁹⁵ deposited WSi_x films from a SiH_4/WF_6 source gas mixture in a high frequency, parallel-plate plasma reactor. Additional hydrogen, as needed to permit W deposition (see Sectio 4.3.5), was not necessary, since an adequate H atom supply for F atom scavenging was provided by the SiH_4 decomposition. Rapid surface reaction kinetics between WF_n and SiH_4 radicals produced high deposition rates of 500 Å min⁻¹, an order of magnitude faster than for deposition of W alone. The W/Si stoichiometry in the film could be controlled over a very wide range by variation of the WF_6/SiH_4 flow ratio. This determined the as-deposited resistivity, with W rich films having lower resistivity. For films of W/Si ratio around unity, PECVD WSi_x films have similar resistivity to those sputter deposited. High temperature annealing reduces this resistivity, although curiously, for films with W/Si close to unity, annealing at temperatures higher than 500°C does not produce any further reduction relative to the value reached at 500°C.

Tabuchi et al.²⁹⁴ added SiH_4 to their $MoCl_5/H_2$ mixture employed for Mo deposition, and deposited films with a wide range of Mo/Si stoichiometry. Resistivities as-deposited were high.

4.3.7 Other Materials. Many other materials have been deposited by PECVD, as shown in Table 4. Those of interest in semiconductor applications will be discussed in this section. This includes various oxides, nitrides (dielectric and metallic), and polymer films. Other materials, such

as various forms of carbon and carbides which are mainly used as anti-wear and anti-corrosion surface coatings, are not considered here, but have been reviewed in some detail by Ohja.¹⁸⁶ In addition, the growth of surface native oxides, nitrides, and carbides, by exposure to oxygen, nitrogen and carbon containing plasmas respectively, is not covered here. These are also covered in the review of Ohja,¹⁸⁶ and the former, which is of great relevance in semiconductor processing, is covered in an earlier chapter of this work.²⁹⁶

In addition to the previously discussed nitride, oxide and oxynitride of silicon, other oxides and nitrides are also of interest for microelectronics applications. In particular, for processing of III-V semiconductors, the use of non-silicon based dielectrics can be advantageous, since Si is an amphoteric dopant in III-V compound semiconductors. This can be a concern for high temperature processing applications, such as implant annealing caps. An appropriate alternative is the use of oxides and nitrides of the group III elements, from which unintentional doping of the III-V semiconductor cannot occur.

The oxides and nitrides of group III elements which have been deposited by PECVD include the oxide and nitrides of boron, aluminum and gallium. A potential advantage of the use of the oxide or nitride of gallium or aluminum is that it is feasible to deposit these materials in the same reactor as that used for the MOCVD growth of the III-V material itself, to which the vapor phase supply of the group III element is already connected. Thus one can deposit the group III oxide or nitride directly onto a freshly grown III-V surface which has not been air exposed or etched in any way. This offers the possibility to study a very clean dielectric/semiconductor interface and to obtain interfaces with low densities of interface states, of great interest for MIS applications.

It is for this last application, on InP, that Meiners²⁹⁷ has studied PECVD Al_2O_3 and, in addition, PECVD AlP_xO_y . The former was deposited from trimethylaluminum, $(\text{CH}_3)_3\text{Al}$, and O_2 . These reactants were only mixed at low pressure within the deposition chamber above the heated substrate. The plasma reactor was of the indirect variety, and only the O_2 reactant was passed through the upstream plasma chamber. Substrate temperatures in the range 300-450°C were employed, and the film resistivity was found to increase with deposition temperature. However, even for a 450°C deposition temperature, the resistivity of $10^{14}\Omega\text{ cm}$ obtained was not sufficiently high for meaningful capacitance-voltage studies from which interface density of states could be obtained. This relatively low resistivity does not appear to be a result of the PECVD technique. AlP_xO_y films were deposited by the same indirect plasma technique, but with the addition of pre-pyrolyzed PH_3 above the substrate. These films were of much improved resistivity, with values in excess of $10^{16}\Omega\text{ cm}$ obtained for a deposition temperature of 375°C. The deposited film composition with regard to the Al/P ratio was Al rich relative to stoichiometric AlPO_4 . These AlP_xO_y films produced the best interface properties with InP yet reported, with interface densities of states within the gap of $1 \times 10^{11}\text{ cm}^{-2}\text{ eV}^{-1}$. Low densities of states for the InP interface with PECVD SiO_2 had previously been reported,^{298,299} but these were rather irreproducible and very dependent of the surface cleaning technique prior to dielectric deposition.²⁹⁸

Al_2O_3 films have also been deposited by PECVD from AlCl_3 and O_2 source gases³⁰⁰ in a low frequency, inductively coupled plasma, at substrate temperatures in the range 400-500°C. AlCl_3 was transported into the plasma from a variable temperature bubbler using the O_2 as carrier gas. Adherent films of up to several microns thick could be deposited on silicon, suggesting that stress in the Al_2O_3 film was low. AlCl_3 was also employed as the Al source in the PECVD of AlN_x films from an AlCl_3/N_2 plasma.³⁰¹ Films were only deposited over the temperature range 800-1200°C in order to compare their properties to CVD films deposited at the same temperature.

Gallium oxide is of interest for use as a dielectric film, particularly on Ga based III-V semiconductors. It has been deposited on GaAs in an attempt to eliminate the detrimental occurrence of elemental As at the interface between GaAs and its native oxide.³⁰² Trimethylgallium in argon and oxygen were fed into a magnetically confined high frequency plasma at a low pressure (≈ 1 mtorr). The optimum deposition temperature was 150°C, and films of Ga_2O_3 stoichiometry were deposited. Fluorine was incorporated into the Ga_2O_3 film by addition of a small amount of CF_4 to the plasma. This was done in the initial stage of deposition in an attempt to compensate dangling bonds and hence interface states produced by oxygen at a GaAs surface. An alternative approach to eliminate the effects of oxygen at the GaAs/dielectric interface is to use a nitride in place of an oxide, and GaN is a natural candidate for this application. Matsushita et al.³⁰³ deposited approximately stoichiometric GaN films by PECVD from trimethylgallium and ammonia, with Ar or N_2 diluents, in an inductively coupled plasma. Properties of the films such as H content, stress, resistivity, etc. were not reported.

Titanium nitride is a technologically interesting material both for microelectronic applications as well as for tribological and decorative coatings. Interest for semiconductor applications arises from its metallic conductivity and its excellent diffusion barrier properties. While the normal method of deposition is by reactive sputtering, TiN_x has also been deposited by PECVD.³⁰⁴ TiCl_4 vapor in H_2 carrier gas and N_2 source gases were fed into an inductively coupled plasma to which the substrate was directly exposed. TiN_x films could be deposited down to 250°C, but temperatures in excess of about 350°C were needed to obtain good adhesion to Ni substrates. Stoichiometry or electrical properties were not reported.

Another interesting dielectric nitride, which is composed entirely of group V elements, is P_3N_5 . This material has been deposited by Veprek and Roos³⁰⁵ from elemental phosphorus and nitrogen sources by a plasma-enhanced chemical transport technique. A high frequency N_2 plasma generated above amorphous phosphorus plasma etched the phosphorus and produced vapor phase PN species in the plasma. These then produced deposition of PN_x films on a heated substrate exposed to the plasma. Stoichiometry of the deposit could be varied up to P_3N_5 . In later work, Veprek et al.³⁰⁶ used conventional PECVD from PH_3 and N_2 source gases in a high frequency plasma to deposit hydrogenated P_3N_5 films. At a N_2/PH_3 flow ratio in excess of 10 and a substrate temperature of $\approx 330^\circ\text{C}$, a film of P_3N_5 stoichiometry but with an additional 13 atomic% H content was deposited. Infra-red spectra showed the existence of both P-H and N-H bonds. However this H content does not appear to modify the dielectric

properties relative to the hydrogen-free P_3N_5 deposited by the plasma-enhanced chemical vapor transport. In addition, the H is tightly bound and reported to be stable up to 700°C.

Polymeric organic materials may also be deposited by PECVD (References 307-309 and references therein). In plasma etching, this can either be a problem to be circumvented or an effect to be employed beneficially. For example, as discussed in Sec. 3.4.4, the addition of H_2 to a CF_4 etching plasma promotes polymer deposition in competition with etching. Since under dynamic equilibrium conditions, the thickness of polymer present on an SiO_2 surface is less than that on a Si surface, the effect of polymer deposition is to enhance the SiO_2/Si etch rate selectivity. However, this can also result in undesirable contamination of the etched surface.

Retajczyk and Gallagher³¹⁰ have studied PECVD polymer films with regard to their dielectric applications in microelectronics. Fluorocarbon polymeric films were deposited in a high frequency plasma in a parallel plate reactor from tetrafluoroethylene ($F_2C=CF_2$) monomer source gas. The deposited polytetrafluoroethylene (PTFE) films had a number of attractive properties. These included a low dielectric constant (beneficial to minimizing capacitance in interlevel dielectric applications), very low stress, excellent adhesion, and general chemical inertness, but with the ability to be patterned by oxygen plasma etching.

4.4 Interface Properties

An ideal interface can be defined as one which is atomically abrupt, i.e., of submonolayer width, with each of the materials at either side existing in a state unperturbed by the presence of the other. This is generally not the case for most practical interfaces. An interface often can be characterized by the extent of interfacial mixing or compound formation, and by the extent of the substrate surface modification. The former is frequently a chemical effect, albeit ion-assisted, (although it can occur by a purely physical process), whereas the latter is a physical effect.

Interfacial mixing or compound formation generally is a function of the chemical identities of the two materials involved, as is the case in thermal CVD. However, in PECVD interfacial reaction rates may be greater or less than in thermal CVD, due to the competing effects of a rate reduction due to the lower bulk substrate temperature and a rate enhancement due to ion bombardment (giving an increased effective surface temperature). This rate enhancement occurs by the same mechanism as that in which ion bombardment increases the rate of heterogeneous reactions involved either in film growth (see, for example, Table 8, in Section 4.3.1 for SiN_x deposition) or in etching (see Section 2.2.2). Intermixing can also occur by the physical mechanism of momentum transfer when incident ion energies are high, as in low frequency PECVD (see Sections 2.2.2 and 4.2.1). The use of a low pressure shifts the incident ion energy distribution to a higher value, and the deposition of a material from a plasma from which the deposition rate to plasma power density ratio is low increases the ion dose received in the interface region. Hence both those deposition conditions will favor the possibility of physical intermixing occurring.

Even if intermixing does not occur, the ion bombardment discussed above can also produce damage in the atomic structure of the substrate surface. In the case of a monocrystalline substrate, this structural damage may be in the form of complete or partial amorphization of a thin surface layer, or in a less severe case, the formation of vacancies and interstitials within the lattice. This type of surface structural damage may be of concern in semiconductor processing applications where the deposited film acts as an etch mask or a diffusion mask.

Both of the above discussed modes of structural deviation from an ideal interface are also expected to modify the electrical and optical properties of the near-surface region of the semiconductor. For example, electrical modification may be observed in the form of an increased density of mid-gap interface states, modified Schottky barrier heights, reduced carrier mobility, surface carrier type conversion and modified Ohmic contact formation. A related, optically observable effect is a reduction in photoluminescence yield due to increased non-radiative recombination rates.

Structural modification of a monocrystalline surface may be observed by Rutherford backscattering (RBS), with the best sensitivity obtained under channeling conditions with detection of nuclei backscattered at grazing exit angle. This technique enables the number of atoms displaced from equilibrium lattice sites to be counted, and this is a direct measurement of the degree of damage. The method has been employed to study the interface between an InP(100) surface and SiO₂ and SiN_x dielectrics deposited by PECVD over a wide range of deposition conditions, and for comparison, films deposited by rf diode sputtering.^{195,196}

In Figure 62, we show the areal density of In atoms producing backscattering measured for a clean, native oxide stripped InP surface and for similarly prepared surfaces subsequently coated by $\sim 50\text{\AA}$ of SiO₂, deposited either by high frequency PECVD, or by rf diode sputtering in an Ar:O₂ atmosphere. It may be seen that there is no measurable increase in the number of off-lattice site In atoms for the PECVD SiO₂ coated surfaces at any plasma power density up to the limit studied (0.32 Wcm^{-2}), at which the SiO₂ deposition rate was 1500\AA min^{-1} . In contrast, diode sputtering of SiO₂ produced significant surface damage and/or interfacial mixing at all power densities, reaching a saturation level of $5.5 \times 10^{15}\text{ cm}^{-2}$ displaced In atoms (equivalent to the total number of In atoms in 30\AA InP) at about 0.25 Wcm^{-2} . Thus the PECVD SiO₂/InP interface is abrupt on a monolayer scale, with no induced structural modification of the InP surface. In later work,¹⁹⁶ we have also looked at InP/PECVD dielectric interfaces for PECVD situations in which the interface region receives a larger and/or more energetic ion dose. These were low frequency PECVD SiO₂ (increased energy), high frequency PECVD SiN_x (increased dose) and low frequency PECVD SiN_x from SiH₄/N₂ (increased dose and energy). Ratios of deposition rate to plasma power density and of incident ion energy distributions for these cases have been given in Sections 4.3.1.1 (inc. Table 8), 4.3.2, and 4.2.1 (inc. Figure 10). In all three cases, a monolayer abrupt, damage-free interface was also produced. However, the low frequency PECVD of SiN_x from a dilute SiH₄ in N₂ plasma appeared to be on the verge of producing

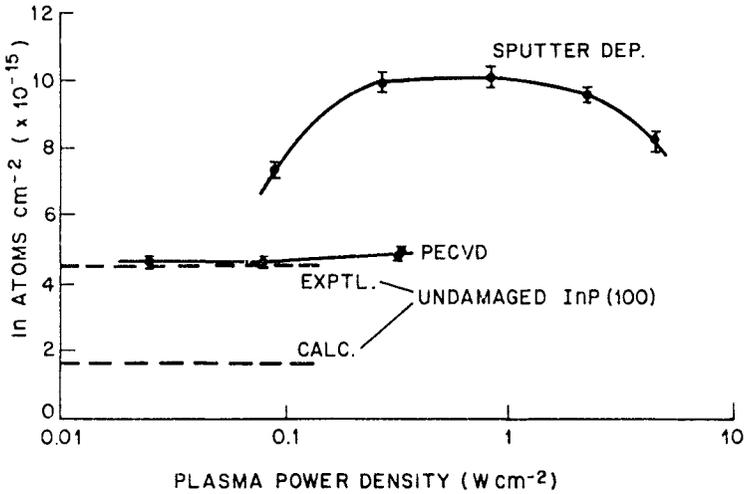


Figure 62: Surface structural damage produced in an InP (100) surface as a function of plasma power density as a result of SiO₂ deposition by PECVD at 13.56 MHz compared with deposition by rf diode sputtering from an SiO₂ target in an Ar:O₂ atmosphere (from Reference 195).

measurable structural modification. Since GaAs and Si have each been shown to be less susceptible than InP to low energy ion bombardment damage,³¹¹ it is reasonable to assume that PECVD will produce sharp, damage-free interfaces on those materials also, except in cases where chemical interactions occur.

On a more sensitive scale, when electrical characterization techniques are employed which are sensitive to atomic displacements at the parts-per-million level (as opposed to the percent level at monolayer sensitivity for the channeling described above), PECVD is generally found to produce modification of a semiconductor surface. As a generality, low frequency PECVD is more damaging than high frequency. For example, low frequency PECVD of SiN_x from SiH₄/N₂ onto GaAs produces interface states in a broad band peaking at about 0.6 eV below the conduction band edge, with the high peak density of $1.5 \times 10^{13} \text{ eV}^{-1} \text{ cm}^{-2}$.³¹² The use of high frequency PECVD SiN_x as an interlayer dielectric on normally-off GaAs FETs produces strong degradation of the FET I-V characteristics.³¹³ However these authors also reported that the PECVD could be made degradation-free merely by isolating the GaAs wafer from the grounded electrode by the use of a quartz spacer. The authors believed that the deposition was made damage-free by elimination of ion bombardment due to elimination of the potential difference between the wafer and the plasma. However it is by no means clear that this was the effect responsible, since even a floating potential surface acquires a negative bias relative to the glow region of the plasma in order to equalize the incident electron and positive ion fluences. Since at the 13.56 MHz employed, the sheath behaves capacitively (see Section

2.1.2.1.3), the additional capacitance of the quartz spacer beneath the wafer relative to the capacitance of the surrounding element of the sheath will act as a voltage divider, so that some reduction in incident ion energies will be achieved. This argument is, however, only applicable if the sheath is not disturbed. This would not be the case if the quartz spacer was thick enough for its capacitance per unit area to be as large as that of the sheath (i.e. a voltage reduction to zero is not possible).

Regarding interface state generation by PECVD of a dielectric onto InP, a number of general observations may be made. First, both chemical and physical effects are important. It is not sufficient merely to eliminate radiation-induced damage. For example, SiN_x is not a good gate dielectric on InP,^{298,299} since it causes the Fermi level to be pinned in the upper half of the band gap. It is assumed that these interface states are the result of the chemical interaction of the InP surface with the reducing, H containing ambient involved in SiN_x deposition, producing P vacancies at the interface. In contrast to GaAs, however, presence of native oxide at the InP/dielectric interface does not pin the Fermi level. Direct PECVD, as in a parallel plate reactor, is probably not suitable for InP MISFET applications, but indirect PECVD of SiO_2 can produce a low density of interface states.²⁹⁸ These results were, however, rather irreproducible, and very dependent on the preceding cleaning and etching of the InP surface. Good results with PECVD SiO_2 were also obtained by Woodward et al.²⁹⁹ even with direct plasma exposure of the InP surface in a system with external electrodes. Since tetraethoxysilane (TEOS) and O_2 were used as source gases, at a substrate temperature of 300°C , it is possible that a few monolayers of deposition occurred by thermal CVD prior to ignition of the plasma, thus protecting the InP surface from the effects of direct plasma exposure. These authors attributed their interface quality (relative to that with SiN_x) to the minimization of P loss in the oxidizing deposition ambient. Finally, the best electrical interface properties reported to date were recently obtained by Meiners²⁹⁷ using indirect PECVD of AlP_xO_y (this is described in Section 4.3.7). Perhaps the presence of PH_3 in the deposition ambient further reduced or eliminated any phosphorus deficiency in the InP surface or in any interfacial native oxide layer.

In summary, in the absence of chemical interaction, PECVD materials form monolayer sharp interfaces with monocrystalline semiconductor substrates, and do not produce structural surface damage even for PECVD with direct plasma exposure at low frequency. In addition, PECVD is applicable to high quality electronic interface formation, but in situations where the plasma is employed only to pre-dissociate the reactants and direct plasma exposure is avoided.

5. SUMMARY AND CONCLUSIONS

We have shown in this chapter how plasmas are utilized in semiconductor processing for both etching and thin film deposition after first discussing the fundamental aspects of glow discharge physics and chemistry and emphasizing unifying principles which pertain to both etching

and deposition. Electrical properties of the plasma were described in terms of equivalent circuit models and some of the shortcomings of these models were given. Plasma chemistry in the gaseous state and the current understanding of plasma-surface interactions were reviewed in depth. The effects of changing interrelated plasma parameters such as frequency, pressure, rf power, loading, feed gas, etc. were qualitatively described with reference to quantitative models wherever possible. In particular, emphasis was placed on the effects of frequency on the energetics and anisotropy of ion bombardment, with discussion of the circumstances under which this produces anisotropy in the etching or deposition process.

Though a detailed understanding of all plasma interactions is clearly lacking, many plasma processes are presently in commercial use. An approach to developing a plasma etching process was given along with a review of standard reactor geometries. Some standard etching processes were also described. An in-depth review of plasma-enhanced chemical vapor deposition (PECVD) has been given from the point of view of its application in semiconductor technology. Generalities of PECVD were presented first; these included the necessary properties of PECVD materials for their applications in microelectronic devices, the types of PECVD reactor in use and their relative suitability for different applications, source gases which can be used, control of deposition uniformity and the attendant loss of one degree of freedom in plasma parameter variation, and the generally applicable methods of influencing PECVD film properties. Following this, the specific PECVD processes and range of process parameters for deposition of the materials (dielectrics, semiconductors, and metals) of interest in microelectronics were discussed in detail. For the well-studied, silicon-based materials, emphasis was placed on recent developments in techniques and on the intercorrelation of film properties, as well as on their dependence on the interrelated PECVD parameters for the specified process. In particular, the PECVD processes for SiO_2 and SiN_x were contrasted, in terms of the differences resulting from rapid heterogeneous reaction rates for SiO_2 deposition, but slow, rate-limiting heterogeneous reaction rates in SiN_x deposition. In addition, the PECVD of numerous other materials of relevance, including elemental and compound semiconductors and III-V dielectrics has been included.

The last ten years have witnessed a rapid growth in the application of plasma processes. The general utility of PECVD has been recognized and it has been applied for more materials than just the silicon-based ones to which it was more-or-less restricted in the mid-seventies. In particular, PECVD has recently been applied to epitaxial growth at reduced temperature, with significant growth expected in this area in forthcoming years. During this preceding decade more effort was frequently expended in developing processes that fitted specific needs than in attempting to understand the dominant mechanisms responsible. More recently there has been a growing interest in unravelling the many interrelated chemical and physical phenomena to gain a better understanding of processes already in use. As this understanding increases, the future should bring innovations, especially in the area of reactor design and feedstock blending, to yield better process control and reproducibility. Other areas of current

research and growth potential include microwave discharges and hybrid plasma-laser-electron beam processes. These techniques are principally aimed at eliminating the damaging effects of exposure to the many forms of radiation present in an rf discharge.

For the present, though a lack of understanding is frequently a limiting factor, plasma processing is still the only known method of achieving, on a commercial scale, desired etching and deposition properties. In view of the continued feature size reduction for VLSI in Si technology, and the expanding III-V technologies, we expect to see a continued increase in the research, development, and commercial exploitation of plasma processing.

REFERENCES

1. C.M. Melliar-Smith and C.J. Mogab, in *Thin Processes*, ed. J.L. Vossen and W. Kern, Academic Press (New York, 1978), p. 497.
2. C.J. Mogab, *Inst. Phys. Conf. Ser.* 53:37 (1980).
3. S. Gourrier and M. Bacal, *Plasma Chem. Plasma Proc.* 1:217 (1981).
4. D.L. Flamm and V.M. Donnelly, *Plasma Chem. Plasma Proc.* 1:317 (1981).
5. J.W. Coburn, *Plasma Chem. Plasma Proc.* 2:1 (1982).
6. D.W. Hess, *Plasma Chem. Plasma Proc.* 2:141 (1982).
7. D.L. Flamm, V.M. Donnelly, D.E. Ibbotson, *J. Vac. Sci. Technol.* B1:23 (1983).
8. H.F. Winters, J.W. Coburn, and T.J. Chuang, *J. Vac. Sci. Technol.* B1:469 (1983).
9. C.J. Mogab, in *VLSI Technology*, ed. S.M. Sze, McGraw-Hill (New York, 1983), p. 303.
10. J.A. Mucha and D.W. Hess, *Microcircuit Processing: Lithography and Dry Etching*, eds. L.F. Thompson and C.G. Willson (American Chemical Society).
11. R.A. Powell, *Dry Etching for Microelectronics*, North-Holland (Amsterdam, 1984).
12. D.L. Flamm, V.M. Donnelly, and D.E. Ibbotson, in *VLSI Electronics 8*: ed. D.L. Brown and N.E. Einspruch, Academic Press (New York, 1984), p. 189.
13. V.M. Donnelly, D.E. Ibbotson, and D.L. Flamm, in *Ion Bombardment Modification of surfaces*, ed. O. Auciello and R. Kelly, Elsevier (Amsterdam, 1984), p. 323.
14. A.J. van Roosmalen, *Vacuum* 34:429 (1984).
15. J.W. Coburn, *Vacuum* 34:157 (1984).
16. J.A. Mucha, *Solid State Technol.* 28:123 (1985).
17. F.F. Chen, *Introduction to Plasma Physics*, Plenum (New York, 1974).
18. G.P. Davis and R.A. Gottscho, *J. Appl. Phys.* 54:3080 (1983).
19. B. Chapman, *Glow Discharge Processes*, John Wiley and Sons (New York, 1980).
20. W.B. Pennebaker, *IBM J. Res. Develop.* 23:16 (1979).
21. R.A. Gottscho and M.L. Mandich, *J. Vac. Sci. Technol.* A3:617 (1985).
22. R.A. Gottscho, R.H. Burton, D.L. Flamm, V.M. Donnelly, and G.P. Davis, *J. Appl. Phys.* 55:2707 (1984).
23. R.W. Gould, *Phys. Lett.* 11:236 (1964).
24. H.R. Koenig and L.I. Maissel, *IBM J. Res. Develop.* 14:168 (1970).
25. J.H. Keller and W.B. Pennebaker, *IBM J. Res. Develop.* 23:3 (1979).
26. C.M. Horwitz, *J. Vac. Sci. Technol.* A1:60 (1982).
27. C.B. Zarowin and R.S. Horwath, *J. Electrochem. Soc.* 129:2541 (1982).
28. C.B. Zarowin, *J. Electrochem. Soc.* 130:1144 (1983).
29. K. Kohler, J.W. Coburn, D.E. Horne, E. Kay, and J.H. Keller, *J. Vac. Sci. Technol.* A3:638 (1985).
30. J.W. Coburn and M. Chen, *J. Appl. Phys.* 51:3134 (1980).
31. R. d'Agostino, F. Cramarossa, S. DeBenedictis, and G. Ferraro, *J. Appl. Phys.* 52:1259 (1981).

32. R. d'Agostino, V. Colaprico, and F. Cramarossa, *Plasma Chem. Plasma Proc.* 1:365 (1981).
33. R. d'Agostino, F. Cramarossa, V. Colaprico, and R. d'Ettole, *J. Appl. Phys.* 54:1284 (1983).
34. R. d'Agostino, F. Cramarossa, and S. DeBenedictis, *Plasma Chem. Plasma Proc.* 2:213 (1982).
35. H.-J. Tiller, D. Berg, and R. Mohr, *Plasma Chem. Plasma Proc.* 1:247 (1981).
36. T.A. Miller, *Plasma Chem. Plasma Proc.* 1:3 (1981).
37. R.A. Gottscho and T.A. Miller, *Pure and Appl. Chem.* 56:189 (1984).
38. K.E. Greenberg, G.A. Hebner, and J.T. Verdeyen, *Appl. Phys. Lett.* 44:299 (1984).
39. R. Walkup, R.W. Dreyfus, and Ph. Avouris, *Phys. Rev. Lett.* 50:1846 (1983).
40. V.M. Donnelly, D.L. Flamm, and G. Collins, *J. Vac. Sci. Technol.* 21:817 (1982).
41. L.A. Farrow, to be published (1985).
42. D.K. Doughty and J.E. Lawler, *Appl. Phys. Lett.* 45:611 (1984).
43. D.K. Doughty, S. Salihi, and J.E. Lawler, *Phys. Lett.* 103A:41 (1984).
44. J.P.M. Schmitt, P. Gressier, M. Krishnan, G. De Rosny, and J. Perrin, *Chem. Phys.* 84:281 (1984).
45. D.K. Doughty, E.A. Den Hartog, and J.E. Lawler, *Appl. Phys. Lett.* (in press).
46. C.A. Moore, G.P. Davis, and R.A. Gottscho, *Phys. Rev. Lett.* 52:538 (1984).
47. R. Walkup, Ph. Avouris, R.W. Dreyfus, J.M. Jasinski, and G.S. Selwyn, *Appl. Phys. Lett.* 45:372 (1984).
48. M.L. Mandich, C.E. Gaebe, and R.A. Gottscho, *J. Chem. Phys.* 83:3349 (1985).
49. R.H. Bruce, *Solid State Technol.* 24:64 (1981).
50. R.H. Bruce, *J. Appl. Phys.* 52:7064 (1981).
51. D.E. Ibbotson, D.L. Flamm, and V.M. Donnelly, *J. Appl. Phys.* 54:5974 (1983).
52. R.H. Bruce, in *Plasma Processing*, R.G. Frieser and C.J. Mogab, eds., The Electrochemical Society (Pennington, NJ, 1981), p. 243.
53. F.H.M. Sanders, A.W. Kofschoten, J. Dieleman, R.A. Haring, A. Haring, and A.E. de Vries, to be published.
54. D.L. Smith and P.G. Saviano, *J. Vac. Sci. Technol.* 21:768 (1982).
55. D.L. Smith and R.H. Bruce, *J. Electrochem. Soc.* 129:2045 (1982).
56. J.W. Coburn and E. Kay, *J. Appl. Phys.* 43:4965 (1972).
57. C.A. Moore and R.A. Gottscho, unpublished results.
58. C.B. Zarowin, *Thin Solid Films* 85:33 (1981).
59. F.-A. Sarott, Z. Iqbal, and S. Veprek, *Solid State Commun.* 42:465 (1982).
60. S. Veprek, *Proc. Mater. Res. Soc. Eur.* Strasbourg, July 1984.
61. D.L. Flamm, *Plasma Chem. Plasma Proc.* 1:37 (1981).
62. W.R. Harshbarger, R.A. Porter, T.A. Miller, and P. Norton, *Appl. Spectrosc.* 31:201 (1977).
63. C.J. Mogab, A.C. Adams, and D.L. Flamm, *J. Appl. Phys.* 49:3796 (1978).
64. G. Smolinsky and D.L. Flamm, *J. Appl. Phys.* 50:4982 (1979).
65. R.H. Burton, C.L. Hollien, L. Marchut, S.M. Alys, G. Smolinsky, and R.A. Gottscho, *J. Appl. Phys.* 54:6663 (1983).
66. H.M. Rosenstock, K. Draxl, B.W. Steiner, and J.T. Herron, *J. Phys. Chem. Ref. Data* 6:Supp. 1 (1977).
67. L.G. Christophorou, *Environ. Health Perspect.* 36:3 (1980).
68. M. Hecq, A. Hecq, and M. Fontignies, *Thin Solid Films* 115:L45 (1984).
69. E.A. Triesdale, G. Smolinsky, and T.M. Mayer, *J. Appl. Phys.* 51:2902 (1980).
70. B.N. Chapman and V.J. Minkiewicz, *J. Vac. Sci. Technol.* 15:239 (1978).
71. W.D. Davis and T.A. Vanderslice, *Phys. Rev.* 131:219 (1963).
72. R.T.C. Tsui, *Phys. Rev.* 168:107 (1968).
73. V.J. Minkiewicz, M. Chen, J.W. Coburn, B.N. Chapman, and K. Lee, *Appl. Phys. Lett.* 35:393 (1979).
74. C.B. Zarowin, *J. Vac. Sci. Technol.* A2:1537 (1984).

75. C.B. Zarowin, *J. Appl. Phys.* 57:929 (1985).
76. A. Günterschulze, *Z. Physik* 36:683 (1926).
77. S. Veprek and V. Marecek, *Solid State Electron.* 11:683 (1968).
78. J.J. Wagner and S. Veprek, *Plasma Chem. Plasma Proc.* 2:95 (1982).
79. J.L. Mauer, J.S. Logan, L.B. Zielinski, and G.C. Schwartz, *J. Vac. Sci. Technol.* 15:1734 (1978).
80. C.J. Mogab, *J. Electrochem Soc.* 124:1262 (1977).
81. D.L. Flamm, D.N.K. Wang, and D. Maydan, *J. Electrochem. Soc.* 129:2755 (1982).
82. J.W. Coburn and H.F. Winters, *J. Appl. Phys.* 50:3189 (1979).
83. S. Veprek and F.-A. Sarott, *Plasma Chem. Plasma Proc.* 2:233 (1982).
84. V.M. Donnelly and D.L. Flamm, *J. Appl. Phys.* 51:5273 (1980).
85. D.L. Flamm, V.M. Donnelly, and J.A. Mucha, *J. Appl. Phys.* 52:3633 (1981).
86. J.A. Mucha, V.M. Donnelly, D.L. Flamm, and L.M. Webb, *J. Phys. Chem.* 85:3529 (1981).
87. D.E. Ibbotson, D.L. Flamm, J.A. Mucha, and V.M. Donnelly, *Appl. Phys. Lett.* 44: 1129 (1984).
88. M.A.A. Clyne and W.S. Nip, in *Reactive Intermediates in the Gas Phase*, ed. D.W. Setser, Academic Press (New York, 1979), p. 1.
89. T.J. Chuang, *J. Appl. Phys.* 51:2614 (1980).
90. J.F. Morar, F.R. McFeely, N.D. Shinn, G. Landgren, and F.J. Himpsel, *Appl. Phys. Lett.* 45:174 (1984).
91. F.R. McFeely, J.F. Morar, N.D. Shinn, G. Landgren, and F.J. Himpsel, *Phys. Rev.* B30:764 (1984).
92. F.R. McFeely, American Vacuum Society National Symposium, Reno, Nevada, December 1984.
93. M.J. Vasile and F.A. Stevie, *J. Appl. Phys.* 53:3799 (1982).
94. H.F. Winters and F.A. Houle, *J. Appl. Phys.* 54:1218 (1983).
95. M.A. Bosch, L.A. Coldren, and E. Good, *Appl. Phys. Lett.* 38:264 (1980).
96. D.M. Brown, B.A. Heath, T. Coutumas, and G.R. Thompson, *Appl. Phys. Lett.* 37: 159 (1980).
97. U. Gerlach-Meyer, J.W. Coburn, and E. Kay, *Surf. Sci.* 103:177 (1981).
98. U. Gerlach-Meyer, *Surf. Sci.* 103:524 (1981).
99. M.W. Geis, G.A. Lincoln, N. Efremow, and W.J. Piacentini, *J. Vac. Sci. Technol.* 19:1390 (1981).
100. B.A. Heath, *Solid State Technol.* 24:75 (1981).
101. G.A. Lincoln, M.W. Geis, L.J. Mahoney, A. Chu, B.A. Vojak, K.B. Nichols, W.J. Piacentini, N. Efremow, and W.T. Lindley, *J. Vac. Sci. Technol.* 20:786 (1981).
102. T.M. Mayer, R.A. Barker, L.J. Whitman, *J. Vac. Sci. Technol.* 18:349 (1981).
103. H. Okano and Y. Horiike, in *Plasma Processing 81-1*, ed. R.G. Frieser and C.J. Mogab, The Electrochemical Society (Pennington NJ, 1981), p. 199.
104. H. Okano and Y. Horiike, *Jpn. J. Appl. Phys.* 20:2429 (1981).
105. Y.Y. Tu, T.J. Chuang, H.F. Winters, *Phys. Rev.* B23:823 (1981).
106. R.A. Barker, T.M. Mayer, R.H. Burton, *Appl. Phys. Lett.* 40:583 (1982).
107. R.A. Haring, A. Haring, F.W. Saris, and A.E. de Vries, *Appl. Phys. Lett.* 41:174 (1982).
108. T.M. Mayer and R.A. Barker, *J. Electrochem. Soc.* 129:585 (1982).
109. T.M. Mayer and R.A. Barker, *J. Vac. Sci. Technol.* 21:757 (1982).
110. H. Okano and T. Horiike, *Jpn. J. Appl. Phys.* 21:696 (1982).
111. R.A. Powell, *Jpn. J. Appl. Phys.* 21:L170 (1982).
112. R.A. Barker, T.M. Mayer, and W.C. Pearson, *J. Vac. Sci. Technol.* B1:37 (1983).
113. E. Ikawa and Y. Kurogi, *Jpn. J. Appl. Phys.* 22:1263 (1983).
114. A.W. Kolfschoten, R.A. Haring, A. Haring, and A.E. de Vries, *J. Appl. Phys.* 55:3813 (1984).
115. S.C. McNevin and G.E. Becker, *J. Vac. Sci. Technol.* B2:27 (1984).

116. R.A. Haring, A.W. Kofschoten, and A.E. de Vries, *Nucl. Instr. Meth.* 2:544 (1984).
117. R.A. Rossen and H.H. Sawin, *Appl. Phys. Lett.* 45:860 (1984).
118. R.A. Rossen and H.H. Sawin, *J. Vac. Sci. Technol.* A3:881 (1985).
119. D.J. Vitkavage, M.S. Ameen, and T.M. Mayer, *J. Vac. Sci. Technol.* A2:492 (1984).
120. E.L. Barish, D.J. Vitkavage, and T.M. Mayer, *J. Appl. Phys.* 57:1336 (1985).
121. S.C. McNevin and G.E. Becker, *J. Vac. Sci. Technol.* B3:485 (1985).
122. S.C. McNevin and G.E. Becker, *J. Vac. Sci. Technol.* B4:1203 (1986).
123. T. Mizutani, C.J. Dale, W.K. Chu, and T.M. Mayer, *Nuc. Instr. Method B* (in press).
124. P. Sigmund, in *Inelastic Ion-Surface Collisions*, ed. N.H. Tolk, J.C. Tully, W. Heiland, and C.W. White, Academic Press (London, 1977), p. 121.
125. D. Edelson and D.L. Flamm, *J. Appl. Phys.* 56:1522 (1984).
126. M.J. Kushner, *J. Appl. Phys.* 53:2923 (1982).
127. M.J. Kushner, *J. Appl. Phys.* 54:4958 (1983).
128. M.J. Kushner, H.M. Anderson, P.J. Hargis, *Mater. Res. Soc. Symp. Proc.* 38:201 (1985).
129. M.T. Mocella, M.W. Jenkins, H.H. Sawin, and K.D. Allen, *Mat. Res. Soc. Symp. Proc.* 38:227 (1985).
130. R.L. Maddox, *Microelectron. J.* 11:4 (1980).
131. A.A. Milgram, *J. Vac. Sci. Technol.*, B1:490 (1983).
132. P.D. Townsend, J.C. Kelly and N.E.W. Hartly, *Ion Implantation, Sputtering and Their Applications*, Academic, New York, 1976.
133. L.I. Maissel and R. Glang, *Handbook of Thin Film Technology*, McGraw-Hill, New York (1970).
134. A.R. Reinberg, *Extended Abstracts Electrochem. Soc. Spring Meeting*, San Francisco, CA, May 12-17, 1974.
135. L.M. Ephrath, *IEEE Trans. on Elec. Dev.*, 28:1315 (1981).
136. D. Maydan U.S. Patent, 4,298,443, Nov. 3, 1981.
137. Y.W. Hu, *Electrochem. Soc. Meeting*, 84-2:abs. 392 (1984).
138. B.J. Curtis, *Sol. State Technol.*, 23:129 (1980).
139. P.A. Heimann and R.J. Schutz, *J. Electrochem. Soc.*, 131:881 (1984).
140. V.M. Donnelly, D.L. Flamm, W.C. Dautremont-Smith, and D.J. Werder, *J. Appl. Phys.* 55:242-52 (1984).
141. C.J. Mogab and H.J. Levinstein, *J. Vac. Sci. Technol.*, 17:721 (1980).
142. R.W. Light, *J. Electrochem. Soc.*, 130:2225 (1983).
143. S.E. Bernacki and B.B. Kosicki, *J. Electrochem. Soc.*, 131:1926 (1984).
144. E.L. Hu and R.E. Howard, *J. Vac. Sci. Technol.* B2, 85 (1984).
145. J.D. Chinn, I. Adesida and E.D. Wolf, *J. Vac. Sci. Technol.* B1:1028 (1983).
146. J.M. Moran and D. Maydan, *Bell Sys. Tech. J.*, 58:1027 (1979).
147. D.L. Flamm, *Sol. State Technol.*, 22:109 (1979).
148. L.M. Ephrath, *J. Electrochem. Soc.*, 126:1419 (1979).
149. R.A.H. Heinecke, *Sol. State Electron.*, 18:1146 (1975).
150. S. Matsuo, *J. Vac. Sci. Technol.* 17:587 (1980).
151. L.M. Ephrath and D.J. DiMaria, *Sol. State Technol.* 24:182 (1981).
152. R.A. Gdula, *IEEE Trans. on Elec. Dev.*, 26:644 (1979).
153. E.H. Nicollian and J.R. Brews, *MOS Physics and Technology*, John Wiley & Sons, New York (1982).
154. R.A. Gdula and P.C. Li, *J. Electrochem. Soc.*, 124:1927 (1977).
155. J.R. Brews, W. Fichtner, E.H. Nicollian and S.M. Sze, *IEEE Electron Devices Lett.* EDL-1:2 (1980).
156. S.P. Murarka, *J. Vac. Sci. Technol.*, 17:775 (1980).
157. A.C. Adams and C.D. Capio, *J. Electrochem. Soc.*, 128:366 (1981).
158. K. Hirata, Y. Ozaki, M. Oda and M. Kimizuka, *IEEE Trans. on Electron Dev.*, 28:1323 (1981).
159. G.C. Schwartz and P.M. Schaible, *J. Electrochem. Soc.*, 130:1898 (1983).

160. R.W. Light and H.B. Bell, *J. Electrochem. Soc.*, 131:459 (1984).
161. F.Y. Robb, *Electrochem. Soc. meeting*, 84-2:abs. 355 (1984).
162. T.P. Chow and G.M. Fanelli, *Electrochem. Soc. meeting*, 84-2, abs. 406 (1984).
163. K.M. Cham and S.-Y. Chiang, *IEEE Electron Dev. Lett.*, EDL-4:303 (1983).
164. T. Morie, K. Minegishi and S. Nakajima, *IEEE Electron Dev. Lett.*, EDL-4:411 (1983).
165. J.A. Bondur and H.B. Pogge, U.S. patent 4,104,086, Aug. 1, 1978.
166. J.A. Bondur and H.A. Clark, *Solid State Technol.*, 23:122 (1980).
167. T.M. Mayer, *J. Electron. Mat.*, 9:513 (1980).
168. D.N.K. Wang, D. Maydan and H.J. Levinstein, *Solid State Tech.*, 23:122 (1980).
169. B. Chapman and M. Nowak, *Semi. Inter.*, 3:139 (1980).
170. K. Hirobe, Y. Kureishi and T. Tsuchimoto, *J. Electrochem. Soc.*, 128:2686 (1981).
171. D.W. Hess, *Solid State Technol.*, 24:189 (1981).
172. M. Oda and K. Hirata, *Jap. J. of Appl. Phys.*, 19:L405 (1980).
173. R.A.H. Heinecke, *Sol. State Electron.*, 19:1039 (1976).
174. Y. Kawamoto, T. Kure, N. Hashimoto and T. Takaichi, *Electrochem. Soc. Meeting*, 84-2:abs. 395 (1984).
175. V.M. Donnelly, D.L. Flamm, and D.E. Ibbotson, *J. Vac. Sci. Technol. A*2:626 (1983).
176. R.H. Burton, R.A. Gottscho, G. Smolinsky, *Dry Etching for Microelectronics*, edited by R.A. Powell, Elsevier (New York, 1984).
177. R.P.H. Chang, C.C. Chang and S. Darack, *J. Vac. Sci. Technol.*, 20:45 (1982).
178. R.C. Weast, *Handbook of Chemistry and Physics*, CRC Press, Boca Raton, Fla. (1984).
179. R.P.H. Chang and S. Darack, *Appl. Phys. Lett.* 38:898 (1981).
180. G. Smolinsky, R.P.H. Chang, T.M. Mayer, *J. Vac. Sci. Technol.* 18:12 (1981).
181. V.M. Donnelly, D.L. Flamm, C.W. Tu, and D.E. Ibbotson, *J. Electrochem. Soc.* 129:2533 (1982).
182. Hollahan, J.R. and Rosler, R.S., in: *Thin Film Processes* (J.L. Vossen and W. Kern, eds.), pp. 335-360, Academic Press, New York (1978).
183. M.J. Rand, *J. Vac. Sci. Technol.* 16:420-7 (1979).
184. A.R. Reinberg, *Ann. Rev. Mater. Sci.* 9:341 (1979).
185. E.P.G.T. van de Ven, *Solid State Technology*, Vol. 24, no. 4, 167-71 (1981).
186. S.M. Ojha, *Physics of Thin Films* 12:237-96 (1982).
187. T.D. Bonfield, in: *Deposition Technologies for Films and Coatings* (R.F. Bunshah, ed.), pp. 365-384, Noyes, Park Ridge, NJ (1982).
188. J.A. Thornton, in: *Deposition Technologies for Films and Coatings* (R.F. Bunshah, ed.), pp. 19-62, Noyes, Park Ridge, NJ (1982).
189. J.A. Thornton, *Thin Solid Films* 107:3-19 (1983).
190. A.C. Adams, in *VLSI Technology* (S.Sze, ed.), pp. 93-129, McGrawHill, New York (1983).
191. A.C. Adams, *Solid State Technology*, Vol. 26, No. 4, 135-9 (April 1983).
192. D.W. Hess, *J. Vac. Sci. Technol. A*2:244-52 (1984).
193. R. Reif, *J. Vac. Sci. Technol. A*2:429-35 (1984).
194. A. Sherman, *Thin Solid Films* 113:135-149 (1984).
195. W.C. Dautremont-Smith, and L.C. Feldman, *Thin Solid Films* 105:187 (1983).
196. W.C. Dautremont-Smith and L.C. Feldman, *J. Vac. Sci. Technol. A*:April-May (1985).
197. J.W. Peters, *Tech. Dig. IEEE Int. Electron Dev. Meeting*, 240 (1981).
198. P.K. Boyer, G.A. Roche, W.H. Ritchie, and G.J. Collins, *Appl. Phys. Lett.* 40:716 (1982).
199. S.D. Allen, *J. Appl. Phys.* 52:6501-5 (1981).
200. E. Ong, and E.L. Hu, *Solid State Technology*, Vol. 27, No. 6, 155-60 (June 1984).
201. A.R. Reinberg, U.S. Patent 3,757,733; Sept. 11, 1973; assigned to Texas Instruments.

202. R.S. Rosler, W.C. Benzing, and J. Baldo, *Solid State Technology*, Vol. 19, No. 6, 45-50 (June 1976).
203. R.S. Rosler, and G.M. Engle, *Solid State Technology*, Vol. 24, No. 4, 172-7 (April 1981).
204. A.D. Weiss, *Semiconductor International*, July 1983, pp. 88-94.
205. L.R. Thompson, J.J. Rocca, K. Emery, P.K. Boyer, and G.J. Collins, *Appl. Phys. Lett.* 43:777-779 (1983).
206. D.C. Bishop, K.A. Emery, J.J. Rocca, L.R. Thompson, H. Zarnani, and G.J. Collins, *Appl. Phys. Lett.* 44:598-600 (1984).
207. J.F. O'Hanlon, *A User's Guide to Vacuum Technology*, Wiley, New York, 1980, pp. 321-7.
208. J.F. O'Hanlon, *Pumping Hazardous Gases*, AVS Monograph, AVS, New York, 1981.
209. A.K. Sinha, and E. Lugujo, *Appl. Phys. Lett.* 32:245 (1978).
210. P.S. Peercy and H.J. Stein, in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
211. F.L. Schuermeyer, in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
212. K.C. Vanner, J.R. Cockrill, and R.S. Butlin, *The Solid Films* 80:161 (1981).
213. A.K. Sinha, H.J. Levinstein, T.E. Smith, G. Quintana, and S.E. Haszko, *J. Electrochem. Soc.* 125:601 (1978).
214. A.K. Sinha, *Solid State Technology*, Vol. 23, No. 4, 133 (1980).
215. K.M. Mar, and G.M. Samuelson, *Solid State Technology* Vol. 23, No. 4, 137 (1980).
216. H.E. Maes, G. Heyns, M. Hinoul, H. Loos, A. Van Calster, and K. Allaert, in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
217. V.S. Nguyen in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
218. H.F. Sterling, and R.C.G. Swann, *Solid State Electronics* 8:653 (1965).
219. Y. Kuwano, *Jpn. J. Appl. Phys.* 7:88 (1968).
220. Y. Kuwano, *Jpn. J. Appl. Phys.* 8:876 (1969).
221. H. Dun, P. Pan, F.R. White, and R.W. Douse, *J. Electrochem. Soc.* 128:1555 (1981).
222. M.J. Helix, K.V. Vaidyanathan, B.G. Streetman, H.B. Dietrich, P.K. Chatterjee, *Thin Solid Films* 55:143 (1978).
223. R. Gereth, and W. Scherber, *J. Electrochem. Soc.* 119:1248 (1972).
224. G.J. Valco and V.J. Kapoor, in *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society (Pennington, NJ, 1983).
225. W.C. Dautremont-Smith, "PECVD of SiN_x", unpublished (1982).
226. M.J. Rand, and D.R. Wonsidler, *J. Electrochem. Soc.* 125:99 (1978).
227. W.A.P. Claassen, W.G.J.N. Valkenburg, F.H.P.M. Habraken, and Y. Tamminga, *J. Electrochem. Soc.* 130:2419 (1983).
228. G.M. Samuelson, and K.M. Mar, *J. Electrochem. Soc.* 129:1773 (1982).
229. H.E. Maes, J. Remmerie, M. Hinoul, R. Van den Berghe, and R. Vlaeminck, in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
230. Chow, R., Lanford, W.A., Ke-Ming, W., and Rosler, R.S., *J. Appl. Phys.* 53:5630 (1982).
231. K. Koyama, K. Takasaki, M. Maeda, and M. Takagi, paper presented at the Electrochemical Society Fall Meeting, Denver, CO, Oct. 1981.
232. F. Martinet, G. Guegan, and J.-C. Jesionka, in: *Proc. Symp. on Silicon Nitride Thin Insulating Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
233. T.F. Retajczyk, and A.K. Sinha, *Thin Solid Films* 70:241 (1980).
234. A.K. Sinha and T.E. Smith, *J. Appl. Phys.* 45:2756 (1978).
235. C.D. Fung, T.E. Nagy, and W.H. Ko, in: *Proc. Symp. on Silicon Nitride Thin Insulating*

- Films*, Vol. 83-8, The Electrochemical Society, Pennington, NJ (1983).
236. W.A. Lanford and M.J. Rand, *J. Appl. Phys.* 49:2473 (1978).
 237. S. Fujita, H. Toyoshima, T. Ohishi, and A. Sasaki, *Jpn J. Appl. Phys.* 23:L144 (1984).
 238. L.L. Ley, H.R. Shanks, C.K. Fang, K.J. Grunts, and M. Kardona, *Phys. Rev. B* 22, 6140 (1980).
 239. A.C. Adams, F.B. Alexander, C.D. Capiro, T.E. Smith, *J. Electrochem. Soc.* 128:1545 (1981).
 240. J.R. Hollahan, *J. Electrochem. Soc.* 126:930 (1979).
 241. D. Kupperts, J. Koenings, and H. Wilson, *J. Electrochem. Soc.* 123:1079 (1976).
 242. S.W. Ing, and W. Davern, *J. Electrochem. Soc.* 111:120 (1964).
 243. W.C. Dautremont-Smith, and J. Lopata., *J. Vac. Sci. Technol.* B1:943 (1983).
 244. L.A. Koszi, V.M. Donnelly, W.C. Dautremont-Smith, and P.A. Barnes, *IEEE Trans. Electron Dev.* ED-31:1910 (1984).
 245. G. Kaganowicz, V.S. Ban, and J.W. Robinson, *J. Vac. Sci. Technol.* A2:1233 (1984).
 246. W.E. Spear, and P.G. LeComber, *Solid State Commun.* 17:1193 (1975).
 247. D.E. Carlson, and C.R. Wronski, *Appl. Phys. Lett.* 28:671 (1976).
 248. C.R. Wronski, D.E. Carlson, and R.E. Daniel, *Appl. Phys. Lett.* 29:602 (1976).
 249. D.E. Carlson, *IEEE Trans. Electron. Dev.* ED-24:449 (1977).
 250. D.E. Carlson, *Sol. Energy Mat.* 3:503 (1980).
 251. W.E. Spear, P.G. LeComber, S. Kinmond, and M.H. Brodsky *Appl. Phys. Lett.* 28:105 (1976).
 252. P.G. LeComber, W.E. Spear, and A. Ghaith, *Electron. Lett.* 15:179 (1979).
 253. P.G. LeComber, A.J. Snell, K.D. Mackenzie, and W.E. Spear in: *Proc. 9th Int. Conf. on Amorphous and Liquid Semiconductors*, Grenoble, July 1981 (B.K. Chakraverty and D. Kaplan, eds.), Paris, 1981.
 254. I. Shimizu, T. Komatsu, K. Saito, and E. Inoue, *J. Non-Cryst. Solids* 35-36:773 (1980).
 255. W.E. Spear, P.G. LeComber, A.J. Snell, and R.A. Gibson *Thin Solid Films* 90:359 (1982).
 256. Y. Hamakawa, (ed.) *Amorphous Semiconductor Technologies and Devices*, OHM and North-Holland, Tokyo and Amsterdam, 1982.
 257. Y. Hamakawa, *Thin Solid Films* 108:301 (1983).
 258. H. Fritzsche, *Thin Solid Films* 90:119 (1982).
 259. T. Hamasaki, M. Ueda, A. Chayahara, M. Hirose, and Y. Osaka, *Appl. Phys. Lett.* 44:600 (1984).
 260. J. C. Knights, R.A. Lujan, M.P. Rosenblum, R.A. Street, and D.K. Biegelsen, *Appl. Phys. Lett.* 38:331 (1981).
 261. B.A. Scott, M.H. Brodsky, D.C. Green, P.B. Kirby, R.M. Plecenik, and E.E. Simonyi, *Appl. Phys. Lett.* 37:725 (1980).
 262. O. Kuboi, M. Hashimoto, Y. Yatsurugi, H. Nagai, M. Aratani, M. Yanokura, S. Hayashi, I. Kohno, and T. Nozaki, *Appl. Phys. Lett.* 45:543 (1984).
 263. F.J. Demond, G. Müller, H. Damjantschitsch, H. Mannsperger, S. Kalbitzer, P.G. LeComber, and W.E. Spear, in: *Proc. 9th Int. Conf. on Amorphous and Liquid Semiconductors*, Grenoble, July 1981 (B.K. Chakraverty and D. Kaplan, eds.), in *J. Phys.* (Paris), Colloq. C4:42 (1982).
 264. M. Hirose, in: *Amorphous Semiconductor Technologies and Devices* (Y. Hamakawa, ed.), OHM and North Holland, Tokyo and Amsterdam, 1982.
 265. J.C. Knights, and R.A. Lujan, *Appl. Phys. Lett.* 35:244 (1979).
 266. R.A. Street, J.C. Knights, and D.K. Biegelsen, *Phys. Rev.* B18:1880 (1978).
 267. J. Shirafuji, H. Matsui, Y. Inuishi, and Y. Hamakawa, *Jpn. J. Appl. Phys.* 22, 775 (1983).
 268. A. Matsuda, S. Yamasaki, K. Nakagawa, H. Ohkushi, K. Tanaka, S. Iizima, M. Matsumura, and H. Yamamoto, *Jpn. J. Appl. Phys. Lett.* 19:L305 (1980).

269. H. Okamoto, Y. Nitta, T. Yamaguchi, and Y. Hamakawa, *Solar Energy Mat.* 2:313 (1980).
270. T. Hamasaki, H. Kurata, M. Hirose, and Y. Osaka, *Appl. Phys. Lett.* 37:1084 (1980).
271. A. Madan, S.R. Ovshinsky, and E. Benn, *Philos. Mag.* B40:259 (1979).
272. H. Matsumura, and S. Furukawa, in: *Amorphous Semiconductor Technologies and Devices* (Y. Hamakawa, ed.), OHM and North Holland, Tokyo and Amsterdam, 1982.
273. Y. Kuwano, M. Ohnishi, H. Nishiwaki, S. Tsuda, and H. Shibuya, *Jpn. J. Appl. Phys.* 20:supplement 20-2, 157 (1981).
274. Y. Kuwano, in Reference 256.
275. Y. Hamakawa, in Reference 256.
276. H. Haruki and Y. Uchida, in Reference 256.
277. J.E. Greene, *CRC Crit. Rev. Solid State and Mat. Sci.*, Vol. 11, 189-227 (1984).
278. F. Morin and M. Morel, *Appl. Phys. Lett.*: 35, 686 (1979).
279. Y. Nagata, and A. Kunioka, *Appl. Phys. Lett.* 38:142 (1981).
280. T.I. Kamins, and K.L. Chiang, *J. Electrochem. Soc.* 129:2326 (1982), and 2331 (1982).
281. W.G. Townsend, and M.E. Uddin, *Solid State Electron.*, 16:39 (1973).
282. S.R. Shanfield, and R. Reif, Abs. 144, *The Electrochemical Society Extended Abstracts*, Vol. 83-1, San Francisco, CA, May 1983.
283. T.J. Donahue, W.R. Burger, and R. Reif, *Appl. Phys. Lett.* 44:346 (1984).
284. R. Reif, *J. Electrochem. Soc.* 131:2430 (1984).
285. R.A. Outlaw and P. Hopson, Jr., *J. Appl. Phys.* 55:1461 (1984).
286. T. Hariu, K. Takenaka, S. Shibuya, Y. Komatsu, and Y. Shibata, *Thin Solid Films* 80:235 (1981).
287. T. Hariu, Y. Matsushita, Y. Komatsu, S. Shibuya, S. Igarishi, and Y. Shibata, *Inst. Phys. Conf. Ser. No. 65*, Chap. 2, p. 141; paper presented at Int. Symp. GaAs and Related Compounds, Albuquerque, NM, 1982.
288. Y. Sato, K. Matsushita, T. Hariu, and Y. Shibata, *Appl. Phys. Lett.* 44:592 (1984).
289. K.P. Pande, Abstract No. 340, *Extended Abstracts*, Vol. 83-1, The Electrochemical Society Spring Meeting, San Francisco, CA, May 8-13, 1983.
290. T. Ito, *Denshi Zairyo (Electronic Parts and Materials)*, 22:31 (1983) (in Japanese).
291. C.C. Tang, J.K. Chu, and D.W. Hess, *Solid State Technology*, p. 125 (Mar. 1983).
292. J.K. Chu, C.C. Tang, and D.W. Hess, *Appl. Phys. Lett.* 41:74 (1982).
293. F. Okyama, *Appl. Phys.* A28:125 (1982).
294. A. Tabuchi, S. Inoue, M. Maeda, and M. Takagi, *Jpn. Semicond. Technol. News*, p. 43 (Feb. 1983).
295. K. Akimoto, and K. Watanabe, *Appl. Phys. Lett.* 39:445 (1981).
296. B.E. Deal, chapter X in this work (1985).
297. L.G. Meiners, *Thin Solid Films* 113:85 (1984).
298. L.G. Meiners, *J. Vac. Sci. Technol.* 19:373 (1981).
299. J. Woodward, D.C. Cameron, L.D. Irving, and G.R. Jones, *Thin Solid Films* 85:61 (1981).
300. H. Katto, and Y. Koga, *J. Electrochem. Soc.* 118:1619 (1971).
301. J. Bauer, L. Biste, and D. Bolze, *Phys. Status Solidi (a)* 39:173 (1977).
302. R.P.H. Chang, C.C. Chang, S. Darack, and M.H. Read, unpublished (1980).
303. K. Matsushita, Y. Matsuno, T. Hariu, and Y. Shibata, *Thin Solid Films* 80:243 (1981).
304. N.J. Archer, *Thin Solid Films* 80:221 (1981).
305. S. Veprek, and J. Roos, *J. Phys. Chem. Solids*, 37:554 (1976).
306. S. Veprek, Z. Iqbal, J. Bunner, and M. Scharli, *Philos. Mag.* B43:527 (1981).
307. M. Millard, in: *Techniques and Applications of Plasma Chemistry* (J.R. Hollahan and A.T. Bell, eds.), Wiley, New York, 1974; Chapter 5.
308. H. Yasuda, in: *Thin Film Processes* (J.L. Vossen and W. Kern, eds.), Academic

- Press, New York 1978; chapter IV-2.
309. E. Kay, and A. Dilks, *J. Vac. Sci. Technol.* 18:1 (1978).
 310. T.F. Retajczyk, Jr. and P.K. Gallagher, *Materials Lett.* 2:23 (1983).
 311. R.S. Williams, *Solid State Commun.* 41:153 (1982).
 312. M.C. Chen, D.V. Lang, W.C. Dautremont-Smith, A.M. Sergent, and J.P. Harbison, *Appl. Phys. Lett.* 44:790 (1984).
 313. Y. Yamane, Y. Ishii, T. Mizutani, *Japanese Journal of Applied Physics*, 22:L350 (1983).
 314. A. Androshuk, et al., U.S. Patent 3,424,661 (1969).
 315. M. Shiloh, et al., *J. Electrochem. Soc.* 124:295 (1977).
 316. D.R. Secrist, and J.D. Mackenzie, *Bull. Am. Ceram. Soc.* 45:784 (1966).
 317. D.A. Anderson and W.E. Spear, *Phil Mag.* 35:1 (1977).
 318. J.H. Alexander, R.J. Joyce, and H.F. Sterling, in *Thin Film Dielectrics* (ed. F. Vratny), (The Electrochemical Society, Princeton, NJ, 1970), p. 186.
 319. C.J. Dell'Oca
 320. S.B. Hyder, and T.O. Yep, *J. Electrochem. Soc.* 123:1721 (1976).
 321. K.C. Wiemer, U.S. Nat. Tech. Inform. Serv. AD Ref. No. 769884/8GA (1973).
 322. H.F. Sterling et al., *Vide* 21:80 (1966).
 323. R.C. Chittick, *J. Noncryst. Sol.* 3:255 (1970).
 324. L. Holland, and S.M. Ojha, *Thin Solid Films* 38:L17 (1976).
 325. D.S. Whitmell, and R. Williamson, *Thin Solid Films* 35, 255 (1976).
 326. F.H. Cocks, P.L. Jones, and L.J. Dimmey, *Appl. Phys. Lett.* 36:970 (1980).
 327. A. E. Hultquist, and M.E. Sibert, *Adv. Chem. Ser.* 80:182 (1969).
 328. R.D. Wales, *Adv. Chem. Ser.* 80:198 (1969).
 329. J.C. Knights and J.E. Mahan, *Solid State Commun.* 21:983 (1977).
 330. W.R. Burger, T.J. Donahue, and R. Reif, in *Proc. 4th European Conf. on CVD* (J. Bloem et al., eds), Eindhoven Druk B.V./GEMA, The Netherlands, 1983, pp. 265-272.
 331. M. Pons, D. Bourgeat, J. Trilke, and B. Bourdon, *Le Vide* 200:3 (1980).
 332. S. Suzuki, and T. Itoh, *J. Appl. Phys.* 54:1466 (1983).
 333. K.R. Ryan and I.C. Plumb, *J. Phys. Chem.* 86:4678 (1982).

6

Physical Vapor Deposition

John A. Thornton

*Department of Materials Science and Coordinated Science Laboratory
University of Illinois
Urbana, Illinois*

1. INTRODUCTION

The term physical vapor deposition is used to denote those vacuum deposition processes such as evaporation and sputtering where the coating material is passed into a vapor transport phase by physical mechanism—i.e., evaporation, sublimation, or ion bombardment. Physical vapor deposition methods are the most universal of the available means for depositing thin films and coatings. Metallic, dielectric, and semiconducting coatings, in some cases with unique properties, can be deposited. Sputtering is particularly effective for providing controlled deposition of materials with complex composition. No attempt is made in this chapter to give a complete discussion of these technologies. The objective is instead to simply highlight those aspects that are of particular importance in semiconductor device fabrication and processing.

The metallization of semiconductor devices constitutes one of the primary applications of physical vapor deposition. Originally most device metallization was done using evaporation. The current trend is toward increased use of sputtering. The reasons are (1) the effectiveness of sputtering for depositing refractory metals and materials of complex composition, such as silicides, which are coming into increased use in device design, and (2) the recent development of sputtering technologies (magnetron sputtering) that minimize substrate radiation damage and offer greatly increased production capabilities. In fact, it has been estimated that about 70% of the 1983 wafer metallization activity involves the use of magnetron sputtering.

Other applications where sputtering plays a prominent role include magnetic thin films for recording applications, thin film resistors and capacitors for hybrid interconnect circuits, as well as microcircuit photolithographic mask blanks and transparent conducting electrodes. Sputtering may also be used to deposit microcircuit insulation layers, although chemical vapor deposition (CVD) or plasma-assisted CVD is a more common method for these particular coatings.

Physical vapor deposition is also used to deposit thin films for piezoelectric transducers, photoconductors, waveguides for integrated optics devices, and luminescent films for display devices. A potentially very large application is the fabrication of thin film photovoltaic devices for direct energy conversion.

A rapidly developing new method called molecular beam epitaxy (MBE) uses evaporation from multiple sources to deposit device quality semiconductors such as GaAs with precisely controlled doping profiles. MBE is finding increasing application for microwave and optoelectronic devices, with the promise of novel future devices which will incorporate specially synthesized superlattice structures having properties not found in homogeneous materials.

This chapter is divided into seven sections. Section two reviews some of the important aspects of the vacuum environment which relate to vacuum deposition. Section three discusses vacuum evaporation, and Section four discusses the application of vacuum evaporation to molecular beam epitaxy. Section five discusses deposition by sputtering. Section six discusses and the growth and properties of thin films and coatings, with particular attention to the influence of deposition parameters, such as concurrent ion bombardment, on coating properties. Considerable attention is also given to internal stresses because of the importance of these stresses in the performance and adhesion of device metallization layers. Finally, Section seven reviews some of the major considerations in the important application of device metallization.

2. THE VACUUM ENVIRONMENT

The unit of measure for pressure in vacuum systems is the Torr or Pascal. The Torr (1 Torr = 1 mm Hg) is a carryover from the time when pressure measurements were made primarily with manometers. Starting about 1975, most technical publications began changing to the International System of Units (SI), where the unit of pressure is the Pascal (Pa). The Pascal is the MKS unit of pressure: 1 Pa = 1 N/m² = 7.5 mTorr (1 mTorr = 0.133 Pa). Most pressure gauges are still calibrated in Torr or microns (1 micron = 1 mTorr). Therefore Torr will be used in this text, although both Torr and Pa are given in many cases.

Vacuum systems can be classified as high vacuum (10⁻³ to 10⁻⁶ Torr), very high vacuum (10⁻⁶ to 10⁻⁹ Torr), and ultrahigh vacuum (below 10⁻⁹ Torr). Figure 1 shows schematic drawings of several typical pumping configurations. The high vacuum pumps may be of the oil diffusion or turbomolecular (Figure 1 a), getter (Figure 1 b), or cryogenic (1 c) type. Pumping systems are

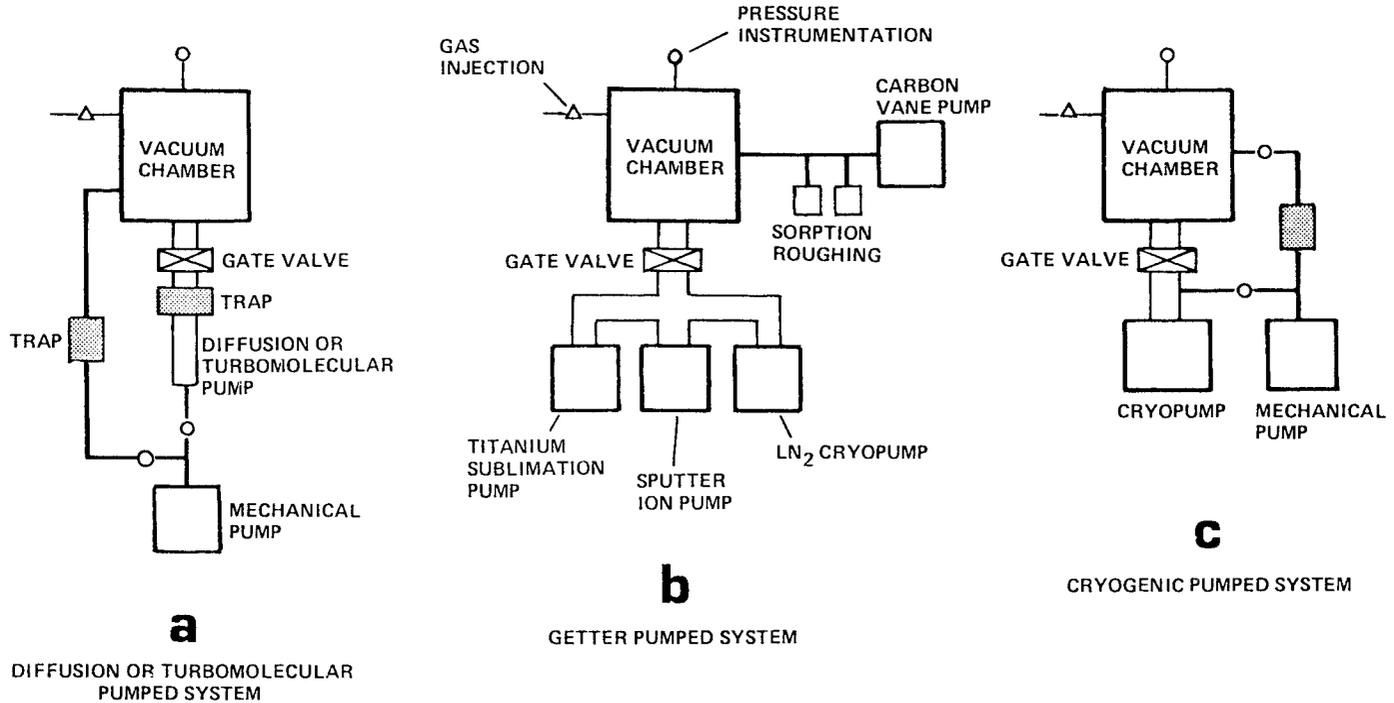


Figure 1: Schematic drawings showing vacuum pumping systems of the various types used in deposition technology: (a) diffusion or turbomolecular pump configuration, (b) getter pump configuration, and (c) cryogenic pump configuration.

discussed in References 1-4. Our concern here is with the “vacuum state” that is achieved and its implications on the deposition process.

Consider a cubic vacuum system with sides 1 m in length. The volume is 1 m^3 . The internal surface area is 6 m^2 . An ideal metal surface contains about 2×10^{15} adsorption sites per cm^2 . When the chamber is exposed to the atmosphere, an even larger density of molecules will become attached onto the walls because of surface irregularities and multilayer adsorption. The number of molecules per cubic centimeter in a room temperature gas is about

$$n = 3.3 \times 10^{16} p \quad \text{particles/cm}^3 \quad (1)$$

where p is the pressure in Torr. Thus when pumping is initiated we have the task of removing $(760 \text{ Torr})(3.3 \times 10^{16} \text{ molecules/cm}^3 \cdot \text{Torr})(10^6 \text{ cm}^3) = 2.5 \times 10^{25}$ molecules from the volume and at least $(2 \times 10^{15} \text{ molecules/cm}^2)(6 \times 10^4 \text{ cm}^2) = 1.2 \times 10^{20}$ molecules from the walls.

Now consider a chamber filled to an initial pressure P_0 with an ideal gas, which has no interactions with the walls other than reflections. When such a chamber is evacuated by a pump of constant volumetric efficiency, the pressure will decrease with time according to the equation.

$$p(t) = P_0 \exp(-t/\tau) \quad (2)$$

where the time constant, $\tau = V/S$, is a function of the chamber volume (V) and pumping speed (S). Suppose that the pumping system for our 1 m^3 chamber has a speed of 1000 liters/sec. Then $\tau = 1$ sec. Typical values are in the range from 0.1 sec to a few seconds.⁵ Thus Equation (2) predicts that the pressure in our chamber will decrease exponentially once pumping has commenced, dropping almost an order of magnitude during each 2 sec interval.

Now consider the removal of atmospheric gases from a chamber. The relationship given by Equation (2) is obeyed, after a few seconds from the start of evacuation and until a pressure of about 10 Torr is reached, as gas is removed from the volume of the chamber. Subsequently, the evacuation will become rate-limited by outgassing from the chamber walls. Under these conditions, the pressure will decrease much more slowly, obeying an equation of the form

$$p(t) = Q(t)/S \quad (3)$$

where $Q(t)$ is the total outgassing rate from the surfaces within the chamber at the time t . The outgassing rate, $Q(t)$, and therefore the chamber pressure, decrease as a function of time, because internal diffusion and surface desorption deplete the reservoirs of stored gas entrapped on the chamber internal surfaces.

The dwell time of an atom or molecule on a surface under vacuum will depend on the binding energy between the molecule and the surface, and on the surface temperature. See Equation 13 in Section 6.1. Physisorbed gases with binding energies of the order of 0.1 to 0.5 eV desorb quickly

during the initial pump-down and do not contribute to Equation 3. At the other extreme, chemisorbed gases with binding energies of 1.5 to 3 eV are released at very small rates, which are persistent but do not contribute significantly to the gas load. The troublesome gases are those which produce significant outgassing on the same time scale that is used in executing the deposition processes. These gases have desorption energies of about 1 eV. The most notable example is water. In some cases water may condense on vacuum surfaces to thicknesses of hundreds of monolayers.⁵

Figure 2 shows the specific outgassing rates (primarily water vapor) for several engineering materials.⁶ The specific outgassing rates typically obey an equation of the form $q = q_0(t_0/t)^n$, where n varies from about 0.5 for elastomers and plastics to about 0.3 to 1 for metals.⁷ Thus the total outgassing rate can be expressed as

$$Q(t) = A q_0 (t_0/t)^n \quad , \quad (4)$$

where A is the chamber surface area. The time t_0 is a reference point where $q = q_0$. For approximate calculations t_0 can be taken as the point at which the high vacuum valve is opened.

Consider the case of our 1 m³ chamber after three hours of pumping. Assume that the chamber is constructed of stainless steel and that the high vacuum pump has a speed of 600 liters/sec. The specific outgassing rate for stainless steel after 180 min. of pumping is seen in Figure 2 to be about 4×10^{-8} Torr-liters/sec-cm². From Equation 3 we estimate the chamber pressure to be $(4 \times 10^{-8} \text{ Torr-liters/sec-cm}^2)(6 \times 10^4 \text{ cm}^2)(600/\text{liters/sec})$ or 4×10^{-6} Torr.

In practical deposition systems unwanted gases are a result of desorption from the deposition sources, substrates, and hot filaments as well as the chamber walls. Back-streaming gases from the pumps also contribute contamination species. After prolonged pumping, the residual gases are typically H₂O, CO, CO₂, O₂ and N₂.¹ For a residual gas pressure p (given in Torr), the impingement rate on a substrate surface is

$$R = 3.51 \times 10^{22} p(\text{mT})^{-1/2} \text{ molecules/cm}^2\text{-sec} \quad (5)$$

where M is the molecular weight in grams and T is the temperature in °K. It is useful to keep in mind that at a residual gas pressure of 10^{-6} Torr the flux of gas incident on a (substrate) surface within the chamber corresponds to the flux associated with a deposition rate of about 1 Å/sec for a material of typical density.

The achievement of vacuums less than about 10^{-7} Torr in room temperature systems generally requires prohibitively long pumping times. Therefore, baking of the walls is used for ultra high vacuum systems. Because of the temperature dependence of the adsorbed atom binding (see Equation 13), this procedure is extremely effective. Thus the number of molecules pumped from a system at 200°C in one second can equal the number that would be pumped in a whole day at 20°C.⁵ It is important to note that the entire system must be heated at one time, because many of the liberated molecules in a partially heated system will recondense on the

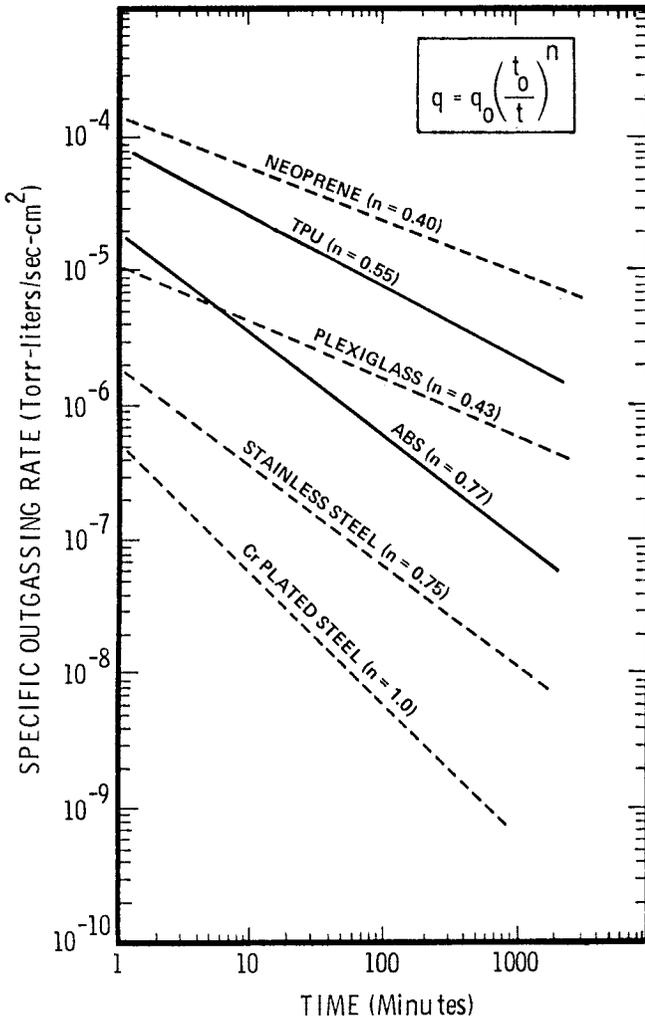
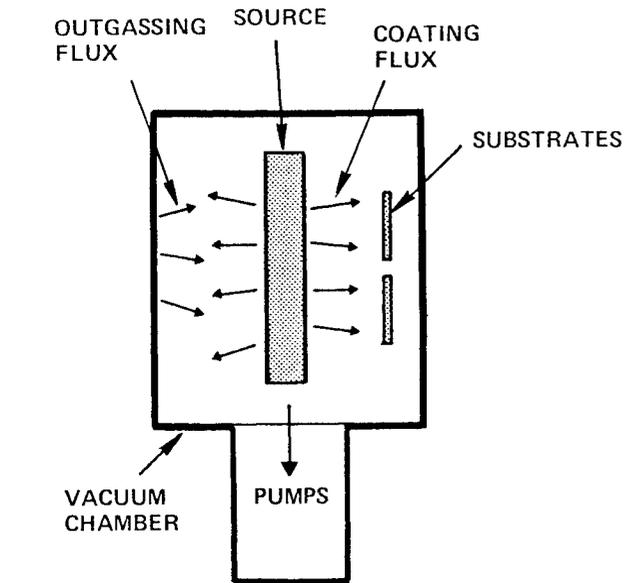


Figure 2: Specific outgassing rate as function of time for various materials. From Reference 6.

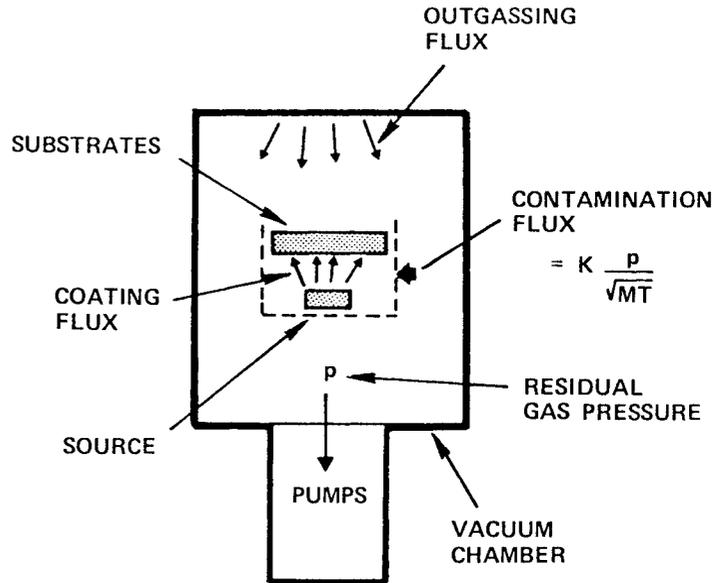
cooler surfaces, from which they can provide a continued source of impurity gases.

The required pumping time will depend on the application. Deposition processes are often started when the chamber pressure has reached an empirically determined value. High pumping speeds are often equated with cleanliness because of the relationship expressed by Equation 3. This is not necessarily correct.⁴ Two extremes are illustrated in Figure 3. In the case shown in Figure 3a, the “coating apparatus” fills much of the chamber and the deposition area is large. The gettering capacity of the coating flux,



$$\text{IMPURITY LEVEL} \sim \frac{\text{TOTAL OUTGASSING FLUX}}{\text{TOTAL COATING FLUX}}$$

a



$$\text{IMPURITY LEVEL} \sim \frac{K p / \sqrt{MT}}{\text{TOTAL COATING FLUX}}$$

b

Figure 3: Schematic illustration showing influence of apparatus geometry on the way in which wall outgassing affects the coating impurity level.

Q_c is large compared to the capacity of the physical pumping system. In such systems the purity of the deposit will be roughly proportional to $Q(t)/Q_c$. Therefore, the deposit purity depends on the pumping time, t , and the deposition rate, but is independent of the pumping speed of the vacuum system. In the case shown in Figure 3b, the deposition process is confined locally and represents a small perturbation to the vacuum system: i.e., the gettered flux is small compared to the physically pumped flux. In this case the flux of residual gas which enters the deposition region is given by Equation 5 and is therefore dependent on the residual gas pressure in the chamber. For a given pumping time, a larger pump will therefore decrease this pressure and improve the purity of the deposits. Actual situations will generally lie between the two extremes shown in Figures 3a and 3b. However, most production deposition systems will tend toward the case shown in Figure 3a.

3. EVAPORATION

3.1 Introduction

In the evaporation process, vapors of the coating material are released from a source because of heating. The source material may be in the liquid or solid state, depending on its vapor pressure relative to its melting point. Almost any conceivable method can be used to heat the source. One of the most common methods is resistive heating, either of the source material itself or of a support containing the material. Other common heating methods involve the use of an electron beam, a laser beam, or an arc discharge to produce local surface heating of the source material.

The evaporation process is usually carried out at a sufficiently low pressure (typically 10^{-5} to 10^{-6} Torr) so that the evaporated atoms undergo an essentially collisionless "line-of-sight" transport to the substrates. In this connection it is useful to remember that the mean free path of gas particles is about equal to

$$\lambda = 5/p_m, \text{ cm} \quad (6)$$

where p_m is the pressure in m Torr. Thus at a pressure of 10^{-4} Torr, λ is of the order of 50 cm and about equal to the size of a typical vacuum chamber. A second reason for using a low pressure is to avoid oxidation of the hot source material and the condensing coating. The substrates are generally unbiased, i.e., electrically isolated or at ground potential.

The advantages of evaporation include the possibilities of high deposition rates and the fact that the source material can be in a relatively simple form. Evaporation is most effective for depositing low melting point materials. Although evaporation can be used for refractory materials, the high temperatures make the process more difficult to execute. Accordingly, it is estimated that about 90% of the commercial applications of evaporation involve the deposition of aluminum. Difficulties of stoichiometry control are encountered in evaporating many alloys and compounds. It is for this reason that sputtering is replacing evaporation for many microcircuit metallization

applications (see Section 1). A host of special techniques have been developed for evaporating multi-component materials. These include flash evaporation, hot-walled evaporation, co-evaporation from multiple sources, and reactive evaporation. Several of these methods have proven very effective for depositing semiconducting coatings of high quality.

References 8 to 10 contain detailed discussions of coating deposition by evaporation.

3.2 Evaporation Rate

The rate at which atoms pass into vacuum from a heated source is given by the Hertz-Knudsen equation

$$W = 3.5 \times 10^{22} \alpha p^* / (MT)^{1/2} \text{ atoms/cm}^2\text{-sec} \quad (7)$$

where p^* is the vapor pressure in Torr, T is the temperature in $^{\circ}\text{K}$, and M is the molecular weight in grams. The parameter α is the evaporation coefficient. It is dependent on the cleanliness of the evaporation surface and can range from unity for clean surfaces to very low values (10^{-3}) for dirty surfaces.⁹ The evaporation coefficient can also be less than unity in the case of materials that evaporate as molecules for which the liquid-to-solid phase change involves a change in degree of freedom.⁹

The vapor pressure, p^* , is a very sensitive function of temperature. This is shown in Figure 4 for several materials. Thus maintenance of a constant evaporation rate requires extreme control over the temperature. This is often a difficult task. This problem is frequently avoided by using an *in situ* thickness monitor to simply indicate when the desired coating thickness has been achieved (see Section 3.7) or by evaporating a fixed charge of source material to completion.

Rough estimates of required source operating temperatures are commonly based on the assumption that vapor pressures of $\sim 10^{-2}$ Torr must be established to produce useful condensation rates. Temperatures, T_0 , that give $p^* = 10^{-2}$ Torr are given in Tables 1 and 2 respectively for typical elements and inorganic compounds of interest in electronics related applications. More detailed tables are given in References 8-10.

The average energy of the evaporated molecules is $3/2 kT$ where T is the source temperature. Thus for the case of gold, where the source temperature is 1400°C at $p^* = 10^{-2}$ Torr, the kinetic energy of the evaporated Au atoms will be about 0.20 eV.

The emission of vapor from a liquid or solid surface obeys the cosine emission law for the case of clean metal surfaces where $\alpha = 1$, and to first order for other cases.⁹ Thus the evaporated flux from a small source, of area A_e and evaporation rate W , which is incident on an elemental substrate area located at an angle ϕ off the perpendicular axis of the source as shown in Figure 5, yields a deposition flux per unit area of the substrate that is given by

$$W_d = \frac{W A_e}{\pi r^2} \text{Cos} \phi \text{Cos} \theta \quad (8)$$

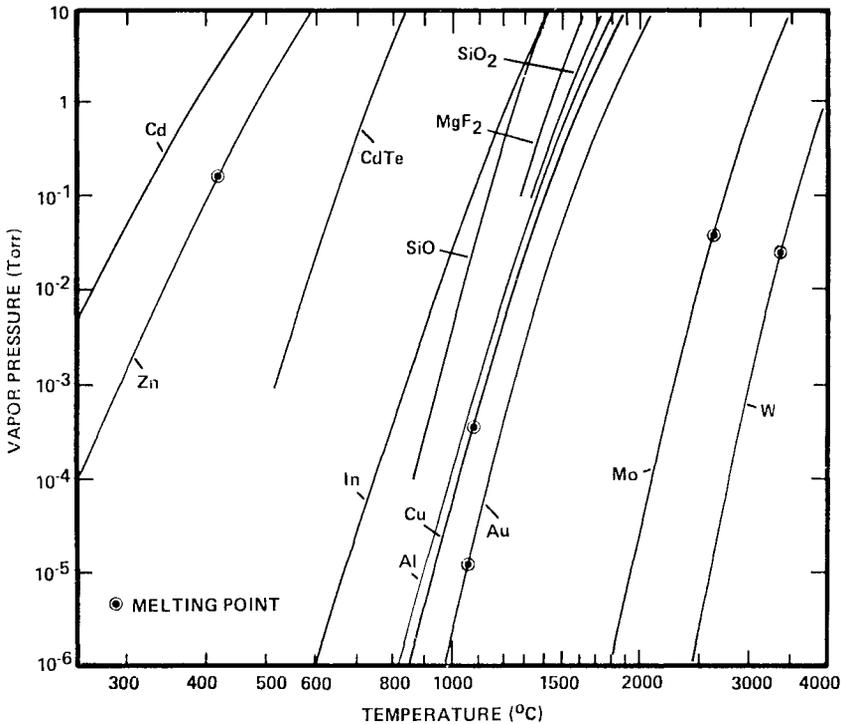


Figure 4: Equilibrium vapor pressures of several elements and inorganic compounds of interest in electronics related applications.

where r is the distance from the source to the substrate. The $\text{Cos}\theta$ term accounts for the fact that the substrate may not be perpendicular to the line of centers connecting the source and substrate. The thickness growth rate of the film is given by

$$D = \frac{M W_d}{\rho N_A} \tag{9}$$

where N_A is Avogadro's number, and M is the molecular weight and ρ is the mass density of the deposit.

Deposition rates for the evaporation process are clearly dependent on the material being evaporated, the type of evaporation source, and the position and orientation of the substrate surface. Consider the case of aluminum evaporation from a 1 cm diameter source onto a substrate located 15 cm directly above the source ($\text{Cos}\phi = 1$) and oriented to face the source ($\text{Cos}\theta = 1$). Thus we have $r = 15$ cm (this is a typical distance), and Equation 8 yields $W_o = W/707$. From Table 1 we see that a source temperature of 1220 °C is required to provide $p^* = 10^{-2}$ Torr. We assume that this temperature is used. Therefore, from Equation 7 with $p^* = 10^{-2}$

Table 1: Temperatures and Support Materials for Evaporating Elements Commonly Used in Electronics Related Processing

Element	Predominant Vapor Species	Melting Temp. (°C) T_M	Temp. (°C) at 10^{-2} Torr T	T/T_M (°K)	Support Materials**
Aluminum	Al	659	1220	1.61	W,C,BN
Antimony	Sb ₄ ,Sb ₂	630	530	0.89*	Mo,Ta,BN,Oxides
Arsenic	As ₄ ,As ₂	820	300	0.52*	Oxides,C
Beryllium	Be	1283	1230	0.97*	Mo,Ta,W,Oxides
Cadmium	Cd	321	265	0.91*	Mo,Ta,W,Oxides
Chromium	Cr	~1900	1400	~0.77*	Ta,W
Copper	Cu	1084	1260	1.13	Mo,Ta,W,C,Al ₂ O ₃
Gallium	Ga	30	1130	4.63	BeO,Al ₂ O ₃
Germanium	Ge	940	1400	1.38	Mo,Ta,W,C,Al ₂ O ₃
Gold	Au	1063	1400	1.25	Mo,W,C
Indium	In	156	950	2.85	Mo,W,C
Lead	Pb	328	715	1.64	Mo,W
Molybdenum	Mo	2620	2530	0.97*	-
Nickel	Ni	1450	1530	1.05	W,Oxides
Palladium	Pd	1550	1460	0.95*	W,Al ₂ O ₃
Platinum	Pt	1770	2100	1.16	W,Oxides
Silicon	Si	1410	1350	0.96*	BeO,ZrO ₂ ,C
Silver	Ag	961	1030	1.06	Mo,Ta,C
Tantalum	Ta	3000	3060	1.02	-
Tellurium	Te	450	375	0.90*	Mo,Ta,W,C,Al ₂ O ₃
Tin	Sn	232	1250	3.02	Ta,W,C,Al ₂ O ₃
Titanium	Ti	1700	1750	1.03	Ta,W,C,ThO ₂
Tungsten	W	3380	3230	0.96*	-
Zinc	Zn	420	345	0.89*	Mo,Ta,W,C,Al ₂ O ₃
Zirconium	Zr	1850	2400	1.26	W

*Materials that can be evaporated effectively from solid state.

**See References 8 and 9 for more detailed specifications.

Torr, $T = 1493^\circ \text{K}$, an aluminum molecular weight of 27 gm, and $\alpha = 1$, we obtain $W = 1.75 \times 10^{18}$ atoms/cm-sec. From Equation 8 we obtain $W_d = 2.5 \times 10^{15}$ atoms/cm²-sec. If we assume a bulk density of 2.7 gm/cm³ for the aluminum deposit, we obtain $D = 4.1 \text{A/sec}$ or 0.41 nm/sec using Equation 9. Rates of 1 to 10 nm/sec are typical for many materials. In practice, for a low melting point material such as aluminum, a p^* of about 10^{-1} Torr and a deposition rate of 4 nm/sec would be more typical for many applications.

Finally we can consider the impurity level which we might expect in our aluminum coating. We assume a p^* of 10^{-1} Torr, so that the deposition flux is 2.5×10^{16} atoms/cm²-sec. We assume that the residual gas pressure during deposition is 10^{-6} Torr. Thus, using Equation 5 and the molecular weight of oxygen, we estimate a residual gas flux of 3.6×10^{14} molecules/cm²-sec reaching the substrate. Since each oxygen molecule contains two atoms, the implied impurity level is about 3%. High purity starting material, high deposition rates, and/or low residual gas pressures are required to assure high purity deposits.

Table 2: Direct Evaporation of Inorganic Compounds Commonly Used in Electronics Related Processing

Compound	Predominant Vapor Species *	Melting Temp. (°C) T_M	Temp. (°C) at 10^{-2} Torr T	T/T_M (°K)	Support Materials **
Al ₂ O ₃	Al ₂ O, AlO Al ₂ O, O ₂ (AlO) ₂	2030	~1800	0.90	Mo, W
In ₂ O ₃	In, In ₂ O, O ₂	-	Vapor species observed at 1100-1450°C ⁹	-	Pt
SiO	SiO	-	1025	-	Mo, Ta
SiO ₂	SiO, O ₂	1730	~1250	0.76	Mo, Ta, W
TiO ₂	TiO, Ti, TiO ₂ O ₂	1840	Low vapor pressure at 2000°C ⁹	-	-
WO ₃	(WO ₃) ₃ , WO ₃	1473	1140	0.81	Pt
ZrO ₂	ZrO, O ₂	2700	-	-	-
ZnS	-	1830 (150 atm)	1000	0.61	Mo
ZnSe	-	1520 (2 atm)	820	0.61	-
CdS	S ₂ , Cd, S, S ₃ S ₄	1750 (100 atm)	670	0.47	C, Mo, Ta, W Al ₂ O ₃
CdSe	Se ₂ , Cd	1250	660	0.61	Al ₂ O ₃
PbS	PbS, Pb, S ₂ (PbS) ₂	1112	675	0.68	Mo
MgF ₂	MgF ₂ , (MgF ₂) ₂ (MgF ₂) ₃	1263	1130	0.91	Mo
CaF ₂	CaF ₂ , CaF	1418	1300	0.93	Mo

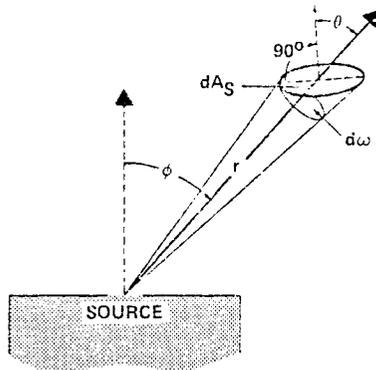
*Given in order of decreasing prominence.

**See References 8 and 9 for more detailed specifications.

3.3 Evaporation Sources

Several common types of evaporation sources are shown schematically in Figure 6.

3.3.1 Wire and Metal Foil Sources. The simplest evaporation sources are resistance-heated wires and metal foils of various types. A wire source is shown in Figure 6a and a metal foil source is shown in Figure 6b. They are commercially available in a variety of materials and sizes, at sufficiently low



EMISSION FLUX FROM SMALL SOURCE OF AREA A_e .

$$W_e = W A_e \cos \phi$$

DEPOSITION RATE PER UNIT AREA

$$W_d = \frac{W A_e}{\pi r^2} \cos \phi \cos \theta$$

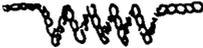
Figure 5: Flux passing from small area source to elemental substrate area dA_S which inscribes solid angle $d\omega$.

prices to be discarded after one use if necessary. The wire or foil supports must be fabricated from materials which have negligible vapor or dissociation pressures at the operating temperatures. These temperatures are typically in the range from 1000 to 2000°C. Wetting of the wire or foil surface by the evaporant is also desirable in order to achieve good thermal contact. Detailed recommendations pertaining to wire or foil support materials for various evaporants are given in References 8 and 9. These recommendations for a few evaporants are summarized in Tables 1 and 2. The most commonly used support materials are tungsten, molybdenum, and tantalum. Suitable wire or foil sources are available to evaporate small charges of nearly all the elements except the refractory metals themselves.

The maximum capacity of wire and foil sources is typically a few grams. The usual approach is to calculate the charge of source material that will provide a given deposit thickness, using a relationship of the form of Equation 8 for the apparatus geometry in question, and then to evaporate the entire charge. The wire/foil approach is in general too time-consuming for most production applications, but it is an effective method, for example, for depositing test electrodes in laboratory studies.

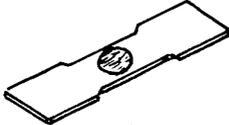
3.3.2 Crucible Sources. Crucible sources are required to support molten metals in quantities of a few grams or more. Since the melt is in contact with the container for prolonged periods of time, the selection of a noncontaminating and thermally stable crucible material is very important. Detailed recommendations for crucible materials are given in References 8 and 9. The non-metallic support materials summarized in Tables 1 and 2 apply strictly to crucibles. Thus it is seen that graphite and the refractory

RESISTANCE HEATED WIRE



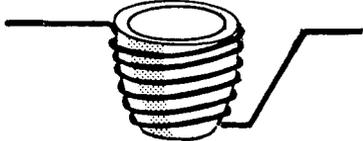
a

METAL FOIL SOURCE



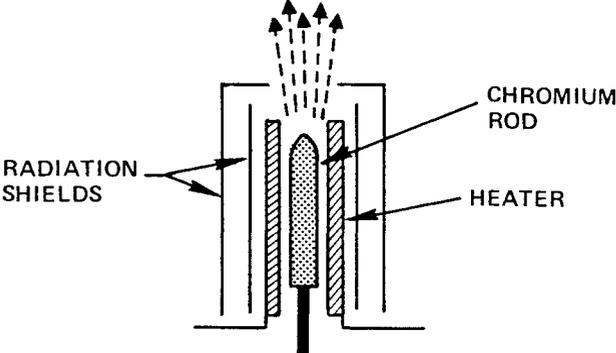
b

RESISTANCE HEATED CRUCIBLE



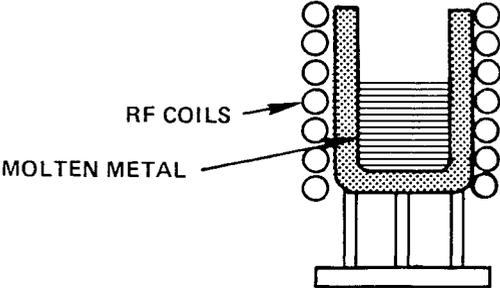
c

SUBLIMATION SOURCE



d

RF HEATED CRUCIBLE



e

Figure 6: Schematic illustrations showing several types of evaporation sources.

oxides are commonly used in addition to the refractory metals as crucible materials.

A wide variety of methods are used to heat the crucibles. The most common methods are radiation heating from a surrounding oven, conduction and radiation heating from a surrounding coil as shown in Figure 6c, and rf induction heating as shown in Figure 6e. The latter method has the advantage that energy is coupled directly into the evaporant metal, so that it is not necessary to produce crucible temperatures in excess of the vaporization temperature in order to produce heat flow.

3.3.3 Sublimation Sources. Figure 4 shows that the vapor pressures of materials increase continuously with temperature and do not show an abrupt change at the melting point. Tables 1 and 2 show, in fact, that many elements and compounds reach a pressure of about 10^{-2} Torr before melting (i.e., $T_c/T_M < 1$ in Tables 1 and 2) and hence can be sublimated at rates which are practical for coating applications. Direct wire and foil evaporation is particularly effective for metals with significant sublimation rates. Figure 6d shows a rod type sublimation source that has been used for chromium.⁹ Sublimation sources remove the problem of contact with foreign support materials.

3.3.4 Baffle Type Sources. In the evaporation of many materials the spontaneous release of absorbed or occluded gases can cause the violent ejection of droplets or, in the case of sublimation, particulates of the evaporation material. These particles can become incorporated into the growing film. To avoid this problem, baffled sources are often used which inhibit direct line-of-sight transmission from the evaporation charge to the substrates. These baffles are typically constructed of Ta and are maintained at sufficiently high temperature to prohibit condensation (see Figure 11 in Section 3.5). Atoms or molecules incident onto such baffles are generally re-emitted in a cosine distribution.⁸

3.3.5 Knudsen Cell Sources. A Knudsen cell is an evaporation crucible with a small exit orifice. The orifice is made small enough so that the evaporation flux passes through it via free molecular flow. The flux is emitted in a near cosine distribution if the thickness of the orifice is negligible.⁸ Formation of a molecular beam occurs under conditions of free molecular flow when the aperture is in the form of a tube. Although the flux is not baffled, the small size of the orifice makes Knudsen cells relatively immune to the spitting described above. A principal advantage of the Knudsen cell is that, when the surface area of the evaporant is an order of magnitude larger than the aperture, control of the evaporant temperature establishes the pressure, p^* , which exists inside the cavity in front of the orifice. Since there is no phase change as material passes through the orifice, Equation 7 with $\alpha = 1$ is closely obeyed.⁸ This removal of uncertainty over α allows effective control of the deposition rate by controlling the source temperature. (Holland draws the analogy between the use of a Knudsen cell and the use of a cavity for obtaining black body radiation from a substrate whose emissivity is less than unity.⁸) Therefore, Knudsen cells are particularly important when film stoichiometry is to be controlled by co-evaporation (see Section 3.5).

The general term "effusion cells" is used to identify the class of cells

with flow orifices that are restricted but not necessarily small enough to satisfy the free molecular flow conditions that are implicit in the particular case of the Knudsen cell. The flow from effusion cells can be theoretically predicted if the fluid dynamics of the flow through the orifice is properly taken into account. However, as a general rule the emission characteristics as a function of temperature from effusion cells, including Knudsen cells, are determined experimentally. Figure 7 shows an array of effusion cell sources used for multi-source deposition of semiconducting coatings of CuInSe_2 .¹¹ Effusion cells play a very important role in the process of molecular beam epitaxy, which is discussed in Section 4.

3.3.6 Electron Beam Sources. Figure 8 shows a schematic diagram of an electron beam evaporation system. Since the beam is concentrated on the evaporating surface, while other portions of the evaporant are maintained at lower temperatures, the evaporant can form its own crucible. Hence, interactions between evaporant and support materials are greatly reduced. Electron energies are typically in the 3 to 10 KW range, with power levels in the range 2 to 50 KW. Therefore, relatively high evaporation rates can be achieved, even for the refractory metals. Rod-fed sources can provide a large inventory of coating material. Therefore, electron beam sources are the most commonly used evaporation sources for large scale production applications. However, two difficulties that must be dealt with are, first, that electron beam sources are vulnerable to spitting because of

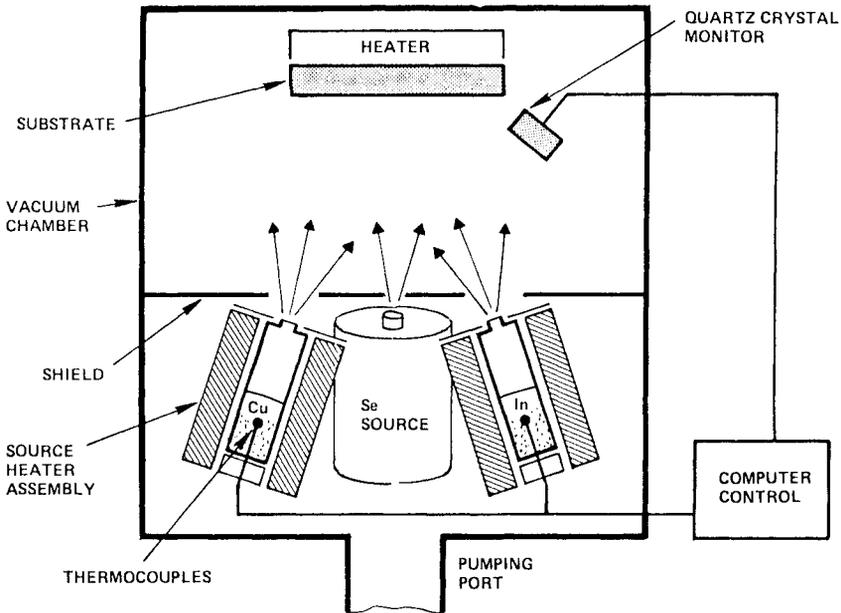


Figure 7: Knudsen cell evaporation sources arranged for co-evaporation to form coatings of CuInSe_2 . See Reference 11.

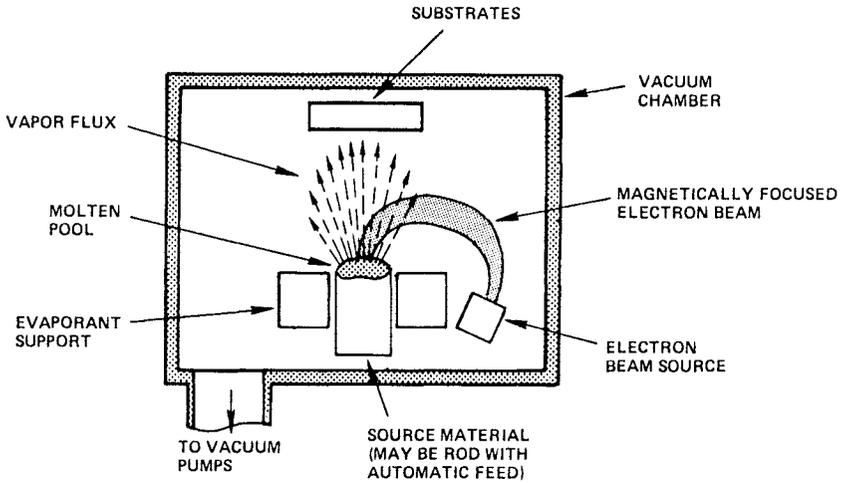


Figure 8: Schematic diagram of electron beam evaporation source.

the high power densities at the point of beam impact, and second, that the deposition flux is nonuniform as discussed in Section 3.4.

3.3.7 Other Types of Evaporation Sources. Several other types of evaporation sources have been developed to deal with the particular problems associated with forming stoichiometric coatings of alloys and compounds. These are discussed in Section 3.5.

3.4 Deposit Thickness Uniformity

It is difficult to maintain a uniform evaporation temperature over large surface areas because of radiation losses.⁹ Consequently, relatively small area sources are nearly always used. Deposition fluxes in the substrate plane are therefore non-uniform, and some type of substrate movement is generally required.

A small source emitting from its surface according to the cosine law will provide uniform deposition over the inside of a spherical surface if the source is placed on the circumference.⁸ This is easily seen from Equation 8 by noting that $\text{Cos}\phi = \text{Cos}\theta$, and $r = 2R \text{Cos}\theta$, for all deposition points on the circumference of a sphere of radius R . Therefore, planetary substrate holders of the type shown in Figure 9, which continuously move the substrates over a hemispherical surface with its center placed about one radius above the source, are commonly used.

Theoretical deposition profiles calculated from Equation 8 are given in References 8 and 9 for evaporation from point sources, small area sources, extended strip sources, cylindrical rod or wire sources, and ring and circular disk sources, all depositing onto plane receivers. The calculations show that the thickness uniformity of coatings deposited from flat filament or crucible sources cannot be improved significantly by enlarging the size of the sources, but that ring sources are particularly effective in providing

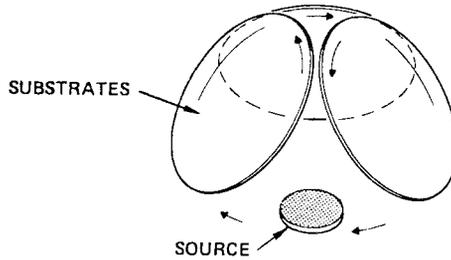


Figure 9: Schematic illustration of planetary type substrate tooling for small area vacuum coating source.

near-uniform deposits over relatively large areas. In particular, the deposit thickness is uniform over an area about equal to the area inside the ring when the substrate is placed above the ring at a distance equal to its radius. This is also an important consideration when using ring type planar magnetron sputtering sources. See Section 5.6.

Theoretical and experimental deposition profile data are also given in References 8 and 9 for several practical evaporation sources. The degree to which wire baskets of the type shown in Figure 6a duplicate a point source depends on the density of the wire winding. A dense winding promotes directional emission from the open ends.⁸ Flat metal strips or shallow dimpled boats of the type shown in Figure 6b have been found to yield near cosine-law emission. The emission patterns from practical effusion cells of the type shown in Figure 7 tend to be more directional, and to have a more pointed maximum in the center than the cosine emission pattern, because the requirement of negligible aperture thickness is usually not satisfied. The emission patterns from crucibles with relatively wide openings, and cylindrical cone shaped side walls which can act as extended emitting surfaces, tend to follow the cosine emission law, although the deposition profiles are slightly more directional than those predicted for cosine emission from a flat surface source. Electron beam sources tend to behave as small area sources, but often depart from the cosine emission law because the high evaporation rates, which are typically achieved, yield high enough vapor densities in the immediate vicinity of the source to cause collisional scattering of the evaporated molecules.

3.5 Evaporation of Alloys, Compounds and Mixtures

3.5.1 Introduction. The constituents which are present in most inorganic compounds, alloys, or mixtures differ in their vapor pressures. Consequently, during evaporation the composition of their vapors, and hence of their condensates, is not the same as that of the source material. This behavior is known as incongruent evaporation.

The available thermochemical data are seldom sufficient for predicting the conditions necessary to achieve coatings of the desired composition. Accordingly, the approach is usually empirical. Often, because of incongruent behavior, coatings of the desired composition cannot be reached

by direct evaporation. This has led to the use of special methods such as flash, two-source, and reactive evaporation. These processes are discussed in Section 3.6.

3.5.2 Evaporation of Alloys. The constituents in alloys evaporate independently of one another, mostly as single atoms. However, the vapor pressures of the individual constituents are not equal to their pure metal values at the temperature in question, because there is a contribution to the chemical potential when one metal is dissolved in another. Most metals evaporate incongruently, and this has led to the use of sputtering where extreme composition control is necessary. An example is the deposition of Nichrome (80%Ni - 20%Cr) to form thin film resistors. However, in the important case of Permalloy (85%Ni - 15%Fe), the evaporation is sufficiently congruent to permit the use of simple single source evaporation for many applications.⁸

Electron beam evaporation can significantly expand the range of materials which can be evaporated with reasonable composition control.¹⁰ This is possible because the electron beam source creates a small molten region, as shown in Figure 10. During an incubation period the molten region becomes deficient in the volatile species. The composition is then rate-limited by the passage of material by diffusion from the solid into the melt, across interface "A" in the figure. If the vapor pressure difference is not too large for the constituent diffusion rates, a steady state is developed, where the composition of the melt is just such as to produce a vapor composition equal to that of the solid. It is reported that reproducible compositions of Ni-20Cr, Ti-6Al, Ag-5Cu, Ag-10Cu, Ag-20Cu, Ag-30Cu, and Ni-xCr-yAl-zY have been successfully achieved by electron beam evaporation.¹⁰

3.5.3 Evaporation of Compounds. In the evaporation of compounds the transition to the vapor phase rarely occurs without changes to the molecular species. Thus evaporation is usually accompanied by molecular dissociation, association, or a combination of both processes. Dissociation represents thermal decomposition and generally makes simple direct evaporation impractical. The species formed in the direct evaporation of a number of compounds are summarized in Table 2.

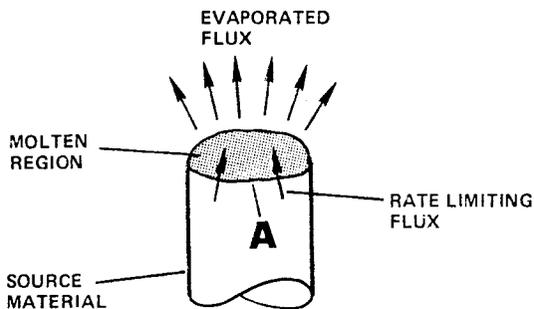


Figure 10: Schematic drawing showing equilibration of molten region during electron beam evaporation.

There are, however, some important compounds that do evaporate as constituent molecules and thereby maintain their composition. Most notable are SiO , MgF_2 and CaF_2 , a fact which explains the successful use of thin evaporated coatings of these materials in the optics industry.¹² Similar behavior is found for B_2O_3 , GaF_3 , and most of the divalent group IV oxides such as GeO and SnO (SiO -like species).⁹

Generally, the tendency to produce a dissociated vapor flux increases with increasing evaporation temperature and decreasing pressure. The evaporation of most oxides requires temperatures in excess of 1500°C . The binary oxides of Be, Mg, Ca, Sr, Ba, and Ni evaporate predominantly by dissociation into metal atoms and oxygen molecules, although their vapors may contain molecular species and in some cases lower oxides. The tendency to form suboxides is stronger among group three metals such as Al and In (see Table 2). Generally, congruent evaporation is more likely to be attained with binary oxides than with sesqui- or dioxides, because the higher oxides tend to lose oxygen at temperatures which are too low for the volatilization of the resulting suboxides.⁹ However, it is reported that Al_2O_3 , SiO_2 , and ThO_2 , as well as MgO and BeO , films have been successfully deposited by direct evaporation using electron beam heating.⁹ The oxides of Ti, Zr, Nb, Ta, Fe, and Cr are examples of materials which do not evaporate congruently and for which difficulties have been encountered in obtaining stoichiometric films by direct evaporation.

The II-VI compound semiconductors are important examples of compounds that undergo complete dissociation on evaporation. However, both the group II and the group VI elements in these compounds are relatively volatile, with the consequence that they are amenable to direct evaporation,¹³ with the coating stoichiometry being controlled by the substrate sticking coefficient.

The III-V compound semiconductors are an example of materials that undergo severely incongruent evaporation. The vapor pressures of the group V constituents, such as P, As, and Sb, are orders of magnitude greater than those of the group III elements, such as Al, Ga, and In. Accordingly, these compounds are difficult to deposit by direct evaporation.^{8,13,14}

Bi, C, Si, Te, P, As, and Sb are examples of materials that undergo association to form polyatomic species. For example, As yields As_2 and As_4 . Other examples are the oxides of molybdenum and tungsten, which can yield $(\text{MO}_3)_n$ and $(\text{WO}_3)_n$ species with n typically equal to two or three.

3.6 Special Evaporation Methods

Several special evaporation techniques have been developed for depositing materials whose constituents have different vapor pressures.

3.6.1 Flash Evaporation. In this technique small quantities of the constituents in the desired ratio are evaporated to completion from a common source using a temperature sufficiently high to evaporate the less volatile component. Often the evaporant is dispensed as a steady trickle onto a hot filament. A wide range of apparatus configurations have been devised for dispensing the source material. See Reference 9. The method is applicable to the evaporation of alloys, metal-dielectric mixtures,

and compounds. In most cases, the vapors impinging on the substrate are highly supersaturated, so that the film composition is not affected by the condensation coefficients. (Condensation coefficients are discussed in Section 6.1.) The most common problem is incomplete evaporation due to particle ejection and deflection. Ni-Cr alloys, Cr/SiO₂ cermets, GaAs, InP, Cu₂S, and BaTiO₃ are examples of materials that have been deposited by flash evaporation.

3.6.2 Hot-Wall Evaporation. In this technique films are grown under conditions that are close to thermodynamic equilibrium.^{15-16a} A schematic drawing of a hot wall evaporation apparatus is shown in Figure 11. The evaporated flux is passed into an enclosure with walls held at a sufficiently high temperature so that condensation is precluded. Accordingly, stoichiometric coatings can be deposited, even on substrates maintained at such high temperatures that one or more of the constituents has a low condensation coefficient. Since wall condensation is prohibited, the vapor pressures of the volatile constituents simply build up until they deposit onto the substrates at steady state rates that are equal to the rates at which they enter the enclosure from the evaporation sources. For example, near-stoichiometric CdTe films have been evaporated from a single CdTe source maintained at 600°C, with a hot-wall temperature of 500°C, and a

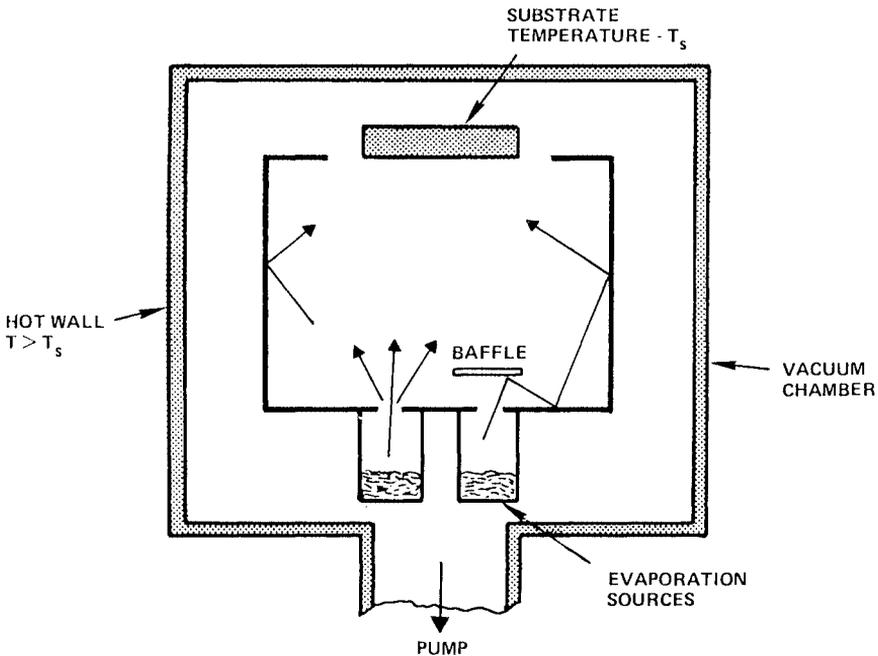


Figure 11: Schematic illustration of hot-wall evaporation system. Baffle of type discussed in Section 3.3 is shown on one of sources. Arrows show coating flux directions and are not meant to imply individual atom trajectories, since atoms are generally re-emitted from walls and baffles in a cosine distribution.

substrate temperature 465°C .¹⁶ If two sources are used, as shown in Figure 11, then the stoichiometry of the coating can be controlled. Thus non-stoichiometric n-type CdTe coatings were deposited using Cd and CdTe sources at the following temperatures: Cd- 330°C , CdTe- 600°C , hot-wall- 550°C , substrate- 515°C .¹⁶

The key advantage of the hot-wall method is that coatings of materials with volatile constituents can be grown with controlled composition at elevated temperatures. Elevated substrate temperatures are often desirable, since coating properties tend to approach bulk values as the substrate temperature is increased. See Section 6.2.

3.6.3 Close-spaced Sublimation. This is another technique in which a net thermal transport of coating material occurs under conditions close to thermodynamic equilibrium. Figure 12 shows a schematic drawing of a close-spaced sublimation apparatus. A flat plate source of coating material and the substrates are maintained in close proximity to one another, being separated by only a few mm. The temperatures of the source and substrates are both maintained at values such that the sublimation rates are significant, but with a temperature difference so that a net transport of coating material occurs from the source to the substrates. The space between the source and substrates is so small compared to the lateral dimensions that little escape of the vapors occurs. Hot walls, such as those shown in Figure 12, may be used to further hinder this escape. Relatively high deposition rates can be achieved. For example, CdTe films have been deposited at 67 nm/sec using a CdTe source temperature of 660°C and a substrate temperature of 600°C , and CdS films at 1000 nm/min using a CdS source temperature of 720°C and a substrate temperature of 550°C .¹⁷

As with the hot-walled method, the advantage of close-spaced sublimation is that near-stoichiometric coatings containing volatile constituents

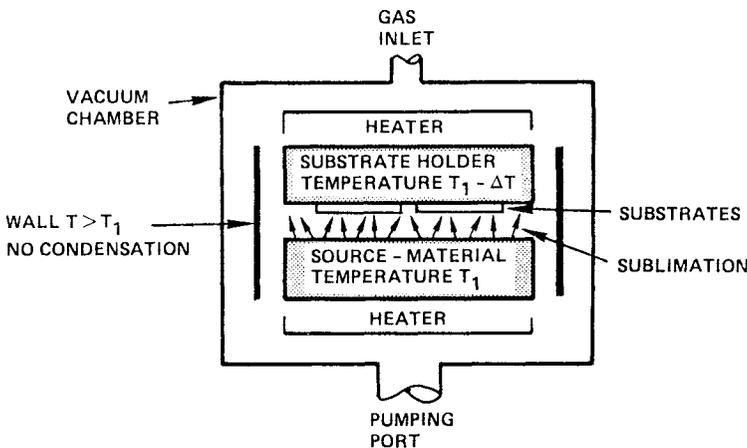


Figure 12: Schematic illustration of close-spaced sublimation type deposition apparatus.

can be deposited at relatively high substrate temperatures. The films can be doped by introducing a gaseous impurity. In a somewhat similar technique called close-spaced vapor transport, a reactive gas is introduced which promotes the formation of volatile species and thereby permits the transport to take place at lower temperatures.¹⁸

3.6.4 Multi-source Evaporation. In this method two or more independent sources are operated simultaneously, thereby permitting the deposition of multiconstituent materials which are not amenable to direct evaporation. The types of sources employed are the same as in single-source evaporation. An apparatus using effusion cells was shown in Figure 7 and one using open crucibles is shown schematically in Figure 13. By controlling the power delivered to the sources it is possible, in principle, to circumvent the problems of fractionation and decomposition encountered in the direct evaporation of most alloys and certain compounds.

There are two central problems in this technique. The first is control of deposition rates in the exact constituent ratio desired. Control is generally achieved through direct measurement of the particle fluxes in the vapor streams. (See Section 3.7.) This feedback method is used for the open crucible Cu-In-Se evaporation shown in Figure 13. The emission rate from effusion cell type sources is generally a reproducible function of the source temperature. Therefore, the effusion rates in multi-source systems using effusion cells can be controlled by feed-forward systems that control the temperatures within the cells, as indicated in Figure 7. The second

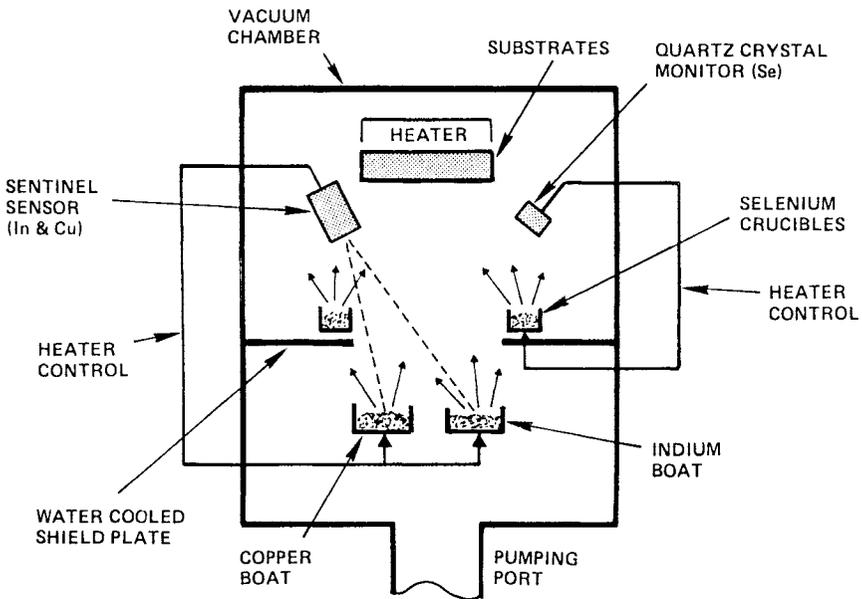


Figure 13: Multi-source evaporation system using open crucibles to deposit CuInSe_2 semiconducting coatings. See Reference 19.

problem encountered in the multi-source evaporation method is the limited deposition area over which coatings have a uniform composition. It is also important that the sources be arranged to minimize deposition flux angle-of-incidence effects. See Section 6.2.

Multi-source evaporation permits the co-deposition of materials that do not form compounds or solid solutions. An example is the co-evaporation of ceramic oxides and metals such as SiO and Cr or Au to form cermet thin film resistors. Other examples are metals such as Nb-Sn and compound semiconductors such as ZnS, CdS, CdSe, BiSe, BiTe, GaAs, InAs, InSb, AlSb, and CuInSe₂, as shown in Figures 7 and 13.

In the case of many compounds the condensation rate is not solely a function of the ratio of the arriving constituent fluxes, but is also determined by surface reactions on the substrates. For some compounds involving constituents of widely differing volatility, a substrate temperature can be selected such that only the stoichiometric compound can survive and grow. In such cases less stringent control of the individual impingement fluxes is required. In fact, one of the important advantages of multi-source evaporation is its applicability to this so-called "three temperature method". See Section 6.2. Multi-source evaporation with extreme control over substrate reactions forms the basis of a special method called "Molecular Beam Epitaxy" which is discussed in Section 4.

3.6.5 Reactive Evaporation. In this technique a metal is evaporated in the presence of a reactive gas in order to form a compound of the metal and the reactive gas. The technique is most commonly used to form coatings of metal oxides that cannot be evaporated directly because of complete or partial decomposition. Table 3 lists deposition conditions for reactive evaporation of some metal oxides.⁹

Deposition conditions are selected so that the reactions occur at the substrate surface. Consequently, the growth process is controlled by the impingement rates of the metal and reactive gas atoms, the condensation coefficients of the two species, and the substrate temperature. The central problem in reactive evaporation is that while the condensation coefficient of the metal constituent is usually near unity, the condensation coefficients of the reactive species become very low as full stoichiometry is approached in the deposit. This is shown for the Si+O₂ case in Figure 14, where an O₂/Si impingement ratio of between 10² and 10³ is required to form coatings having a composition approaching SiO₂.

It is generally found that the use of high reactive gas pressures, to provide high impingement ratios, has detrimental effects on film properties: for example, reduced hardness and refractive index. See Section 6.2. Accordingly, as can be seen in Table 3, deposition rates are usually kept low, in the few angstrom per second range, to provide the required high reactive-gas-molecule to metal-atom impingement ratios without requiring gas partial pressures above 10⁻⁴ Torr. Elevated substrate temperatures are also generally used in order to promote the surface reactions, to lessen the detrimental effects of high reactive gas pressures, and to provide significant improvements in structure. Coatings deposited at elevated substrate temperature exhibit improved crystallinity, density, hardness, optical constants, and dielectric properties. By contrast, films deposited at

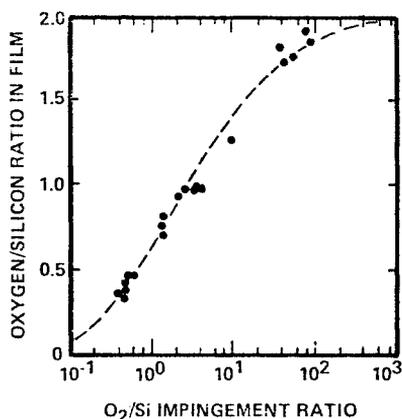


Figure 14: Film stoichiometry versus oxygen-to-silicon impingement ratio for reactively evaporated films. Data from Reference 20.

Table 3: Reactive Evaporation of Metal Oxides

Evaporated Metal	Desired Metal Oxide	Oxygen Pressure (Torr)	Deposition Rate ($\text{\AA}/\text{s}$)	Substrate Temperature ($^{\circ}\text{C}$)
Al	Al_2O_3	$10^{-5} - 10^{-4}$	~ 1	400-500
Cr	Cr_2O_3	2×10^{-5}	~ 2	300-400
Ta	Ta_2O_5	$10^{-4} - 10^{-3}$	~ 2	700-900
Ti	TiO_2	10^{-4}	-	300
Ba+Ti	BaTiO_3	10^{-2}	2-8	770-1025

Data from Ref. 9.

low substrate temperatures tend to have amorphous or poorly crystallized structures (see Section 6).

The low reactive gas condensation coefficient occurs because the reactive gas molecules must undergo dissociative chemisorption on the surface of the growing film, with the resulting reactive atoms being incorporated into the coating. The reaction process can be stimulated, and high deposition rates can be achieved, by using a plasma to dissociate the reactive gas molecules and/or to produce other species that promote the surface reactions involved in compound formation. The process is generally referred to as plasma assisted or activated reactive evaporation (ARE). The reactive gas may be passed through a hollow cathode discharge,²¹ or an

external electrode may be used to create a plasma discharge in the deposition chamber as shown by the ARE apparatus in Figure 15. The ARE process has been successfully used, for example, to produce oxide, carbide, nitride and sulfide films of various metals.^{10,23}

3.7 Deposition Rate and Flux Monitors

One of the disadvantages of evaporation, as compared, for example, to sputtering, is that the rate of passage of material into the vapor phase is a very nonlinear function of the power delivered to the source. Thus as a general rule, the evaporation flux must be monitored to control the deposit thickness. Deposition rate and species flux monitors also play a very important role in the multi-source evaporation process. Several types of rate monitors are available. They are discussed in detail in Reference 9.

- 1) *Ionization Gauge Rate Monitors.* These devices are similar to hot cathode ionization gauges. They monitor the atom density in the vapor phase by ionizing the vapors and measuring the ion current. The obvious difficulty is residual gas contributions to the current.
- 2) *Mass Spectrometers.* Quadrupole mass spectrometers are small enough so that they often can be conveniently arranged

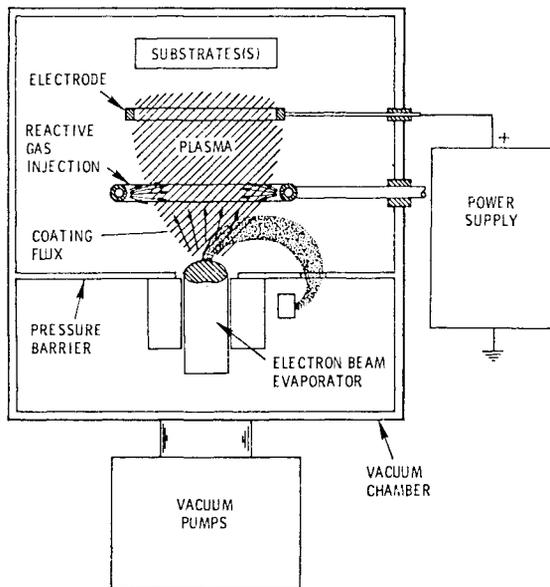


Figure 15: Schematic drawing of apparatus configuration for activated reactive evaporation. From Reference 22. See Reference 23.

- to intercept the vapor flux. This method is frequently used in molecular beam epitaxy (see Section 4).
- 3) *Electron Impact Spectroscopy*. The vapor flux is passed through an electron beam which produces an atomic line emission that is monitored, usually with a narrow band filter and photomultiplier. An electron impact spectroscopy type sensor is used to monitor the Cu and In fluxes in the apparatus shown in Figure 13.
 - 4) *Microbalances*. The balance is used to measure the accumulated coating mass on a thin vane.
 - 5) *Crystal Oscillators*. This method utilizes the piezoelectric properties of quartz and the fact that the resonant frequency is influenced by the accumulation of a mass of coating material on the surface of the crystal. A quartz crystal monitor for controlling the Se evaporation rate is shown in Figure 13.

3.8 Evaporation Source Material

An advantage of the evaporation process as compared, for example, to sputtering, is the simple form of the starting material; i.e., the material does not have to be in the form of a sputtering target. However, the need for purity cannot be overemphasized. Impure source material not only results in impure coatings, but also can lead to the difficulties of "spitting" discussed in Section 3.3. These problems are particularly severe for electron beam evaporation because of the high power levels involved. Therefore, vacuum melted, rather than power metallurgy, metal rod sources should be used in electron beam evaporation.

4. MOLECULAR BEAM EPITAXY

4.1 Introduction

Molecular beam epitaxy (MBE) is a multi-source evaporation process, of the type discussed in Section 3.6, which is done with extreme control over the deposition parameters in order to exploit the kinetic processes of film growth that are discussed in Section 6.1. MBE has been applied primarily to the growth of single crystal films of compound semiconductors. Thermal molecular beams of each constituent of the film are directed to converge on a single crystal substrate under conditions suitable for epitaxial growth. Deposition rates are low (typically about 0.1 nm/sec). The low deposition rates reduce the temperature required to achieve epitaxial growth (Figure 47 in Section 6.2). The low growth rates are made feasible by the use of ultra-high vacuum systems which have base pressures in the 10^{-10} to 10^{-11} Torr range and thereby reduce the residual gas contamination flux incident on the substrates, as discussed in Section 2.

The slow growth rates permit very precise control of layer thicknesses in the nm range. Shields are used to provide abrupt initiation or cessation of the molecular beam fluxes and thereby to create sharp interfaces or

precisely controlled doping profiles. The reduced growth temperatures minimize the disturbance of these built-in composition profiles because of bulk diffusion.

An important advantage of MBE over most other forms of epitaxy is the ability to include *in situ* facilities for process monitoring and control and for surface analysis. These diagnostic tools include (1) quadrupole mass spectrometers to monitor the composition of the incident beam and the residual background gases; (2) high energy electron diffraction (HEED) systems to monitor the surface structure of the substrate and coating; and (3) Auger electron spectroscopy (AES) systems to monitor the composition at the substrate surface before and during coating.

Excellent reviews of MBE are given in References 24 to 31.

4.2 Apparatus Configuration

Figure 16 shows a schematic diagram of a typical MBE deposition chamber. The systems generally have three to eight evaporation sources. Resistance or electron beam heated sources are usually used. However, gas phase sources may be used.³² This arrangement is sometimes referred to as chemical beam epitaxy. Crucibles are typically made from pyrolytic boron nitride because of its inertness to metals such as Ga and Al. Source temperatures are generally controlled to within about 0.2°C , using thermocouples inserted within the crucibles.²⁶ Precise temperature control of the individual cells is essential because of the strong temperature dependence of the evaporation rate (see Section 3.2). Thus, temperature fluctuations of $\pm 1^{\circ}\text{C}$ can result in molecular beam flux variations ranging from ± 2 to 4%.²⁸ Shutters are provided, as noted previously, to abruptly control the flux from

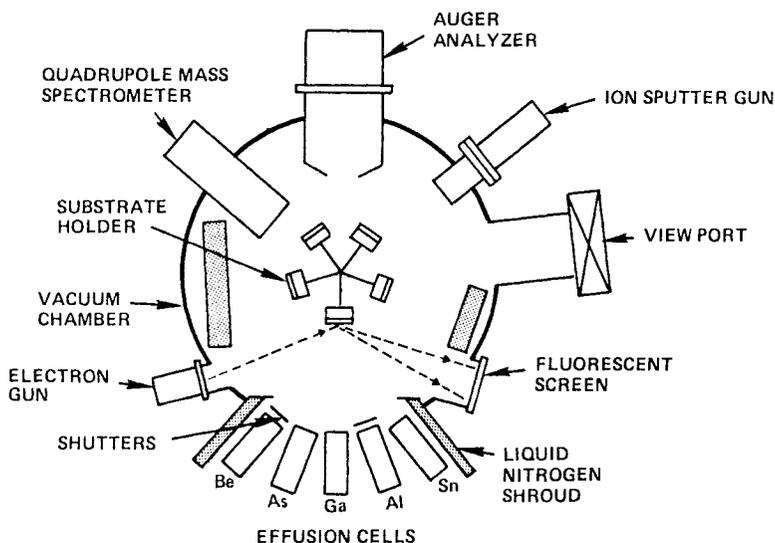


Figure 16: Schematic illustration of MBE deposition chamber with sources configured for depositing GaAs type films.

a source which in turn is controlled to provide a given steady state emission.

MBE pumping systems typically use carbon vane, turbomolecular, or sorption pumps for the initial chamber evacuation. Titanium sublimation, ion pumps, and liquid nitrogen cryopanel are generally used for high vacuum pumping, although oil diffusion pumps have been used. Most MBE apparatuses have load locks for introducing the substrates. This is an important consideration, since long pumpdown times and chamber heating are generally required to reach ultra-high vacuums in an apparatus that has been exposed to the atmosphere.

An important aspect of MBE systems is that liquid nitrogen cooled shrouds are usually designed to cover as much of the wall surface as possible in the deposition chamber. See Figure 16. In addition to impurity species such as water vapor and CO, these cryogenic surfaces tend to trap any coating atoms that are re-emitted from the substrate (sticking coefficient less than unity). Thus the flux of a given species arriving at the substrate can be accurately controlled by the power delivered to its source. The goal is generally to keep the partial pressures of residual gases with a high sticking coefficient below 10^{-14} Torr in the vicinity of the substrate.²⁸ A general guideline is to provide sufficient liquid-nitrogen-cooled surface area in the deposition chamber to yield a pumping speed for water that exceeds 20,000 liters/sec.²⁷ In addition, liquid-nitrogen-cooled shrouds are usually provided in the vicinity of the evaporation sources to prevent intercontamination and thermal cross talk.

The schematic drawing in Figure 16 shows a number of characterization tools, including an Auger spectrometer and an ion sputter gun. A common practice is to configure MBE apparatuses with separate chambers for sample analysis and deposition. A diagram of a typical commercial system of this design is shown in Figure 17. A photograph of the apparatus is shown in Figure 18. Such systems consist typically of (1) a sample entry or load-lock chamber with a turbomolecular pump, (2) a sample analysis chamber with an ion pump, Ti sublimation pump and cryopump, and (3) a deposition chamber with similar pumping. The deposition chamber shown in Figure 17 is equipped with a quadrupole mass spectrometer and a HEED system. The test chamber is configured to incorporate an ion sputter gun with provisions also for a range of surface analysis tools such as Auger electron spectroscopy, X-ray photoelectron spectroscopy, or secondary ion mass spectroscopy.

The quadrupole mass spectrometer is probably the single most important analytical tool in an MBE system.²⁷ The second most widely used technique is HEED. These instruments are placed in the deposition chamber, as shown in Figure 16, and are also recommended for production systems.²⁸ Ion gauges are often used for neutral beam monitoring. Auger electron spectroscopy is the most commonly used method of composition analysis.

Load lock systems introduce substrates in short time periods (~ 15 min) while the growth chamber is maintained in the 10^{-10} Torr range. In a typical machine designed for small production runs, a batch of five two-inch wafers is introduced into a preparation and analysis chamber, where they are stored and sequentially cycled through the growth process, while

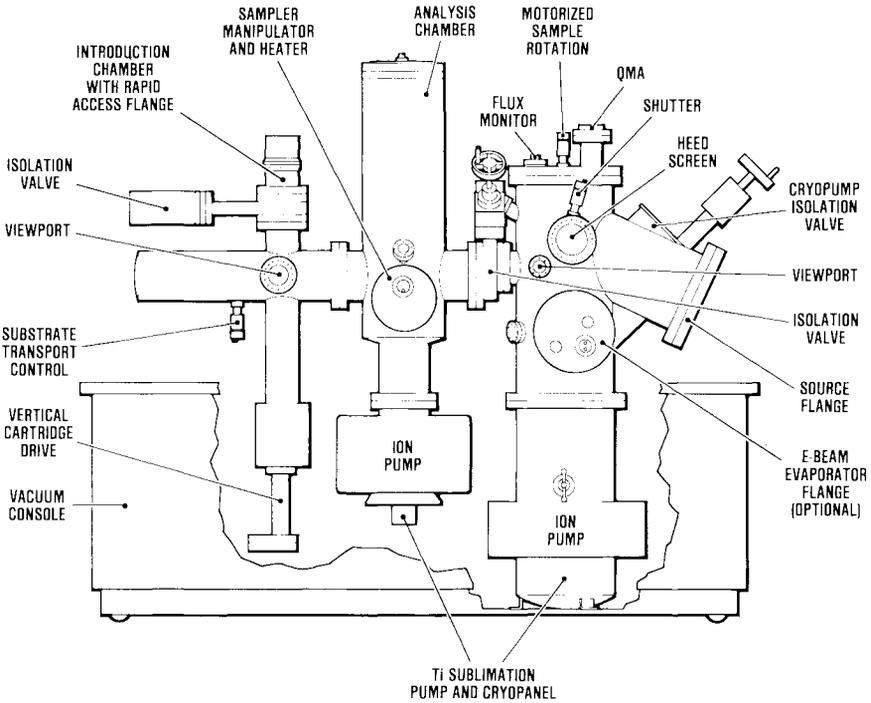


Figure 17: Commercial MBE deposition system (PHI Model 425). Drawing courtesy of Physical Electronics Division of Perkin-Elmer, Eden Prairie, MN.

five freshly prepared substrates are inserted into the sample entry chamber and evacuated.³⁴ Loadlock systems of this type have been designed which can operate for more than 100 hrs without exposure of the growth environment to atmosphere. Such systems have wafer processing rates of 3 to 4 wafers per hour. Typical systems can provide epitaxial films with thickness uniformity of better than 5% over substrates 2 inches in diameter, with extremely uniform epitaxial layers over about 15 cm² of each wafer.³⁴ New systems can handle 3 inch wafers, with a general trend to greater production capabilities.

The substrates in Figure 16 are shown mounted on a carousel. Such carousels incorporate substrate heaters and thermocouples such that temperatures up to about 700°C can be maintained within about 0.2°C.²⁶ A precision manipulator permits the substrates to be accurately positioned for deposition and HEED analysis.

4.3 Deposition Procedure

Substrate surface preparation is a critical part of MBE. A typical preparation procedure will be described in some detail, since it applies in a general sense to all of the vacuum deposition methods. A starting substrate

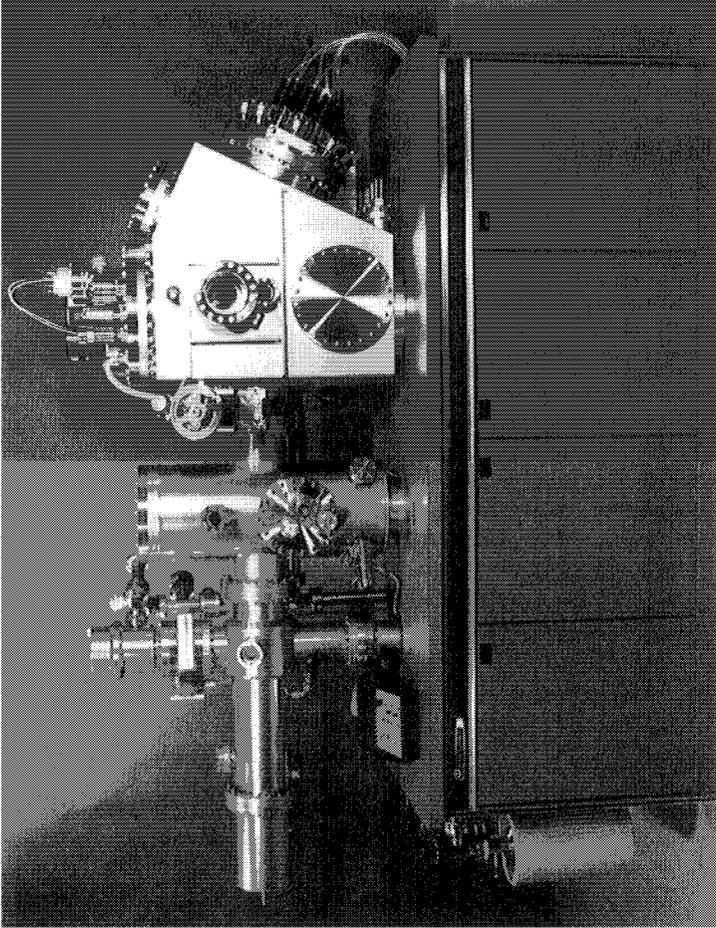


Figure 18: Photograph of MBE deposition apparatus (PHI Model 425B). Photo courtesy of Physical Electronics Division of Perkin-Elmer, Eden Prairie, MN.

is typically covered with a contamination layer consisting of carbonaceous compounds and oxides which must be removed. The most detailed MBE work has involved the deposition of GaAs onto single crystal GaAs substrates. The (100) plane is chosen for the technologically important reason that it has orthogonal (110) type cleavage planes.²⁷ A typical preparation procedure involves polishing with a diamond paste followed by etch-polishing on an abrasive-free lens paper soaked with a sodium hypochlorite or bromine/methanol solution.²⁸ The substrate is then rinsed in trichloroethylene, methanol, and distilled water, following which it is boiled in hydrochloric acid, free etched in sulfuric acid, again rinsed in distilled water, and finally soldered with In to a Mo backing plate. At this point the substrates, which are generally contaminated with both oxygen and a small amount of carbon, are introduced into the MBE chamber.

A typical *in situ* process within the MBE chamber involves the use of ion bombardment to remove foreign materials from the surface, and Auger spectroscopy to verify that the surface is clean. The sample is then passed into the deposition chamber and annealed to remove the surface damage created by the sputter cleaning. In many cases thermal desorption rather than ion bombardment may be used to avoid composition changes due to preferential sputtering. It is well established that temperatures in the 525–535°C range evaporate the passivating oxide film on GaAs without causing surface composition changes because of incongruent evaporation of the GaAs.²⁸ In any event, the HEED system is used to verify the crystalline quality of the substrate surface prior to deposition.

The sources are then adjusted to temperatures that provide the desired deposition rates. The shutters are closed during this operation to prevent deposition on the substrates. The substrate temperature is adjusted to the desired value and the appropriate shutters are opened to commence deposition. The HEED system can be used to provide periodic verifications that epitaxy is proceeding. The quadrupole mass spectrometer is used to monitor the beam fluxes and the residual gas partial pressures. Finally the coated substrates are withdrawn through the load-lock system.

4.4 Coating Growth

The general principles that govern the growth of evaporated and sputtered coatings are discussed in Section 6. MBE creates an environment in which remarkable control can be exerted over these growth processes. Table 4, from Reference 26, lists some representative semiconductors which have grown by MBE. Most of the detailed studies of growth kinetics have been done on GaAs. However, sufficient work has been carried out on other binary compounds such as InAs, InP, and AlAs to show that similar behavior is observed for most combinations of Al, Ga, and In with P, As and Sb.³⁵ The essential feature is that at the substrate temperatures used, the column III elements (Al, Ga, In) have near unity sticking coefficients on the growing coating surfaces, while the sticking coefficients of the column V elements (P, As, Sb) are dependent on the density of column III atoms which are available on the surface to react. Typical substrate temperatures are $\sim 600^\circ\text{C}$. The excess column V species

Table 4: Representative Semiconductors Grown by MBE
(From Reference 26)

<u>Group IV</u>	<u>III-V</u>	<u>IV-VI</u>	<u>II-VI</u>
Silicon	<u>Binary</u>	<u>Binary</u>	<u>Binary</u>
Germanium	GaAs	PbTe	ZnTe
	GaP	PbS	ZnSe
	GaSb	PbSe	CdTe
	InAs	SnTe	CdS
	InP		
	InSb		
	AlAs		
	<u>Ternary</u>	<u>Ternary</u>	<u>Ternary</u>
	GaAlAs	PbSnTe	ZnSeTe
	GaAsP	PbSnSe	CuInSe ₂
	GaAsSb	PbSSe	
	GaInAs		
	GaInP		
	GaSbAs		

are lost by re-evaporation. Thus, near-stoichiometric coatings can be grown under a range of deposition conditions, provided that there is an excess flux of the column V element arriving at the substrate. This type of growth is described in Section 6.3.

It should be noted that even at the UHV pressures of 10^{-10} to 10^{-11} Torr used during MBE deposition, doping levels of 10^{17} to 10^{18} cm^{-3} would result if all the impurities which are incident on the substrate were incorporated into the film and were electrically active.³⁶ (See Section 2.) Fortunately, the sticking coefficients for residual gases on III-V compounds are sufficiently low so that the electrically active impurity concentrations are in the 10^{14} to 10^{15} cm^{-3} range. The main residual impurity in MBE GaAs is carbon, which appears to be a shallow acceptor with an energy of around 26 meV.^{27,28} Therefore, undoped GaAs films are generally found to be p-type. The question of residual impurities and their effects is one that should be considered in evaluating any new MBE materials and/or applications.

The incorporation of dopants is a key consideration in MBE. Dopant incorporation is governed largely by reaction kinetics rather than by equilibrium thermodynamics.²⁶ Doping levels are typically less than 10^{19} cm^{-3} . Many of the dopants which are useful in other forms of epitaxy do not behave well with MBE. The actual behavior is strongly dependent on the substrate temperature and the surface reconstruction which occurs during growth.²⁶ It is difficult to dope with an element having a high vapor pressure, since such materials are desorbed prior to incorporation. Therefore, dopant incorporation is a contemporary research area in MBE. Considerable attention is being given to the use of ion beams.^{28,30,37-40} In this case, the dopant flux is directed at the substrate in the form of an ion beam with energies in the 0.2 to 1.5 keV range. The kinetic energy of the incident

ions appears to permit them to penetrate into the lattice of the growing coating to a sufficient degree so that the incorporation probability is increased.

The complexity of the doping problem is illustrated by the GaAs technology. Commonly used dopants are Sn, Si, and Ge for n-type and Be for p-type GaAs.^{27,28} Tin, which is the most commonly used n-type dopant, illustrates the complexities that can occur. Surface segregation causes the Sn to accumulate on the surface in a concentration which is several orders of magnitude larger than that in the bulk. The rate of incorporation is controlled by the Sn surface concentration and the Ga vacancy concentration within the GaAs.²⁷ The Sn segregated on the surface precludes the formation of abrupt changes in doping concentration, since its concentration cannot be reduced to zero by simply closing the shutter at the Sn source. Another source of complexity occurs because many of the dopants are amphoteric. For example, under As-rich conditions Ge tends to be incorporated as a donor, while under Ga-stabilized conditions (see Section 6.3) Ge is incorporated predominantly as an acceptor.²⁷ These examples illustrate how the dopant incorporation is dependent on relative substrate arrival rates of As and Ge atoms as well as the doping flux itself.

The difficulties in forming p-type GaAs are even more severe. The conventional acceptor dopants for GaAs such as Zn and Cd have high vapor pressures and therefore low incorporation coefficients. Beryllium has shown the most promising p-type doping properties, but is an extreme toxicity hazard.²⁷ The ion beam technique has been successfully used to incorporate Zn⁺ and provide carrier concentrations in the 10^{19}cm^{-3} range.^{28,37} Manganese has been used but causes adverse surface degradation.²⁷ Ion beam deposition has also been proven effective in Si MBE.^{40,41}

4.5 Applications

MBE is particularly effective when control over thickness, composition, and doping profiles are critical to device performance. Such requirements are often encountered in microwave and optoelectronics devices. These needs have stimulated the development of MBE technology in general and GaAs technology in particular. MBE has been used not only to produce state-of-the-art performance in conventional structures, but also to produce totally new types of thin film devices. Table 5 lists some devices which contain epitaxial structures grown by MBE.

GaAs field effect transistors are typically used as low-noise microwave signal detectors and microwave signal generators. Both low noise and high-power field-effect transistors require n-type layers less than 1000 nm thick. Low-noise FET's have been reported using 100 nm thick, heavily doped, MBE GaAs layers.²⁷ The linearity of power FET's can be improved by tailoring the doping profile, a requirement that can be achieved by MBE. Microwave varactors, mixer diodes, and IMPATT diodes are other examples of devices in which controlled doping profiles are required and therefore where MBE is useful.²⁷

The formation of low-resistance contacts to n- and p-type GaAs, as well as Schottky barrier diodes on GaAs, is also of great technological importance

Table 5: Devices Whose Epitaxial Structures Have Been Grown by MBE
(From Reference 34)

Discrete microwave devices

Low noise FETS
Power FETS
Novel FET structures
IMPATT diodes
Mixer diodes
Varactor diodes
Gun diodes

Discrete optoelectronic devices

Laser diodes
Waveguides
Integrated optics
Taper couplers
Light-emitting diodes
Photodetectors

Other devices

Diodes
MIS capacitors
Superlattices
Tunnel triodes
Solar cells

in microwave and optoelectronic device fabrication. The unique capabilities of MBE are well suited to first growing the epitaxial layer structure needed for device operation, and then growing the required metal film onto the freshly deposited semiconductor surface in the same growth cycle without breaking vacuum. The semiconductor surface is never exposed to the atmosphere or solvents. MBE *in situ* metallization makes the reproducible production of ideal metal-semiconductor interfaces feasible.²⁹ This is a particularly important consideration as very small circuit dimensions are contemplated.

The high electron mobility transistor (HEMT) and the multiquantum well (MQW) laser are examples of novel thin film devices which have been produced by MBE. In the case of the HEMT device, a thin $\text{Ga}_{1-x}\text{Al}_x\text{As}$ layer is grown in the middle of a GaAs active channel. As a result of the greater electron affinity of the lower bandgap GaAs, the donor electrons from the $\text{Ga}_{1-x}\text{Al}_x\text{As}$ layer thermalize in the GaAs active layers, where they are no longer scattered by the ionized parent donors. Such devices can have very fast switching times (in the 10 ps range).³⁴

The double heterojunction injection lasers consist essentially of a thin active layer at the interface of a p-n junction.³⁴ The higher refractive index of the active layer with respect to the confinement layers forms an optical waveguide that, with cleaved mirror faces, can define a resonant optical

cavity. The band energies of the active layers are such that the charge carriers injected under forward bias are trapped in the active region. In recent work, lasers with unique performance have been fabricated by making the active layers have the form of quantum well superlattices consisting of alternate layers of materials with different bandgaps and with layer thicknesses less than the Debye length. Thus, in one example, fourteen GaAs quantum-well active layers, only ~ 14 nm thick, were sandwiched between $\text{Al}_{0.27}\text{Ga}_{0.73}\text{As}$ confinement layers ~ 13 nm thick.^{34,36} Injection MQW laser diodes are of great importance in fiber optics communication systems, because laser operation can be tailored to emit at frequencies well above the standard lasing frequencies of the host material, and thereby to control the losses in the fiber optics. The deposition of superlattice structures with properties not found in homogeneous materials is an active area of current research which can be expected to yield a host of applications in the future.^{42,43}

5. SPUTTERING

5.1 Introduction

Sputtering is a process whereby material is dislodged and ejected from the surface of a solid or a liquid due to the momentum exchange associated with surface bombardment by energetic particles. A source of coating material called the target is placed into a vacuum chamber along with the substrates, and the chamber is evacuated to a pressure typically in the range 5×10^{-4} to 5×10^{-7} Torr. The bombarding species are generally ions of a heavy inert gas. Argon is most commonly used. The sputtered material is ejected primarily in atomic form. The substrates are positioned in front of the target so that they intercept the flux of sputtered atoms.

The most common method of providing the ion bombardment is to backfill the evacuated chamber with a working gas in the 1 to 100 mTorr pressure range and to ignite an electric discharge with the target serving as the cathode or negative electrode. Such an apparatus configuration is shown schematically in Figure 19. Applied potentials are typically between 500 and 5000V. Direct currents are generally used when the target material is a good electrical conductor. Radio frequencies are used when the target material is poorly conducting or an insulator. Deposits of poorly conducting metallic compounds can also be formed by dc sputtering the metallic component while injecting other constituents in the gas phase. This is known as reactive sputtering. A voltage bias may be applied to the substrates so that they are at a negative potential relative to the plasma and therefore subject to an ion bombardment that can influence coating properties. This is known as bias sputtering.

The most striking characteristic of the sputtering process is its universality. Since the coating material is passed into the vapor phase by a mechanical (momentum exchange) rather than a chemical or thermal process, virtually any material is a candidate coating. Films containing almost every element in the periodic table have been prepared by sputter-

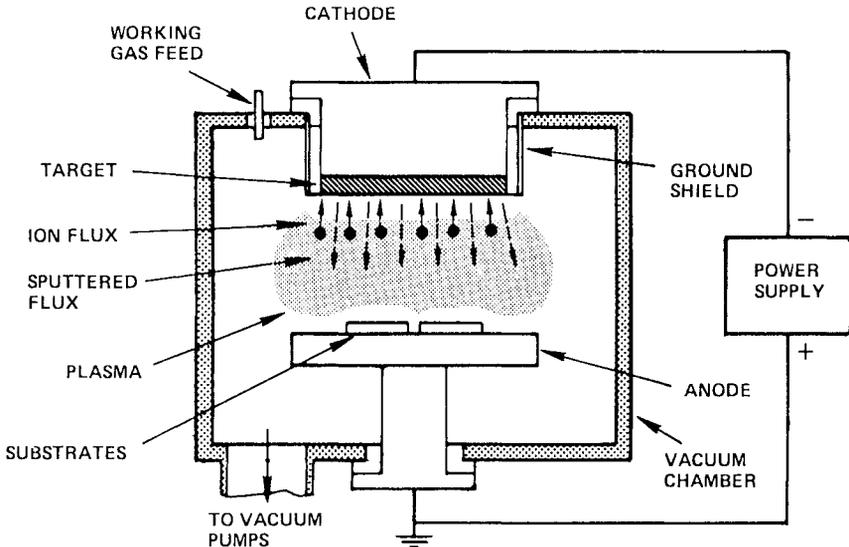


Figure 19: Schematic drawing showing glow discharge sputtering apparatus of the planar diode type.

ing. Alloys and compounds can generally be sputter-deposited while preserving their composition. For example, organic bone has been sputtered, with the deposits having an amorphous microstructure rather than the crystalline structure of the target material, but an identical composition.⁴⁴ PTFE (Teflon) has been sputtered to produce films having many of the properties of the starting material.^{45,46} However, most applications involve metals and more common compounds such as aluminum oxide. One of the primary applications of sputtering is for interconnect metallization in integrated circuits. Other areas attracting increasing attention are films for magnetic⁴⁷ and optical data storage.⁴⁸

A series of review papers in References 49-59 trace the recent development of the physics and technology of sputtering.

5.2 Basic Sputtering Mechanisms

In sputter deposition we are concerned primarily with what is termed physical, as opposed to chemical, sputtering. In physical sputtering, the bombarding particle transfers kinetic energy to the target atoms and incurs the subsequent ejection through the target surface of those atoms which acquire sufficient kinetic energy to overcome the local binding forces. In chemical sputtering, chemical reactions induced by the impinging particles produce an unstable compound at the target surface, which subsequently passes into the gas phase.⁶⁰ Chemical sputtering is particularly important in plasma etching applications. See chapter on plasma etching.

The fundamental event in physical sputtering is an atomic collision. It is therefore useful to review the case of a simple binary hard-sphere elastic collision, in which an incident particle of mass M_i and velocity V_i impacts on a line of centers (zero impact parameter) with a target particle of mass M_t which is initially at rest as shown in Figure 20a. Three observations can be made: (1) the target particle is driven deeper into the target by the momentum exchange, (2) the momentum passed to the target particle is greatest when the two particles are of identical mass, and (3) the bombarding particle will be reflected if its mass is less than that of the target particle.

From the first observation listed above we see that a single binary collision will not, in general, produce sputtering. The ejection of a sputtered particle, due to a bombarding particle incident normal to the target surface, requires a sequence of collisions so that a component of the initial momentum is changed by more than 90° . Sputtering should therefore be envisioned as a statistical process that occurs as a result of a collisional cascade which is initiated by the incident energetic particle. A general picture of the collisional events within the cascade has been provided by theoretical and computer modeling studies.⁶¹⁻⁶⁶ At the bombarding energies of interest in sputter deposition, the sputter ejection is believed to result primarily from the scooping action as a low energy knock-on passes underneath an adjacent atom, or from a primary knock-on or reflected ion which approaches the surface from within the target and dislodges a surface atom by striking it on the under side. Both cases are shown in Figure 20b. Although the energy from the impinging particle is partitioned over a region of target material that may extend 5 to 10 nm below the surface,⁶⁷ the particular collisions that give rise to sputtering occur primarily within about 1 nm of the surface.^{66,67} It is interesting to note that recent

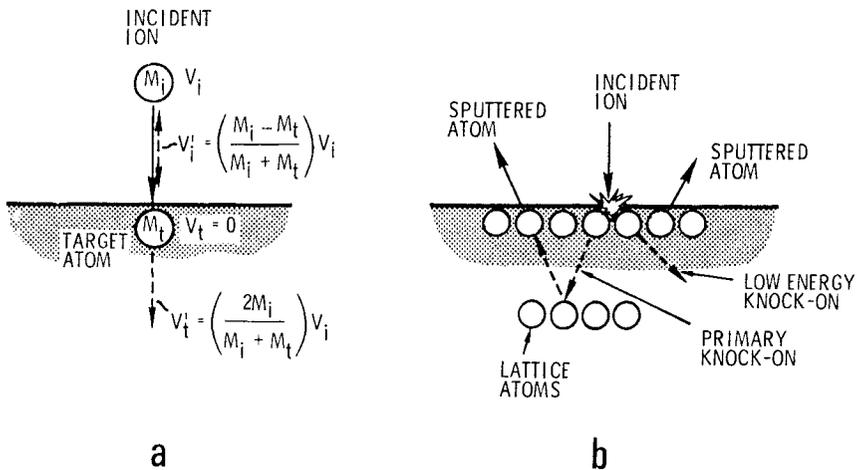


Figure 20: Schematic diagram showing some of the momentum exchange processes that occur during sputtering; V_i is the ion velocity, V_t is the target atom velocity, and the prime denotes velocities after the collision. From Reference 59.

computer simulations indicate that some cascades produce no sputtered atoms while others produce two orders of magnitude more ejected atoms than the average measured value.⁶⁶ Thus the majority of the sputtered atoms originate from a small fraction of the cascades. Therefore, these simulations suggest that one should not think of the average cascade, which may produce one or two sputtered atoms, as being the typical sputter-producing cascade. The majority of the sputtered atoms originate from cascades producing perhaps 10 or more ejected atoms.

The general nature of collisional cascades in “knock-on” sputtering have been classified into (1) the single knock-on regime, (2) the linear cascade regime, and (3) the spike regime.⁶⁴ In the single knock-on regime, recoil atoms, resulting from the collisions between the incident particles and the target atoms, receive enough energy to be sputtered but not enough to produce recoil cascades. In the linear cascade regime, recoil atoms produced by the interaction of the bombarding particle with the target have sufficient energy to generate recoil cascades. However, only a small fraction of the target atoms within the cascade volume are set in motion. In the spike regime, the density of recoil atoms is so high that the majority of the atoms within the spike volume are set in motion.

The bombarding particle masses and energies that are used in sputter deposition produce behaviors that fall into single knock-on (low and medium eV range) and linear cascade (keV range) regimes. The single knock-on regime also appears to apply to a process called ion-induced desorption, which is very important in sputter cleaning and bias sputtering (see Section 6.4). Thus theoretical studies indicate that the sputter desorption of nitrogen from tungsten occurs via mechanisms where (1) the bombarding Ar ions are reflected from the tungsten lattice and dislodge the nitrogen atoms by striking them from beneath, and (2) the nitrogen atoms receive momentum directly from the incident ions and escape by being reflected from the tungsten lattice.⁶⁸ The spike regime applies to heavy ions or molecules that strike the target at high energies.

The second observation concerning the basic momentum exchange shown in Figure 20a is that the momentum exchange between the bombarding particle and the target particles is greatest when their masses are equal. The actual sputtering process is complicated, of course, by the fact that it involves a complex sequence of collisions of various types involving the bombarding particle and the various species that compose the target, as well as the target atoms interacting among themselves. However, as a general rule for a given material, the sputtering rate will be highest when there is a good match between the masses of the bombarding particle and the atoms within the target.^{69,70} Accordingly, argon is generally used as a sputtering working gas because of its inertness, its mass compatibility with materials of engineering interest, and its low cost.

The third observation from Figure 20a is that the bombarding particle may be reflected backward in a single collision if its mass is less than that of the target atom. The energy of the reflected particle may be a significant fraction of its initial energy. For a 180° reflection (line of centers impact) this fraction is $(M_t - M_p)^2 / (M_t + M_p)^2$.² Thus for a 500 eV Ar particle, mass 40, undergoing a 180° reflection on a Mo target, mass 96, the reflected particle

will have an energy of 85 eV. If $M_i > M_t$, a 180° reflection of the bombarding particle requires more than one collision.

In most sputtering applications the majority of the bombarding particles are ions. An ion approaching a clean conducting surface has a high probability of being neutralized by an emitted electron prior to impact.^{71,72} Consequently, the reflected particle will be neutral and unaffected by the electric field in the vicinity of the negatively biased target. Therefore, when $M_i < M_t$ the sputtering process produces a flux of energetic reflected working gas species as well as sputtered particles. At low working pressures these reflected atoms can reach the substrates with a substantial fraction of their initial energy. The bombardment of a growing coating by these reflected working gas atoms can have an important influence on coating properties and can result in the reflected atoms becoming entrapped in the coating.⁷³⁻⁷⁸ Figure 21 shows the concentration of entrapped working

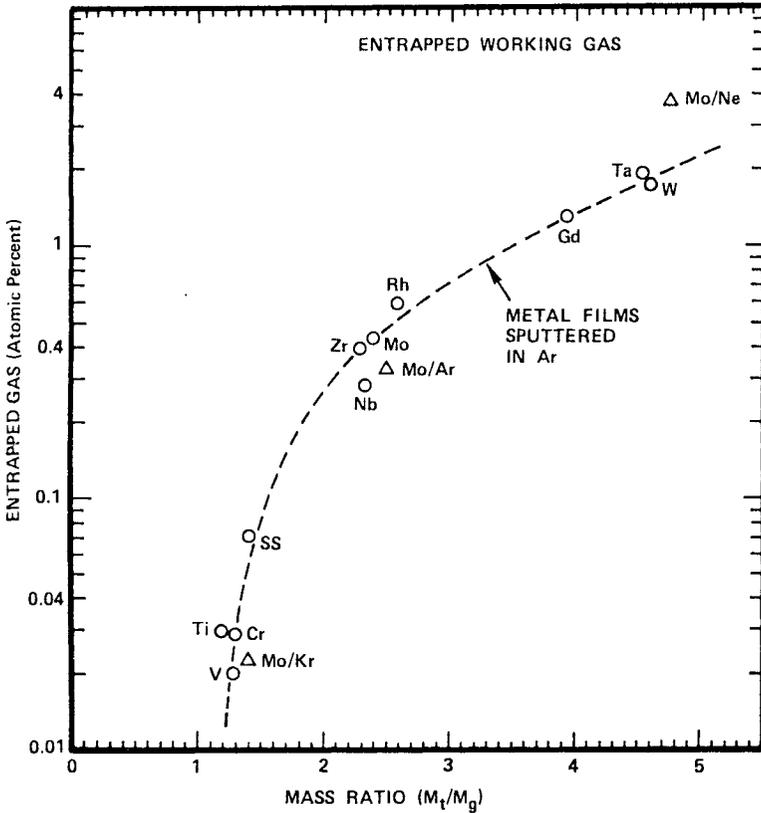


Figure 21: Entrapped working gas in metal coatings deposited at low pressures using cylindrical-post magnetron sputtering sources.⁷⁷ Circular data points refer to metals sputtered in Ar. Triangular data points from study in which Mo was sputtered in Ne, Ar and Kr.⁷⁸

gas as a function of the target-to-working gas mass ratio for metallic coatings sputter-deposited at low working pressures. Bombardment by these reflected atoms is an important consideration in controlling coating properties and is discussed in Section 6.5.

It is clear from our previous discussion that the processes induced within the target involve a much larger number of particles, and a much greater amount of energy, than is represented by the actual sputter ejection process. The binding energy of a surface atom (sublimation energy) is typically in the range from 5 to 10 eV.⁷⁹ The energy, E_d , to displace an atom from its regular lattice site in the bulk is about 25 eV.⁸⁰ For a typical material, an incident 500 eV bombarding particle will produce, on the average, about one sputtered atom with a kinetic energy of about 20 eV. Therefore, most of the particle energy, about 450 eV neglecting reflection, is dissipated within the target. This will result in the displacement of about $450/(2E_d) = 9$ target atoms.⁸¹ In metals a very efficient process of recovery sets in so that the final defect density is much smaller.⁸⁰ However, particularly in covalently bonded materials, high incident particle fluxes can cause an amorphous surface layer to develop.⁸² In addition, the probability of the incident particles being trapped in the target increases rapidly above a threshold energy of about 100 eV.⁸³ Thus an inert gas density, which depends on a balance between the rates of implantation and release, will develop near the target surface. At very high bombarding particle fluxes the density of entrapped gas can be large enough to influence the sputtering process.⁸⁴ Thus the impacting particles can have a very great effect on the target material in addition to simply producing surface erosion. It is for this reason that ion bombardment of a coating during deposition can be used effectively to modify its properties (see Section 6.4).

5.3 Sputtered Species

The sputtered species are primarily atoms.^{53,55} However, atomic clusters and molecular fractions have been observed as well as positive and negative ions. Relatively little experimental data is available on the ejection of sputtered material as molecules or clusters.

Mass spectroscopy measurements indicate that the proportion of material ejected as dimers (two-atom clusters) is a few percent or less for Ar-sputtering of metals.⁸⁵⁻⁸⁷ The fraction of trimers is an order of magnitude less. Computer modeling indicates that metal atom clusters are not sputtered as a unit, but form in flight because the sputtered atoms come into close proximity as they leave the target surface.⁸⁸

The formation of molecular fractions is much more common in the sputtering of compounds. Measurements for alkali halides and oxides indicate that molecular fractions can account for a significant proportion of the total sputtered flux.⁸⁹⁻⁹² For oxide targets the MO dimer fraction increases with the strength of the M-O bond.⁶⁰ Al_2O_3 and AlO molecules have been identified from the Ar^+ sputtering of Al_2O_3 ,^{91,92} PrO from Pr_2O_3 ,⁹⁰ and SnO from SnO_2 .⁹⁰

Molecular fractions have also been observed in reactive sputtering. Thus Ar^+ bombardment of oxidized W was found to yield WO and WO_2

species.⁹³ Similarly, O_2^+ bombardment of Mo produced MoO and MoO_2 as well as Mo,⁹³ and O_2^+ bombardment of Ti and Zr produced TiO and ZrO species.⁹⁴ The mechanism for the formation of molecular fractions from compounds, or in reactive sputtering, may be similar to the pure metal case, with the bonds forming just as the atoms leave the target surface.

The fraction of positive ions in the sputtered flux is generally less than one percent and is relevant mainly to surface analysis methods such as SIMS.^{55,95,96} In a glow discharge sputtering source, the electric field over the target surface prevents the escape of positive ions. However, negative ions can be accelerated to very high energies in this field and ultimately impact on the substrates. The yield of negative ions can be very high for targets which contain the combination of one constituent with a low ionization potential and another with a high electron affinity.^{97,98} For materials such as SmAu, it has been found that the flux of negative ions reaching the substrate can be large enough to cause a significant reduction in the deposition rate via backspattering.⁹⁷

The sputtered atoms are ejected from the target surface with considerable kinetic energy—for example, 50-100 times higher than in vacuum evaporation.^{53,55,59} The energy distribution is approximately Maxwellian, with a most probable energy of about one half the surface binding energy, and a slightly overpopulated high energy tail, so that the average energy is of the order of 10-40eV.⁵³ See Figure 22. Increasing the bombarding ion energy increases the population of the high energy tail. However, the average energy of the ejected particles ceases to increase significantly for ion energies above about 1 keV.⁵³ The atoms ejected from most metals ($Z > 20$) under Kr bombardment have average velocities which lie in a relatively narrow range (4 to 8 km/sec), so that the average kinetic energies increase with the ion mass as shown in Figure 22. Ejection velocities under Ar

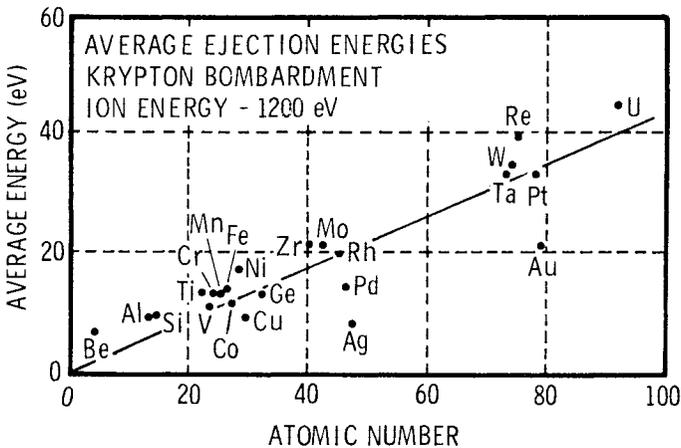


Figure 22: Average energies of sputtered atoms produced by 1.2 KeV krypton ion bombardment. The line corresponds to a velocity of 6 Km/s. Data from Reference 53.

bombardment are about 20% lower than for Kr bombardment,³⁷ and thus in the range of 3 to 6 km/sec. Several theories have been developed to describe the energy distribution of atoms sputtered from a target.⁹⁹ A relatively straightforward model by Thompson permits a calculation of the energy distribution of the sputtered atoms as a function of the incident ion energy, the ion and target atom atomic masses, the interatomic potential, the surface atom binding energy and the nearest neighbor spacing of the target atoms.¹⁰⁰

Atoms sputtered from polycrystalline or amorphous targets under perpendicular ion incidence at typical working energies (1 to 3 keV) are ejected in nearly random directions as a consequence of the multiple collisions that occur within the target, and therefore have near-cosine distributions.⁵³ At low ion energies (~ 1 keV) the distribution may be slightly under cosine (more emission at large angles) and at higher energies (~ 3 keV) over cosine.⁵² Under oblique incidence the target atoms are sputtered in the forward direction from smooth surfaces. However, the roughness of most practical targets causes a trend toward random emission. This is particularly true for polycrystalline targets, where the difference in yield for different crystallographic directions can lead to an increase in surface roughness as sputtering proceeds. Thus a cosine distribution is generally a good approximation for calculating deposition profiles. Particles sputtered from single crystal targets, or from targets with a high degree of preferred orientation, are preferentially ejected along crystallographic directions.⁵³ This apparently occurs because low energy atomic interactions in solids involve collective types of motion that may be more or less columnated in specific directions.⁶⁴

5.4 The Sputtering Yield

The sputtering process is quantified in terms of the sputtering yield, defined for a mono-element target as the average number of target atoms ejected per incident particle. For multi-element targets one has the *partial sputtering yield* of component i , defined as the average number of sputtered atoms of component i per incident ion, and the *component sputtering yield*, defined as the partial sputtering yield of component i divided by the equilibrium surface concentration of species i during sputtering.¹⁰¹

The yield depends on the target species and its surface topography, and on the bombarding species and its energy and angle of incidence. It is insensitive to the target temperature.⁵³ The yield is also independent of whether the bombarding species is ionized or not, as ions have a high probability of being neutralized by a field-emitted electron prior to impact as discussed in Section 5.2. However, ions are generally used because of their ease of production and acceleration in a glow discharge. At the bombarding energies used for sputter deposition, molecular bombarding species behave as if the atoms of the molecule arrived separately with the same velocity as the molecule and initiated their own sputtering events.⁵³ High energy molecular bombardment in the spike regime can induce nonlinear behavior and a sputtering yield that is substantially higher than twice the single particle yield.⁶⁴

Sputtering yields are generally determined experimentally. Figure 23 shows experimental yield versus ion energy data for several materials sputtered with Ar ions under normal ion incidence. Additional data are given in Table 6 and in Reference 105. The yield dependence on the bombarding ion energy is seen to exhibit a threshold of about 10-30 eV, followed by a near-linear range which may extend to several hundred eV. At higher energies the dependence is less than linear. The sputtering process is most efficient from the standpoint of energy consumption when the ion energies are within the linear range.

Considerable progress has been made in theoretically predicting the sputtering yield for elemental materials. Sigmund's linear cascade theory⁶¹⁻⁶⁴ correlates well with experimental values in the near-linear energy range¹⁰⁶ and has become the most widely used. The analysis expresses the sputtering yield as

$$S = K \frac{M_i M_t}{(M_i + M_t)^2} \left(\frac{E}{U_o} \right) \alpha(M_t/M_i) \tag{10}$$

where K is a constant in the range from 0.1 to 0.3, M_i and M_t are the masses of the incident ion and target atom respectively, E is the energy of the incident ion, and U_o is the binding energy of the target atom (usually taken

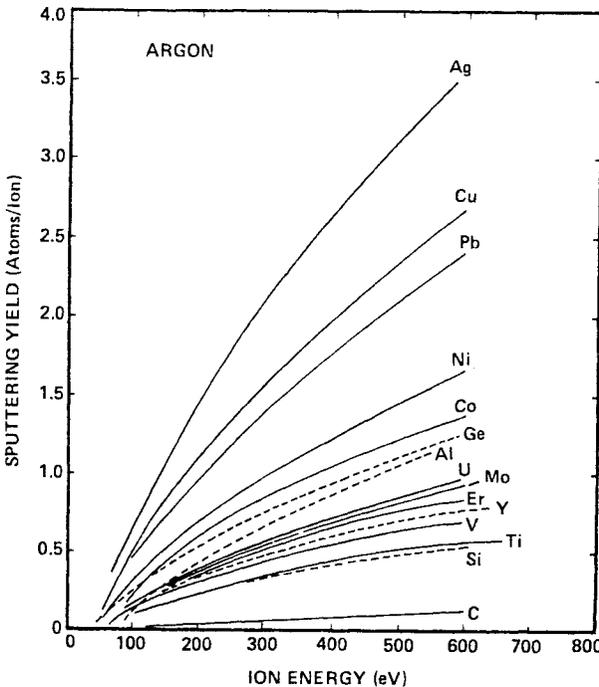


Figure 23: Variation of sputtering yield with ion energy for Ar⁺ ions at normal angle of incidence. Data from References 52 and 102-104.

as the sublimation energy). $\alpha(M_i/M_j)$ is a monotonically increasing function of M_i/M_j , which is tabulated in References 61 and 64. At typical mass ratios ($M_i/M_j \sim 2$) the value of α is about 0.3.

The Sigmund formulation has been correlated with existing yield measurements and then used to estimate yield values for elements such as Cd, Zn, Sn, and Sb where low energy experimental data are not generally available.¹⁰⁶ Accordingly, theoretical values are listed on Table 6 where experimental values are not available.

Table 6: Sputtering Yield for Various Materials Under Argon Bombardment*

Target	Ion Energy (eV)	Yield	Reference
Au	600	2.8	54
Bi	500	6.64	56
Cd	500	7.20	106
Cr	600	1.3	54
Dy	500	0.88	56
Er	500	0.77	56
Eu	500	5.02**	106
Gd	500	0.83	56
Hf	500	0.70	56
In	500	3.25	106
Ir	500	1.01	56
Fe	600	1.3	54
Mn	500	1.9	56
Nb	600	0.65	54
Nd	500	2.65**	106
Os	600	0.95	54
Pb	500	4.81**	106
Pd	600	2.40	54
Pr	500	2.40**	106
Pt	600	1.60	54
Rb	500	9.20**	106
Re	600	0.40	54
Rh	600	1.50	54
Sb	500	2.83	56
Se	500	3.35**	106
Sm	500	0.80	56
Sn	500	1.20	56
Ta	600	0.60	54
Tb	500	2.25**	106
Te	500	4.34**	106
Th	600	0.70	54
W	600	0.60	54
Zn	500	5.07**	106
Zr	600	0.75	54

*See Figure 23 for elements not given in this table.

**Theoretical prediction.

The general dependence of the sputtering yield on the ion angle of incidence is indicated in Figure 24.¹⁰⁷ The relationship provides another example of the surface nature of the sputtering process. An ion which is incident on the target surface at an angle θ will, to first order, have its path length increased by a factor $\text{Sec } \theta$ before it leaves the depth, d , where the primary sputtering momentum exchange occurs. At large angles of incidence, ion reflection dominates and the yield decreases. In glow discharge sputtering devices the ions generally approach the target in a direction normal to the target surface. The relationship shown in Figure 24 is of particular significance when the target surface is highly irregular. This will be discussed in Section 5.5.

Note in Figure 23 and Table 6 that the yields of most metals are about unity and within an order of magnitude of one another. This is in contrast, for example, to evaporation, where the rates for different materials at a given temperature can differ by several orders of magnitude. See Section 3.2. It is this universality that makes sputtering such an attractive process for many applications.

Figure 25 shows the dependence of the sputtering yields for Ag, Cu, and Ta on the species of the bombarding ion.¹⁰⁸ Although the ion energies are considerably above those used for sputter coating technology, the

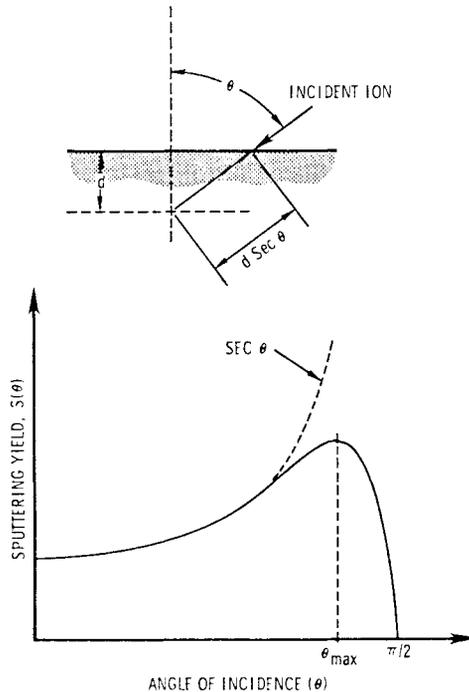


Figure 24: Schematic diagram showing variation of sputtering yield with ion angle of incidence. Ion energy is constant. From Reference 59.

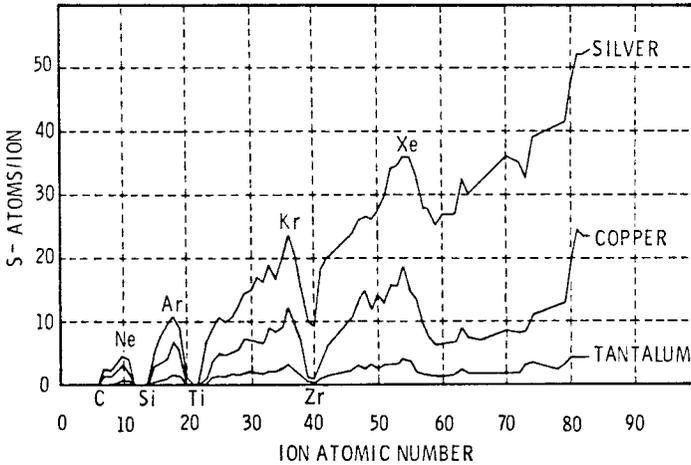


Figure 25: Sputtering yields for various ions impacting at normal incidence on silver, copper, and tantalum surfaces at high energies (45 keV). Data from Reference 108.

figure does illustrate the trends. It is seen that the noble gas ions give the highest yields. Of particular interest is the fact that yields vary much more with ion species, a factor of 100 or more, than they do with atom species, a factor of 10.⁵³ This occurs because the bombarding ions form alloys or compounds with lower sputtering yields on the surface of the target. Note that the yields are particularly low for active species such as Be, C, Mg, Si, Ti, and Zr. The formation of target surface compounds with reduced sputtering yields is commonly encountered in reactive sputtering. See Section 5.9.

It was mentioned in Section 5.2 that inert working gas species become entrapped in the target. The amount of gas entrapped in a target can be large enough to influence the sputtering yield,⁸⁴ although this is not a problem in most practical applications. The mechanism of release is still unresolved.¹⁰⁹ Measured equilibrium argon densities in tungsten imply that if the release is by argon sputtering, then the inert gas yield is an order of magnitude larger than that of the host tungsten lattice.⁸³

An important consideration in reactive sputtering and sputter cleaning is the sputter desorption of reactive gas species that have become chemisorbed on the surface of a target or substrate. Calculations and measurements indicate that such species are sputtered as atoms with yields that are of the same order of magnitude or higher than those of elemental materials.⁶⁸ See Section 5.2.

Sputtering apparatuses are generally calibrated to determine the deposition rate under given operating conditions. However, yield data of the type described above are often used in projecting rate changes when changing coating materials and in estimating the amount of material removed during sputter cleaning and bias sputtering. The erosion rate is given by

$$R = (0.10) JS M_A / \rho \text{ nm/sec} \quad (11)$$

where J is the ion current density in mA/cm^2 , S is the sputtering yield in atoms/ion, M_A is the atomic weight in grams, and ρ is the density in gm/cm^3 of the target material.

5.5 Sputtering Alloys and Compounds

An important advantage of the sputtering process is that the composition of a sputtered film tends to be the same as that of the target, provided that (1) the target is maintained sufficiently cool to avoid diffusion of the constituents, (2) the target does not decompose, (3) reactive contaminants are not present, (4) the target-to-substrate transport is the same for all of the constituents, and (5) the sticking coefficients at the substrate are the same for all of the constituents.⁹⁵ Composition control of alloys and compounds is a very important consideration for many electronics-related applications, as discussed in Section 5.1, and has led to wide usage of sputtering.

In sputtering multi-component materials, one finds in general that the different components are sputtered at different rates.^{82,99,110,111} This creates a surface or "altered" layer which has a different composition from the bulk. However, if conditions (1) and (2) above are satisfied, conservation of mass requires that one sooner or later reaches a situation where the sputtered flux leaving the target has the same composition as the bulk. In this equilibrated condition the composition of the altered layer is just such that the effective sputtering yield times the surface concentration for each species is proportional to its composition in the target. The thickness and composition of the altered layer will depend on the target material and sputtering conditions. Typically thicknesses are similar to the range of the bombarding species, with values of 30-100 Å.^{82,99,110,112} A change in sputtering conditions will in general require an adjustment of the altered layer. Thus one must allow for suitable equilibration times when sputtering multi-component materials. It is important to note that the effective sputtering yield of a constituent in an alloy or compound will not be the same as that of the constituent by itself, because of the different binding energies and the different atomic masses involved in the collision sequence within the alloy or compound.^{62,63,111}

Failures to achieve coatings having the same composition as the target can generally be traced to a violation of one of the points listed above. The most straightforward case is that of homogeneous single phase alloy targets. Suitably cooled and equilibrated targets will yield stoichiometric fluxes. However, the angular distribution of ejection for the various sputtered constituents may not be the same.⁸² This is particularly true for constituents with significantly different masses. Thus failure to achieve stoichiometric coatings from such targets is generally the result of a failure to satisfy conditions (4) or (5) cited above.

The case of multiphase alloy or pressed powder targets can involve the development of an altered layer, with a modified surface topography as well as composition. For example, when the phases in a multiphase alloy

target have significantly different sputtering yields, the inhomogeneous sputtering yield over the target will cause an irregular surface topography to develop.^{107,113-116} This development is exacerbated by the angular dependence of the sputtering yield (see Figure 24). The altered layer can therefore assume a macroscopic character, in which the surface area exposed to grains or phases with high sputtering yields is shrunk so that the surface is enriched with low sputtering yield material. In such cases, the equilibration time will increase with grain or powder size. The process is complicated by surface diffusion or gas phase backscattering of sputtered atoms, both of which can cause a mixing on the target surface of atoms from the various phases.¹¹⁷ For example, low sputtering yield material from protruding grains can be sputtered onto recessed grains of high yield material. The low yield material can then serve as seed species for the development of a surface topography consisting of a forest of densely packed cones on the high yield material.^{82,116,118-123} The cone-covered, or otherwise modified, surface topographies can develop a very low effective sputtering yield because atoms are sputtered back and forth many times before they clear the surface.^{124,125} Nevertheless, the important point is that despite these aberrations, after a suitable equilibration time conservation of mass dictates that the sputtered flux match the bulk composition, provided that conditions (1) and (2) cited at the beginning of this section are satisfied.

Uniform compound targets of materials such as carbides and silicides should behave as the single phase alloys described above and yield stoichiometric erosion fluxes as soon as an equilibrated altered layer is formed. However, particular caution should be exercised when using pressed-powder targets. High contamination levels may be present throughout such targets because of the large surface area contained in the starting powder.^{126,127}

Compound targets incorporating volatile constituents such as oxides, nitrides, and halides often yield deposits that are deficient in the more volatile constituents.^{128,129} The fact that volatile species may evaporate rather than be sputtered from the target surface will not alter the overall material balance, provided that the loss is limited to the surface. Thus an equilibrated target should yield stoichiometric emission flux. The problem is that the volatile species are likely to have low substrate sticking coefficients, with the consequence that some of the evaporated material is pumped away, thereby yielding non-stoichiometric deposits. The problem is generally solved by a form of reactive sputtering in which "make-up gas" containing the volatile species is injected into the sputtering chamber. See Section 5.9.

Composite targets consisting of relatively large regions of different materials are sometimes used. The sputtering rate from each region will depend on the sputtering yield and electron secondary emission coefficient as well as the degree to which gas scattering of the sputtered flux causes mixing between the target regions. Therefore, an empirical approach is generally required.

Caution must also be exercised when using targets composed of compounds having poor electrical and thermal conductivity. Cracking

often limits allowable current densities. The problem is particularly important for planar magnetrons where concentrated heating occurs under the plasma ring.¹³⁰ Poor thermal conductivity leads to high surface temperatures and may result in the loss of volatile constituents by evaporation or sublimation. The high electric field in a poorly conducting target can act in concert with the high temperature and promote diffusion within the target.

The sputtered flux must pass through the working gas to reach the substrates. Significant gas scattering of the sputtered flux occurs in those apparatuses which operate at elevated pressures. See Section 5.6. The scattering is dependent on the mass of the sputtered atoms. The consequences of scattering depend on apparatus geometry. Thus the composition of the flux of sputtered atoms which arrive at the substrate in a planar electrode apparatus may be quite different from that which leaves the target. By contrast, gas scattering has a much lesser effect on the composition in a cylindrically symmetric apparatus.

Any substrate that is in contact with a plasma will be subject to resputtering, if the potential difference between the substrate and the plasma exceeds the threshold of the sputtering yield for some of its constituents. Accordingly, coatings deposited by bias sputtering, see Section 6.4, are vulnerable to the loss of high yield constituents. In some cases, controlled resputtering is used to deposit coatings with a range of compositions from a given target.¹³¹

5.6 Glow Discharge Sputtering Apparatuses

A key problem in industrially implementing the sputtering process is to provide a uniform and copious supply of ions over the surface of the target. The low pressure glow discharge has proven to be the most cost-effective source of ions. Glow discharge plasmas are discussed in the sputtering context in References 57 and 132. A wide range of glow discharge apparatus geometries have been used in attempts to (1) increase the ion supply and thus the sputtering rate, (2) increase the target area and thus the available deposition area, (3) reduce the plasma heating of the substrates, (4) permit a lowering of the working gas pressures, and (5) facilitate the coating of particular substrate shapes.⁵⁹ The emphasis here is on those apparatus types which are commonly used in the electronics industry.

5.6.1 Planar Diodes. The planar diode shown in Figure 19 is the simplest of the sputtering apparatus configurations. One electrode, typically 10 to 30 cm in diameter, is configured to mount the target. The substrates are mounted on a table, which is typically spaced 5 to 10 cm from the target and generally serves as the second electrode. When operated with dc power, the target electrode serves as the cathode and the substrate table as the anode as shown in Figure 19. If ion bombardment during coating growth is desired (bias sputtering) the substrate table is also biased as a cathode and the chamber wall or an auxiliary electrode will serve as an anode (such a system is strictly a triode—three electrodes—but is still generally referred to as a diode). When operated with rf power, the electrodes change cathode/anode roles on each half cycle. Consequently, it is necessary that the substrate electrode be made considerably larger than the

target electrode, so that the sputtering is concentrated at the target. See Section 5.8. Sometimes this can be accomplished by connecting the chamber wall and substrate table together as a common electrode. Bias sputtering in the rf case is often accomplished by controlling the external impedance, to adjust the relative power input to the target and substrate electrodes.^{133,135} Ground shields such as those shown in Figure 19 are typically used to suppress current flow from all except the desired surfaces of the electrodes. The target electrode is generally water cooled. The substrate table may include provisions for heating or cooling the substrates.

The condition for sustaining a discharge in an apparatus of the type shown in Figure 19 is that the rate of production of ions in the plasma volume be adequate to balance the fluxes of electrons and ions that pass to the electrode and chamber wall surfaces. First, consider the dc case. The current in such a discharge is carried in the vicinity of the cathode primarily by positive ions and in the vicinity of the anode by electrons. Because of the relatively low mobility of the ions compared to the electrons, most of the electrical potential that is applied between the electrodes by the power supply is consumed in a "cathode dark space," or sheath region adjacent to the cathode.⁵⁷ Dark space thicknesses are typically 1 to 4 cm, depending on the pressure and current density. Accordingly, strong electric fields are formed, and ions passing from the plasma volume to the cathode are accelerated by these fields and on impact at the cathode produce the desired sputtering. A small number of "secondary electrons" are emitted from the cathode as a consequence of the ion bombardment¹³⁶ (about one for every ten ions in the case of argon ions impacting on a metal cathode¹³⁷). These electrons are accelerated in the cathode dark space to energies approaching the applied potential, and enter the plasma volume (negative glow) where, known as "primary electrons," they exchange energy and produce the volume ionization required to sustain the discharge.^{57,132} Unfortunately, the energy exchange between the primary electrons and the working gas is relatively inefficient, so that it is difficult to sustain a high current density plasma discharge in the planar diode electrode geometry. Thus working gas pressures are relatively high, 3 to 10 Pa (20 to 75 m Torr), and current densities are low, typically less than 1 mA/cm². The energy exchange in the rf case is more efficient, so that considerably lower pressures, 0.5 to 2 Pa (5 to 15 m Torr) can be used. See Section 5.8.

There are therefore three factors that characterize dc sputtering with planar diode sources: (1) the cathode current densities and sputtering rates are low, (2) the working pressures are high, and (3) the substrates are in contact with the plasma. Typical operating conditions for metal sputtering are: cathode current density—1 mA/cm², discharge voltage—3000V, Ar pressure—10 Pa, cathode-to-substrate separation—4 cm, deposition rate—0.7 nm/sec.⁵² (See Figure 28).

Because of the high gas pressure, the motions of both the ions and the sputtered atoms are dominated by collisions. Thus, Ar ions passing through the dark space undergo charge exchange collisions with neutral Ar atoms that produce fast neutrals and "slow" ions.^{138,139} Consequently, the target is not bombarded by a current of ions having an energy equal to the

potential drop across the dark space, but instead a much larger number of ions and atoms having energies that are often less than 10% of the potential difference across the cathode dark space.¹⁴⁰ This redistribution of energy can actually prove to be an advantage. Because of the nonlinear dependence of the sputtering yield on the bombarding particle energy (see Figure 23), ten 300 eV particles can, for example, produce more sputtered species than one 3000 eV particle.

Gas scattering of the sputtered particles has several important consequences. At the pressures used for dc planar diode sputtering, the transport of sputtered species from the target to the substrates is largely by diffusion.¹⁴¹ The deposition rate is therefore reduced because a significant fraction of the sputtered particles diffuse back to the target or to the chamber walls. It is estimated that about 10% of the sputtered material reaches the substrates in a well designed planar diode.¹⁴² Because of the diffusion nature of the transport, surfaces that are adjacent to a substrate, but do not shield it optically from the target, can still rob it of coating flux. The complexity of this diffusion transport, along with the charge exchange processes in the cathode sheath, makes it necessary to determine deposition rates experimentally for each set of operating conditions. Another important consequence of the collision-dominated transport of the sputtered atoms is that their initial high energies of ejection (see Section 5.3) are reduced to near thermal values by the time they reach the substrates under typical planar diode dc sputtering conditions.^{55,143,144,145,146}

The fact that substrates in a planar diode are in contact with the plasma means that they are subjected to bombardment by ions and electrons from the plasma.¹³² The energies and relative fluxes will depend on the potential of the substrates relative to the plasma potential. The substrates are also subject to bombardment by the energetic primary electrons,¹⁴⁷ particularly at the lower operating pressures, and to electromagnetic radiation from the plasma. The plasma bombardment can be beneficial to the structural properties of the coatings. See Section 6.4. However, bombardment by energetic species can cause damage to semiconductor devices. See Section 7.0.

As noted above, planar diode systems can be operated at lower pressures when rf power is used. However, the general behavior differs from that described above only in degree. Thus the charge exchange and gas scattering of the sputtered flux will be less, while the substrate bombardment by primary electrons will be greater.

Substrate heating rates are relatively high in planar diodes. Typical rates are in the range from 100 to 300 eV/atom deposited.^{148,149} The major sources of heating are bombardment by the primary electrons and species from the plasma. Uncooled substrates typically reach temperatures in the 300 to 500°C range.

Planar diodes were the most commonly used sputtering apparatus for many years, but are being replaced by magnetrons for most metal deposition applications. The present applications of planar diodes are primarily for rf sputtering of various poorly conducting compounds. They are also used for magnetic materials because of the limitations of magnetrons for this application.

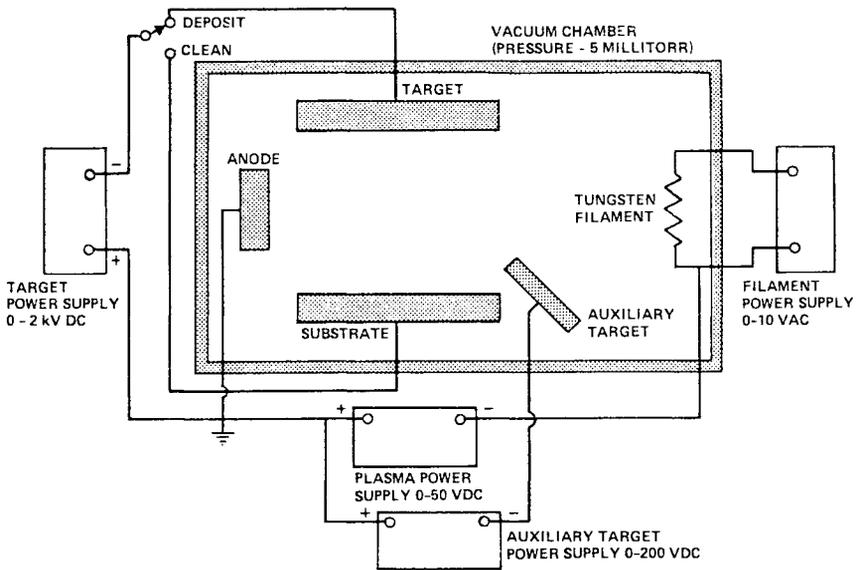


Figure 26: Schematic drawing of enhanced thermionically supported discharge system with a tungsten filament hot cathode. This apparatus, developed by Battelle Northwest Laboratories,¹⁵⁰ uses a unique auxiliary target to continually deposit a fresh low work function material such as thorium onto the cathode filament. This significantly reduces the required filament power.

5.6.2 Assisted-Discharge Devices, Triodes. In assisted or supported discharge apparatuses an electrode system that is independent of the target is provided for sustaining the glow discharge. The most common such configuration is the hot cathode triode. Such an apparatus is shown in Figure 26.¹⁵⁰ Electrons are emitted at the cathode surface by thermionic emission rather than ion bombardment. This largely uncouples the volume ionization requirement for sustaining the discharge from the secondary-electrons. Consequently, hot cathode triodes can be operated at low pressure (0.5 to 1 mTorr). Driving voltages for the thermionic discharges are typically only 50-100V, although the current may be several amperes. A magnetic field may be used to confine the plasma. The targets may be driven by dc or rf power.

Typical operating conditions for metal sputtering are as follows: target current density 1-20 mA/cm², target voltage 500-1500V, Ar pressure 1 mTorr (0.13 Pa) cathode to substrate spacing 5-10 cm, deposition rate 1-20 nm/sec.

The substrates may or may not be in direct contact with the plasma. However, in the absence of a magnetic field they will be subjected to bombardment by energetic electrons originating from the target. Triodes operate at sufficiently low pressures so that the sputtered atoms may undergo a near-collisionless transport to the substrates and preserve

much of their initial kinetic energy. Furthermore, energetic neutral working gas atoms, which are generated at the cathode by the neutralization and reflection of ions as discussed in Section 5.2, can also pass to the substrates with little loss of kinetic energy.¹⁴⁶ These atoms can become trapped in the growing coating, particularly if the substrate temperature is low. Thus the trapped argon content for coatings deposited with a triode has been found to be larger than that for similar coatings deposited with a planar diode.^{75,76}

5.6.3 Magnetrons. The recent development of high performance magnetron sputtering sources that provide (1) relatively high deposition rates, (2) large deposition areas, and (3) low substrate heating, is revolutionizing the sputtering process by greatly expanding the range of feasible applications.^{59,151}

Magnetron sputtering sources can be defined as diode devices in which magnetic fields are used in concert with the cathode surface to form electron traps which are so configured that the EXB electron drift currents can close on themselves.^{152,153} Magnetic field strengths are sufficient to confine the electrons but not the ions. However, the ions tend to be confined electrostatically in the vicinity of the electron component of the plasma. Typical magnetic field strength values are a few hundred gauss. The secondary electrons which are emitted from the cathode surface as a consequence of the ion bombardment, and are accelerated in the dark space to become energetic primary electrons, are trapped in the vicinity of the cathode by the magnetic field. The trapping of the primary electrons in a well designed magnetron is sufficiently effective so that these electrons are able to transfer most of their energy to the plasma, and thereby to cause the production of large numbers of ions, before they are lost from the system.¹⁵⁴ Furthermore, the ions are produced in the immediate vicinity of the cathode, so that they have a high probability of making their way to the cathode and therefore participating in the current flow that produces sputtering.

Magnetron sources can be configured in many different forms. Several of the more common types are shown in Figure 27. Properly designed magnetrons of all the types yield comparable total sputtering rates when operated at common currents.¹⁵⁴ The discharge current-voltage (I-V) characteristic provides a very revealing signature identifying the nature of the discharge process. I-V characteristics for a cylindrical-post magnetron (Figure 27A) and a planar magnetron (Figure 27H) are shown in Figure 28 and compared to a planar diode. The low operating voltages, and the nearly flat I-V characteristics for the magnetrons (small voltage change for large increase in current) are a signature of efficient energy exchange processes within the plasma discharge.¹⁵⁴ High impedance, constant current, power supplies are generally used for magnetrons.

The cylindrical-post magnetrons (Figures 27A, B, D), planar magnetrons (Figure 27H), and gun type magnetrons (Figure 27I) are the most commonly used types in the electronics industry. The cylindrical magnetrons shown in Figures 27A and 27B use uniform magnetic fields generated by coils located external to the vacuum chamber.⁵⁹ The magnetic fields for the plasma ring type devices shown in Figures 27D, E, G, and I are usually generated by permanent magnets positioned behind the target. Small

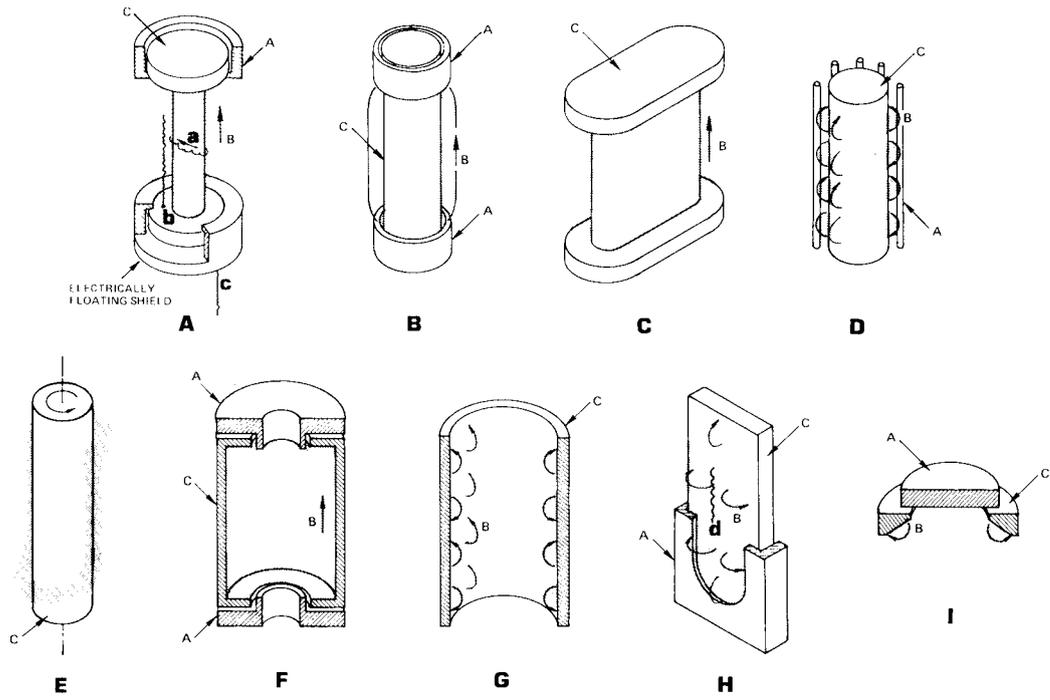


Figure 27: Schematic illustrations of various types of magnetron sputtering sources. In each illustration A is anode, B is magnetic field direction, C is cathode. (A) cylindrical-post magnetron with electrostatic end confinement,¹⁵³ (B) cylindrical-post magnetron with magnetic end confinement,¹⁵⁴ (C) rectangular post magnetron,¹⁵⁴ (D) ring discharge post magnetron,¹⁵⁵ (E) spiral discharge magnetron,¹⁵⁶ (F) cylindrical hollow magnetron,¹⁵³ (G) ring discharge hollow magnetron,¹⁵⁷ (H) planar magnetron,¹⁵⁸ (I) "gun type" magnetron.¹⁵⁹

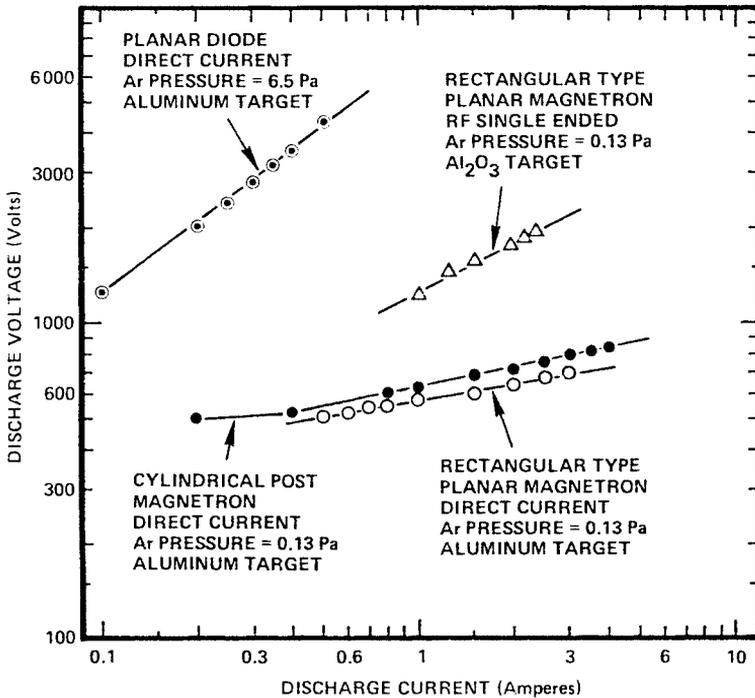


Figure 28: Current-voltage characteristics for planar and cylindrical magnetron sputtering sources compared to those for a planar diode.

planar and gun type magnetrons are particularly suitable for retrofitting existing vacuum chambers. Cylindrical-post and rectangular planar magnetrons have the advantage that they can be scaled to long lengths to facilitate large deposition areas.^{151,153,160}

Cylindrical-post magnetrons are most commonly used for batch processing, with the substrates arranged surrounding the source as shown in Figure 29. For example, this configuration is used in the manufacture of chromium photomask blanks for patterning semiconductor devices. The current density is uniform over the cathode of a properly designed cylindrical magnetron of the type shown in Figure 27A. Typical values are about 20 mA/cm² which yield cathode erosion rates of about 20 nm/sec for metal targets. See Equation 11. Total discharge currents are generally in the range 1 to 50A. Operating pressures are typically about 1 mTorr (0.13 Pa) with discharge voltages in the 800V range. Because of the low pressures the sputtered atoms undergo a near-collisionless, line-of-sight transport to the substrates.¹⁴⁶ Accordingly, deposition flux profiles at various radial substrate positions can be predicted (a cosine emission is assumed).¹⁵³ At a typical substrate radial position equal to half the cathode length, a thickness uniformity of about 10% can be achieved over an axial length that is about half the cathode length. Thus a production coating machine

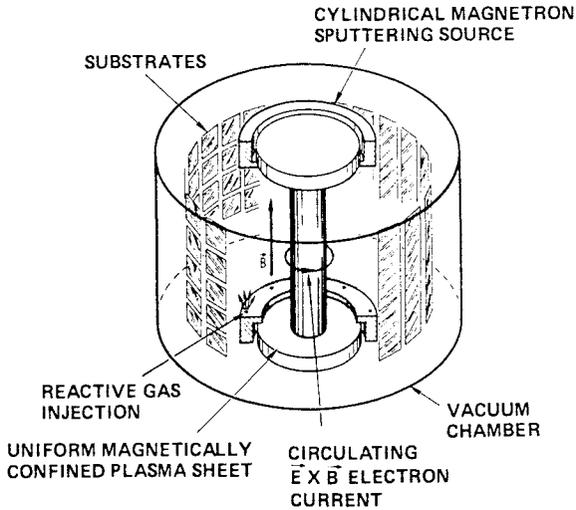


Figure 29: Typical arrangement of substrates for batch processing with a cylindrical-post magnetron sputtering source.

with a 1 m long cathode may yield acceptably uniform deposition over an area of about 1.5m^2 . Deposition rates are typically in the 1 to 3 nm/sec range. An advantage of cylindrical magnetrons is that they can be configured with thick walled targets that provide a large inventory of coating material, if the material is non-magnetic.^{151,153} Furthermore, the target material is used efficiently because of the uniform erosion along the cathode length. A disadvantage is that cylindrical target fabrication from complex materials may be difficult. See Section 5.10.

Planar and gun type magnetrons do not provide uniform deposition over large areas. Therefore, substrate motion is generally used with these devices. In the case of gun type devices, planetary systems of the type shown in Figure 9 are commonly used. An important advantage of planar and gun type devices, particularly for electronic related applications, is their adaptability to apparatus configurations which incorporate several sources for depositing multilayer coatings. Figure 30 shows a typical commercial batch coating system which can be configured with three rectangular planar magnetron sources. Up to 76 three-inch diameter wafers, mounted on a vertical carousel, can be coated with a thickness uniformity of about $\pm 5\%$. Rectangular planar magnetrons are particularly well suited to in-line systems in which the substrates are transported in a direction perpendicular to the long axis of the target. Axial uniformity can be obtained by using cathodes of sufficient length or apertures. Figure 31 shows deposition rate profiles for a typical rectangular planar magnetron. Figure 32 shows a schematic drawing of an aperture configuration.

In-line planar magnetron sputtering systems with automated cassette-to-cassette wafer handling are becoming widely used for semiconductor

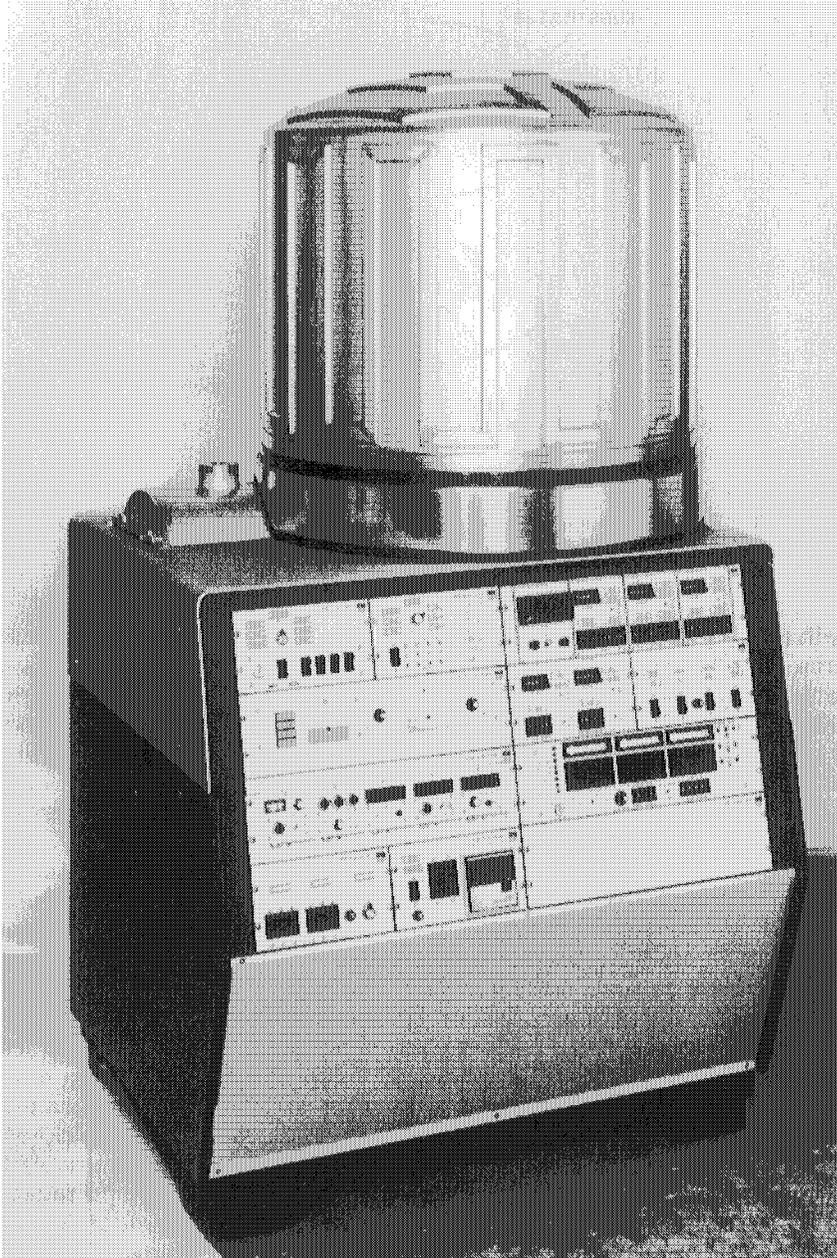


Figure 30: Typical batch type sputtering apparatus which mounts three planar magnetron sputtering sources. (Photo courtesy of CHA Industries, Menlo Park, CA.)

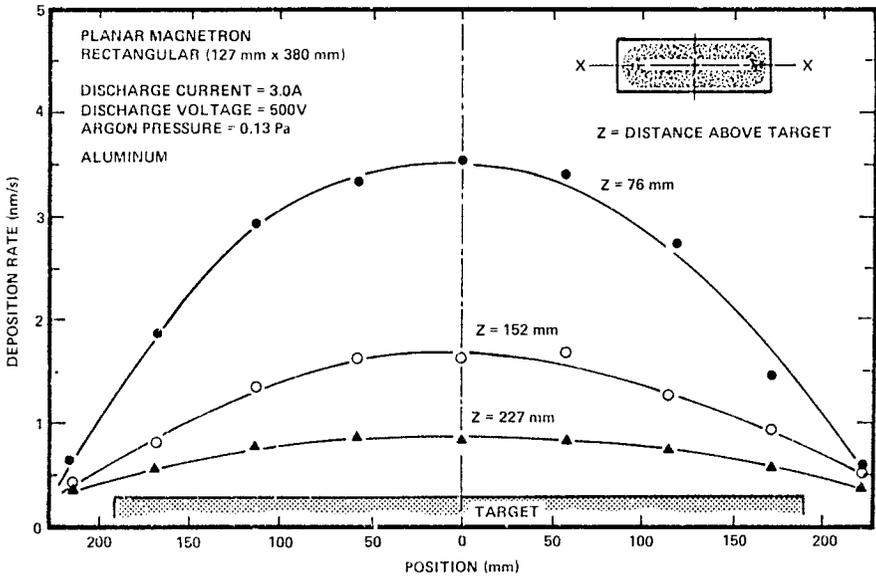


Figure 31: Deposition rate profile for rectangular type planar magnetron sputtering source on the long axis (A-A) at various distances from the cathode surface.

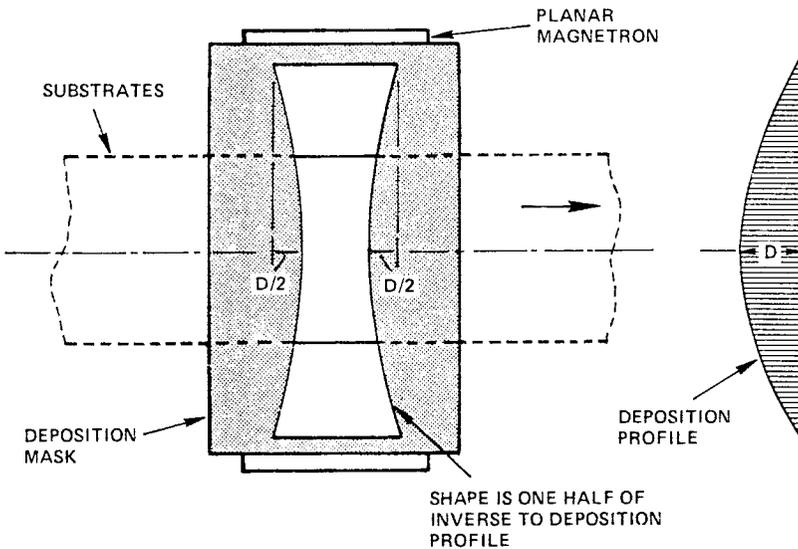


Figure 32: Schematic illustration showing the use of an aperture to improve the axial uniformity of the deposition flux which reaches the substrates from a planar magnetron sputtering source.

metallization in wafer fabrication lines. Figure 33 shows a schematic drawing of a typical system, which includes provisions for substrate heating, sputter cleaning, and the deposition of a multilayer metallization. Figure 34 shows a commercial microcomputer controlled in-line sputtering system which incorporates rectangular planar magnetron sources. Most wafer processing systems mount the sputtering sources and wafers vertically in order to sputter sideways and thereby avoid the accumulation of debris on the target and deposition surfaces. Vertical substrate mounting is used for the apparatuses shown in Figures 29, 30, and 34.

Planar magnetrons are typically operated at the same total currents as cylindrical magnetrons, 1 to 50A. However, the planar magnetron current densities are higher, 50 to 300 mA/cm², because the plasma is magnetically confined to a limited region of the target surface. Substrates are generally placed within 2 to 6 cm of the source so that metal deposition rates are typically 10 to 30 nm/sec. Thus planar magnetrons coat substrates sequentially and at high deposition rates, while cylindrical magnetrons coat larger numbers of substrates simultaneously at a slower rate but with a comparable total sputtered flux. Planar magnetron working gas pressures are typically in the range from 1 to 5 mTorr (0.13 to 0.65 Pa), with discharge voltages in the 400 to 800V range. Targets for planar magnetrons are limited to a thickness of about 1 cm by the requirement that the magnetic field over the front surface be of sufficient strength to confine the plasma. In addition, the target material is eroded only under the plasma ring and is therefore used inefficiently. Typically only about 25 to 30% of a rectangular planar magnetron target is consumed before the target must be replaced. Sometimes relative motion between the target and the magnetic field pattern is used to improve the efficiency of target consumption.¹⁶¹ Another approach uses target inserts under the plasma ring.

The substrates in magnetron systems are generally located beyond the magnetically confined plasma. If bias sputtering is to be used, the usual procedure is to move the substrates into the plasma or to use an auxiliary plasma generated in the vicinity of the substrates.¹⁶² However planar magnetron magnetic field configurations have recently been examined which are particularly effective in providing high fluxes of low energy ions at the substrate surface, while preserving the basic performance of the magnetron source.^{162a}

The substrates in cylindrical magnetrons are usually free from plasma or primary electron bombardment. However, substrate bombardment by neutralized and reflected ions is especially important in cylindrical-post magnetrons, because atoms undergoing reflections of considerably less than 180°, and therefore having relatively high kinetic energies, pass in the direction of substrates.¹⁵³ See Section 5.2. Thus coatings deposited at low pressures, where these atoms lose little kinetic energy in transport, can incorporate a considerable concentration of working gas. See Figure 21.

The primary sources of substrate heating in cylindrical magnetrons are (1) the heat of condensation, (2) the sputtered atom kinetic energy, (3) plasma radiation, and (4) energetic working gas atoms which are reflected at the cathode.¹⁶³ The contributions from the kinetic energy of the sputtered atoms and the reflected working gas atoms increase with the atomic mass.

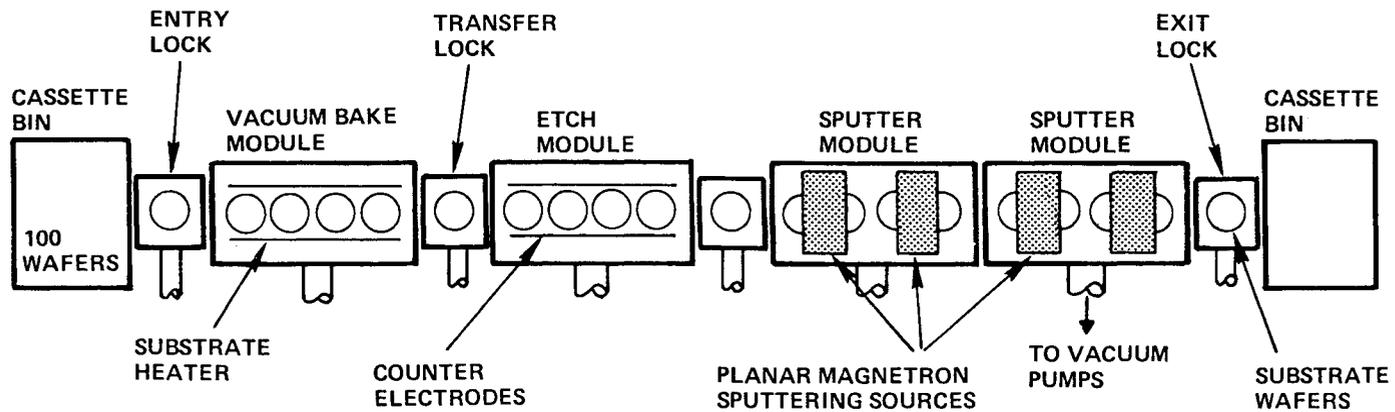


Figure 33: Schematic illustration of in-line sputtering system configured for depositing multilayer metallization onto semiconductor wafers.



Figure 34: Cassette-to-cassette in-line sputtering system designed for semiconductor wafer processing. (Photo courtesy of Materials Research Corporation, Orangeburg, NY.)

Typical substrate heating fluxes vary from 15 to 25 eV/atom for light metals to over 50 eV/atom for heavy metals with moderate sputtering yields. See Table 7. Substrates in radiation equilibrium with the heating flux are projected to reach temperatures in the 100-200°C range at a typical deposition rate of 1.7 nm/sec.¹⁵¹

Substrates in planar magnetrons are generally placed close to the sputtering sources and therefore may be subject to modest plasma bombardment. Energetic electrons moving along magnetic field lines may bombard the substrates if these field lines intersect the substrate plane.^{153,154} Substrates positioned in front of planar magnetrons are subjected to less bombardment by energetic reflected working gas atoms, because the most energetic atoms (those reflected at angles considerably less than 180°) pass to the side and miss the substrates.¹⁵⁴ Accordingly, planar magnetron substrate heating rates typically have a greater contribution from the plasma and a lesser contribution from the reflected atoms.¹⁶⁴ However, typical planar magnetron heating rates for sputtered metals are comparable to those for cylindrical magnetrons. See Table 7.

Magnetrons have become the primary sputtering method for depositing metallic coatings. Magnetron sources can be used to sputter magnetic materials. However, when a sputtering target composed of a magnetic material is used, it must be saturated magnetically so that its magnetic behavior is suppressed and a field of the desired shape can be maintained over its surface.¹⁵³ Thus, as a general rule, the allowable magnetron target thicknesses are limited for magnetic materials. Reference 165 describes several magnetron configurations specifically designed for sputtering magnetic materials.

Magnetron sources driven by rf power are also becoming widely used for depositing poorly conducting and insulating materials. However, the magnetron is essentially a dc concept, and its performance as an rf sputtering source is limited compared to its behavior as a dc source.¹⁵⁴ See Section 5.8.

Table 7: Representative Substrate Heating Rates for Cylindrical and Planar Magnetron Sputtering Sources

Metal	.Heating Flux per Deposition Atom (eV/atom).	
	Cylindrical Magnetron (Reference 163)	Planar Magnetron (Reference 164)
Aluminum	13	11
Chromium	20	16
Copper	17	12
Molybdenum	47	42
Indium	20	15
Tantalum	68	—
Tungsten	73	98
Platinum	—	48
Gold	23	—

5.7 Ion Beam Sputtering

Glow discharge sputtering technology is limited, in the sense that the target current density and voltage cannot be independently controlled except by varying the working pressures. An exception is magnetron systems in which the magnetic field is provided by an electromagnet and the voltage can be varied at a fixed current by varying the field strength.

Ion beam sputtering permits independent control over the energy and current density of the bombarding ions.¹⁶⁶ A typical ion beam sputtering system is shown schematically in Figure 35. The sputtering target is arranged to obliquely intersect an ion beam of given energy and flux density that is created by an independent ion source. Substrates are suitably placed to receive the sputtered flux, as shown in the figure. In addition to independent control over the ion current and voltage, ion beam sources permit sputtered coatings to be deposited at very low working gas pressures, ≈ 0.1 mTorr, onto substrates which are not in contact with the plasma. Suitable neutralization of the ion beam permits insulating targets to be sputtered without the use of an rf potential. Furthermore, since no potential is applied to the target, secondary charged particles (electrons and negative ions) are not accelerated away from the target and toward the substrates. However, ions will be neutralized and reflected at the target. See Section 5.2. Accordingly, substrate placement is important, since the ratio of the sputtered to the reflected ion flux will depend on the angular position of the substrates relative to the target.

The ion source shown in Figure 35 is one of the simplest and most commonly used configurations.¹⁶⁷ A plasma discharge is sustained be-

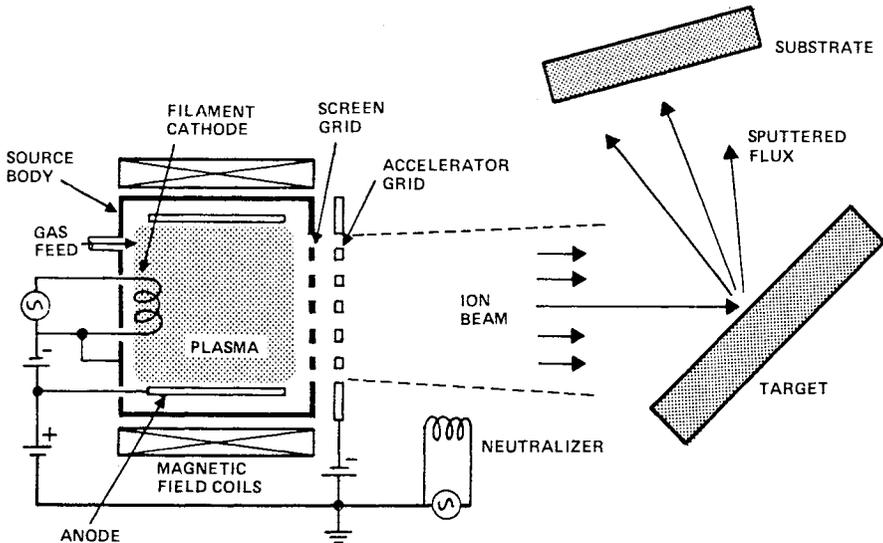


Figure 35: Schematic drawing of ion beam deposition system with discharge chamber having axial magnetic field.

tween the thermionic cathode and the anode. Electrons emitted from the cathode must cross the axial-magnetic field to reach the anode. As in the magnetron case, the magnetic field strength is made large enough so that the electron cyclotron radius is small compared to the distance from the cathode to the anode. Thus the electrons suffer many collisions with the working gas atoms, and high ionization rates can be achieved. Axial losses of electrons are minimized by maintaining the screen and chamber walls at cathode potential. Consequently, ion (electron) densities in the 10^{10} cm^{-3} range can be maintained at low, ≈ 1 mTorr, working pressures in the chamber. Anode-cathode voltages are typically about 40V. The plasma potential is typically a few volts above the anode potential. The anode is maintained above the ground potential of the vacuum coating chamber by an amount about equal to the desired ion energy. Typical values are +500 to 1000V. The accelerator grid is maintained at a potential that is about 100V negative relative to the ground potential. The accelerator grid controls beam divergence and provides a negative potential barrier which prevents the passage of electrons from the beam plasma backwards into the positive discharge plasma. A thermionically emitting hot filament supplies electrons to neutralize the positively charged beam of ions.

The ions are extracted from the plasma by the electric field which is produced because of the potential difference between the screen and accelerator grids. The accelerator grids are typically fabricated from graphite, with grid hole diameters of about 2 mm. Screen-to-accelerator grid spacings are typically 1 to 2 mm. Typical beam current densities are 1 to 2 mA/cm², with 500 eV ion energies and beam sizes of 5 to 10 cm.

A large number of other apparatus configurations have been developed to improve certain aspects of the performance.¹⁶⁷ In the multipole magnetic field configuration, the walls of the discharge chamber are covered with anodes located behind localized magnetic fields produced by permanent magnets. The bulk of the discharge chamber is field free, so that the ion production and density are relatively uniform throughout the chamber. Multipole devices are effective for producing large diameter beams, 15 to 30 cm, with relatively uniform ion current densities. In the single grid configuration the screen grid is omitted and the acceleration distance in this case is simply the thickness of the plasma sheath. Current densities of about 1 mA/cm² can be achieved in such systems at ion energies as low as 20 eV.

Ion beam sputtering systems cannot, in general, compete with magnetrons as large scale production sources. However, the control that they provide makes them very attractive for research studies and for special applications. Ion beam sources are also used for providing controlled ion bombardment during coating growth. In this case, an ion beam sputtering or evaporation source provides a source of coating flux.¹⁶⁸⁻¹⁷⁰ The secondary ion beam is arranged to bombard the growing coating in order to modify its properties as discussed in Section 6.4. In another method, called primary ion beam deposition, an ion beam of the depositing material itself is directed at the substrate. This technique provides a high degree of control over the depositing film and has been used to produce unusual film properties such as diamond-like carbon,¹⁷¹ and to implement semiconductor

doping in deposition by molecular beam epitaxy. See Section 4.4. In reactive ion beam sputtering the target is bombarded with a beam of reactive ions. Thus, for example, silicon nitride coatings can be produced by bombarding a silicon target with nitrogen ions.¹⁷² A nitride layer forms on the target, as discussed in Section 5.9, and species sputtered from this are deposited onto adjacent substrates. Ion beam sources also play an important role in microcircuit patterning via ion beam etching. Recent reviews of ion beam deposition and etching are provided in references 173 through 175.

5.8 RF Sputtering

Direct current methods cannot be used to sputter nonconducting targets because of charge accumulation on the target surface.⁵⁷ This difficulty can be overcome by using radio frequency (rf) sputtering. A single rf sputtering apparatus can be used to deposit conducting, semiconducting, and insulating coatings. Consequently, rf sputtering has found wide application in the electronics industry. Examples of nonconducting and semiconducting materials which have been deposited by rf sputtering are given in Table 8.

Many of the phenomena which occur in glow discharge plasmas are a consequence of the large difference in mass between the electrons and ions. Thus, in the absence of strong magnetic fields, the electron flux from a plasma to a surface will tend to be significantly higher than the ion flux because of the larger thermal velocity of the electrons. Consequently, an insulating surface will accumulate a negative charge which causes it to float at a potential that is negative with respect to the plasma.⁵⁷⁻¹³² The variation in potential between the conducting plasma and the floating surface will occur in a sheath region adjacent to the surface. The magnitude of this floating potential will be just sufficient to retard the electron flux to the point where it matches the ion flux. Values depend on the electron temperature and the ion/electron mass ratio, but are typically in the range -5 to -30 V and too low to produce a significant sputtering rate.

Now let an electrode be placed behind the insulator and let an rf potential be applied to the electrode. Two things happen. First since the capacitive impedance varies inversely with frequency, an rf-current can now pass through the insulator. However, since the dc current must still be zero, the total electron and ion current flux during each complete cycle must balance to zero. Accordingly, the second consequence of applying such an rf potential is that the insulator will develop a voltage bias that is negative with respect to the plasma potential. The situation is very similar to the floating potential case described above. However, in the rf case the bias potential that develops is just sufficient to equalize the accumulated electron and ion currents which pass to the electrode during each complete cycle. The magnitude of the voltage bias will approach the zero-to-peak voltage of the applied rf power, with values that are typically several hundred volts. Accordingly, ions passing across the sheath to the surface will accumulate sufficient energy to cause sputtering. This is the basis of the rf sputtering method.^{59,128,193-197}

Table 8: Typical Nonconducting and Semiconducting Compounds Which Have Been Deposited by RF Sputtering

Target	Sputtering Source	Reference
SiO ₂	planar diode	176
	planar magnetron	177
Al ₂ O ₃	planar magnetron	178
	cylindrical magnetron	179
Nb ₂ O ₅	planar diode	180
Ta ₂ O ₅	planar diode	181
ZnO	planar diode	182
In ₂ O ₃ /SnO ₂	planar diode	183
LiNbO ₃	planar diode	184
BaPb _{1-x} Bi _x O ₃	planar diode	185
Si ₃ N ₄	planar diode	186
SiC	planar diode	187
Si	planar diode	188
GaAs	planar diode	189
(Hg, Cd) Te	triode	190
CdSiAs ₂	planar diode	191
PvZrO ₃ -PbTiO ₃	planar magnetron	192

Figure 36 shows a schematic drawing of an rf planar diode sputtering system. The target is attached to one electrode and the substrates are placed on the other one. The electrodes reverse cathode-anode roles on each half cycle. The discharge is operated at a frequency that is sufficiently high so that the ion charge accumulation during each cycle is not large enough to significantly influence the voltage.^{53,57} Frequencies greater than about one MHz are required. Most apparatuses are operated at a frequency of 13.560 MHz, since this is the frequency in the 10 to 20 MHz range that has been allocated by the Federal Communications Commission for industrial-scientific-medical purposes. Operation at another frequency will require careful shielding to assure compliance with FCC regulations on radio interference.

At MHz operating frequencies the massive ions tend not to follow the temporal variations in the applied potential; however, the electrons do. Thus the cloud of electrons that constitute the electron component of the glow discharge plasma can be pictured as moving back and forth at the applied frequency in a sea of relatively stationary ions. As the electron cloud approaches one electrode, it uncovers ions at the other electrode to form a positive ion sheath. This sheath takes up nearly the entire applied voltage, the same as in the dc case. A large electron current flows to a given

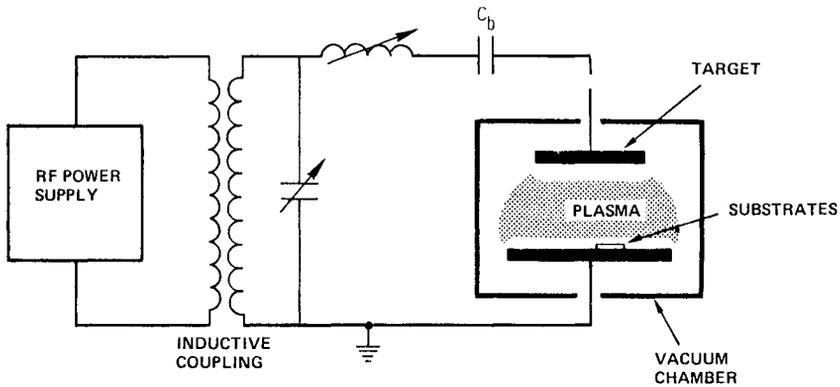


Figure 36: Schematic drawing of rf planar diode sputtering system with single-ended drive showing impedance matching network.

electrode as the electron cloud makes contact. Thus the electron cloud need approach a given electrode for only a small fraction of a half cycle for purposes of supplying sufficient electrons to fulfill the anode requirement; i.e., to balance the entire ion flux through the cycle. Accordingly, in the steady state both electrodes develop a negative dc bias relative to the plasma potential (become anodes) only for very short portions of their rf cycle,⁵⁹ as discussed previously. Because of their inertia, the motion of the ions can be approximated as if they follow the dc potential and pass to both electrodes throughout the cycle. The electron cloud spends most of its time near the center position between the electrodes. Visually, the discharge looks like a dc discharge with a cathode dark space over each electrode. Functionally, sputtering occurs very much as in the dc case, but at both electrodes.

RF discharges in planar diode apparatuses can be operated at considerably lower pressures than can dc discharges. Typical operating pressures are 5 to 15 mTorr (0.67 to 2 Pa). There are several reasons for this. Collisional interactions between the oscillating electron cloud and the working gas provides an energy exchange mechanism, not present in the dc case, which can contribute to the production of ionization. Secondary electrons capable of producing plasma ionization are generated at two electrodes, rather than just one. Finally, the positive space charge sheaths, which are present at both electrodes during most of the rf-cycle, tend to prevent the loss of at least the low and modest energy electrons.

Effective power transfer from the rf power supply to the plasma discharge requires that the load impedance be adjusted to match the output impedance of the power supply. Most 13.56 MHz power supplies are designed to operate into 50 ohm resistive loads. The impedance introduced by the plasma discharge and the sputtering target is primarily capacitive. In fact, the development of an rf self bias on the target requires

that there be no dc current flow. Thus a “blocking capacitor” (C_b in Figure 36) is generally added, to make the system insensitive to variations in the target capacitance, or to allow metal targets to be used.⁵⁷ A matching network is used to match the capacitive impedance of the load to the power supply. A typical matching network consists of a variable inductor in series with the load, and a variable capacitor in parallel with it, as shown in Figure 36. These elements provide the two degrees of freedom necessary to convert any load to a specific value of series resistance.^{133,134} In combination with the load, the matching network forms a resonant circuit.¹⁹⁸ When the load is matched, the circuit is in a state of resonance, with large circulating reactive currents. The matching network should therefore be placed as close as possible to the discharge chamber to avoid excessive power losses from these circulating currents. Many commercial sputtering sources monitor the “reflected power” from the load as an index to how effectively the matching network is adjusted. The reflected power should be minimized. Often this is done automatically. RF electrical systems are discussed in References 133, 134, 135, 195.

It has been noted in the above discussion that both electrodes in an rf planar diode apparatus have positive space charge sheaths throughout most of the rf cycle and that sputtering tends to occur continually at both. Thus, referring to Figure 36, an asymmetry must be built into the system so that sputtering occurs at one electrode and deposition at the other. This is usually done by making the substrate or “counter” electrode much larger than the target electrode. The counter electrode is often grounded as shown in Figure 36. Systems of this configuration, with the power supply connected between a “powered” or “driven” target electrode and a grounded chamber or counter electrode, are said to be “single-ended”. Sputtering systems are occasionally configured in a double-ended arrangement, where two identical target-mounting electrodes are positioned to sputter material onto electrically floating or grounded substrates.¹⁹⁶ Such systems are stabilized by placing the center tap of the rf power supply inductive coupling (shown in Figure 36) at ground potential. Since the center tap is at zero potential relative to the rf voltage, no rf currents can flow to grounded elements.⁵⁹

An approximate theoretical analysis¹⁹⁷ predicts that the voltage division between the sheaths on two planar electrodes of areas A_1 and A_2 will obey the relationship $V_1/V_2 = (A_2/A_1)^{1/2}$. The analysis is based on the assumptions (1) that the current densities to both electrodes are uniform and equal, and proportional to $V^{3/2}/d^2$ (Child Langmuir Law where d is sheath thickness); (2) that the sheath capacitances are proportional to A/d ; and (3) that the rf voltage is capacitively divided between the two sheaths so that $V_1/V_2 = C_2/C_1$.⁵⁷ However, experimental results suggest that for most real systems the relationship is closer to $V_1/V_2 \sim A_2/A_1$, and that special care must be exercised to minimize the sputtering on counter electrode surfaces.¹⁹⁹ A recent and excellent discussion of the influence of the electrode areas is given in Reference 200.

It is important to realize that the asymmetry of an rf discharge system is not influenced by the fact that one electrode is grounded. The variation in potential between two identical equal-area electrodes on a single-ended

rf system is shown in Figure 37A. The plasma potential, V_p , tends to be slightly more positive than that of the most positive surface in contact with the plasma.¹³² Thus the plasma potential relative to ground varies with time such that the time-average potential difference between the plasma and each electrode is equal, and equal sputtering occurs at both electrodes. In the case of unequal electrode areas shown in Figure 37B, the average sheath voltage variations (potential variations between the plasma and the electrodes) are unequal, as discussed in the previous paragraph. Typically, the grounded electrode is made the larger one, with a size sufficient so that the sheath voltage drop is below the sputtering threshold. Consequently, the plasma potential remains close to ground, and the self-bias discussed previously assumes the form of a near-dc voltage offset with respect to ground on the smaller electrode as shown in Figure 37B. This dc offset voltage is generally measured with an oscilloscope and is one of the primary parameters which is used to specify deposition conditions in rf sputtering. In some cases circuit elements are placed between the substrate electrode and the grounded end of the power supply, so that bias sputtering bombardment can be induced at the substrates.^{133,134,135}

The above discussion has been presented in the context of planar diode sputtering systems. Magnetron sputtering sources can also be operated with rf power. However, some problems are encountered.¹⁵⁴ Magnetron sputtering technology is basically a dc concept with specific configurations and orientations of the cathode and anode with respect to the magnetic field. Effective double-ended rf magnetrons can be provided for some geometries. These configurations provide independent traps for both electrodes but allow magnetic coupling between them so that the electrons leaving one trap can diffuse freely to the vicinity of the other.¹⁵⁴ However, most magnetron configurations must be driven with single-ended arrangements. These magnetrons operate in hybrid modes with current-voltage characteristics which are not symptomatic of true magnetron behavior. This can be seen in Figure 28 by comparing the dc and rf planar magnetron I-V characteristics. RF driven magnetrons yield deposition rates that are typically a factor of three greater than are obtained with rf planar diodes, but far below the factor of twenty-to-thirty improvement in deposition rate which dc magnetrons provide over planar diodes when sputtering metals.

It should be noted that the performance of rf-driven planar diode and magnetron systems for sputtering poorly conducting materials is limited not by the capabilities of the plasma, since high rates have been achieved with planar diodes²⁰¹, but by the power densities that can be passed into the target without causing damage. Ring type magnetrons, such as the planar magnetrons shown in Figures 27D, G, H, and I, are particularly vulnerable in this respect because the power input is concentrated in the region of the plasma ring. Therefore, RF driven magnetrons are generally used not to provide high deposition rates, but to achieve large area deposition and to reduce substrate electron bombardment and heating. Special precaution should be taken in magnetron systems to avoid unwanted sputtering at counter electrodes, since the magnetic confinement produces gradients in plasma density which can yield sheaths with elevated voltage drops at the counter electrodes.

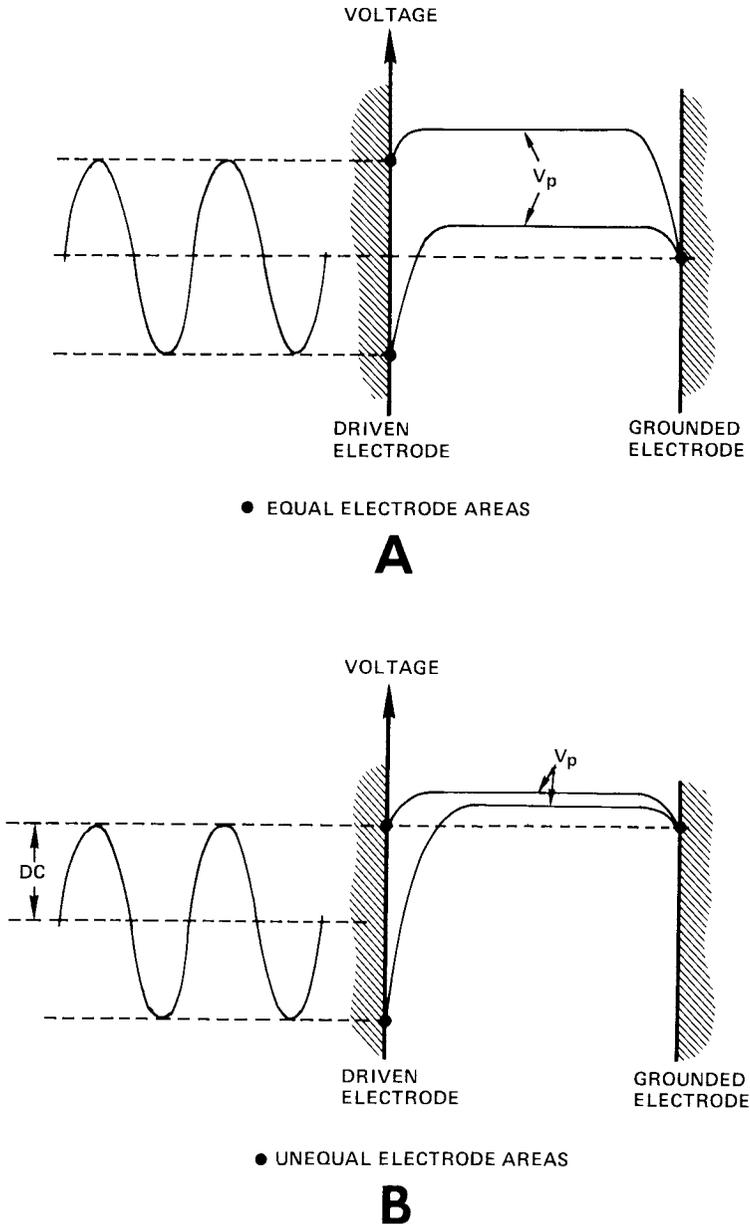


Figure 37: Schematic representation of the potential variation between the electrodes of rf glow discharges with equal electrode areas (A) and unequal electrode areas (B). A nonconducting target or blocking capacitor is assumed to be present in the unequal area case so that unequal space charge "sheath" voltage drops develop.

5.9 Reactive Sputtering

Reactive sputtering is that process where at least one of the coating species enters the system in the gas phase. Thus reactive sputtering is similar to the process of reactive evaporation which was discussed in Section 3.6. Examples of some reactive sputtered compounds, that are important for electronic-related applications, are listed in Table 9.

The advantages of reactive sputtering are that (1) many complex compounds can be formed using relatively easy-to-fabricate metallic targets; (2) insulating compounds can be deposited using dc power supplies, although rf power is sometimes used; and (3) coatings with graded compositions can be easily formed. The difficulty in the reactive sputtering process is the complexity which accompanies its versatility.

Reactive sputtering differs from reactive evaporation in that, in addition to the reactions which occur at the substrate, reactions can also occur on the target surface, following which the reacted material is sputtered. At high working gas pressures reactions can occur in the gas phase, but this situation is unusual.

When sputtering with a reactive-gas/argon mixture, one should be aware that the relationship between the film properties and the reactive gas injection rate is very nonlinear, and strongly dependent on the apparatus geometry and the history of its operation. There are two reasons for this. First, the condensing film can be considered as an additional pump for the reactive gas, and the sticking coefficient or speed of this getter pump

Table 9: Examples of Compounds Which Have Been Deposited by Reactive Sputtering

Target	Reactive Gas	Film	Sputtering Source	Reference
Al	O ₂	Al ₂ O ₃	planar magnetron	202
Ti	O ₂	TiO ₂	ion beam	203
			planar magnetron	204
Ta	O ₂	Ta ₂ O ₅	planar magnetron	204
Cu	O ₂	Cu ₂ O	planar magnetron	205
Zn	O ₂	ZnO	planar magnetron	206
In-Sn	O ₂	In ₂ O ₃ /SnO ₂	planar magnetron	207
			cylindrical magnetron	208
Al	N ₂	AlN	planar magnetron	209, 210
Ti	N ₂	TiN	planar magnetron	211, 212
Zr	N ₂	ZrN	planar diode	213
Si	N ₂	Si ₃ N ₄	planar diode	214
Cu	H ₂ S	Cu ₂ S	cylindrical magnetron	215
Cd	H ₂ S	CdS	cylindrical magnetron	216
Cu-In	H ₂ Se	CuInSe ₂	planar magnetron	217

depends in a complex way on the growth rate, composition, structure, and temperature of the growing film. Second, reactions at the target surface can change the rate at which the metal species are sputtered. In the case of diode systems, where the target is also the cathode, target surface reactions can also influence the plasma discharge.

The overall performance of a reactive sputtering system is dependent on the deposition area compared to the size of the chamber, and on the getter pumping effect of the sputtered flux compared to the speed of the physical pumps. These relationships are identical in principle with those discussed in Section 2 from the standpoint of wall outgassing. See Figure 3. Thus if the getter pumping effect dominates, the reactive gas injection rate is a particularly relevant parameter. This is the situation in most production coating systems. If the physical pumping dominates, the reactive gas partial pressure is the more useful parameter. It should be noted that most papers on reactive sputtering describe the process in terms of the reactive gas partial pressure prior to ignition of the plasma discharge. This pressure is proportional to the flow rate and is not the true partial pressure during the reactive sputtering.

Figure 38 shows the composition dependence of the sticking coefficient for N_2 incident on a growing Ti film.²¹⁸ As the number of gettered N_2 molecules per Ti atom approaches 0.5 (stoichiometric TiN), the sticking coefficient drops by more than two orders of magnitude. This is a general trend which applies for all materials, because the number of unoccupied surface absorption sites decreases as a stoichiometric composition is approached. Thus in the Si/ O_2 reactive evaporation case discussed in Section 3.6, an oxygen impingement rate about two orders of magnitude larger than the Si impingement rate was required to form SiO_2 coatings. See Figure 14.

In reactive sputtering the relatively high reactive gas partial pressures that are needed to produce the required high impingement rates at the substrates also generally cause high impingement rates at the target. The target reactions are stimulated by the fact that many of the incident ions are reactive gas species. In addition, active radicals, such as dissociation products, reach the target from the adjacent plasma where they are produced. Consequently, target surface compounds with reduced sputtering yields can form, as discussed in Section 5.4. These target reactions can also limit the range of coating compositions that can be achieved.

The general concepts discussed above are illustrated in Figure 39. The figure shows the variation in discharge voltage and deposition rate with O_2 injection rate, at a constant discharge current, for Cr- O_2 reactive sputtering in a cylindrical magnetron sputtering system in which the getter pumping capacity of the sputtered flux is large compared to the physical pumping system.⁶ At low O_2 injection rates virtually all of the O_2 is getter pumped by the condensing Cr coating. Consequently, the O_2 partial pressure remains relatively low, and the cathode process remains primarily one of simple Ar sputtering of Cr. The coatings deposited under these conditions are metallic in nature, but possess an oxygen content that increases almost linearly with the O_2 injection rate and may exceed the equilibrium solubility of oxygen in Cr. As the O_2 injection rate approaches

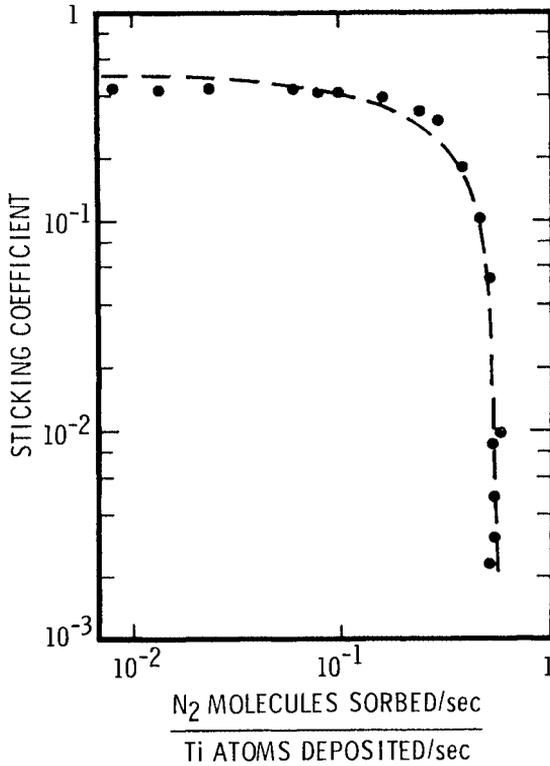


Figure 38: Sticking coefficient measured during the continuous deposition of titanium as a function of the ratio of the gettered nitrogen flux to the titanium deposition flux. Data from Reference 218.

that required to produce a stoichiometric chromium oxide, the O_2 partial pressure rises because of the reduced getter pumping rate, analogous to the example shown in Figure 38. As a consequence, the target cathode develops a surface oxide, and the sputtering process undergoes an abrupt transition into a mode in which the metal sputtering rate, and therefore the reactive compound deposition rate, are reduced. Beyond this transition, coatings of Cr_2O_3 are deposited. In some cases the transitions are very abrupt while in other cases they appear as a long-term, tens of minutes, drift in operating parameters. The reactive gas injection rate which produces the transition tends to increase linearly with the discharge current. The phenomenon of forming such a surface layer on the target is often referred to as target poisoning.

The nature of the surface layer on a poisoned target depends on the metal-gas combination and the reactive sputtering conditions. The reduction in sputtering rate results because the atom binding energies are often higher in these surface layers and because the mass of the reactive gas ions may make them less effective than Ar as sputtering agents. In addition,

the surface layers often have higher electron secondary emission coefficients, which result in a reduction in both the discharge voltage, as seen in Figure 39, and the ion component of the discharge current, for discharges driven at constant currents. Thus in the case of a Au cathode, where no surface oxide forms, the discharge voltage and the deposition rate are not significantly influenced by the O_2 injection rate. See Figure 39. Very pronounced losses in deposition rate are seen for oxide reactive sputtering of materials such as Al (36X), Ti (14X), and Cr (4X), which form surface compounds with high interatom binding energies and increased electron

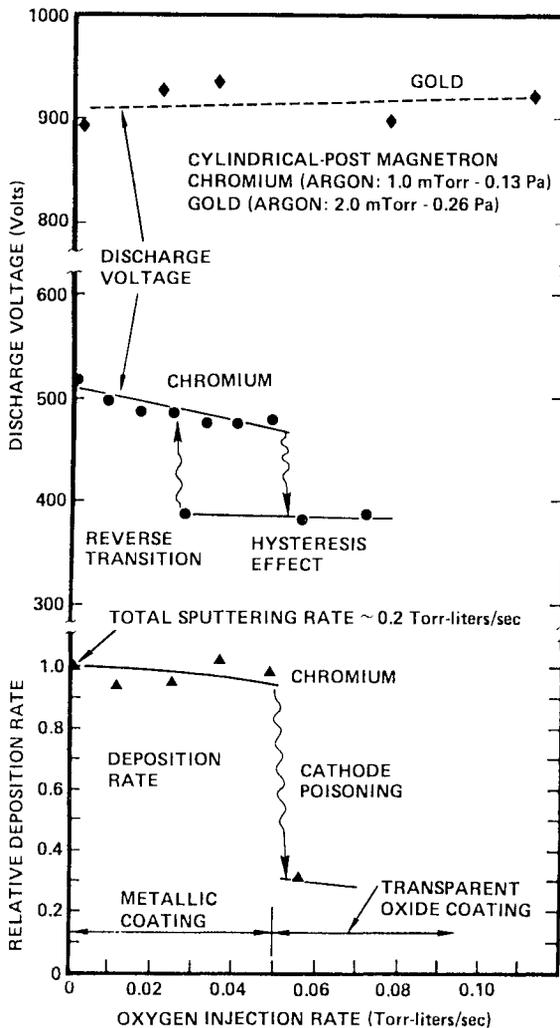


Figure 39: Transition in steady state operating mode of cylindrical-post magnetron sputtering source because of cathode poisoning. Data from Reference 6.

secondary emission coefficients.¹⁵⁴ The decrease in deposition rate is generally less for materials such as nitrides (2X for NbN) and sulfides. The decrease in deposition rate which occurs with poisoning is relatively independent of the cathode geometry.¹⁵⁴

The cathode surface processes in the poisoned mode produce an energetic flux of highly reactive gas atoms and molecular fractions (reflected atoms, sputtered chemisorbed atoms and sputtered surface compounds) which accompany the sputtered metal atoms to the substrates. It is this large flux of reactive species that promotes the formation of stoichiometric coatings and makes the reactive sputtering process so effective for producing a wide range of compounds. Thus sputtering sources can often be operated in the poisoned-target mode to produce stoichiometric coatings under conditions that are not particularly sensitive to the reactive gas injection rate.

Once formed, the target surface layer will remain until the working gas is made sufficiently lean in the reactive species so that a net sputter removal of the layer can occur. This is the origin of the hysteresis effect, which is shown in Figure 39 for the discharge voltage, but which also applies to the deposition rate. These target surface layers are often very similar to the materials being deposited, with thicknesses that depend on the compound and the sputtering conditions, i.e. current density and reactive gas partial pressure, but which are typically in the range from 5 to 100 nm.^{202,219,220} The hysteresis effect can therefore be used as a method for controlled and reproducible deposition of ultrathin metal oxide layers.^{220,221} Typical uses are for metal-insulator-semiconductor or Josephson tunneling junctions. The method consists of first operating the sputtering source under given conditions for sufficient time to generate a poisoned-target surface layer of given thickness, with the substrates shielded. The target is then sputtered in pure Ar with the substrates exposed. The surface layer is therefore sputtered from the target and deposited onto the substrates, with a smooth *in situ* transition being made from the ultrathin compound to the metal as the target surface is returned to the pure metal state.

The poisoning effect introduces two practical problems. One is the loss in deposition rate. The second is that during the poisoning transition the material being deposited often passes abruptly from a metal to a nearly full-stoichiometric compound. Intermediate materials such as suboxides therefore become difficult to deposit. The losses in deposition rate can be very large (X 30 for Al, as noted previously). Often coatings with the desired stoichiometry and physical properties are formed right at the transition. Examples are NbN coatings with the maximum superconducting transition temperature,²²² $\text{In}_2\text{O}_3/\text{SnO}_2$ coatings with the optimum combination of optical and electrical properties,²⁰⁷ and VO_2 coatings which exhibit a maximum resistivity change in the thermally driven semiconductor-metal transition.²²³

Consequently, considerable effort has been directed toward developing methods for operating sputtering sources right at or very near to the transition point. For many years these attempts met with little success. However, considerable progress has been achieved over the past few

years using (1) computer driven control systems, and/or (2) apparatus configurations which provide the maximum possible separation between the target sputtering process and the film growth reactions at the substrates. Ring type plasma devices such as planar magnetrons (Figure 27H) have proven to be particularly effective, as compared to uniform current density devices. The higher current density under the plasma ring makes these devices capable of operating at higher reactive gas partial pressures without the occurrence of target poisoning. Target surface layer formation begins at the edges of the sputtering region underneath the plasma ring and closes in on the region as the reactive gas partial pressure increases.²²⁴

The difficulty of the control problem depends on the nature of the surface layer.²²⁵ Thus transition point control is generally more difficult for oxide than nitride deposition. Figure 40 summarizes the performance of a planar magnetron source which was used to deposit AlN coatings by sputtering an Al target in an Ar/N₂ working gas.^{209,210} A microprocessor was used to provide feedback control to parameters such as the N₂ flow rate and the discharge current, voltage, and power. The figure shows data for cases in which (a) the flow rate was controlled at a constant power, (b) the power was controlled at a constant flow rate, and (c) the voltage was controlled at a constant flow rate. Cases (a) and (b) yielded abrupt cathode poisoning transitions and hysteresis effects, with metallic films formed from F-D and nitride films from A-C. The voltage hysteresis in case (a) is essentially identical to that shown in Figure 39, which corresponds to similar operating control. In case (c) the computer adjusted the current to maintain the voltage at the programmed value. No abrupt transitions or hysteresis effects were encountered. Stable operation was continuously obtained at all degrees of target surface coverage. The effectiveness of voltage control has also been found in depositing VO₂ coatings with large semiconductor-metal resistivity transitions.²²⁶

In most cases of oxide deposition, special efforts must be made to create controlled conditions at the cathode and substrates. The most common approach is to use one or more of the following: (1) a planar magnetron with a high current density as the sputtering source, (2) a gas baffle surrounding the target or substrate to increase the reactive gas partial pressure gradient between the target and substrate regions, and/or (3) an auxiliary plasma discharge at the substrate to enhance the reaction by creating active species.^{226a}

Figure 41 shows a schematic drawing of an apparatus configuration with a reactive gas baffle. The sputtered metal flux that condenses within the baffle housing acts as a getter pump which reduces the reactive gas partial pressure at the target. The flow impedance introduced by the baffle permits maintenance near the substrate of reactive gas partial pressures which are several times the values for the case without a baffle. This is shown by the data given in Figure 42a. The transition to an oxidized target occurs at a lower flow rate when a baffle surrounds the target, compare Figures 42b and 42c, apparently because the baffle reduces the pumping speed provided in the vicinity of the target by the combination of the getter effect and the vacuum system diffusion pump.²²⁷

The exact reaction mechanisms that are promoted at the substrate by

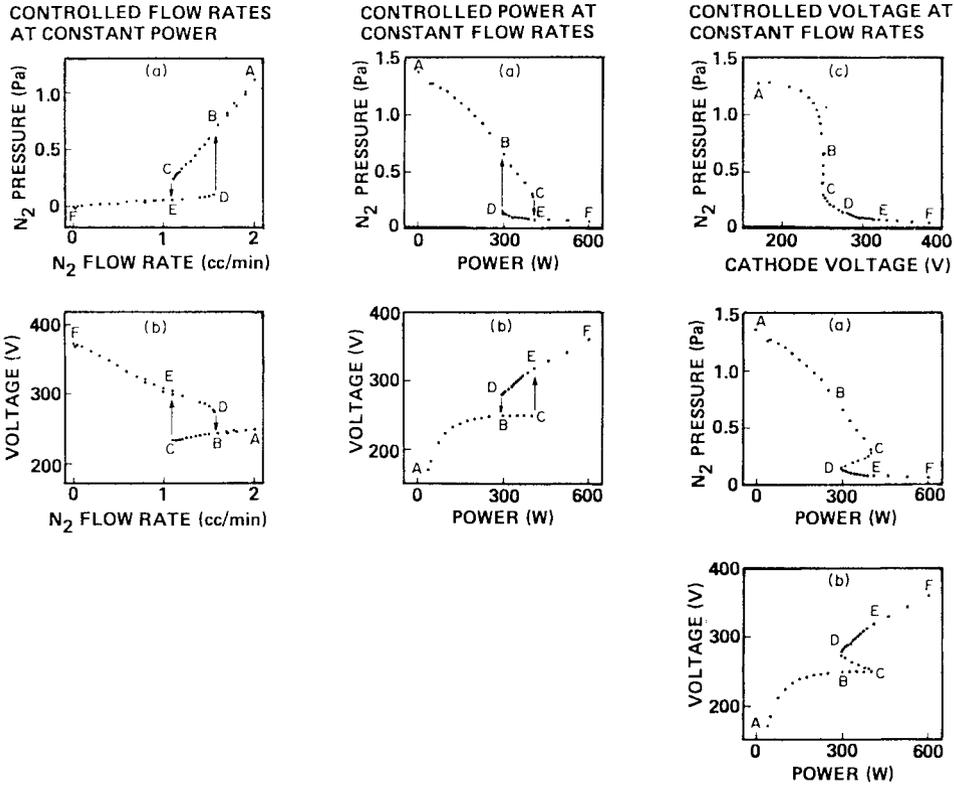


Figure 40: Variation of Al/N₂ reactive sputtering parameters during various modes of computer controlled operation using planar magnetron sputtering source. Data from Reference 209.

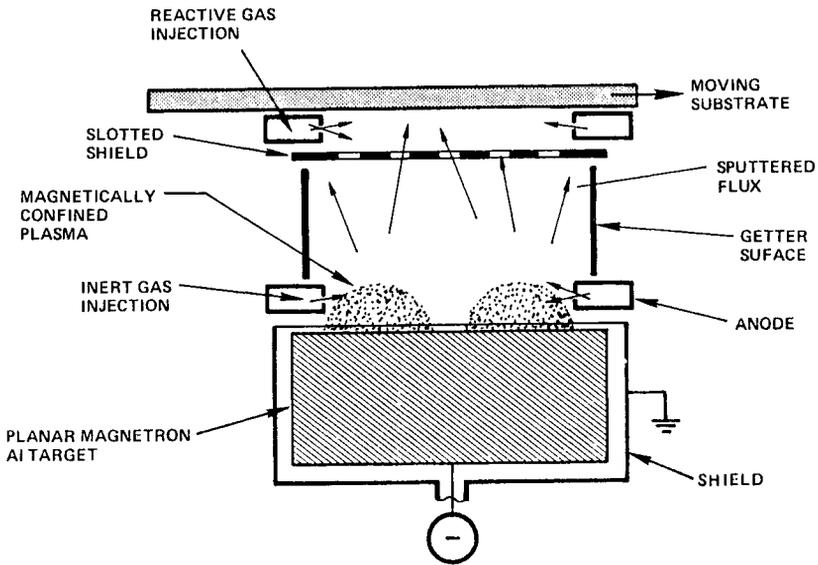


Figure 41: Schematic drawing of planar magnetron reactive sputtering apparatus with gas baffle to assist in isolating the reactive processes at the target and the substrate.

a plasma discharge are still a subject of research.²²⁸ However, the role of such discharges in reactive sputtering is undoubtedly very similar to their role in the plasma-assisted reactive evaporation case discussed in Section 3.6. Auxiliary discharges are obviously not required in apparatuses such as planar diodes, which place the substrates within the sputtering plasma. However, a substrate bias may be used in such cases.

The baffle reduces the metal flux reaching the substrate. However, this reduction is more than compensated for, if films of the desired composition can be deposited in the absence of cathode poisoning. Thus reactive sputter deposition of Ta_2O_5 at 18 nm/sec,²²⁴ TiO_2 at 11 nm/sec,²²⁴ and Al_2O_3 at 2.5 nm/sec²²⁹ has been achieved using this method. The use of baffles and auxiliary substrate discharges has proven very useful in depositing transparent conducting oxide coatings with optimum combinations of optical transmission and resistivity on low temperature substrates. Examples are ZnO ,²²⁷ Cd_2SnO_4 ,²²⁹ and In_2O_3 doped with Sn.²⁰⁷ In these cases it was necessary to operate very close to the poisoning transition point. Control of the discharge is often facilitated in such cases by a voltage maximum which forms at reactive gas injection rates just below the values that produce poisoning, because the target compound layer closes in on the plasma ring.^{224,230} The baffle technique has also been used to deposit high quality Al_2O_3 optical waveguides by reactive sputtering at reasonable deposition rates.²³¹

In considering the reactive sputtering process it is important to recognize the importance of the overall apparatus surface chemistry in achieving

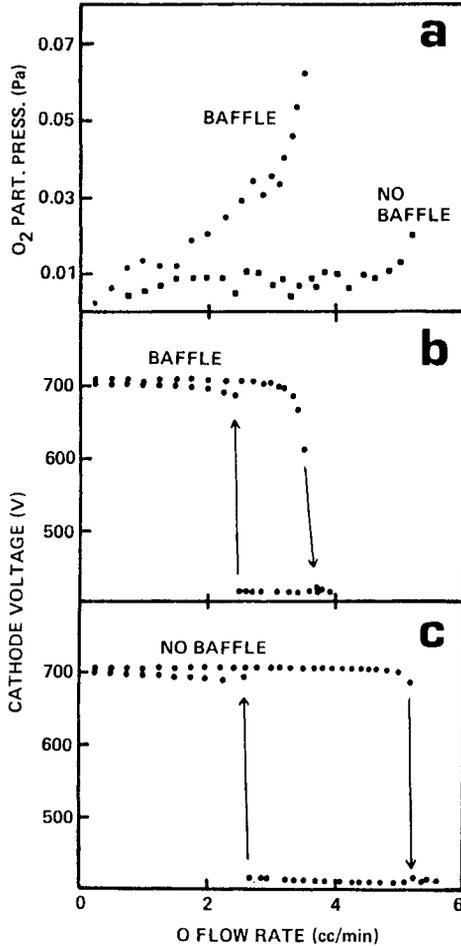


Figure 42: Oxygen partial pressure and cathode voltage vs oxygen flow rate for Zn-O₂ reactive sputtering, both with and without a baffle surrounding the target. Argon partial pressure was 0.74 Pa. Data from Reference 227.

steady state operation. Clear evidence has been reported, for both Ta/O₂²²⁰ and Cu/H₂S²³² reactive sputtering, which shows that the equilibration time for steady state operation at a given reactive gas injection rate depends on the initial state of the substrate and chamber wall surfaces as well as the surface of the target. Equilibration times of several ampere-hours of target operation have been encountered.²¹⁷ Consider the following case history. A series of experiments were conducted to develop a reactive sputtered coating. In each experiment a few test substrates were positioned surrounding a cylindrical-post magnetron source of the type shown in Figure 29. Once operating conditions which yielded the desired coating properties had been determined, a trial production run was made with the entire

chamber filled with substrates. Deposition under the prescribed conditions in this run produced coatings with very different properties than had been obtained with the test substrates. This occurred because the load of fresh substrates had significantly changed the state of the overall chamber deposition surfaces. Reliable production deposition conditions were subsequently established in a series of experiments in which a fresh load of substrates was introduced for each deposition run.

The problem of low substrate sticking probabilities for many reactive gases can also affect the composition of coatings deposited by sputtering from compound targets. Thus, as discussed in Section 5.5, coatings sputtered from oxide targets are often found to be deficient in oxygen, because oxygen released from the target by sputtering or thermal decomposition is not incorporated into the growing coating and is lost in the chamber pumps. A common practice in such cases is to introduce “make-up” oxygen in the gas phase. This, then, is another form of reactive sputtering.

Finally it should be noted that when a batch processing chamber is brought up to air for the purpose of introducing substrates, a thin oxide layer will form on the target. Experiments with a Cr cathode indicate that this layer is very similar in its sputter-removal characteristics to the one that forms during reactive sputtering in the poisoned mode.⁶ Thus a target that has been exposed to the atmosphere must undergo a target clean-up “reverse” transition similar to that shown in Figure 39 before metal coatings can be deposited. This requires that the chamber be pumped for sufficient time so that the outgassing flux entering into the plasma and deposition region is less than the critical flux for permitting the reverse transition at the target current density being used. If the outgassing flux exceeds the critical value, the target will not “clean up,” and a dielectric coating will be deposited. For materials such as aluminum, the critical flux is relatively low. This has led inexperienced workers, using devices operating at low cathode current densities, to conclude that aluminum cannot be sputtered.

5.10 Target Fabrication

Targets can be formed by an almost unlimited variety of methods, ranging from vacuum casting of metals, to hot pressing of powders, and to forming composite targets by placing wires, strips, or discs of one material to cover a portion of the surface of another material. However, caution must be exercised. Poor coating quality or consistency can often be traced to an inadequate target.

Target cooling is important, particularly at the high current densities used with magnetrons. If the target material is a metal of good mechanical integrity, direct water cooling can usually be provided on the rear surface. Mechanically weak or nonconducting targets must be attached to a metal backing structure which serves as the vacuum seal and, for nonconducting targets, as an electrode. The attachment is usually made with a low-temperature metal or alloy, or a conducting epoxy. If target purity is critical, it is important to assure that the attachment medium does not diffuse into and contaminate the target material. In the case of nonconducting targets, the sputtering rate is limited by heat transfer within the target.²⁰¹ In some

planar magnetron designs the backing plate is a permanent part of the cathode structure. Ring type metal targets for gun-type magnetrons (Figure 271) are designed to make contact with the support structure by undergoing thermal expansion.

A number of suppliers offer a wide selection of high purity metal and compound targets, particularly in the disk form used by planar diode and small ring type magnetron sputtering sources. The compound targets are generally formed from powders by hot pressing. As noted previously, hot-pressed targets are particularly vulnerable to contamination. It has, in fact, been recommended that sintered targets be avoided if coatings with bulk-type properties are desired.²³³ For critical applications it is wise to measure the composition of new targets prior to placing them in service. It is also important to note that the purity levels quoted for most metal targets do not generally include gaseous impurities such as oxygen.

In the cylindrical magnetron case, targets of common engineering materials can be easily fabricated from commercially available tube and rod stock. The tubular form possesses sufficient strength so that high purity targets have been cast using materials as soft as Cd. Very soft metals such as In, or compounds such as CdS, are generally attached to tubular sections that contain the vacuum seals.

Reference 234 contains lists of suppliers of both sputtering equipment and sputtering targets as well as general vacuum equipment and evaporation systems.

6. THIN FILM GROWTH AND PROPERTIES

6.1 Coating Nucleation and Growth

The literature on nucleation and growth is very extensive. The discussion here is necessarily selective and brief. The reader is encouraged to consult References 95 and 235-240 for more detailed reviews.

6.1.1 Condensation. The two main processes involved in condensation are the arrival of the coating atoms, or molecules, at the substrate surface and the motion of these atoms on the surface. Coating atoms incident on the surface of a substrate or growing coating can (1) bounce off the surface, (2) adsorb for a finite time, or (3) adsorb and stick permanently. The probability of reflection is low for most practical cases of interest. Thus, even energetic sputtered atoms will generally transfer sufficient energy to the substrate on their initial encounter so that they become loosely bonded species known as adatoms.⁹⁵ The adatoms diffuse over the surface, exchanging energy with the substrate atoms and other adsorbed species, until they either are desorbed, by evaporation or sputtering, or become trapped at low energy sites as indicated schematically in Figure 43.

Two coefficients are used to describe the condensation process. The accommodation coefficient provides a measure of the energy exchange process and is defined as

$$\alpha = (T_i - T_e)(T_i - T_s) \quad (12)$$

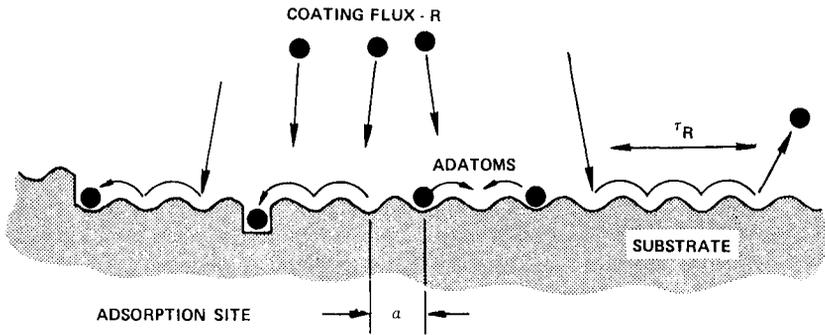


Figure 43: Schematic illustration of condensation process that occurs during film growth.

where T_i is an effective temperature describing the kinetic energy of the incident atoms, T_s is the substrate temperature, and T_e is the temperature of atoms evaporating from the substrate. For deposition by vacuum evaporation, and probably also for most cases of sputter deposition,⁹⁵ the accommodation process is so fast that there is negligible probability of an atom re-evaporating before it reaches equilibrium, and α can be taken as unity. The second coefficient is the sticking or condensation coefficient. It is defined as the fraction of incident atoms which adhere and remain on the substrate. Sticking coefficients are commonly less than unity, and in cases where the adsorption energy is low and/or the substrate temperature is high, can be very low. It should be noted that at high substrate temperatures atoms accommodated to the substrate temperature can have a high probability of loss by evaporation. Thus the sticking coefficient can be zero even though the accommodation coefficient is near unity.²⁴⁰

The mean residence time for desorption by evaporation depends on the adsorption binding energy, E_a , and substrate temperature, T (absolute degrees), and is given by

$$\tau_R = \tau_o \exp(E_a/kT) \quad (13)$$

where τ_o is a characteristic time approximately equal to the lattice vibration time, $\sim 10^{-13}$ sec.

Under steady state conditions, where the rates of adsorption and desorption are equal, an incident rate R of coating atoms per unit surface area per second leads to an adatom surface density n_A :

$$n_A = R \tau_R \quad (14)$$

The exponential term in Equation 13 makes τ_R and therefore n_A strongly dependent on the adsorption binding energy and substrate temperature. A typical deposition rate of 2 nm/sec (1200 Å/min) with unity condensation coefficient corresponds to a flux of about 10^{16} atoms/cm³.

sec for a representative material. The atom or site density on a metal surface is about 10^{15} atoms/cm². An E_a of 1 eV on a surface at 200°C leads to a residence time of about 5×10^{-3} sec and a flux of 10^{16} atoms/cm²-sec yields an equilibrium adatom density that covers about 5% of the surface sites. By contrast, an E_a of about 0.15 eV (typical of Cu on clean glass) on a surface at 200°C leads to a τ_R of about 5×10^{-12} sec and an equilibrium surface density of only about 5×10^4 atoms/cm². These calculations neglect the interactions between the adatoms themselves and between the adatoms and the surface sites (see next section). They are given simply to provide a feeling for the magnitudes involved and the importance of E_a and T .

6.1.2 Nucleation. As the coating process progresses, the condensed adatoms must combine to form a coating. This nucleation process is controlled by three parameters: E_a , the adsorption energy between the adatoms and the substrate or growing coating; E_d , the surface diffusion energy required to transfer an adatom to an adjacent adsorption site; and E_b , the binding energy between coating atoms. Actually there are a set of E_b values depending on the degree to which the coating atom in question is surrounded by, and bonded to, other coating atoms. In the following generalized discussion, E_b will refer to the overall set of binding energies unless specifically stated otherwise.

Thus the evaporation rate from a condensed deposit is given by

$$R_e = \Psi_e \exp(-E_{bb}/kT) \quad (15)$$

where E_{bb} is the sublimation or average surface atom binding energy for the bulk condensate and Ψ_e is a rate constant (typical units-atoms/cm²-sec).

The adatom site-to-site hopping frequency is given by

$$\nu = \frac{1}{\tau_o} \exp(-E_d/kT) \quad (16)$$

The number of sites visited by an isolated adatom during the residence time, τ_R , that it is adsorbed on the surface is therefore

$$n_s = \tau_R \nu \quad (17)$$

A random-walk diffusion distance before desorption can be written as

$$x = a_o(n_s)^{1/2} \quad (18)$$

where a_o is the mean distance between adsorption sites (see Figure 43). However, it is important to realize that the adatoms do not necessarily move uniformly over a real surface, since the migration rates are dependent on the substrate crystallographic directions and surface topography.²³⁹

The collision rate between adatoms migrating over the surface is given by²³⁷

$$J \propto (n_A)^2 \nu = R^2 \tau_o \exp[(2E_a - E_d)/kT] \quad (19)$$

Various regimes can be defined in terms of the parameters listed above. If one assumes an ideal substrate with N_o (cm⁻²) absorption sites of

equal energy, then for $R > N_o v$, the arriving atoms stay essentially where they arrive.²³⁹ If the atoms have directional bonds the layer will tend to be amorphous. For materials with nondirectional bonds, such as metals, this condition will lead to very fine-grained polycrystalline films. This regime of low adatom mobility is encountered at low substrate temperatures.

At elevated substrate temperatures and low R , such that $R < R_e$, no film buildup can occur, although a monolayer of coating material may form on the substrate for the case where $E_a > E_b$. The most common regime of film growth ($R > R_e$) is one in which the substrate temperature is high enough to produce considerable surface diffusion. The substrate is usually different from the coating material, so that E_a is considerably different from the various values of E_b . If $E_a > E_b$, the first arriving atoms will condense as a single monolayer. If $E_a < E_b$ and τ_R is small, growth can occur on a uniform substrate only through adatoms combining to form nuclei, which will tend to be three dimensional as shown in Figure 44. There is said to be a nucleation barrier to growth.²⁴⁰ At low deposition fluxes, J (Equation 19) may be too small to permit nuclei formation. Thus one has the concept of a critical condensation flux to produce coating growth under given conditions.^{8,95} The stability of the nuclei depends on a balance between the surface and volume free energies.²³⁶ Thus nuclei that reach a critical radius, such that subsequent growth decreases the free energy, survive and can grow together to form grains as shown in the figure. The critical cluster size varies inversely with E_b .²³⁹ The condition $E_a < E_b$ is commonly encountered for metal condensing on insulators.

Most engineering substrates are characterized by a heterogeneous distribution of sites of preferred nucleation. On such substrates the collision rate of adatoms with these nucleation sites is generally higher than the collision rate among adatoms themselves [J in Equation 19]. Therefore, these sites tend to dominate the nucleation process. In fact at high temperatures or low coating fluxes, such that J is small, nucleation will essentially be impossible except at these sites. Thus, controlling the density and distribution of the nucleation sites is an important consideration in controlling coating properties such as grain size.

Once a continuous coating is formed, or when the condensate and substrate are a common material, E_a and the various E_b values will depend on the number of nearest neighbors at the atom site in question. However, E_a (isolated adatom) will always be less than E_b . For conditions that yield a low surface adatom density, n_A , the growth rate will be limited by the nucleation process as discussed above. For conditions that yield a high n_A , there will be no impediment to condensation, but the atoms will tend to group as clusters, since the atoms in the cluster have more nearest neighbors (higher E_b) than isolated adatoms.

All stages of the growth described above are greatly affected by impurity atoms. Thus, for example, the deposition of Cd on clean W represents a case where $E_a > E_b$ and layer growth occurs with no nucleation barrier. However, the presence of less than one monolayer of oxygen on the tungsten has been found to reduce E_a by a factor of more than two and to yield evidence of a nucleation barrier.²³⁹ Energetic particle bombardment, such as ion bombardment, can dramatically affect the nucleation

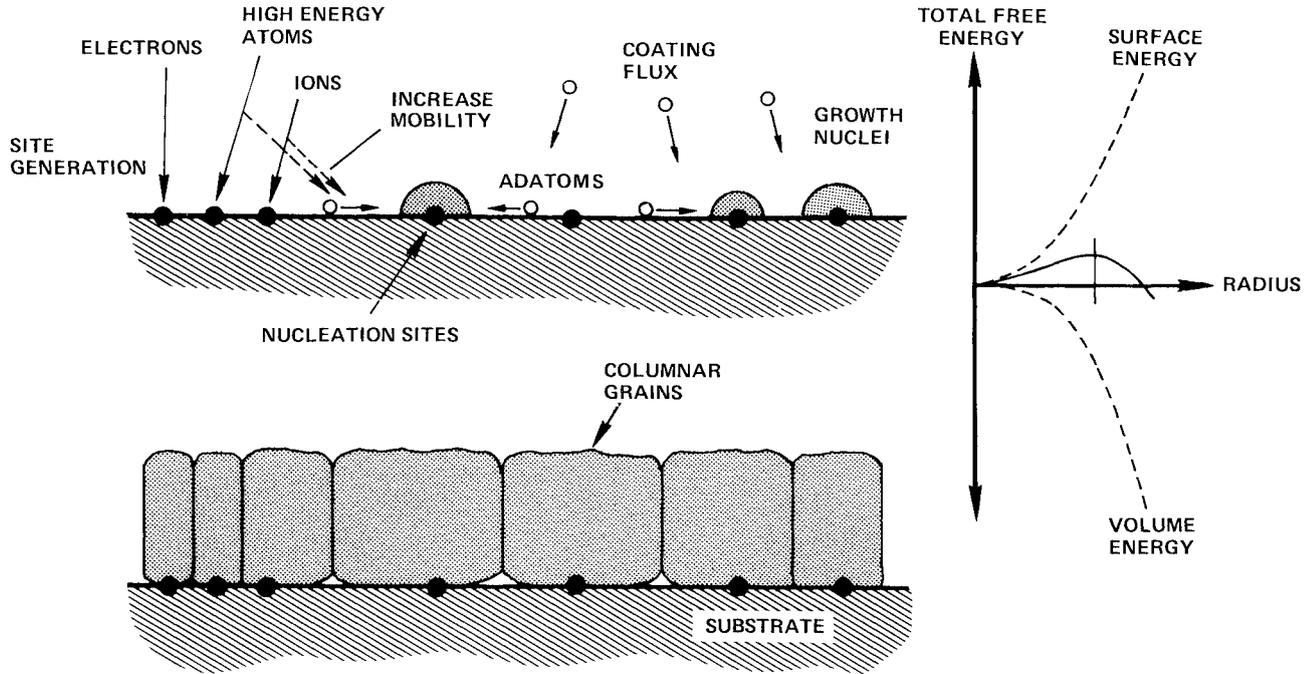


Figure 44: Schematic illustration of film nucleation and growth.

and growth process through the removal of impurity atoms, the creation of nucleation sites, and perhaps even by influencing the effective mobility of the adatoms and nuclei.^{38,241} (See Section 6.4)

A special case of importance for many device applications is the growth of single crystal films by the phenomenon of epitaxy.^{95,236,238,239} The term epitaxy means the oriented or single-crystalline growth of one material upon another, such that there exists a crystallographic relation between the overgrowth and the substrate.²³⁵ Epitaxy may be classed as either isoepitaxy where the coating and substrate are the same material, or heteroepitaxy, where the coating and substrates are different materials.

On single crystal substrates, crystallographic edges and steps of atomic dimensions offer preferred growth sites. Layer-by-layer growth from these sites yields epitaxial coatings. Growth can also occur via the nucleation of clusters of the type described previously. If the clusters can adjust to a common orientation during coalescence, a single crystal film can be produced.^{95,236,239} Epitaxial growth conditions involve clean surfaces, modest deposition rates, and elevated substrate temperatures. See Section 6.2. Such conditions are apparently required to suppress the formation of unoriented nuclei or growth at unfavorable sites. There is evidence that energetic particle bombardment improves epitaxy. Possible mechanisms include removing impurity atoms, causing a higher density of nucleating sites and hence a relatively slow growth rate of each cluster, and increasing the cluster mobility.²³⁹ However, this is a controversial subject of continuing research. See Section 6.4.

6.2 Evolution of Microstructure

Once a continuous coating is formed the growth continues, as described above, but with the surface of the growing coating serving as the substrate. Thus the properties of vacuum deposited coatings are determined largely by adatom surface diffusion processes which are made evolutionary by the way the state of the coating surface changes as the coating grows.²⁴¹ However, there are two additional processes that can affect the evolution and growth of the coating microstructure. These are bulk diffusion and atomic shadowing. Bulk diffusion affects the coating structure at elevated temperatures, because atoms incorporated into the coating can readjust their positions within the lattice by bulk diffusion processes. Any process that causes a systematic nonuniformity in the arriving coating atom flux over the substrate surface can have a drastic effect on the evolutionary growth process. Shadowing, a simple geometric interaction between the roughness of the growing surface and the line-of-sight directions of the arriving coating atoms, provides such an effect.^{241,242} The shadowing effect is most pronounced at low substrate temperatures. At higher temperatures the shadowing can be compensated for by surface diffusion; i.e., if the surface diffusion is large enough, the point of arrival of a coating atom loses its significance.

For many pure metals the adatom binding energies and the activation energies for both surface and bulk diffusion are related and proportional to the melting point.²⁴³ Thus the basic mechanisms of coating growth, i.e., surface diffusion, bulk diffusion and shadowing, can be expected to dominate over

different ranges of T/T_m , and to manifest themselves as differences in the resulting coating structures (where T is the substrate temperature and T_m is the coating material melting point—both in absolute degrees). Such is the basis for the structure zone models.

One of the first and most important zone models was developed by Movchan and Demchishin from a study of thick evaporated coatings of Ti, Ni, W, ZrO_2 , and Al_2O_3 .²⁴⁴ The model, which related coating microstructure to T/T_m , has been extended to sputtering, in the absence of ion bombardment, by adding an additional axis to account for the effect of the sputtering gas.^{241,245} See Figure 45. Recent studies have been found to be in basic agreement with the model.^{246-249a} The numbered zones correspond to those suggested by Movchan and Demchishin. The transition zone (Zone T) was not specifically reported by Movchan and Demchishin for their evaporated coatings.

The Zone 1 structure results when adatom diffusion is insufficient to overcome the effects of shadowing. It therefore forms at low T/T_m and is promoted by an elevated working gas pressure which causes coating atoms to be gas-scattered and therefore to arrive at oblique angles. Oblique deposition and substrate surface roughness also contribute to the shadowing effects. The Zone 1 structure usually consists of tapered crystals, with domed tops, which are separated by voided boundaries and tend to point in the direction of the arriving coating flux.²⁵⁰ The crystal diameter increases with T/T_m . Shadowing introduces open boundaries, because high points on the growing surface receive more coating flux than valleys do, particularly when a significant oblique flux is present. The coating surface roughness can result from the shapes of initial growth nuclei, from preferential nucleation at substrate inhomogeneities, from

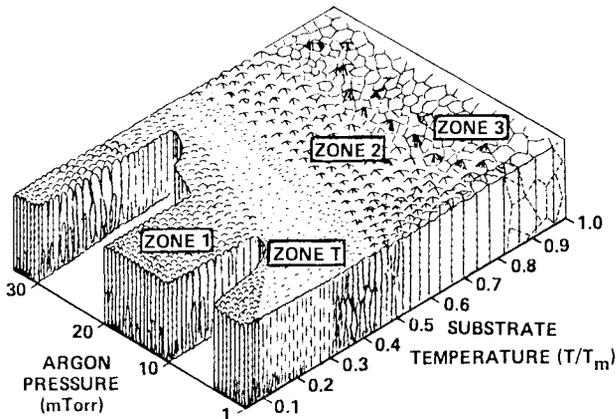


Figure 45: Schematic representation of the influence of substrate temperature and argon working pressure on the structure of metal coatings deposited by sputtering using cylindrical magnetron sources. T is the substrate temperature and T_m is the melting point of the coating material in absolute degrees. See References 241 and 245.

substrate roughness (most common), and from preferential growth. The Zone 1 crystals are generally much larger than the crystallographic grain size. In fact, Zone 1 structures can occur in amorphous as well as crystalline deposits.^{242,251,252} The internal structure of the crystals is poorly defined, with a high dislocation density (see below).²⁴⁴

The transition structure, Zone T, consists of a dense array of poorly defined fibrous grains. It is a columnar Zone 1 structure with crystal sizes that are small and difficult to resolve.²⁴¹ It forms the internal structure of the Zone 1 crystallites. Coatings with structures approaching the Zone T form grow under normal incidence on relatively smooth homogeneous substrates, at T/T_m values that permit the adatom diffusion to largely overcome the roughness introduced by the substrate and the initial nucleation.^{241,253}

The Zone 2 region is defined as that range of T/T_m where the growth process is dominated by adatom surface diffusion.²⁴¹ The structure consists of columnar grains separated by distinct dense intercrystalline boundaries. Dislocations are primarily in the boundary regions.²⁴⁴ Grain sizes increase with T/T_m and may extend through the coating thickness at high T/T_m . Surfaces tend to be faceted. Coatings with platelet structures and whisker growth can also grow under Zone 2 conditions of high surface diffusion.²⁴¹

The Zone 3 region is defined as that range of T/T_m where bulk diffusion has a dominant influence on the final structure of the coating.²⁴¹ The grains may be columnar or equiaxed. Coatings tend to grow with columnar grains, as indicated in Figure 44. Recrystallization into equiaxed grains may be expected if points of high lattice strain energy are generated throughout the coating during deposition. Large columnar grains can grow from columnar as-deposited grains by strain-induced boundary motion and grain growth.²⁴¹ In general, coatings adopt bulk-like behavior as T/T_m is increased into the Zone 3 region.

The general concept of the zone model shown in Figure 45 appears to apply to a wide range of materials and deposition processes. However, the most detailed use of the model has been in the deposition of metals. Thus reference will be made to the model in formulating guidelines for semiconductor device metallization. See Section 7. Alternative versions of the zone model are given in References 247 and 255.

The conical nature of the low-mobility Zone 1 structure is a consequence of the evolutionary nature of the coating growth process. In addition to the evolution of dominant physical structure there is an apparent repetition of honeycomb-like features throughout the film thicknesses.²⁴⁷⁻²⁵¹ This self-similarity in the structural evolution, and the universality of the physical structure of various materials, suggest a common origin, and it has been suggested that the self similarity may be fractal in nature.²⁵⁴

It has been observed that an optimum occurs in the structure-sensitive properties, surface smoothness, crystallographic order, and charge mobility, for coatings of many compounds of electronic interest at a substrate temperature that is within a few percent of $1/3 T_b$, where T_b (absolute degrees) is the boiling point of the compound.^{256,257} See Figure 46. It has been estimated that $T/T_b \sim 1/3$ is that point at which evaporation from amorphous intergrain regions becomes significant. Thus it has been proposed that film quality improves with increasing substrate temperature

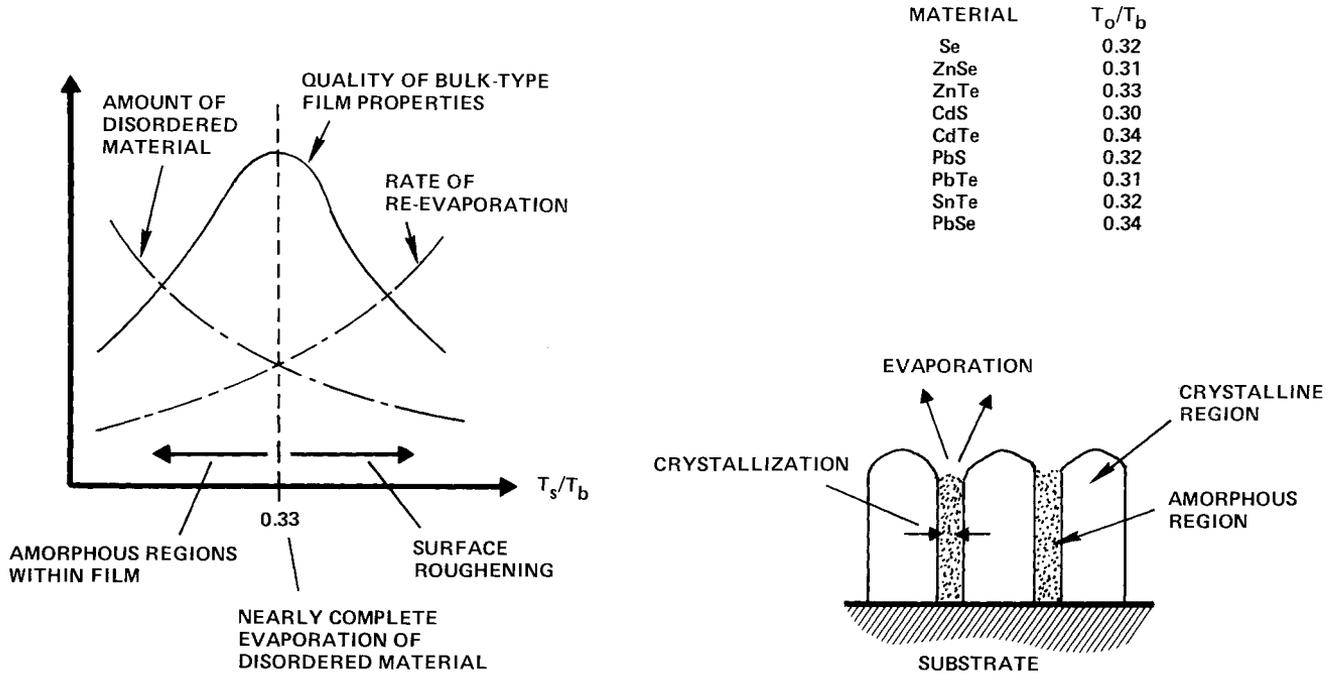


Figure 46: Schematic illustration of Vincett-Barlow-Roberts structure/property relationship for vacuum deposited coatings. See References 256 and 257.

up to $1/3 T_b$ because the existence of such amorphous regions decreases and because slight evaporation from these regions creates vacancies and promotes crystallization and recrystallization.²⁵⁶ At higher temperatures the more complete evaporation from the amorphous regions can make the films rough and voided. This, coupled with a loss of volatile species, leads to degraded coating properties. Thus an optimum structure occurs at $T/T_b \sim 1/3$.

Covalently bonded materials, such as tetrahedral semiconductors, yield coatings with amorphous structures at low substrate temperatures, polycrystalline structures at higher temperatures, and epitaxial single crystal structures under some conditions of high temperature deposition. Figure 47 shows a type of zone diagram which is useful for delineating the T/T_m and deposition rate conditions under which these phases are found.²⁵⁸

6.3 Growth of Compound Semiconductors from Multicomponent Vapors

Compound semiconductors, such as the III-V materials, are composed of constituents having widely different vapor pressures, so that direct evaporation is difficult.^{13,14} Thus, as was discussed in Section 3.6, multi-source evaporation is commonly used. A method called three-temperature evaporation is particularly effective. The method, which is shown schematically in Figure 48, was originally conceived for two-component materials, with the two evaporation sources maintained at different temperatures and the substrate at a third temperature.²⁵⁹ However, the features of the process are far more general than the name three-temperature-evaporation would imply, and form the basis for much of the technology used in Molecular Beam Epitaxy. See Section 4.

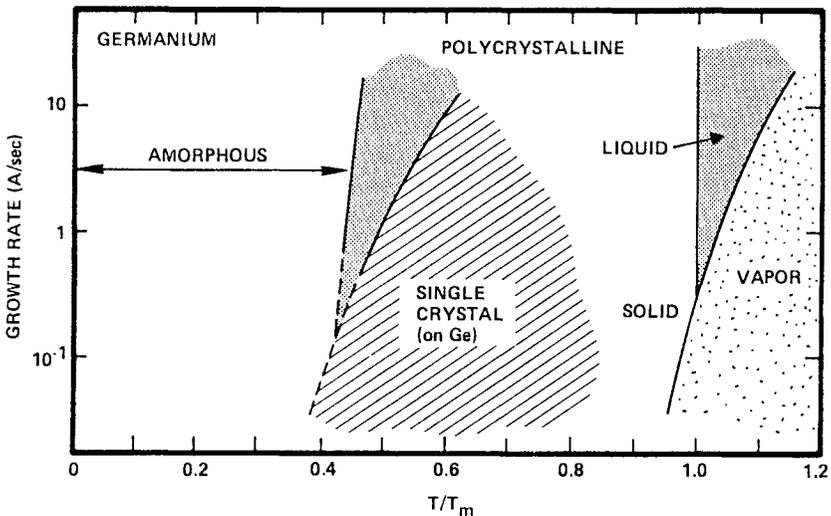


Figure 47: Phase and crystallographic-order transitions in germanium. Data from Reference 258.

Figure 48 shows a schematic plot for the deposition of a compound AB by the three-temperature method. The figure shows the total substrate condensation flux as a function of the incident flux of type A constituent, with the flux of type B constituent maintained constant and with the substrate maintained at a sufficiently elevated temperature so that the B flux cannot condense. Thus, at low fluxes of the type A material, no condensation occurs because both the A and B fluxes are below the critical values for nucleation. See Section 6.1. However, as the flux of type A material is increased, a point is reached where a reaction of A and B adatoms to form AB nuclei results. This occurs before the flux of type A constituent is large enough to condense as elemental type A material. Thus there exist deposition conditions where the only condensate that can survive on the substrate is the compound AB. Subsequent increases in the A flux increase the condensation rate until it becomes rate-limited by the available B flux. If the A flux is further increased to the point where A nuclei form on the growing AB surface, then a two-phase structure consisting of elemental A and AB can result, as indicated in the figure.

The most important present application of the general three-temperature principle is in molecular beam epitaxy. Figure 49 shows a schematic illustration of the substrate processes that are believed to occur in the MBE growth of GaAs.²⁶⁰ The process will be described in some detail, since it illustrates the complexity that can accompany the surface interaction between the vapor fluxes. Ga is supplied as the monomer by evaporation from the liquid. As can be supplied as As_2 or As_4 . The GaAs surface composition is defined as being "As-stabilized" or "Ga-stabilized" according to whether the lattice is terminated by As or Ga atoms respectively.²⁶¹ Most MBE GaAs growth is done under conditions in which the GaAs surface is Ga-stabilized. It is this case that is considered in Figure 49, where the upper drawing refers to growth from Ga and As_2 fluxes, and the lower drawing to growth from Ga and As_4 fluxes.

The Ga atoms have a binding energy of about 2.5 eV and a near unity sticking coefficient at typical substrate temperatures of 600°C. The As_2

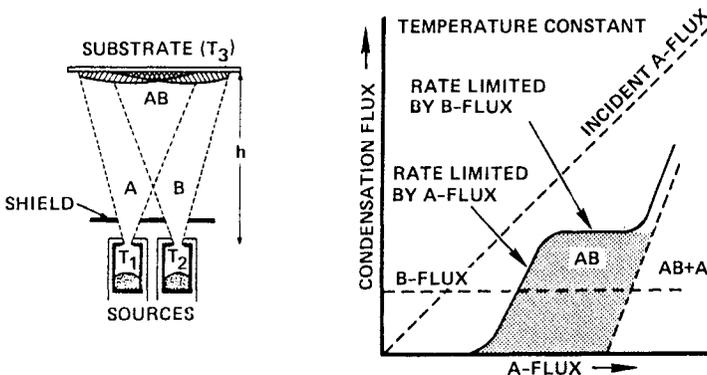


Figure 48: Schematic illustration of three-temperature evaporation method. See Reference 259.

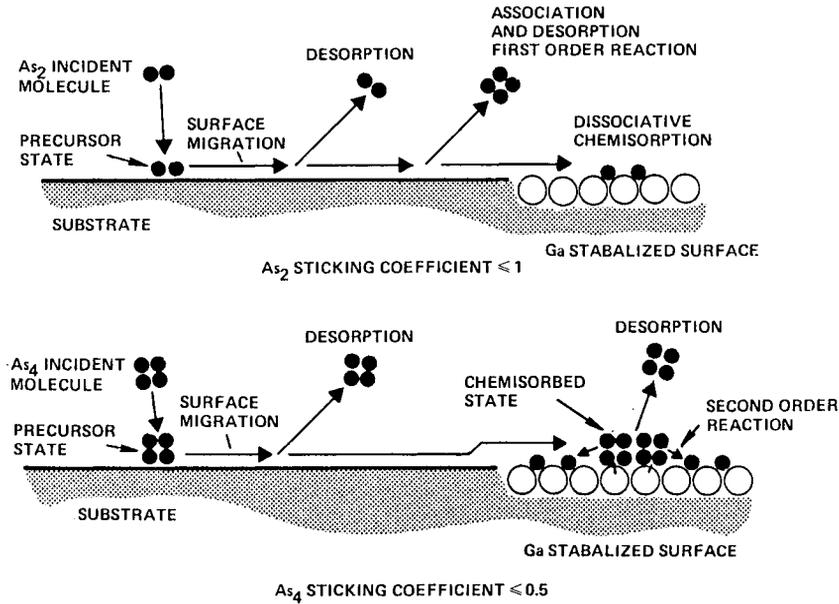


Figure 49: Schematic illustration of MBE growth of GaAs from Ga and As₂ fluxes (upper) and Ga and As₄ fluxes (lower). See Reference 260.

molecules are adsorbed in a weakly bonded, physisorbed, adatom precursor state. They diffuse over the surface, and desorb after a short lifetime ($\tau_R < 10^{-5}$ sec) unless they encounter a Ga atom. The As₂ molecules can also be lost by an association reaction which produces As₄ molecules which subsequently desorb. The As₄ desorption is a first order reaction. This implies that the rate-limiting step is not the surface diffusion of pairs of As₂ molecules to adjacent sites, since this would lead to a second order process, but rather the actual associative interaction or the desorption step itself.²⁶² As₂ molecules encountering Ga atoms undergo a first-order dissociative chemisorption. The sticking coefficient of As₂ molecules is therefore simply proportional to the Ga flux and tends toward unity on Ga-rich surfaces. Excess As₂ is lost by re-evaporation, and stoichiometric GaAs is grown.

As₄ molecules also are adsorbed in a weakly bonded precursor state. The bond energy is ~ 0.4 eV. Under typical deposition conditions the As₄ molecules have a short residence time ($\tau_R < 10^{-5}$ sec) and become chemisorbed only if they encounter a Ga atom under the right conditions.²⁶³ The maximum sticking coefficient is only 0.5. In particular, pairs of As₄ molecules must undergo a second order reaction on adjacent Ga sites to yield four chemisorbed As atoms and a desorbed As₄ molecule, as indicated in Figure 49. This requirement for a pairwise interaction can result in a higher concentration of point defects for films grown with As₄.²⁶⁰

Similar growth behavior is observed for most III-V combinations of Al, Ga, and In with P, As, and Sb.²⁶⁰

The stoichiometric (AB) region is shown as a single phase region in Figure 48. However, it should be divided into two sub-regions, one where the growth is rate-limited by the available A flux, and one where the growth is rate-limited by the available B flux. The detailed microstructure and electrical properties of the coatings grown in these two sub-regions can differ. Consider the case of CdS deposition onto elevated temperature (250°C) substrates by reactive sputtering from a Cd target in an H₂S working gas.²¹⁶ Data are given in Figure 50. The general three-temperature model applies, since at the Cd flux and substrate temperature used, no Cd condensation occurs in the absence of H₂S injection. The S flux rate-limited region yields a film characterized by S vacancies, while the Cd rate-limited region yields a film characterized by Cd vacancies. In doping was effective in the S rate-limited region, but of limited utility in the Cd rate-limited region because of compensation by Cd vacancies.²¹⁶

References 264 to 266 provide a more detailed review of semiconductor crystal growth by sputtering. Compound semiconductors which have been successfully deposited epitaxially include III-V compounds such as InSb, GaSb and GaAs; II-VI compounds such as CdS; IV-VI compounds such as Pb_{1-x}Sn_xTe; nitrides such as AlN; and special materials such as Bi₂Te₃. Other semiconductors produced in polycrystalline form include InN, GaN, (CdZn)S, ZnS, CdSe, CdTe, Cd_xHg_{1-x}Te, ZnO, Cu₂S, CuInS₂ and CuInSe₂.

6.4 The use of Ion Bombardment for Substrate Cleaning and to Influence Coating Growth

Sputter cleaning is that process where the surface of a substrate is subjected to ion bombardment for purposes of sputter-removing oxide layers, or other surface contamination, before depositing a coating. Often from 20 to 500 nm of material are removed using current densities of 0.5 to 5 mA/cm² and ion energies in the 100 to 1000 eV range. Sputter cleaning is commonly used in sputter deposition processes and occasionally in evaporation processes.^{267,268} The usual method is to create a glow discharge plasma at the substrates and to provide movable shields to collect the sputtered flux. Sputter cleaning is easily executed in planar diode sputtering sources of the type shown in Figure 19, because the substrates are in contact with the plasma. It is more difficult to implement in magnetron devices, because of the plasma confinement, and in evaporation systems. In such cases an independent plasma discharge must be sustained at the substrates. In some circumstances these discharges can be implemented by designing the substrate holder so that, in concert with the substrate, it forms an efficient magnetron electrode.¹⁶² See Section 5.6. Rf power is required if the substrates are nonconducting. Ion beam sources are commonly used for sputter cleaning in MBE systems. See Section 4.3.

The controlled use of ion bombardment of a growing film during deposition has been shown to affect all stages of crystal growth, ranging from the initial stages of nucleation discussed in Section 6.2 to the microstructure evolution discussed in Section 6.3. Low energy ion bombardment has proven very effective in removing impurities from the surface of a growing coating.²⁶⁹⁻²⁷¹ There is considerable evidence that low energy ion bombardment also affects the effective surface mobility of adatoms

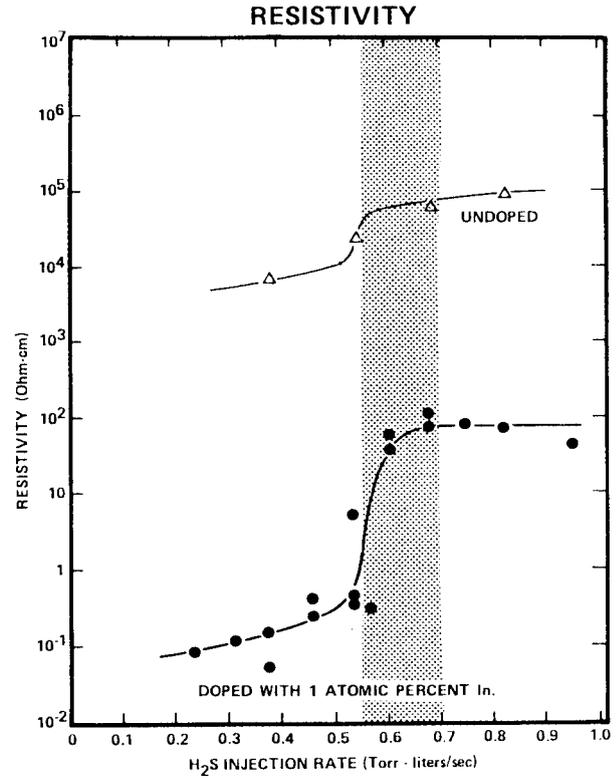
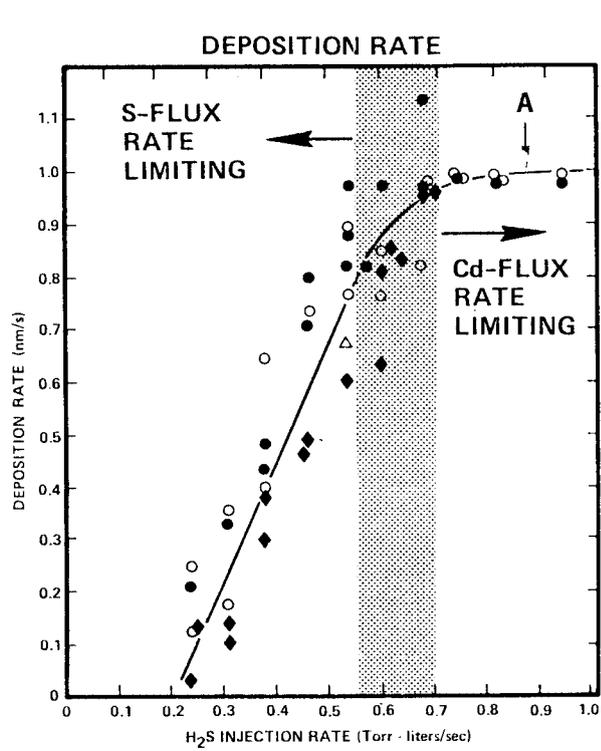


Figure 50: Deposition rate and resistivity for CdS undoped and In-doped films deposited by reactive sputtering from Cd and Cd-In cylindrical magnetron targets using a H₂S/Ar working gas. Data from Reference 216.

and nuclei and may contribute to lowering the epitaxial temperature.^{38,239,272} In addition, high energy ion bombardment can create nucleation sites.^{267,268} Accordingly, the effect of ion bombardment on coating properties, such as the grain size, appears to be strongly dependent on the ion energy. The parameter T/T_m is also important, since crystallographic damage produced by ion bombardment can be removed through the diffusion processes that are active at elevated substrate temperatures. It has, in fact, been concluded that any chosen film property, whether physical, chemical or electrical, will be modified to some extent by contact with a plasma and application of a bias.²⁷³

The term bias sputtering is used to refer to the specific process of maintaining a negative bias on substrates in order to induce ion bombardment during sputter deposition. Bias sputtering is particularly effective when the sputter cleaning process is simply continued, often at reduced bias potential, after the deposition process has been initiated. The term "ion plating" is also used, generally to refer to the case where concurrent ion bombardment is used during film growth in evaporation systems.²⁶⁷ However, an appreciation for the control over coating properties that can be achieved through the use of ion bombardment is increasing, so that current applications extend far beyond simple evaporation and sputtering systems. Examples are in MBE (Section 4) and in ion beam deposition (Section 5.7).

The uses of ion bombardment to modify the properties of concurrently deposited coatings can be divided generally into cases of low energy ion bombardment, where the objective is to remove impurity species, or otherwise influence the coating composition, and higher energy bombardment, where the objective is to modify the coating structure. The effectiveness of low energy (<200 eV) ion bombardment in removing loosely bonded impurity atoms^{269,271} is a consequence of the fact that such atoms have a relatively high sputtering yield,⁶⁸ as discussed in Section 5.2.

Figure 51 shows the oxygen content as a function of substrate bias for GaAs coatings which were deposited from a GaAs target using an rf driven planar diode sputtering source of the general type shown in Figure 19.^{274,275} Data are also shown for room temperature Hall mobilities. The figure clearly shows the effect of modest substrate bias ($\approx 150V$) in reducing oxygen contamination in the films and thereby raising the mobility into the range found in bulk GaAs crystals. Higher biases ($\approx 200V$) caused trapping of accelerated oxygen containing ions, or forward sputtering of oxygen species adsorbed on the surface, with the consequence that the oxygen content of the films increased and the electron mobility decreased. Increasing elemental incorporation probabilities is, in fact, another important application of ion bombardment.³⁸ Figure 52 shows an example of using substrate bias to increase the incorporation of sulfur dopant in GaAs films deposited using the same planar diode apparatus. At low biases ($\approx 50V$) the ion bombardment suppressed the doping process by sputter removing S atoms from the surface of the growing film before they could be incorporated. However, at a higher bias ($\approx 200V$), the sulfur content of the GaAs film was significantly increased.

The sputter removal of contamination may be one of the primary mechanisms by which ion bombardment reduces epitaxial temperatures.²³⁹

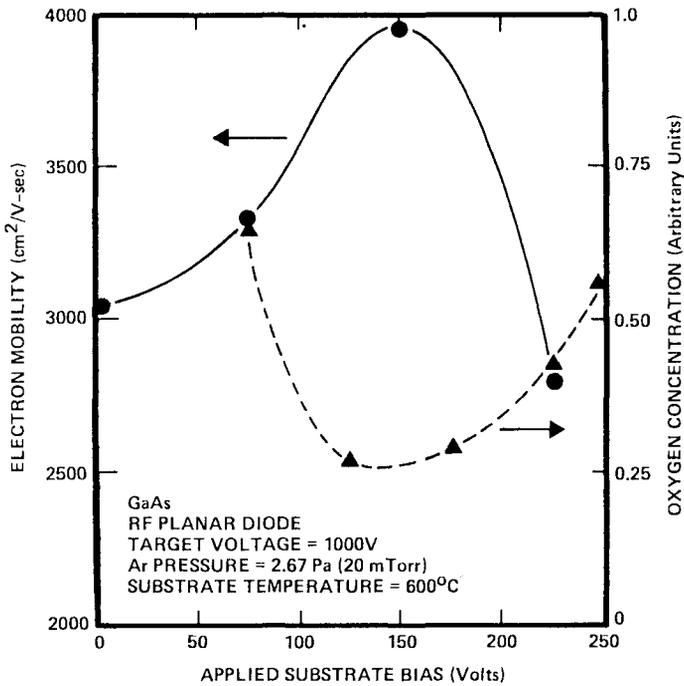


Figure 51: Room temperature electron mobilities and relative oxygen concentrations in GaAs films as function of applied substrate bias during film growth. Data from Reference 275.

Thus in the sputter deposition of Ge films, there is evidence that the use of a substrate bias suppressed the detrimental effects of an oxygen background pressure and permitted epitaxial temperatures comparable to those achieved in ultra high vacuum systems.²⁷⁶ However, it should be noted that even in ultra high vacuum MBE systems it has been found that the growth required to obtain single crystal Si films on both Si and sapphire substrates was reduced when small fractions of the incident Si beam were ionized and accelerated.^{277,278}

One of the exciting new developments in sputter deposition is the growth of single crystal metastable semiconductors such as $(\text{GaAs})_{1-x}\text{Si}_x$, $(\text{GaSb})_{1-x}\text{Ge}_x$, and $\text{InSb}_{1-x}\text{Bi}_x$, which have unique physical properties.^{275,279} The growth of these materials in very controlled environments could be considered to be a sputtering analogue to MBE. It appears that a critical aspect of the growth of such materials is the use of low-energy ion bombardment of the growing film to simultaneously modify the elemental sticking probabilities and effective adatom surface mobilities. For example, one way that ion bombardment apparently assists in the formation of these materials is in inhibiting the formation of second phases by preferential sputtering of incipient precipitates. Thus, for both $(\text{GaSb})_{1-x}\text{Ge}_x$ and $(\text{InSb})_{1-x}\text{Bi}_x$ films it was found that ion bombardment increased the growth temperature at which single phase epitaxial alloys could still be obtained.²⁷⁵

A recent investigation has been made of the mechanisms that characterize the growth of epitaxial films of GaAs by rf sputter deposition in the presence of a substrate bias.²⁸⁰ The sputtering apparatus was of the same rf planar diode type, with a GaAs target, that provided the data shown in Figures 51 and 52. An effusion cell was arranged to deliver an excess As₄ flux at the substrates, which were single crystals of GaAs. The investigation showed that the film growth rate depends not just on the Ga beam flux, as in the MBE growth described in Section 6.3, but also on the bias voltage, the As₄ flux (J_{As_4}), and the substrate temperature (T_s). The growth rate is limited by resputtering of Ga from the substrate surface and is coupled to J_{As_4} and T_s because these parameters determine the As₄ surface coverage and thus the average binding energy and sputtering yield of the Ga atoms on the GaAs surface. It is important to note that, despite the difference in growth mechanisms, these sputtered coatings, which were grown at substrate temperatures of $\approx 600^\circ\text{C}$ with substrate biases of typically from 100 to 200V, yielded carrier mobilities comparable to those obtained using liquid phase epitaxy and MBE. See Figure 51.

Intense substrate ion bombardment during deposition can suppress the development of open Zone 1 coating structures at low T/T_m . See Figure 45. This has been demonstrated for both conducting^{281,282} and nonconducting deposits.²⁸³ The ion bombarded deposits produced at low T/T_m have structures similar to the Zone T type and appear to be free of nodular defects. Ion bombardment on uncooled substrates yields structures typical of high T/T_m .²⁸⁴ As a general rule, the suppression of Zone 1 structures at low T/T_m requires that the ion flux be of such magnitude and energy that it is capable of back-sputtering a significant fraction of the arriving coating flux (i.e., 30-60%).²⁸⁴

Figure 53 schematically illustrates the kinetic processes by which concurrent ion bombardment is believed to influence the microstructure of vacuum deposited coatings. Thus ion bombardment may cause the erosion of surface roughness peaks and the redistribution of material into valleys. The general knock-on behavior discussed in Section 5.2 can cause a large number of atoms within the coating to be relocated by recoil displacement reactions. These processes often lead to the formation of coatings with significant compressive internal stresses. See Section 6.5. Point defects can be generated which promote rapid internal diffusion within the coating or between the coating and the substrate.³⁸ Inert gas ions with energies greater than about 100 eV tend to become trapped in the growing coating, as discussed in Section 5.2. At high ion fluxes, the entrapped inert gas can have concentrations of several atomic percent and may cause blistering in subsequent annealing.^{284,286} Inert gas incorporation is generally greater in amorphous than in crystalline materials. Concentrations of Ne, Ar, and Kr of from a few to 30 atomic percent have been obtained in amorphous GdCo and CdCoMo films deposited using an rf substrate bias.²⁸⁷ Inert gases generally evolve during deposition at T/T_m greater than about $1/2$ and are therefore not entrapped.

Caution must be exercised when using ion bombardment during the deposition of any multicomponent material, since the composition may be modified by preferential sputtering, as discussed in Section 5.5.

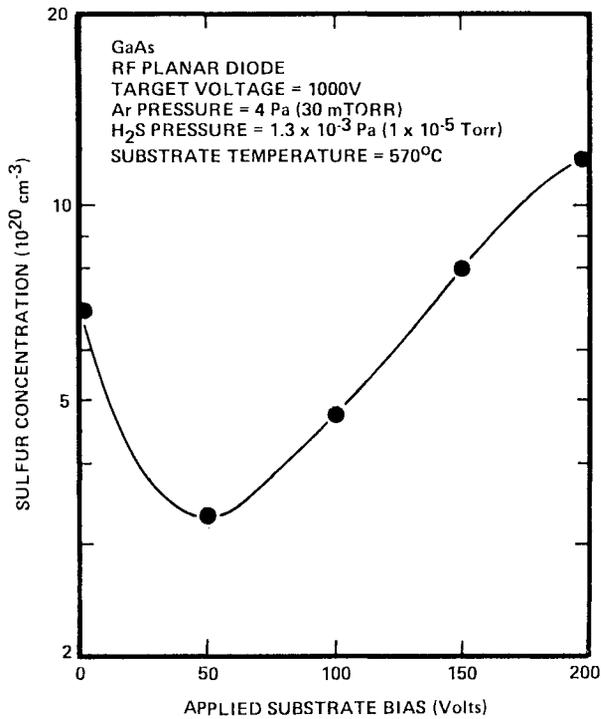


Figure 52: Sulfur concentrations in as-deposited GaAs films as function of the applied substrate bias during film growth. Data from Reference 275.

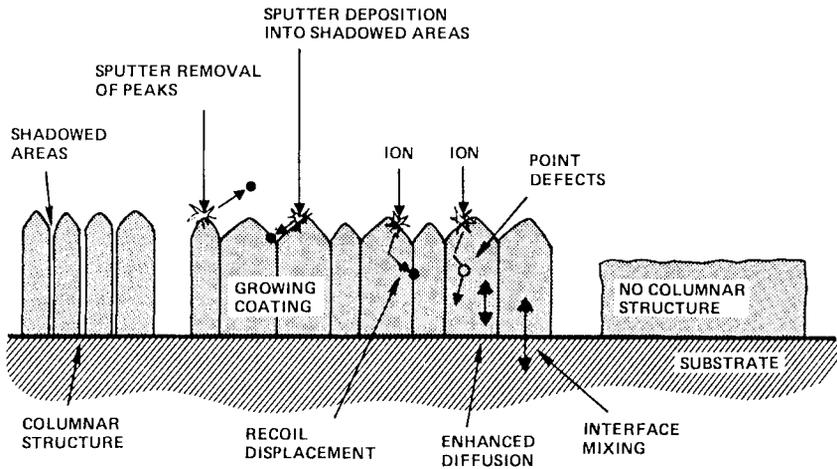


Figure 53: Schematic illustration of kinetic processes by which concurrent ion bombardment influences the microstructure of a vacuum deposited coating. Sufficiently intense ion bombardment can remove the shadow-induced columnar structure as implied by going from left to right in the figure.

It has been noted above that ion bombardment during deposition can promote interdiffusion within a film or between a film and substrate. In a new process called ion-beam mixing, post-deposition bombardment of a thin film structure by very high energy ions is used to cause interdiffusion and reactions within the structure.²⁸⁸ The reaction-mixing mechanism involves collision cascades of the type discussed in Sec. 5.2. In many cases, the process proves to be remarkably efficient, with the number of mixed atoms greatly exceeding the number of bombarding ions. Two examples are illustrated schematically in Figure 54. In the case shown at the top, multiple layers of Ag and Cu, 20 nm thick, were first deposited on SiO₂ with the thickness adjusted so that the average compositions varied between Ag₂₀Cu₈₀ and Ag₈₀Cu₂₀. The films were then mixed using 300 keV Xe⁺ ions, at a dose about equal to the surface atom density (2 × 10¹⁵ cm⁻²), with the samples held at liquid nitrogen temperatures. The resulting coatings were fcc cubic metastable solid solutions of Ag and Cu atoms with lattice parameters which obeyed Vegard's law for ideal solid solutions. In the case shown at the bottom of Figure 54, a 30 nm thick Ni film was caused to react with a Si substrate by bombardment with 300 keV Kr⁺ ions.²⁹⁰ The reacted zone consisted of a Ni₂Si phase which increased in thickness with the square root of the ion dose. Ion doses larger than those required to consume all of the metal yielded amorphous structures. This is in contrast to the thermal diffusion case, where the Ni₂Si transforms to NiSi after all of the metal has been consumed.

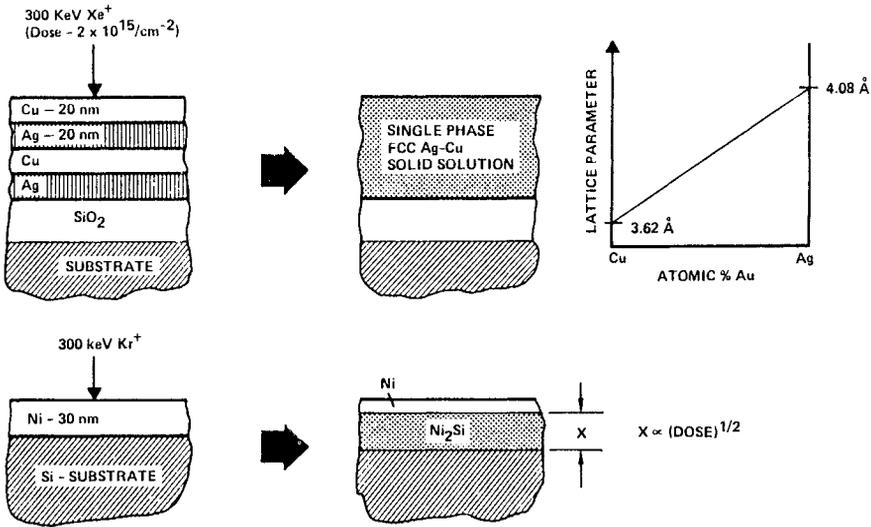


Figure 54: Schematic illustration of interfacial compound formation by ion beam mixing. See References 288 and 290.

6.5 Internal Stresses

Virtually all coatings deposited by evaporation and sputtering are in a state of stress. The total stress is composed of a thermal stress, due to the difference in the thermal expansion coefficients between the coating and substrate materials, and an intrinsic stress due to the cumulative effects of atomic forces generated throughout the coating volume by atoms which are out of position with respect to the minimums in the interatomic force fields.²⁹¹

The parameter T/T_m is particularly important in determining the state of internal stress in a coating. At low T/T_m , the intrinsic stress dominates over the thermal stress. At T/T_m exceeding ≈ 0.3 (Zone 2 in Figure 45) the intrinsic stresses are significantly reduced by recovery²⁹² during coating growth. The thermal stresses are dominant for coatings deposited at high T/T_m and cooled to room temperature. Thermal stresses can also be important for coatings deposited at low T/T_m and then subjected to processing or use at a higher temperature.

The thermal stress in thin films with thicknesses less than 10^{-4} times that of the substrate, where plastic flow within the substrate can be neglected, is given in a one-dimensional approximation (neglecting Poisson effect) by^{95,293}

$$\sigma_{th} = E_f (\alpha_f - \alpha_s) (T_s - T_a) \quad (20)$$

where E_f is Young's modulus, α_f and α_s are the average coefficients of thermal expansion for the film and substrate, T_s is the substrate temperature during deposition, and T_a is the temperature at the time of measurement. A positive value of σ_{th} corresponds to a tensile stress. Table 10 gives elastic constants and thermal expansion coefficients for several metals commonly used in electronics related processing. Large thermal stresses can be induced by typical processing conditions. Consider an Al coating ($\alpha_f = 2.4 \times 10^{-5} \text{ }^\circ\text{C}^{-1}$) deposited on a Si substrate ($\alpha_s = 7.6 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$) at a temperature of 150°C and then annealed at 400°C as part of a device processing step. A compressive stress of about 0.26 GN/m^2 , which is larger than the yield strength of Al (0.11 GN/m^2), is predicted.

The intrinsic stress has been defined as that component of the total stress which cannot be attributed to the thermal stress.²⁹¹ Intrinsic stress magnitudes can approach the yield strength for the coating material. The coating/substrate bond must withstand the interfacial forces associated with the accumulated internal stresses throughout a coating. In contrast to the thermal stress, the contribution of the intrinsic stress to this force tends to increase with the coating thickness. Therefore, premature interface cracking and poor adhesion, for coatings with thicknesses exceeding values as low as 100 nm , can often be attributed to high intrinsic stress levels.

For low melting point materials such as Al, T/T_m exceeds 0.3 for typical deposition conditions (see Table 10), and both the thermal and intrinsic stresses tend to relax by internal diffusion.²⁹⁴ Accordingly, what we encounter is not the mechanical consequences of the stress, such as adhesion failures, but the consequences of the stress driven diffusion. The

Table 10: Properties of Several Metallic Coating Materials Commonly Used in Electronics Related Processing

<u>Materials</u>	<u>Melting Temp. (°C)</u>	<u>T/T_m *</u>	<u>Young's Modulus GPa</u>	<u>Yield Strength GPa</u>	<u>Thermal Expansion Coef. (°C⁻¹) **</u>
Aluminum	660	0.40	63	0.11	2.4x10 ⁻⁵
Titanium	1668	0.19	113	0.14	0.85x10 ⁻⁵
Chromium	1875	0.17	260	0.16	0.68x10 ⁻⁵
Copper	1082	0.28	120	0.32	1.68x10 ⁻⁵
Molybdenum	2610	0.13	300	0.84	0.49x10 ⁻⁵
Palladium	1552	0.20	115	0.31	1.18x10 ⁻⁵
Tantalum	2996	0.11	190	0.35	0.67x10 ⁻⁵
Tungsten	3410	0.10	351	1.8	0.43x10 ⁻⁵
Platinum	1769	0.18	150	0.16	0.90x10 ⁻⁵
Gold	1663	0.19	81	0.21	1.43x10 ⁻⁵
Lead	327	0.62	16	0.009	2.94x10 ⁻⁵

resulting material flow can lead to the formation of hillocks or holes, depending on whether the stresses are compressive or tensile. Hillocks have been observed to grow extensively in Al,²⁹⁵ Au,²⁹⁶ and Pb²⁹⁷ films under conditions of high compressive stress, such as in the numerical Al/Si example given above. Surface growths of this type present obvious problems in the deposition of subsequent coating layers during device processing.

In the case of high melting point materials, T/T_m is generally less than 0.3 at typical deposition temperatures. See Table 10. Accordingly, intrinsic stresses usually dominate in these materials. Therefore, intrinsic stresses are particularly important in the refractory-metal device metallization discussed in Section 7. The intrinsic stresses, in contrast to the thermal stresses, are strongly dependent on the deposition conditions. Therefore, intrinsic stresses will be the subject of the remaining discussion.

The following generalizations can be made regarding the intrinsic stresses in metal coatings.^{95,222,293} (1) The intrinsic stress is not strongly dependent on the substrate material. (2) The intrinsic stress is relatively constant throughout the coating thickness for thin films (<500 nm thick). (3) The intrinsic stresses in evaporated metal coatings are almost exclusively tensile, and most evaporated dielectric coatings also exhibit tensile stresses. (4) Coatings subjected to significant energetic particle bombardment during or following deposition generally exhibit compressive stresses.

Therefore, sputtered coatings are often in a state of compressive stress.^{154,298,299}

Figures 55 and 56 illustrate the effect of ion bombardment on internal stress. Figure 55 shows the internal stresses, as a function of substrate bias voltage, for Mo coatings deposited using an rf planar diode.²⁹⁸ The stresses are seen to be much larger than the estimated thermal stress for the substrate temperature, and to be tensile at low bias voltages and compressive at higher bias voltages. Note that the maximum stresses are comparable to the yield strength of Mo ($\approx 0.8 \text{ GPa/m}^2$). Figure 56 shows the results of experiments in which the influence of concurrent ion bombardment on the internal stresses in evaporated Cr coatings was investigated.³⁰¹ Coatings deposited at low ion doses were in tension, as would be expected for evaporated coatings. Above a critical dose of about 45 eV/Cr atoms, the coatings were in a state of compression.

Tensile intrinsic stresses are generally explained by some type of void network, such as grain boundaries, which place the coating in an underdense state.^{254-257,302-305} It is believed that the compressive stresses are generated by an atomic peening mechanism where incident energetic particles strike the growing coating and drive surface atoms into the interior by recoil displacement reactions. Thus the coating is densified, and the surface is smoothed, via the mechanisms discussed in Section 6.4 and shown in Figure 53. The upper curve in Figure 56 shows that the coating reflectance does indeed undergo a significant increase at the ion bombardment dose that produces compressive stresses. Thus tensile stresses are

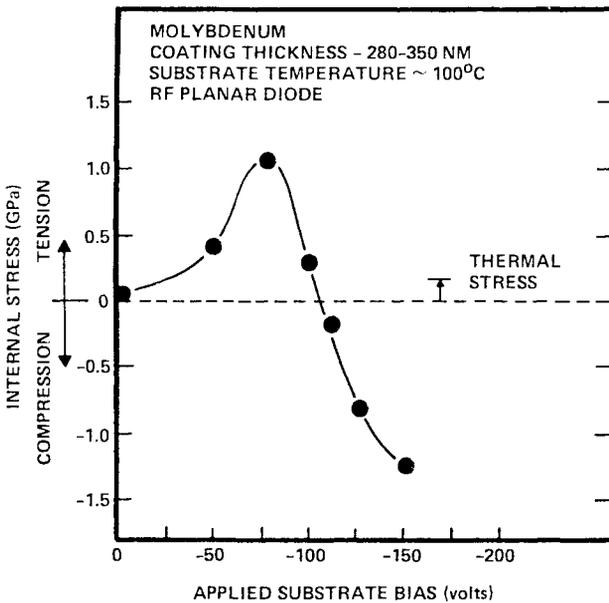


Figure 55: Variation of internal stress with substrate bias for molybdenum coatings deposited using planar diode sputtering source. Data from Reference 300.

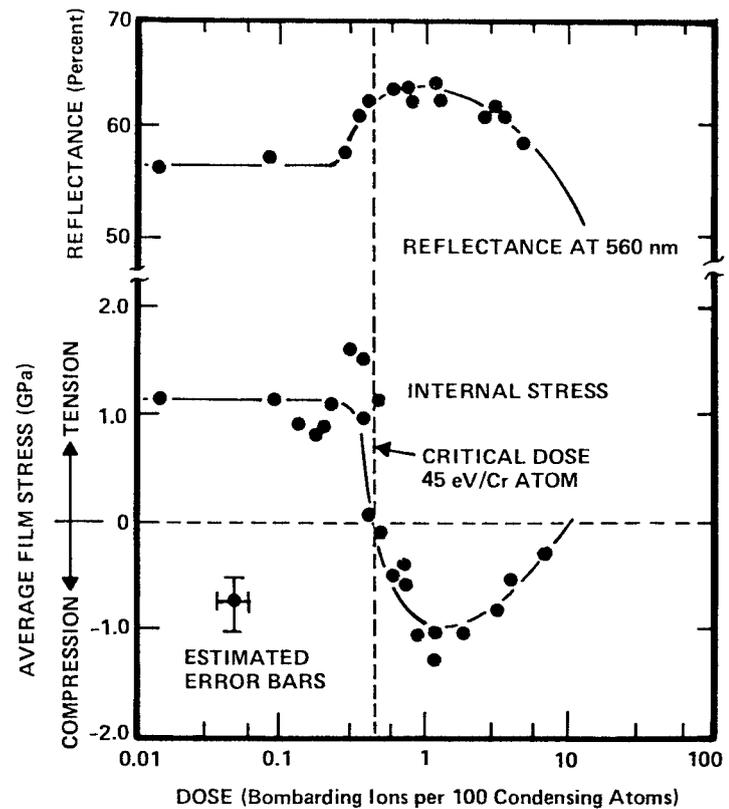
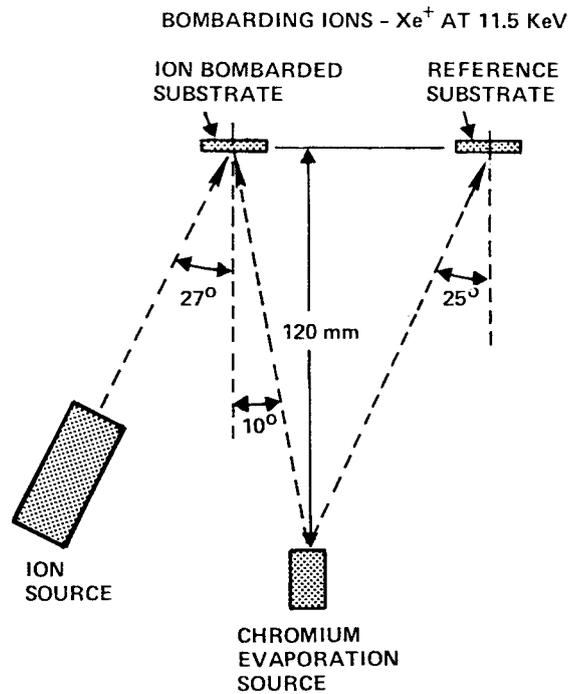


Figure 56: Influence of concurrent ion bombardment on the internal stress in evaporated chromium coatings. Data from Reference 301.

generally found in the low density Zone 1 region of Figure 45 and compressive intrinsic stresses in the relatively dense Zone T region.²²²

In the case of triode and magnetron sputtering sources operating at low working gas pressures, the growing coatings are subjected to bombardment by energetic working gas atoms which originate at the target as ions that are neutralized and reflected. See Section 5.2. These species are as effective as ion bombardment in producing compressive stresses.^{154,298} The stresses in such coatings therefore depend in a seemingly complex way on those deposition parameters that determine the flux and energy of the reflected atoms which are incident on the substrate.

Figure 57 shows the influence of the argon working gas pressure on the interface force per unit width (integrated stress) that developed in 200 nm thick coatings of several materials deposited at normal incidence on near-room-temperature substrates using cylindrical magnetron sources. This general behavior has been observed for more than ten metals ranging in mass from Al to W³⁰⁶ and for amorphous Si.⁷⁷ It has also been observed for coatings deposited using rectangular planar magnetrons³¹⁰ and small gun and ring type planar magnetrons.³¹¹⁻³¹⁴

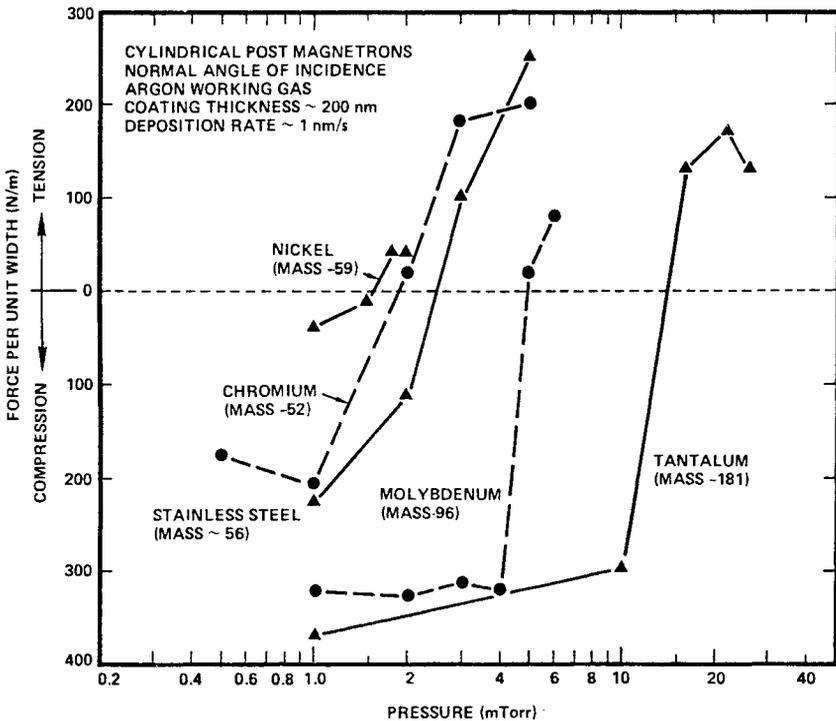


Figure 57: Force per unit width at coating/substrate interface as a function of argon pressure for metallic coatings of various atomic mass which were deposited using cylindrical-post magnetron sputtering sources. Data from References 306 to 309.

It is believed that the transitions from compressive to tensile stress shown in Figure 57 can be explained in terms of a competition between the bombarding effects of the reflected atoms and those factors that tend to promote an open Zone 1 structure. The flux and energy of the reflected atoms increase with the atomic mass of the target relative to that of the working gas.⁷⁸ Increasing the argon working gas pressure at a fixed discharge current causes a reduction in discharge voltage and therefore in the energy of the neutralized and reflected ions. An increase in Ar pressure also leads to gas phase collisions which both attenuate the energy of the reflected species that reach the substrate, and scatter the sputtered atom flux so that the coating atoms approach the substrate at oblique angles. The latter effect promotes the formation of a Zone 1 structure as discussed in Section 6.2. The energy attenuation of the reflected species reduces the effectiveness of these atoms in suppressing the formation of a low density Zone 1 structure. Figure 58 shows the influence of the average coating flux angle-of-incidence, measured from

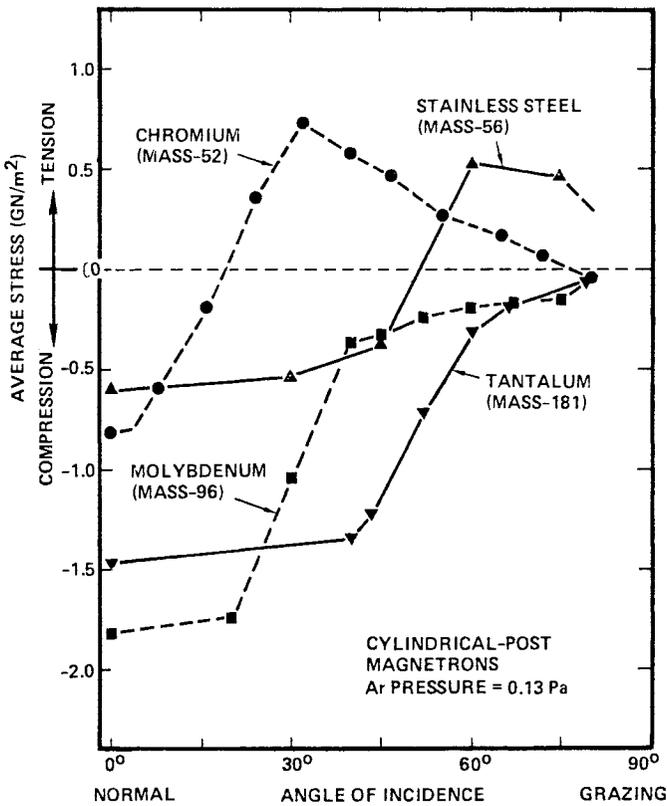


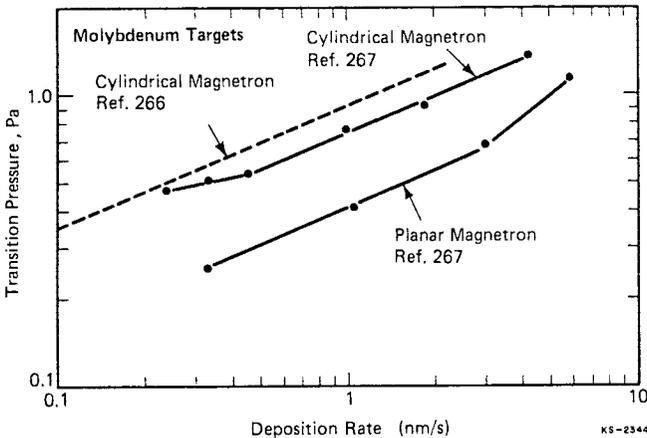
Figure 58: Coating stress as a function of the coating flux angle-of-incidence for coatings of various atomic mass which were deposited using cylindrical-post magnetron sputtering sources. Data from Reference 315.

the substrate normal, on the internal stresses in 200 nm thick coatings of the materials shown in Figure 57. The coatings were deposited onto near-room-temperature substrates, using cylindrical-post magnetron sources operated in Ar at 0.13 Pa (1 mTorr).³¹⁵ Again, it is seen that the presence of an oblique coating flux promotes an open structure and tensile stresses. The oblique flux effect is suppressed, and compressive stresses are formed, up to flux arrival angles which appear to depend on the target mass and therefore on the flux and energy of the neutralized and reflective ions.

Figure 59 shows the transition pressure, below which Mo coatings are in a state of compressive stress, as a function of the deposition rate. Data are given for coatings deposited using a planar magnetron of the type shown in Figure 27H, and cylindrical magnetrons of the uniform plasma type shown in Figure 27A and the ring plasma type shown in Figure 27E.^{316,317} In all cases the transition pressure exhibits a general increase with deposition rate. The cause is believed to be associated with the discharge voltage, and therefore with the energy of the energetic working gas atoms which originate at the target as neutralized and reflected ions. The transition pressure is seen in Figure 60 to be a very sensitive function of the discharge voltage.

Figure 61 shows the transition pressure as a function of the target-to-working-gas mass ratio. The data appear to support the proposition, stated above, that large mass ratios produce large energy fluxes of reflected species and thereby delay the trend toward forming a low density Zone 1 structure at higher pressure, or larger oblique angles, Figure 58.

It will be noted in Figures 59 and 61 that the transition to tensile stress occurs at lower working gas pressures for planar than for cylindrical-post magnetrons. This is believed to occur because the reflected atoms leave the target surface at different angles than do the sputtered species.



EFFECT OF DEPOSITION RATE

Figure 59: Argon transition pressure, below which sputtered molybdenum coatings deposited with cylindrical-post and planar magnetron sputtering sources are in a state of compressive stress, as a function of the deposition rate.

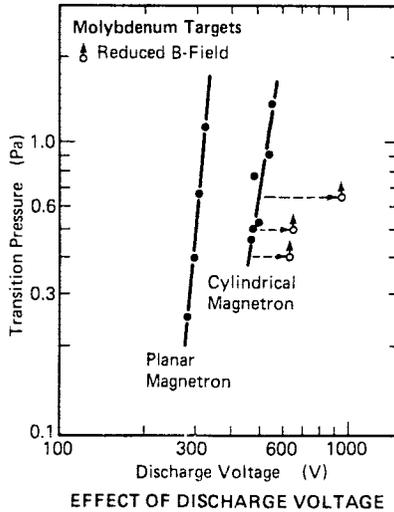


Figure 60: Variation of transition pressure with discharge voltage for cylindrical and planar magnetron cases. The open circle data points refer to cylindrical magnetron cases where the magnetic field was decreased in order to increase the discharge voltage from initial operating conditions near the transition value, as indicated by the dashed lines. The arrows indicate that resultant coatings were all in high compression, signaling that the transition pressures were significantly higher than the pressures (indicated in the figure) under which the experiments were conducted. Data from Reference 317.

Reflection at angles less than 180° , i.e., small angle scattering, is most probable. Substrates placed directly in front of a planar magnetron will be bombarded only by atoms reflected at approximately 180° , while atoms which are scattered at smaller angles and possess larger energies will pass off to the side. The symmetry of the cylindrical-post magnetron case causes the substrates to be bombarded by the atoms that undergo small angle reflections. Therefore, it is believed that the substrates in cylindrical-post magnetrons are subject to bombardment by a larger energy flux of reflected atoms.¹⁵³ Support for this point of view has been provided by experiments in which special shield configurations were used to isolate the particle flux which left cylindrical-post magnetron sources at oblique angles, from the flux which left at more normal angles.^{77,78,309} Coatings that were deposited in the presence of the oblique flux exhibited larger compressive stresses, and contained more entrapped working gas, than coatings grown in the presence of only the normal flux.

It was noted in Section 5.2 that the energetic reflected working gas species can become entrapped in sputtered coatings deposited at low pressures. The amount of entrapped working gas exhibits the same general trend as the stress data, increasing with the target-to-working-gas mass ratio as shown in Figure 21, and decreasing with increased pressure.

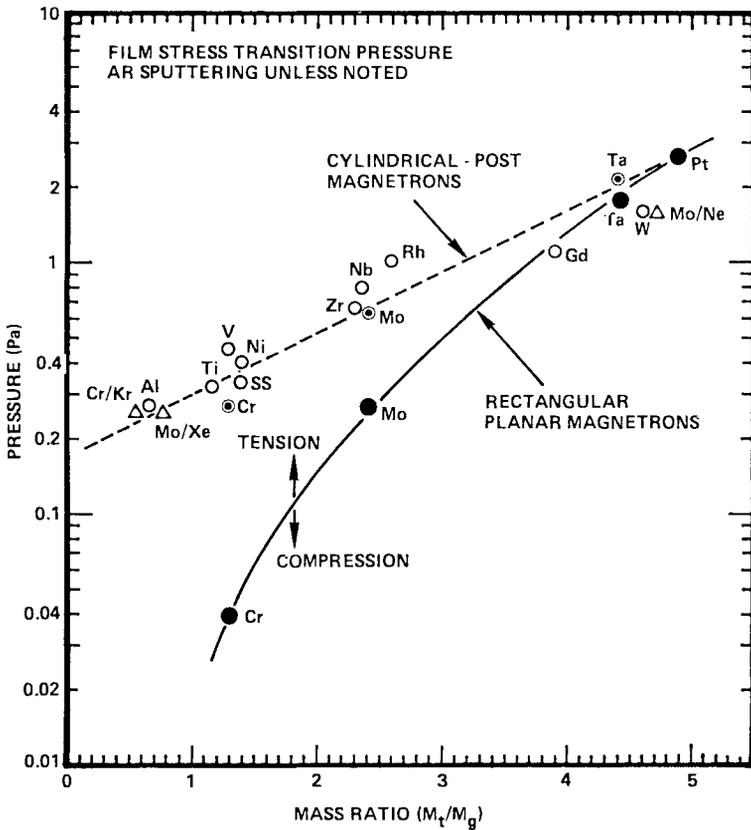


Figure 61: Argon transition pressure, below which sputtered coatings deposited with cylindrical-post and planar magnetron sputtering sources are in a state of compressive stress, as a function of the target-to-working-gas atomic mass ratio. Data from References 306 to 310.

However, the compressive strains within the coatings remained invariant over large changes in the amount of entrapped working gas. Therefore, it has been concluded that the entrapped gas is a consequence of the atomic peening produced by the reflected species, but not the cause of the compressive stress, at least for metal coatings deposited using magnetron sources.³⁰⁹

The above experiments illustrate three important points: (1) the formation of coatings with compressive stresses and entrapped working gas is a common occurrence in sputter deposition at low pressures where the neutralized and reflected ions do not lose their energy via collisions in their passage to the substrates; (2) the angular dependence of the reflected ion flux emitted from the target makes the substrate bombardment dependent on the shape of the target³⁰⁹ and the positioning of the substrates relative to source;^{103,300} and (3) the dependence of the stress-producing mechanism

on the energy of the incident particles makes the tendency to produce compressive stresses a strong function of the discharge voltage. Of course, these effects lose their significance for coatings deposited at high T/T_m .

7. METALLIZATION OF SEMICONDUCTOR DEVICES

7.1 Introduction

The metallization of semiconductor devices is one of the primary applications of physical vapor deposition. Metallization for typical devices must provide (1) metal to silicon contacts, (2) gate contacts, (3) interconnects between various points on the chip, and (4) compatible bonding pads for packaging and assembly operations. Aluminum is the most widely used metallization material. It can be used to form low resistance ohmic contacts to highly doped p^+ and n^+ -type Si and also to meet the interconnect requirements in both bipolar and MOS integrated circuits.

In early circuit manufacture much of the Al was deposited by evaporation. The present trend is toward using Al-Cu-Si alloys containing typically 2 to 4 wt.%Cu and 1 to 2 wt.%Si. The Cu improves the electromigration resistance and the Si improves the stability of the Al/Si interface. Therefore, sputtering is coming into increased use because of its capabilities for reliably depositing such multicomponent materials with controlled stoichiometry.

The more exacting requirements of VLSI will necessitate new metallization systems that permit shallower junctions, reduced contact resistances, the use of dry etching processes, and multilevel interconnection schemes. These requirements will necessitate the increased use of refractor metals and their silicides, i.e., materials that can be deposited effectively by sputtering. Therefore, the trend toward sputtering, which has resulted in its use for perhaps 95% of the new wafer fabrication lines started in the past few years,³¹⁸ is expected to continue.

The reader is referred to References 319 to 323 for more detailed discussions of the metallization of semiconductor devices. An excellent multi-volume bibliography on metallization materials and techniques for silicon devices has been prepared by J.L. Vossen of the RCA Laboratories and is available from the American Vacuum Society.³²⁴

7.2 Metallization Materials Considerations

Ohmic contacts of Al on Si are generally formed by depositing the Al onto the Si, and then annealing at a temperature in the 450 to 525°C range for 15 to 30 minutes in N_2 or H_2 , to consume the native oxide on the Si and form an Al/Si alloy at the junction.²⁷⁰ The solid solubility of Si in Al at the annealing temperature is about 1 at.%. Therefore, Si diffuses into the Al, leaving pits in the Si which are filled by Al to form spikes which can penetrate deep enough to short shallow p-n junctions. One solution is to add Si to the Al metallization material to form an alloy with reduced solubility for externally supplied Si. This suppresses the spiking problem. However, another problem is introduced. During cooling to room temperature the Al becomes supersaturated with Si. The Si can precipitate at the junction and result in

the formation of a layer of p-type Al-doped Si. This leads to erratic barrier heights on Schottky diodes and non-ohmic contacts on n-type substrates. An alternative solution is to deposit a silicide film, such as PtSi or Pd₂Si, between the Al and Si to stabilize the junction.³²⁵ Unfortunately, Al reacts with such silicides at 400°C and is able to penetrate through the silicide and reach the Si to repeat the spiking problem.³²² Therefore, a diffusion barrier is needed in addition to the silicide. Examples are Cr³²² and Ti-W.³²⁰ The general application of such barrier layers is reviewed in Reference 326.

Silicide layers such as PtSi can be formed by evaporating or sputtering Pt films onto the Si and annealing at 400-600°C to cause the silicide reaction. However, when shallow contacts are desired the Si is added externally. This can be done by depositing Si-Pt bi-layers by evaporation or sputtering, and then annealing them to cause silicide formation. In another approach homogeneous layers can be formed directly, by co-evaporating or co-sputtering from Pt and Si sources, or sputtering from a Pt-Si target, and then annealing to form the desired silicide.

Polycrystalline Si, deposited by chemical vapor deposition, has been the primary interconnect material used at the gate level. However, the resistivity of polysilicon is too high for the cross sectional areas that are available in VLSI circuits with their shrinking line widths. Consequently, excessive RC time constants and losses in circuit speed result.³¹⁹ Refractory metal-silicides, such as TiSi₂, MoSi₂, TaSi₂, and WSi₂, offer low resistivities, low contact resistances to Si, stability against oxidation in contact with SiO₂, and stability against reactions with Al and Al alloys. Accordingly, these materials are being investigated as gate metallization layers.

It is projected that the shrinking lateral features of VLSI will require the use of anisotropic plasma etching methods such as ion beam or reactive ion etching.³¹⁹ Such etching processes tend to produce radiation damage that cannot be annealed out at temperatures compatible with Al metallization (see Section 7.4). A possible solution is to form contacts using refractory materials capable of withstanding annealing temperatures in the 600°C range. Promising candidates are again the refractory metals such as Ti, Mo, Ta, and W and their silicides. Since the refractory metals and silicides do not meet requirements for wire bonding to the chips, it is envisioned that the uppermost layer would again be a material such as Al. In some cases a diffusion barrier such as TiN may be required between the refractory base layer and the aluminum top layer.³²⁶

Refractory metals are also used on some conventional devices to eliminate the corrosion problems associated with Al.

As noted previously, Cu is added to Al to improve the electromigration resistance. The Cu apparently improves the electromigration resistance by segregating in the grain boundaries.³²⁷ The composition, resistivity, and microstructure of these films depends strongly on the deposition technique and affects the reliability of their performance as contacts and interconnections.³²⁸ It has been argued that the larger grains produced by electron beam evaporation yield greater electromigration resistance,³²⁹ particularly for the small line widths used in VLSI.³³⁰ However, others have argued that electron beam evaporation is not well suited to controlling the

composition and microstructure, and that stable homogenous small grain size films of a type easily produced by sputtering are required to fabricate reliable VLSI interconnections.^{328,331}

Multilevel metallization interconnections will be used in VLSI to improve circuit performance by minimizing signal paths and enhancing the utilization of chip area. These consist of two or more metallization layers separated by suitable dielectric layers. The principal candidates for interlevel dielectrics are oxide films such as SiO_2 and Al_2O_3 prepared by low pressure or plasma-assisted CVD. A typical VLSI multilevel system is



where PtSi forms the ohmic and Schottky contacts, Ti-W acts as a diffusion barrier, Al-Cu provides the first-level metallization, SiO_2 serves as the interlevel insulator, and Al serves as the second level interconnection. It has been suggested that second-level metallization layers will be deposited primarily by bias sputtering to provide step coverage, as discussed in the next section.³¹⁹

7.3 Step Coverage

Step coverage is a critical consideration for most device metallization applications. The basic problem is the geometric-shadowing and angle-of-incidence effects on coating growth and microstructure that were discussed in Section 6.2. Thus when coating a step, even deposition conditions produce a dense Zone T structure on a flat surface, yield a porous Zone 1 structure on the step side walls, and an open boundary or cusp emanating from the base of the step as a result of self-shadowing. See Figures 62 and 64a.

The techniques used to improve step coverage can be classified into the following broad categories: (1) source geometry adjustments, (2) step

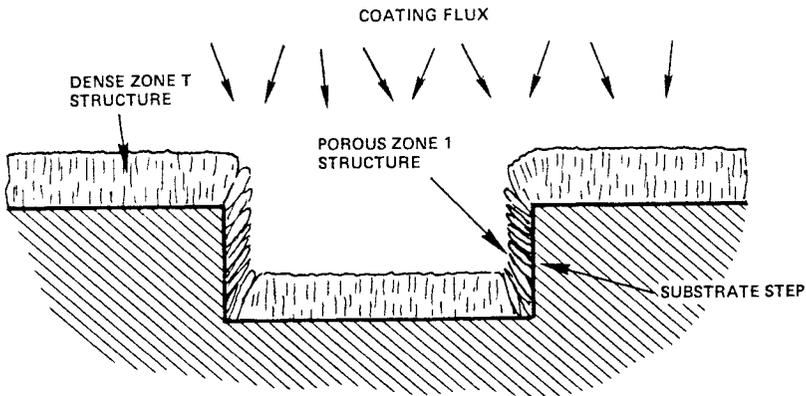


Figure 62: The effect of substrate surface steps on the microstructure of coatings deposited at low T/T_m .

geometry shape adjustments, (3) substrate temperature control, and (4) bias sputtering.

Source geometry considerations are based on the fact that coating atoms approach a substrate surface in directions that are dependent on the geometric configuration of the deposition apparatus. In the case of evaporation, the atoms pass essentially line-of-sight from the source to the substrates. Even in the case of sputtering at elevated pressures, the atoms approach the substrate line-of-sight from the point of last collision, which is nominally one mean free path from the substrate. See Equation 6. A typical mean free path at 50 mTorr is of the order of 1 mm and therefore much larger than the step sizes. Thus the apparatus geometry in the case of high pressure sputtering can be approximated as an extended source, with cosine emission, located one mean free path from the substrate. Since the coating atom mean free paths are long compared to local variations in the substrate topography, these variations can affect the coating distribution and structure by shadowing.

Small source systems, such as electron beam or filament evaporators, or small ring type magnetrons of the planar or gun-type, are generally configured to use wafer motion to provide increased throughput capacity and thickness uniformity, as discussed in Section 3.4. Often planetary tooling of the type shown in Figure 9 is used. As the ratio of the translation or rotation velocity to the deposition rate increases, the effect of the motion approximates an extended-target sputtering system,³³² and thickness uniformity on the top surface improves. However, the extreme oblique component in coating flux that is produced by an extended source or by planetary motion can actually exacerbate the open self-shadow boundary or cusp that emanates from the base of the step.³³² That is, the tooling that provides the most uniform coating over a flat wafer may not provide the best step coverage.³³³

Large source magnetron systems are of increasing interest because of their capacity for large production volumes and their compatibility with in-line processing systems. See Section 5.6. In a recent study, measured erosion rates from a rectangular planar magnetron were used with computer modeling to stimulate step coverage for substrates that were rotated on a single axis aligned with the long axis of the magnetron.³³⁴ The deep cracks at the base of the steps, that are typical of small sources with planetary tooling, were not predicted. The reason was traced to the shape of the vapor emission profile from the magnetron source. Measured step coverages, for Al-2 wt.% Cu sputtered at a pressure (1.5 mTorr) that yielded line-of-sight transport, were in good quantitative agreement with the calculated deposition profiles. More modeling studies of this type can be expected in the future as the step coverage capabilities of large area magnetron sources are explored.

Step geometry shape control is one of the most effective methods for reducing the geometric shadowing problems associated with step coverage. For example, a step slope of 30° from the normal, and suitable substrate motion, can eliminate cusp formation because of shadowing.³³² Unfortunately, the shrinking lateral size features in VLSI will leave no space available to taper steps.³¹⁹

Increased substrate temperature increases the adatom surface diffusion which tends to negate the effects of self-shadowing, as discussed in Section 6.2. Thus, Zone 2 structures, in which surface diffusion dominates over shadowing, can be produced on smooth substrates at $T/T_m \approx 0.5$. However, for the more extreme case of overcoming the shadow boundaries induced by substrate steps, $T/T_m \geq 0.6$ is recommended.³³² This corresponds to a temperature of about 300°C for aluminum.

The combination of planetary or other substrate motion, tapered steps, and substrate heating has proven adequate for providing acceptable aluminum metallization on most conventional devices. However, as noted above, tapered steps cannot be used for VLSI. In addition, it is projected that VLSI metallization will require refractory metal layers.³¹⁹ See Section 7.2. The condition $T/T_m \geq 0.6$ cannot be satisfied for refractory metals. For example, the melting point of W is 3410°C (3683°K). $T/T_m = 0.6$ corresponds to a temperature of 2209°K or 1936°C, which is above the melting point of Si (1410°C). See Table 10. Thus it is anticipated that bias sputtering will be required to provide step coverage on VLSI devices.³¹⁹

In the bias sputtering method, improvements in surface coverage are achieved by resputtering material previously deposited.^{335,336} Thus some of the material deposited at the bottom of a step is resputtered at small angles and redeposited on the side walls of the step, as shown schematically in Figure 63a. The process can also be enhanced by the "forward sputtering" produced by ions striking at an oblique incidence relative to the surface of the step, as shown in Figure 63b. The bias voltage and current must be selected to provide a proper balance between the deposition and resputtering processes. Figure 64, drawn from SEM micrographs,³³⁶ shows the effect of bias voltage in contouring SiO₂ films at a step. At -60V the bias has successfully eliminated crevice formation from the base of the step. The sputtering rate is greater at oblique ion incidence, as discussed in Section 5.4. Thus at a bias of -120V, the fast erosion of the angular surface over the step is clearly apparent. Recent experiments have demonstrated that

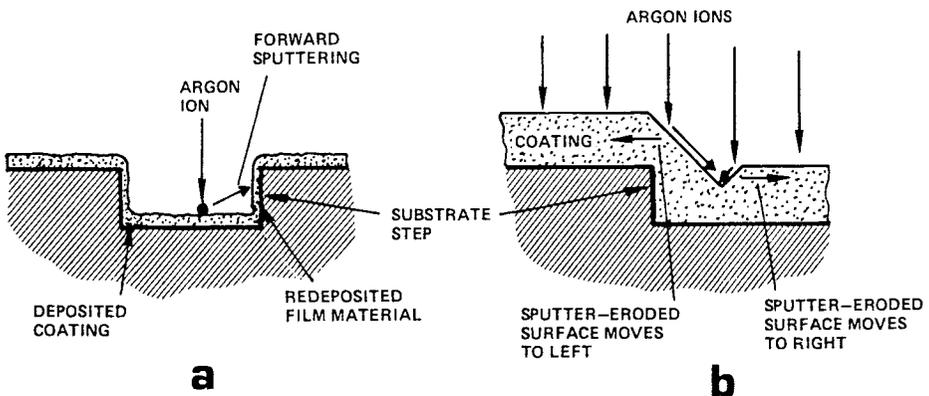


Figure 63: Schematic drawing showing effect of ion bombardment, due to substrate bias, on coating thickness distribution over substrate step.

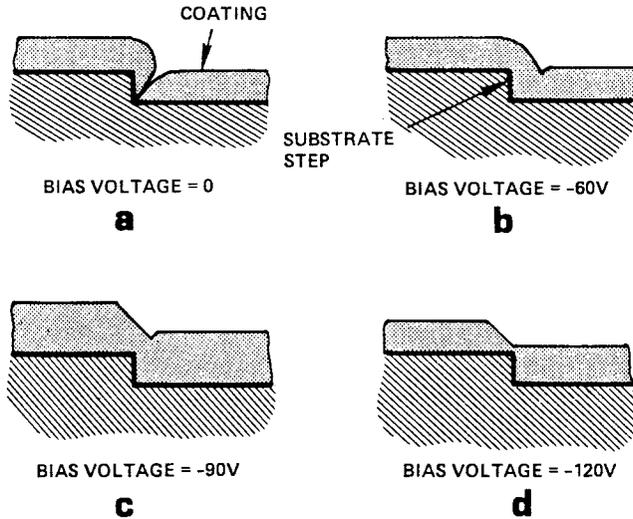


Figure 64: Drawing, based on SEM micrographs from Reference 336, which shows the effect of substrate bias on the thickness distribution of sputtered SiO_2 over a substrate step.

excellent step coverage of SiO_2 over straight edge profiles can be obtained with a combination of rf sputtering from a planar magnetron source and an rf bias applied to the substrates.^{337,338} Furthermore, it has now been shown that proper programming of the deposition and re-sputtering will permit surface insulator layers to be leveled in anticipation of subsequent layers of metallization.^{339-340a}

It is important to note that the use of bias sputtering can eliminate porosity and improve the coating microstructures, as well as improve the step coverage. See Section 6.4. However, it should also be remembered that bias sputtering of a multi-component material can cause composition changes because of preferential sputtering.

7.4 Radiation Damage

Electrically active defects can be produced in semiconductor surfaces by (1) ion or other heavy particle bombardment, (2) electron bombardment, and (3) x-ray and ultraviolet radiation. All three of these elements are present in conventional planar diode sputtering sources, and numerous examples of anomalous junction behavior have been reported for sensitive semiconductor interfaces formed using sputter-related technologies.³⁴¹⁻³⁴⁴ The significance of the defects depends on the nature of the electronic device. MOS devices are particularly sensitive to such defects, and the fabrication of MOS capacitors has been one of the primary methods for evaluating the tendency of various processes for producing radiation damage.

Filament- and induction-heated evaporation sources (see Figure 6) have the advantage that they produce none of the radiation effects cited above, but the disadvantage that the controlled deposition of refractory metals and stoichiometric alloys such as Al-Si-Cu is difficult. Therefore, the use of these sources for device production has become limited. However, they are effective for fabricating test devices that can serve as comparative references in judging the damage produced by other deposition methods.

In the case of electron beam evaporation, reflected electrons can cause device damage. Thus, it is reported that electron beam evaporation of metal layers in the fabrication of MOS capacitors introduced defects in SiO_2 that yielded threshold voltage instabilities which could not always be removed by annealing.³⁴⁵

Substrate bombardment by electrons and ions should be largely eliminated in properly designed dc driven magnetrons. There may be some bombardment by reflected and neutralized ions, particularly with cylindrical-post magnetrons. However, detailed evaluations of this mechanism as a source of device damage have not been reported. Studies of radiation damage to MOS capacitors metallized using gun-type magnetrons (Fig. 27I) have shown that trapping levels are created that cannot be completely annealed out at temperatures up to 500°C .^{346,347} The cause was concluded to be uv radiation. Photons with energies that are greater than the SiO_2 band gap (≈ 8 eV), and therefore that can produce radiation damage, are emitted by the plasma. The damage could be eliminated by precoating the wafers with a 300Å thick layer of evaporated Al using an induction heated source. Because of the strong attenuation of uv photons in such layers, it has been concluded that this radiation should not present a problem in many device metallization applications.³⁴⁶ In fact, it has been reported that post-metallization annealing to remove damage was not required for MOS structures that were metallized in magnetron systems at substrate temperatures of about 300°C .³⁴⁸ A recent study reports damage introduced during dc magnetron sputter deposition of Ti-W onto n-Si.^{348a}

The primary occurrence of heavy damage is during sputter cleaning and dry etching processes such as reactive ion etching. Defects created in the Si surface layer during ion beam etching can arise from three sources: (1) radiation damage arising from the impact of the ions in the beam, (2) implantation of the ions in the beam, and (3) implantation of impurities in the beam.³⁴⁹ Impurities are a particular problem because they cannot be removed by annealing. A study of damage induced in Si by Ar ion milling and reactive ion etching indicates that the interface states are strongly dependent on the bombarding ion energy and the etching gas.³⁵⁰ Recent studies have identified a trivalently bonded Si defect as the probable source of interface states at the SiO_2/Si interface of MOS capacitor structures fabricated on ion beam etched and reactive ion etched Si.³⁵¹ Clearly this is a contemporary area of research. However, as noted in Section 7.2, it is generally agreed that VLSI circuit fabrication using dry etching methods will require annealing, and that the temperatures will be in the 600°C range, or perhaps higher.

REFERENCES

1. R. Glang, R.A. Holmwood, and J.R. Kurtz in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang ed., McGraw-Hill, New York (1970) pp. 2-1 to 2-142.
2. J.F. O'Hanlon, *A User's Guide to Vacuum Technology*, Wiley, New York (1980).
3. G.L. Weissler and R.W. Carlson, editors, *Vacuum Physics and Technology- Vol. 14, Methods of Experimental Physics*, Academic Press, New York (1979).
4. J.F. O'Hanlon, *J. Vac. Sci. Technol.*, A1:228 (1983).
5. W.M. Brubaker, in *Vacuum Physics and Technology- Vol. 14, Methods of Experimental Physics*, G.L. Weissler and R.W. Carlson ed., Academic Press, New York (1979) pp. 183 to 192.
6. J.A. Thornton, "Considerations Relating to Pumping Requirements for Sputtering Metallic Coatings onto Plastic Substrates," Proceedings 20th Annual Technical Conference of the Society of Vacuum Coaters, Atlanta, GA (April 1977) p. 5.
7. B.B. Dayton, in *Handbook of Vacuum Physics*, A.H. Beck ed., MacMillan, New York (1964) p. 3 to 32.
8. L. Holland, *Vacuum Deposition of Thin Films*, Chapman and Hall, London (1966).
9. R. Glang, in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang ed., McGraw-Hill, New York (1970) pp. 1-3 to 1-130.
10. R.F. Bunshah, in *Deposition Technologies for Films and Coatings*, R.F. Bunshah et. al., Noyes Publications, Park Ridge, NJ (1982) pp. 83-169.
11. R.W. Birkmire, R.B. Hall, J.E. Phillips, and J.D. Meakin, "Cadmium Sulfide/Copper Sulfide Cell Research," Final Report Contract No. XS-9-8309-1, Institute of Energy Conversion, University of Delaware, Newark, DE 19711.
12. J.T. Cox and G. Hass, *Physics of Thin Films*, 2:239 (1964).
13. M.H. Francombe and J.E. Johnson, *Physics of Thin Films*, 5:143 (1969).
14. H.H. Wieder, *Intermetallic Semiconducting Films*, Pergamon Press, New York (1970).
15. A. Lopez Otero and L.D. Hass, *Thin Solid Films*, 23:1 (1974).
16. A.M. Hermann and L.B. Fabrick, *J. Crystal Growth*, 61:658 (1983) and L.B. Fabrick, *J. Crystal Growth*, to be published.
- 16a. S. Fuke, S. Kawarabayashi, K. Kwahara, and T. Imai, *J. Appl. Phys.* 60:2368 (1986).
17. Y.S. Tyan and E.A. Perez-Albuerne, in *Proceedings 16th IEEE Photovoltaic Specialists Conf.*, San Diego, CA (Sept. 1982) p. 794.
18. K. Mitchell, A.L. Fahrenbruch and R.H. Bube, *J. Vac. Sci. Technol.*, 12:909 (1975).
19. R.A. Mickelson and W.S. Chen, in *Proceedings 15th IEEE Photovoltaic Specialists Conf.*, Orlando, FL (May 1981) p. 800.
20. E. Ritter, *J. Vac. Sci. Technol.*, 3:225 (1966).
21. W. Heitman, *Appl. Opt.*, 10:2419 (1971).
22. J.A. Thornton, "Designing the Surface for the Job - Thin Film Application," Proceedings 19th National SAMPE Symposium, Buena Park, CA (April 1974).
23. R.F. Bunshah and A.C. Raghuram, *J. Vac. Sci. Technol.*, 9:1385 (1972).
24. A.Y. Cho and J.R. Arthur, *Prog. Solid State Chem.*, 10:157 (1975).
25. L.L. Chang and R. Ludeke, in *Epitaxial Growth, Part A*, ed. by J.W. Matthews, Academic Press, New York (1975) p. 37.
26. R.Z. Bachrach, "Molecular Beam Epitaxial Evaporative Growth," in *Crystal Growth*, B.R. Pamplin ed., Pergamon, New York (1980) p. 121.
27. C.E.C. Wood, *Physics of Thin Films*, 11:35 (1980).
28. K. Ploog, *Ann. Rev. Mater. Sci.*, 11:171 (1981).
29. K. Ploog, *Ann. Rev. Mater. Sci.*, 12:123 (1982).

30. J.C. Bean, in *Impurity Doping*, ed. F.F.Y. Wang, North Holland, Amsterdam (1981). p. 177.
31. J.C. Bean and S.R. McAfee, *J. de Physique*, 43:C-5 (1982).
32. W.T. Tsang, *J. Vac. Sci. Tech.* B3:666 (1985).
33. D. Huet, L. Lambert, D. Bonnevie, and D. Dufresne, *J. Vac. Sci. Technol.* B3, 823 (1985).
34. P.E. Luscher, *Thin Solid Films*, 83, 125 (1981).
35. C.T. Foxon, *J. Vac. Sci. Technol.* B1, 293 (1983).
36. J.R. Arthur and W.C. Johnson, *Industrial Research/Development*, (Jan 1979) p. 107.
37. N. Matsunaga, T. Suzuki, K. Takahashi, *J. Appl. Phys.*, 49:5710 (1978).
38. J.E. Greene and S.A. Barnett, *J. Vac. Sci. Technol.*, 21:285 (1982).
39. J.C. Bean, *J. Vac. Sci. Technol.*, A1:540 (1983).
40. Y. Ota, *J. Appl. Phys.*, 51:1102 (1980).
41. Y. Ota, *J. Vac. Sci. Technol.*, A2:393 (1984).
42. A.C. Gossard, in *Treatise on Materials Science and Technology*, Vol. 24, *Preparation and Properties of Thin Films*, ed. by K.N. Tu and R. Rosenberg, Academic Press, New York (1982) p. 13.
43. L.L. Chang, *J. Vac. Sci. Technol.*, B1:120 (1983).
44. B.J. Shaw and R.P. Miler, U.S. Patent 3,918,100, Nov. 11, 1975.
45. T. Robertson and D.T. Morrison, *Thin Solid Films*, 27:19 (1975).
46. D.T. Morrison and T. Robertson, *Thin Solid Films*, 15:87 (1973).
47. J.K. Howard, *J. Vac. Sci. Technol.*, A3:657 (1985).
48. W.Y. Lee, *J. Vac. Sci. Technol.* A3:640 (1985).
49. G.K. Wehner, *Advances Elec. and Elec. Phys.*, 7:239 (1955).
50. E. Kay, *Advances Elec. and Elec. Phys.*, 17:245 (1962).
51. L.I. Maissel, *Physics of Thin Films*, 3:61 (1966).
52. E. Kay, in *Techniques of Metals Research*, R.F. Bunshah, ed., Vol. 1, Part 3, Interscience, New York (1968) p. 1269.
53. G.K. Wehner and G.S. Anderson, in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang, ed., McGraw-Hill, New York (1970) p. 3-1.
54. L.I. Maissel, in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang, ed., McGraw-Hill, New York (1970) p. 4-1.
55. W.D. Westwood, *Progress in Surface Sci.*, 7:71 (1976).
56. J.L. Vossen and J.J. Cuomo, in *Thin Film Processes*, J.L. Vossen and W. Kern, ed., Academic Press, New York (1978) p. 11.
57. B. Chapman, *Glow Discharge Processes: Sputtering and Plasma Etching*, Wiley, New York (1980).
58. R. Behrisch, ed., *Topics in Applied Physics*, Vol. 47, *Topics in Applied Physics*, Vo. 52, R. Behrsh, ed., Springer-Verlag, New York (1983).
59. J.A. Thornton, in *Deposition Techniques for Films and Coatings*, R.F. Bunshah, et.al., Noyes Press, Park Ridge, NJ (1982) p. 170.
60. J.R. Roth in *Topics in Applied Physics*, Vol. 52, R. Behrisch, ed., Springer-Verlag, New York (1983) p. 91.
61. P. Sigmund, *Phys. Rev.*, 184:383 (1969).
62. P. Sigmund, in *Inelastic Ion-Surface Collisions*, ed. by N.H. Tolk, J.C. Tully, W. Heiland and C.W. White, Academic Press, New York (1977) p. 121.
63. P. Sigmund, *J. Vac. Sci. Technol.*, 17:396 (1980).
64. P. Sigmund, in *Sputtering by Particle Bombardment 1*, Vol. 47, *Topics in Applied Physics*, R. Behrisch, ed. Springer-Verlag, New York (1981) p. 9.
65. D.E. Harrison, Jr., N.S. Levy, J.P. Johnson III, and H.M. Effron, *J. Vac. Sci. Technol.*, 39:3742 (1968).
66. R.P. Webb and D.E. Harrison, *Vacuum* 34:847 (1984).
67. T. Ishitani and R. Shimizu, *Phys. Lett.*, 46A:487 (1974).

68. H.F. Winters and P. Sigmund, *J. Appl. Phys.*, 45:4760 (1974).
69. H. Oechsner, *Z. Phys.*, 37:261 (1973).
70. P.K. Rol, D. Onderlindern, and T. Kistemaker, *Trans. 3rd Int. Vac. Congr.*, Pergamon, Oxford (1966) p. 75.
71. G.M. McCracken, *Rep. Prog. Phys.*, 38:241 (1975).
72. H.D. Hagstrum, in *Inelastic Ion-Surface Collision*, ed. by N.H. Tolk, J.C. Tully, W. Heiland, and C.W. White, Academic Press, New York (1977) p. 1.
73. H.R. Winters and E. Kay, *J. Appl. Phys.*, 38:3928 (1967).
74. I. Brodie, L.T. Lamont, Jr. and R.L. Jepson, *Phys. Rev. Lett.*, 21:1224 (1968).
75. W.W. Lee and D. Oblas, *J. Vac. Sci. Technol.*, 7:129 (1970).
76. W.W.Y. Lee and D. Oblas, *J. Appl. Phys.*, 46:1728 (1975).
77. J.A. Thornton and D.W. Hoffman, *J. Vac. Sci. Technol.*, 18:203 (1981).
78. D.W. Hoffman and J.A. Thornton, *J. Vac. Sci. Technol.*, 17:380 (1980).
79. O. Kubaschewski and C.B. Alcock, *Metallurgical Thermochemistry*, 5th Ed., Pergamon Press, New York (1979).
80. R. Miranda and J.M. Rojo, *Vacuum*, 34:1069 (1984).
81. G.H. Kinchin and R.S. Pease, *Rep. Prog. Phys.* 18:1 (1955).
82. G.K. Wehner, *J. Vac. Sci. Technol.* A1:487 (1983).
83. E.V. Kornelsen, *Can. J. Phys.*, 42:364 (1964).
84. P. Blank and K. Wittmaack, *J. Appl. Phys.*, 50:1519 (1979).
85. H. Oeschner and W. Gerhard, *Surf. Sci.*, 44:480 (1974).
86. W. Gerhard and H. Oechsner, *Z. Physik*, B22:41 (1975).
87. W. Gerhard, *Z. Physik*, B22:31 (1975).
88. D.E. Harrison, Jr. and C.B. Delaplain, *J. Appl. Phys.*, 47:2252 (1976).
89. J.W. Coburn, E.W. Eckstein, and E. Kay, *J. Vac. Sci. Technol.*, 12:151 (1975).
90. J.W. Coburn, E. Taglauer, and E. Kay, *Japan. J. Appl. Phys. Suppl.*: 2:501 (1974).
91. D.M. Gruen, S.L. Guadoso, R.L. McBeth, and J.L. Kerner, *J. Chem Phys.*, 60:89 (1974).
92. D.M. Gruen, P.A. Finn, and D.L. Page, *Nucl. Technol.*, 29:309 (1976).
93. A.R. Krauss and D.M. Gruen, *Appl. Phys.*, 14:89 (1977).
94. C. Steinbruchel and D.M. Gruen, *Surf. Sci.*, 93:299 (1980).
95. K.L. Chopra, *Thin Film Phenomena*, McGraw-Hill, New York (1969).
96. J.M. Schroerer, R.N. Rhodin, and R.C. Bradley, *Surface Sci.*, 34:571 (1973).
97. J.J. Cuomo, R.J. Gambino, J.M.E. Harper, J.D. Kuptsis, and J.C. Webber, *J. Vac. Sci. Technol.*, 15:281 (1978).
98. J.J. Hanak and J.P. Pellicane, *J. Vac. Sci. Technol.*, 13:406 (1976).
99. H. Oechsner, *Appl. Phys.* 8:185 (1975).
100. M.W. Thompson, *Philos. Mag.* 18, 377 (1968).
101. G. betz and G.K. Wehner, in *Topics in Applied Physics*, Vol. 52, R. Behrish ed., Springer-Verlag, New York (1983) p. 11.
102. R.V. Stuart and G.K. Wehner, *J. Appl. Phys.*, 33:2351 (1962).
103. D. Rosenberg and G.K. Wehner, *J. Appl. Phys.*, 33:1842 (1962).
104. R. Behrish, *Ergeb. Exakt. Naturw.*, 35:295 (1964).
105. H.H. Anderson, in *Topics in Applied Physics*, vol. 47, R. Behrish ed, Springer-Verlag, New York (1981) p. 145.
106. M.P. Seah, *Thin Solid Films*, 81:279 (1981).
107. B. Navinsek, *Progress in Surface Sci.*, 7:49 (1976).
108. O. Almen and G. Bruce, *Transac. 8th National Vac. Sym.* New York, Pergamon Press (1962) p. 245.
109. W.A. Grant and G. Carter, *Vacuum*, 15:477 (1965).
110. J.W. Coburn, *J. Vac. Sci. Technol.*, 13:1037 (1976).
111. J.W. Coburn, *Thin Solid Films*, 65:371 (1979).
112. E. Gillam, *J. Phys. Chem. Solids*, 11:55 (1959).
113. M.J. Nobes, J.S. Colligon, and G. Carter, *J. Matl. Sci.*, 4:730 (1969).

114. G. Carter, J.S. Colligon, and M.J. Nobes, *J. Matl. Sci.*, 6:115 (1971).
115. P. Sigmund, *J. Matl. Sci.*, 8:1545 (1973).
116. G. Carter, B. Navinsek, and J.L. Witton, in *Topics in Applied Physics*, R. Behrisch ed., Springer-Verlag, New York (1983) p. 231.
117. S.D. Dahlgren and S.D. McClanahan, *J. Appl. Phys.*, 42:1514 (1972).
118. A.D.G. Stewart and M.W. Thompson, *J. Matl. Sci.*, 4:56 (1969).
119. M.H. Witcomb, *J. Matl. Sci.*, 9:551 (1974).
120. J.E. Greene, B.R. Natarjan, and F. Sequeda-Osorio, *J. Appl. Phys.*, 49:417 (1978).
121. G.K. Wehner and D.J. Hajicek, *J. Appl. Phys.*, 42:1145 (1971).
122. M.L. Tarnng and G.K. Wehner, *J. Appl. Phys.*, 43:2268 (1972).
123. G.K. Wehner, *J. Vac. Sci. Technol.* A3:1821 (1985).
124. T. Oohashi and S. Yamanaka, *Japan. J. Appl. Phys.*, 11:1581 (1972).
125. R. Shimizu, *Japan. J. Appl. Phys.*, 13:228 (1974).
126. J.L. Vossen, *J. Vac. Sci. Technol.*, 8:751 (1971).
127. D.R. Wheeler and W.A. Brainard, *J. Vac. Sci. Technol.*, 15:24 (1978).
128. J.L. Vossen and J.J. O'Neill, Jr., *RCA Rev.* 29:149 (1968).
129. D.B. Fraser and H. Melchior, *J. Appl. Phys.*, 43:3120 (1972).
130. A.B. Cistola, *J. Vac. Sci. Technol.*, 17:420 (1980).
131. J.J. Cuomo and R.J. Gambino, *J. Vac. Sci. Technol.*, 12:79 (1975).
132. J.A. Thornton, in *Deposition Technologies for Films and Coatings*, R.F. Bunshah, et al., Noyes Publications, Park Ridge, NJ (1982) p. 19.
133. J.S. Logan, N.M. Mazza, and P.D. Davidse, *J. Vac. Sci. Technol.*, 6:120 (1969).
134. J.S. Logan, *IBM J. Res. Develop.*, 14:172 (1970).
135. J.L. Vossen, *J. Vac. Sci. Technol.*, 8:S12 (1971).
136. E. Taglauer and W. Heiland eds., *Inelastic Particle-Surface Collisions*, Springer-Verlag, New York (1981).
137. E.S. McDaniel, *Collision Phenomena in Ionized Gases*, Wiley, New York (1964).
138. S.C. Brown, *Basic Data of Plasma Physics*, MIT Press, Cambridge, MA (1959).
139. E.W. McDaniel, *The Mobility and Diffusion of Ions in Gases*, Wiley, New York (1973) p. 132.
140. W.D. Davis and T.A. Vanderslice, *Phys. Rev.*, 131:219 (1963).
141. A.J. Stirling and W.D. Westwood, *J. Appl. Phys.*, 41:742 (1970).
142. J.E. Greene, University of Illinois, private communication.
143. W.D. Westwood, *J. Vac. Sci. Technol.*, 15:1 (1978).
144. K. Meyer, I.K. Schuller, and C.M. Falco, *J. Appl. Phys.*, 52:5803 (1981).
145. A. Gras-Marti, and J.A. Valles-Abarca, *J. Appl. Phys.*, 54:1071 (1983).
146. R.E. Somekh, *Vacuum*, 34:987 (1984).
147. D.J. Ball, *J. Appl. Phys.*, 43:3047 (1972).
148. L.T. Lamont, Jr. and A. Lang, *J. Vac. Sci. Technol.*, 7:198 (1970).
149. S.S. Lau, R.H. Mills, and D.G. Muth, *J. Vac. Sci. Technol.*, 9:1196 (1972).
150. E.D. McClanahan and R. Moss, "Enhanced Thermionically Supported Discharge Sputtering," paper presented at American Vacuum Society Thin Film Division Topical Symposium on Sputtering, San Diego, April 6-8, 1984.
151. J.A. Thornton, *Metal Finishing*, 77:45 (1979).
152. J.A. Thornton, *J. Vac. Sci. Technol.*, 15:171 (1978).
153. J.A. Thornton, and A.S. Penfold, in *Thin Film Processes*, J.L. Vossen and W. Kern, ed., Academic Press, New York (1978) p. 75.
154. J.S. Thornton, *Thin Solid Films*, 80:1 (1981).
155. N. Hosokawa, T. Tsukada, and T. Misumi, *J. Vac. Sci. Technol.*, 14:143.
156. D.W. Hoffman, *Thin Solid Films*, 96:217 (1982).
157. U. Heisig, K. Goedicke, and S. Schiller, *Proceedings 7th Int'l Sym. Electron and Ion Beam Science and Technology*, Washington, D.C. 1976, Electrochemical Society, Princeton, NJ (1976) p. 129.

158. R.K. Waits, in *Thin Film Processes*, J.L. Vossen and W. Kern, ed., Academic Press, New York (1978) p. 131.
159. D.B. Fraser, in *Thin Film Processes*, J.L. Vossen and W. Kern, ed., Academic Press, New York (1978) p. 115.
160. A.S. Penfold, *Metal Finishing*, 77:33 (1979).
161. R.K. Waits, *J. Vac. Sci. Technol.*, 15:179 (1978).
162. J.A. Thornton, U.S. Patent 4,126,530 (1978).
- 162a. B. Window and N. Savvides, *J. Vac. Sci. Technol.*, A4:196 (1986).
163. J.A. Thornton, *Thin Solid Films*, 54:23 (1978).
164. J.A. Thornton and J.L. Lamb, *Thin Solid Films*, 119:87 (1984).
165. B. Window and F. Sharples, *J. Vac. Sci. Technol.* A3:10 (1985).
166. J.M.E. Harper, in *Thin Film Processes*, J.L. Vossen and W. Kern, ed., Academic Press New York (1978) pp. 175-206.
167. H.R. Kaufman, J.J. Cuomo, and J.M.E. Harper, *J. Vac. Sci. Technol.*, 21:725 (1982).
168. P.J. Martin, R.P. Netterfield, W.G. Sainty, and C.G. Pacey, *J. Vac. Sci. Technol.* A2:341 (1984).
169. J.M.E. Harper, H.T.G. Hentzell, and J.J. Cuomo, *J. Vac. Sci. Technol.* A2:405 (1984).
170. G.A. Al-Jumaily, J.J. McNally, J.R. McNeil, and W.C. Herrmann, Jr., *J. Vac. Sci. Technol.* A3:651 (1985).
171. C. Weissmantel, K. Bewilogua, K. Breuer, D. Dietrich, U. Ebersbach, H.-J. Erler, R. Rau, and G. Reisse, *Thin Solid Films*, 96:31 (1982).
172. D. Bouchier, G. Gautherin, C. Schwebel, A. Bosseboeuf, B. Agius, and S. Rigo, *J. Electrochem. Soc.*, 130:638 (1983).
173. J.M.E. Harper, J.J. Cuomo, and H.R. Kaufman, *J. Vac. Sci. Technol.*, 21:737 (1982).
174. J.M.E. Harper, J.J. Cuomo, and H.R. Kaufman, *Ann. Rev. Mater. Sci.*, 13:413 (1983).
175. J.M.E. Harper, J.J. Cuomo, and H.R. Kaufman, *J. Vac. Sci. Technol.*, A1:337 (1983).
176. W.A. Pliskin, P.D. Davidse, H.S. Lehman, and L.I. Maissel, *IBM J. Res. Develop.*, 11:461 (1967).
177. R.C. Mosely, C.A. Sans, S.D. Campbell, and L.T. Lamont, Jr., *J. Vac. Sci. Technol.* A3:596 (1985).
178. R.S. Nowicki, *J. Vac. Sci. Technol.*, 14:127 (1977).
179. J.A. Thornton and J. Chin, *Am. Ceramic Soc. Bul.*, 56:504 (1977).
180. M.C. Chen, *J. Electrochem. Soc.*, 119:887 (1972).
181. P.L. Young, F.P. Fehler, and A.J. Whitman, *J. Vac. Sci. Technol.*, 14:176 (1977).
182. J.O. Barnes, D.J. Leary, and A.G. Jordan, *J. Electrochem. Soc.*, 127:1636 (1980).
183. W.W. Molzen, *J. Vac. Sci. Technol.*, 12:99 (1975).
184. G.H. Hewig, K. Jain, F.O. Sequeda, R. Tom, and P.W. Wang, *Thin Solid Films*, 88:67 (1982).
185. M. Suzuki, Y. Enomoto, T. Murakami, and T. Inamura, *J. Appl. Phys.*, 53:1622 (1982).
186. J.L. Vossen, *J. Vac. Sci. Technol.*, 8:751 (1971).
187. H. Matsunami, H. Masahiro, and T. Tanaka, *J. Electronic Materials*, 8:249 (1979).
188. M.H. Brodsky, H.S. Title, K. Weiser, and G.D. Pettit, *Phys. Rev.* B1:2632 (1970).
189. S.A. Barnet, G. Bajor, and J.E. Greene, *Appl. Phys. Lett.*, 37:734 (1980).
190. R.H. Cornely, L. Suchow, M. Mulligan, T.O. Chan, R. Haq, and C. Mehta, *J. Vac. Sci. Technol.*, 18:190 (1981).
191. A.F. Carroll, C.F. Smith, L.C. Burton, and L.H. Slack, *J. Electrochem. Soc.*, 128:1769 (1981).
192. S.B. Krupanidhi and M. Sayer, *J. Vac. Sci. Technol.* A2:303 (1984).
193. P.D. Davidse, *Vacuum*, 17:139 (1967).
194. B.S. Probyn, *Vacuum*, 18:253 (1968).
195. G.N. Jackson, *Thin Solid Films*, 5:209 (1970).
196. I. Brodie, L.T. Lamont, Jr., and D.O. Meyers, *J. Vac. Sci. Technol.*, 6:124 (1969).

197. H.R. Koenig and L.I. Maissel, *IBM J. Res. Develop.*, 14:168 (1970).
198. F.E. Terman *Electronics and Radio Engineering*, McGraw Hill, New York (1955).
199. C.M. Horwitz, *J. Vac. Sci. Technol.* A1:60 (1983).
200. K. Köhler, J.W. Coburn, D.E. Horne, E. Kay and J.H. Keller, *J. Appl. Phys.*, 57:59 (1985).
201. D.H. Grantham, E.L. Paradis, and D.J. Quinn, *J. Vac. Sci. Technol.*, 7:343 (1970).
202. S. Maniv and W.D. Westwood, *J. Vac. Sci. Technol.*, 17:743 (1980).
203. R.N. Castellano, *Thin Solid Films*, 46:312 (1977).
204. S. Schiller, U. Heisig, K. Steinfelder, and J. Strümpfel, *Thin Solid Films*, 63:369 (1979).
205. V.F. Drobny and D.L. Pulfrey, *Thin Solid Films*, 61:89 (1979).
206. S. Maniv, W.D. Westwood, and E. Colombini, *J. Vac. Sci. Technol.*, 20:162 (1982).
207. S. Maniv, C.J. Miner, and W.D. Westwood, *J. Vac. Sci. Technol.* A1:1370 (1983).
208. J.A. Thornton and V.L. Hedgoth, *J. Vac. Sci. Technol.*, 13:117 (1976).
209. R. McMahon, J. Affinito, and R.R. Parsons, *J. Vac. Sci. Technol.*, 20:376 (1982).
210. J. Affinito, N. Fortier, and R.R. Parsons, *J. Vac. Sci. Technol.* A2:316 (1984).
211. A. Mumtaz and W.H. Class, *J. Vac. Sci. Technol.*, 20:345 (1982).
212. J.P. Noël, D.C. Houghton, G. Este, F.R. Shepherd, and H. Plattner, *J. Vac. Sci. Technol.* A2:284 (1984).
213. M. Östling, S. Nygren, C.S. Petersson, N. Nordström, P. Wiklund, R. Buchta, H.O. Blom, and S. Berg, *J. Vac. Sci. Technol.*, 16:200 (1979).
ed. McGraw Hill, New York (1970) pp. 8-3 to 8-43.
214. C.J. Mogab, P.M. Petroff, and T.T. Sheng, *J. Electrochem. Soc.*, 122:815 (1975).
215. A.D. Jonath, W.W. Anderson, J.A. Thornton, and D.G. Cornog, *J. Vac. Sci. Technol.*, 16:200 (1979).
216. J.A. Thornton, D.G. Cornog, and W.W. Anderson, *J. Vac. Sci. Technol.*, 18:199 (1981).
217. J.A. Thornton, D.G. Cornog, R.B. Hall, S.P. Shea and J.D. Meakin, *J. Vac. Sci. Technol.* A2:307 (1984).
218. D.J. Harra and W.H. Haywood, *Supplemento al Nuovo Cimento*, 5:56 (1967).
219. S. Maniv and W.D. Westwood, *Surface Science*, 100:108 (1980).
220. T.M. Reith and P.J. Ficalora, *J. Vac. Sci. Technol.* A1:1362 (1983).
221. J.H. Greiner, *J. Appl. Phys.*, 42:5151 (1971).
222. J.A. Thornton, in *Proceedings of the Sixth Symposium on Ion Sources and Ion-Assisted Technology*, Tokyo, Japan, June 7-9, 1982, p. 363. Also see J.A. Thornton, *Rev. Brasil. Apl. Vac.*, 2:89 (1982).
223. C.H. Griffiths and H.K. Eastwood, *J. Appl. Phys.*, 45:2201 (1974).
224. S. Schiller, U. Heisig, K. Goedicke, K. Schade, G. Teshner, and J. Henneberger, *Thin Solid Films*, 64:455 (1979).
225. J. Affinito and R.R. Parsons, *J. Vac. Sci. Technol.* A2:1275 (1984).
226. Telic Company, Santa Monica, CA, unpublished research.
- 226a. J.A. Thornton, *Surface Engineering*, 2:283 (1986).
227. M.J. Brett, R.W. McMahon, J. Affinito, and R.R. Parsons, *J. Vac. Sci. Technol.* A1:352 (1983).
228. M.J. Brett, Ph.d. Dissertation, Department of Physics, University of British Columbia, Vancouver BC, June 1985.
229. J.A. Thornton and J.L. Lamb, *Investigation of Continuous Deposition of AMA Type Coatings onto Strip and Sheet by Sputtering*, Final Report Contract No. DE-AC0480AL13116, Report No. TELIC 82-3, U.S. Dept. of Energy, Albuquerque, New Mexico, December 1982. (Available from National Technical Information Service, U.S. Department of Commerce, Springfield, VA).
230. S. Schiller, U. Heisig, K. Steinfelder, J. Strümpfel, W. Sieber, *Vakuum-Technick*, 30:1 (1981).
231. G. Este and W.D. Westwood, *J. Vac. Sci. Technol.* A2:1238 (1984).

232. J.A. Thornton, D.G. Cornog, R.B. Hall, and L.C. DiNetta, *J. Vac. Sci. Technol.*, 20: 296 (1982).
233. M.B. Francombe, in *Epitaxial Growth, Part A*, J.W. Matthews, ed., Academic Press, New York (1975) p. 109-181.
234. *Physics Today*, 37:B6-172 (1984).
235. C. Neugebauer, in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang, ed. McGraw Hill, New York (1970) pp. 8-3 to 8-43.
236. I.H. Khan, in *Handbook of Thin Film Technology*, L.I. Maissel and R. Glang, ed. McGraw Hill, New York (1970) pp. 10-1 to 10-65.
237. J.W. Geus, in *Chemisorption and Reactions on Metallic Films*, Vol. 1, J.R. Anderson ed., Academic Press, New York (1971) p. 129-224.
238. M.J. Faktor and J. Garrett, *Growth of Crystals from the Vapor*, Wiley, New York (1974).
239. J.A. Venables and G.L. Price, in *Epitaxial Growth, Part B*, J.W. Matthews, ed., Academic Press, New York (1975) p. 381-436.
240. B. Lewis and J.C. Anderson, *Nucleation and Growth of Thin Films*, Academic Press, New York (1978).
241. J.A. Thornton, *Ann. Rev. Mater. Sci.*, 7:239 (1977).
242. A.G. Dirks and H.J. Leamy, *Thin Solid Films*, 47:219 (1977).
243. J.H. Brophy, R.M. Rose, and J. Wulff, *The Structure and Properties of Materials*, Vol. 2, Wiley, New York (1964) p. 192.
244. B.A. Movchan and A.V. Demchishin, *Phys. Met. Metallogr.*, 28:83 (1969).
245. J.A. Thornton, *J. Vac. Sci. Technol.*, 11:666 (1974).
246. S. Craig and G.L. Harding, *J. Vac. Sci. Technol.*, 19:205 (1981).
247. R. Messier, A. Giri, and R.A. Roy, *J. Vac. Sci. Technol.* A2:500 (1984).
248. R. Messier and R.C. Ross, *J. Appl. Phys.*, 53:620 (1982).
249. R.C. Ross and R. Messier, *J. Appl. Phys.*, 54:5744 (1983).
- 249a. J.A. Thornton, *J. Vac. Sci. Technol.*, A4:3059 (1986).
250. J.W. Patten, *Thin Solid Films*, 75:205 (1981).
251. A.I. Shaldervan and N.C. Nakhodkin, *Sov. Phys. Solid State*, 11:2773 (1970).
252. F. Jansen, *Thin Solid Films*, 78:15 (1981).
253. J.A. Thornton *J. Vac. Sci. Technol.*, 12:830 (1975).
254. J.E. Yehoda and R. Messier, *Applications of Surface Science*, 22/23:118 (1985).
255. H.T.G. Hentzell, C.R.M. Grovenor and D.A. Smith, *J. Vac. Sci. Technol.* A2:218 (1984).
256. P.S. Vincett, W.A. Barlow, and G.C. Roberts, *J. Appl. Phys.*, 48:3800 (1977).
257. P.S. Vincett, *J. Vac. Sci. Technol.*, 21:972 (1982).
258. K. Behrndt, in *Techniques of Metals Research*, Vol. 1, Part 3, "Techniques of Materials Preparation and Handling, R.F. Bunshah ed., Interscience, New York (1968) p. 1191.
259. K.G. Gunther, in *The Use of Thin Films in Physical Investigations*, J.C. Anderson, ed., Academic Press, New York (1966) p. 213.
260. C.T. Foxon, *J. Vac. Sci. Technol.* B1:293 (1983).
261. J.R. Arthur, *Surface Science*, 43:449 (1974).
262. C.T. Foxon and B.A. Joyce, *Surface Science*, 64:293 (1977).
263. C.T. Foxon and B.A. Joyce, *Surface Science*, 50:434 (1975).
264. J.E. Greene, in *Handbook of Semiconductors*, ed. by T.S. Moss and S.P. Keller, Vol. 3, North-Holland, Amsterdam (1980) p. 499.
265. J.E. Greene, *CRC Critical Reviews of Solid State and Materials Science*, 11:47 (1983).
266. J.E. Greene, *CRC Critical Reviews of Solid State and Materials Science*, 11:189 (1984).
267. D.M. Mattox, in *Deposition Technologies for Films and Coatings*, R.F. Bunshah, et.al., Noyes Press, Park Ridge, NJ (1982) p. 224.

268. D.M. Mattox, "Thin-Film Adhesion and Adhesive Failure-A Perspective," in Adhesion Measurements of Thin Films, Thick Films and Bulk Coatings, ASTM STP 640, K.L. Mittal, ed., American Society for Testing Materials (1978), p. 54.
269. L.I. Maissel and P.M. Schaible, *J. Appl. Phys.*, 36:237 (1965).
270. A.G. Blackman, *Metall. Trans.* 2:699 (1971).
271. A.G. Blackman, *J. Vac. Sci. Technol.*, 10:299 (1973).
272. M. Marinov, *Thin Solid Films*, 46:267 (1977).
273. R.E. Hurley, *Thin Solid Films*, 86:241 (1981).
274. S.A. Barnett, G. Bajor, and J.E. Greene, *Appl. Phys. Lett.*, 37:734 (1980).
275. J.E. Greene, S.A. Barnett, K.C. Cadien, and M.A. Ray, *J. Crystal Growth*, 56:389 (1982).
276. I.H. Kahn, *J. Appl. Phys.*, 44:14 (1973).
277. T. Narusawa, S. Shimizu, and S. Komiya, *J. Vac. Sci. Technol.*, 16:366 (1979).
278. S. Shimizu and S. Komiya, *J. Vac. Sci. Technol.*, 17:489 (1980).
279. J.E. Greene, *J. Vac. Sci. Technol.* B1:229 (1983).
280. S.A. Barnett and J.E. Greene, *Surface Science*, 128:401 (1983).
281. R.D. Bland, G.J. Kominiak, and D.M. Mattox, *J. Vac. Sci. Technol.*, 11:671 (1974).
282. D.M. Mattox and G.J. Kominiak, *J. Vac. Sci. Technol.*, 9:528 (1972).
283. D.M. Mattox and G.J. Kominiak, *J. Electrochem. Soc.*, 120:1535 (1973).
284. J.A. Thornton, *Thin Solid Films*, 40:335 (1977).
285. F.B. Koch, R.L. Meek, and D.V. McCaughan, *J. Electrochem. Soc.*, 121:558 (1974).
286. W.D. Sproul and M.H. Richman, *J. Vac. Sci. Technol.*, 12:842 (1975).
287. J.J. Cuomo and R.J. Gambino, *J. Vac. Sci. Technol.*, 14:152 (1977).
288. J.M. Poate, in *Treatise on Materials Science and Technology, Vol. 24, Preparation and Properties of Thin Films*, ed. by K.N. Tu and R. Rosenberg, Academic Press, New York (1982) p. 213.
289. B.Y. Tsaur, J.W. Mayer, and K.N. Tu, *J. Appl. Phys.*, 51:5326 (1980).
290. B.Y. Tsaur, PhD Thesis, California Institute of Technology (1980).
291. R.W. Hoffman, in *Physics of Thin Films, Vol. 3*, G. Hass and R.E. Thun, ed., Academic Press, New York (1966) p. 211.
292. P.G. Shewmon, *Transformations in Metals*, McGraw Hill, New York (1969).
293. R.W. Hoffman, in *Physics of Non-Metallic Thin Films*, H.S. Dupry and A. Cachard eds., Plenum Press, New York (1974).
294. M. Murakami, T.S. Kuan, and I.A. Blech, in *Treatise on Materials Science and Technology, Vol. 24, Preparation and Properties of Thin Films*, ed. by K.N. Tu and R. Rosenberg, Academic Press, New York (1982) p. 163.
295. C.J. Santoro, *J. Electrochem. Soc.*, 116:361 (1969).
296. W.B. Pennebaker, *J. Appl. Phys.*, 40:394 (1969).
297. S.K. Lahiri, *J. Appl. Phys.*, 41:3172 (1970).
298. P.R. Stuart, *Vacuum*, 19:507 (1969).
299. R.S. Wagner, A.K. Sinha, T.T. Sheng, H.J. Levinstein, and F.B. Alexander, *J. Vac. Sci. Technol.*, 11:582 (1974).
300. R.N. Castellano, M.R. Notis, and G.W. Simmons, *Vacuum*, 27:109 (1977).
301. D.W. Hoffman and M.R. Gaerttner, *J. Vac. Sci. Technol.*, 17:425 (1980).
302. E. Klokholm and B.S. Berry, *J. Electrochem. Soc.*, 115:823 (1968).
303. W. Buckel, *J. Vac. Sci. Technol.*, 6:606 (1969).
304. R.W. Hoffman, *Thin Solid Films*, 34:185 (1976).
305. H. Oikawa and Y. Nakajima, *J. Vac. Sci. Technol.*, 14:1153 (1977).
306. D.W. Hoffman and J.A. Thornton, *Thin Solid Films*, 40:355 (1977).
307. J.A. Thornton and D.W. Hoffman, *J. Vac. Sci. Technol.*, 14:164 (1977).
308. D.W. Hoffman and J.A. Thornton, *Thin Solid Films*, 45:387 (1977).
309. J.A. Thornton, J. Tabock, and D.W. Hoffman, *Thin Solid Films*, 64:111 (1979).
310. D.W. Hoffman and J.A. Thornton, *J. Vac. Sci. Technol.*, 20:355 (1982).
311. D.W. Hoffman and P.I. Goodsmith, "Decorative Metallizing by Magnetron Sput-

- tering: Effects of Geometry on Film Properties," paper #790217 presented at Society of Automotive Engineers winter meeting, Detroit, MI (1979).
312. K.W. Bieg, *Thin Solid Films*, 90:161 (1982).
 313. K.W. Bieg and C.R. Hills, *J. Vac. Sci. Technol.* A1:508 (1983).
 314. C.T. Wu, *Thin Solid Films*, 64:103 (1979).
 315. D.W. Hoffman and J.A. Thornton, *J. Vac. Sci. Technol.*, 16:134 (1979).
 316. D.W. Hoffman, *Thin Solid Films*, 107:353 (1983).
 317. J.A. Thornton and D.W. Hoffman, *J. Vac. Sci. Technol.* A3, 576 (1985).
 318. R. Iscoff, *Semiconductor International*, 5:57 (1982).
 319. J.L. Vossen, *J. Vac. Sci. Technol.*, 19:761 (1981).
 320. P.B. Ghate, *Thin Solid Films*, 93:359 (1982).
 321. P.B. Ghate, in *Deposition Technology for Films and Coatings*, R.F. Bunshah, et al., Noyes Press, Park Ridge, NJ (1982) p. 514.
 322. K.N. Tu, in *Treatise on Materials Science and Technology*, Vol. 24, *Preparation and Properties of Thin Films*, ed. by K.N. Tu and R. Rosenberg, Academic Press, New York (1982) p. 237.
 323. S.P. Murarka, *Silicides of VLSI Applications*, Academic Press, New York (1983).
 324. J.V. Vossen, *Bibliography on Metallization Materials and Techniques*, Vols. I-IV (1974-1983). Available from American Vacuum Society, 335 East 45th St. New York, NY 10017.
 325. K.N. Tu and J.M. Mayer, in *Thin Films, Interdiffusion & Reactions*, ed. by J.M. Poate, K.N. Tu, and J.M. Mayer, Wiley, New York (1973) p. 359.
 326. M. Wittmer, *J. Vac. Sci. Technol.*, A2:273 (1984).
 327. A.J. Learn, *J. Electrochem. Soc.*, 123:894 (1976).
 328. P.B. Ghate, *Thin Solid Films*, 83:195 (1981).
 329. S. Vaidya, D.B. Fraser, and W.S. Lindenberger, *J. Appl. Phys.*, 51:4475 (1980).
 330. E. Kinsbron, *Appl. Phys. Lett.*, 36:968 (1980).
 331. P.B. Ghate, *Proc. 19th Annu. Reliability Physics Symp.*, IEEE, New York (1981) p. 243.
 332. T.C. Tisone and J.B. Bindell, *J. Vac. Sci. Technol.*, 11:72 (1974).
 333. D.B. Fraser, W. Fichtner, and R. Knoell, "Step Coverage and VLSI," paper presented at 30th National Symposium of American Vacuum Society, Boston, Nov. 1-4, 1983.
 334. I.A. Blech and H.A. Vander Plas, *J. Appl. Phys.*, 54:3489 (1983).
 335. J.L. Vossen, G.L. Schnable, and W. Kern, *J. Vac. Sci. Technol.*, 11:60 (1974).
 336. T.N. Kennedy, *J. Vac. Sci. Technol.*, 13:1135 (1976).
 337. A. Mumtaz, J. Drapkin, and R. Mathew, *J. Vac. Sci. Technol.* A3:596 (1985).
 338. R.C. Mosely, C.A. Sans, S.D. Campbell, and L.T. Lamont, Jr., *J. Vac. Sci. Technol.* A3:596 (1985).
 339. H.P. Bader and M.A. Lardon, *J. Vac. Sci. Technol.*, A3:2167 (1985).
 - 339a. H.P. Bader and M.A. Lardon, *J. Vac. Sci. Technol.*, B4:1192 (1986).
 340. Y. Homma and S. Tsunekawa, *J. Electrochem. Soc.* 132:1466 (1985).
 - 340a. T. Mogami, M. Morimoto, H. Okabayashi, and E. Nagasawa, *J. Vac. Sci. Technol.*, B3:857 (1985).
 341. F.F. Mullins and A. Brunnschweiler, *Solid State Electron.*, 19:47 (1976).
 342. W.G. Thompson and R.L. Anderson, *Solid State Electron.*, 21:603 (1978).
 343. E. Grusell, S. Berg, and L.P. Anderson, *J. Electrochem. Soc.*, 127:1573 (1980).
 344. A. Ashok, S.J. Fonash, R. Singh, and P. Wiley, *IEEE Electron Devices Lett.*, EDL-2:184 (1981).
 345. H. Lee, *IEEE Trans. Electron Devices*, 25:795 (1978).
 346. J.L. Vossen, J.J. O'Neill, Jr., G.W. Hughes, F.A. Taft, and R. Shedeker, *J. Vac. Sci. Technol.*, 17:400 (1980).
 347. N.A. Bojarczuk, *J. Vac. Sci. Technol.*, 18:890 (1981).
 348. S. Alexandrova, K. Kirov, and A. Szekeves, *Thin Solid Films*, 75:37 (1981).

454 Semiconductor Materials

- 348a. F.D. Auret, M. Nel, and N.A. Bojarczuk, *J. Vac. Sci. Technol.* B4:1168 (1986).
- 349. R. Singh, S.J. Fonash, S. Ashok, P.J. Caplan, J. Shappiro, M. Hage-Ali, and J. Ponpon, *J. Vac. Sci. Technol.* A1:334 (1983).
- 350. S.W. Pang, D.D. Rathman, D.J. Silversmith, R.W. Mountain, and P.D. DeGraff, *J. Appl. Phys.*, 54:3272 (1983).
- 351. R. Singh, S.J. Fonash, P.J. Caplan, and E.H. Poindexter, *Appl. Phys. Lett.* 43:502 (1983).

Diffusion and Ion Implantation in Silicon

Richard B. Fair

*Microelectronics Center of North Carolina
Research Triangle Park, NC*

INTRODUCTION

As it becomes necessary to achieve higher levels of integration in IC manufacturing, so too does it become necessary to better understand the interrelationships among processing steps. Knowledge of the mechanisms involved in ion implantation and diffusion are basic to obtaining this understanding.

The focus of this chapter is on impurity diffusion and ion implantation in silicon. Because of the large solid solubility of Group III and V doping impurities in silicon, diffusion proceeds by interactions with point defects in the silicon-vacancies and silicon self-interstitials. Each high temperature processing step has the potential for changing the number of vacancies and self-interstitials, and therefore, the properties of impurity diffusion. We can understand these effects at two levels—the atomic level and the continuum level. Thus, discussion of diffusion from both points of reference is provided.

The process of introducing impurities into silicon is called predeposition. Chemical predeposition is described in terms of a solution to the diffusion equation and also in terms of ion penetration into silicon, distributions of implanted impurities, lattice damage, etc. Finally, useful curves for designing implanted junctions with a single annealing step are provided to assist in the selection of implant dose, energy and dopant.

CONTINUUM THEORY

According to the Continuum Theory of diffusion, matter will flow in a manner which will decrease all concentration gradients in an inhomogeneous single phase alloy undergoing annealing. The physics governing diffusion are described in two equations. The first is Fick's First Law which states that there will be a flux of atoms whenever a concentration gradient exists, and that flux is related to a constant that is a fundamental material property. In one dimension Fick's First Law becomes¹

$$J_x = -D \frac{dC}{dx} \quad (1)$$

As an example, Figure 1 shows a plot of the concentration of an impurity distributed in a host lattice as a function of x —the depth into the material. From Equation 1 the flux of this impurity during diffusion is related to the concentration gradient. Figure 1b shows the flux calculated at 2 points— x_1 and x_2 . From Figure 1c, if Δx is small,

$$J_1 - J_2 = -\Delta x \frac{\partial J}{\partial x} \quad (2)$$

The net increase in matter in the volume is

$$J_1 - J_2 = \Delta x \frac{\partial C}{\partial t} = -\Delta x \frac{\partial J}{\partial x} \quad (3)$$

Differentiating Equation 1 and substituting into Equation 3 gives Fick's Second Law:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial C}{\partial x} \right) \quad (4)$$

For the special case where D is constant and the surface concentration of the impurity that is diffusing is fixed, then Equation 5 results:

$$D \frac{\partial^2 C}{\partial x^2} = \frac{\partial C}{\partial t} \quad (5)$$

Fick's Second Law is a continuity equation which describes the time rate of change of the impurity concentration. The diffusion constant D is in units of $\text{cm}^2/\text{sec.}$ and the concentration C is usually in units of atoms/ cm^3 .

Special Cases

Predeposition. Under the special case where D is constant, the surface concentration of the diffusing impurity is fixed, the concentration of the impurity at $x = \infty$ is $C(\infty, t) = 0$ for all time, and the concentration at any point in the crystal at $t = 0$ is $C(x, 0) = 0$, then under these conditions the solution to Equation 5 is given as:²

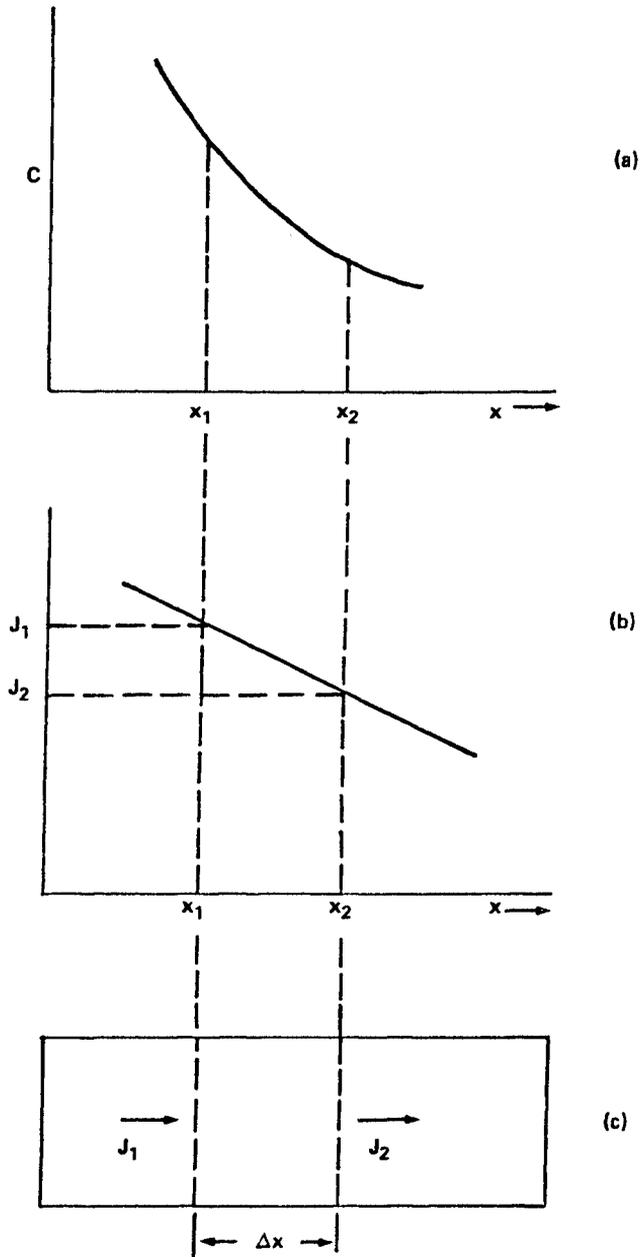


Figure 1: (a) Plot of an assumed concentration as a function of depth, (b) the flux $J(x)$ for this plot and (c) the element of volume with the flux J_1 entering and J_2 leaving.

$$C(x,t) = C_0 \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right). \tag{6}$$

If $C(\infty,t) = C_B$ and $C(x,0) = C_B$, then

$$C(x,t) = C_0 \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right) \pm C_B. \tag{7}$$

where C_B is the background doping concentration in the semiconductor. Thus, for the boundary conditions described above the impurity concentration as a function of space and time is given by a complementary error function whose argument is $x/\sqrt{4Dt}$. The complementary error function is a tabulated function and is described in Equations 8 and 9. A plot of the complementary error function is given in Figure 2. If we form the inverse complementary error function of Equation 6 and set the concentration equal to the background concentration in the host lattice we can solve for the depth of a junction with that background concentration. This is given by Equation 10.

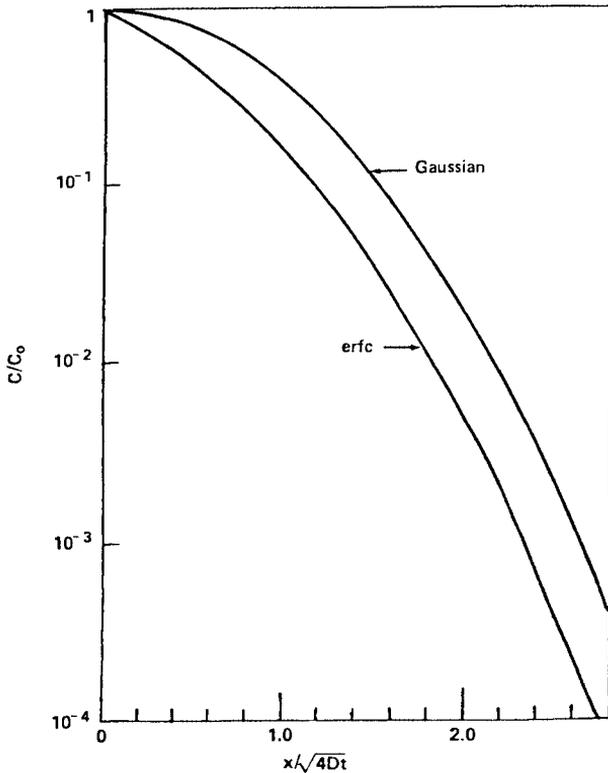


Figure 2: Normalized Gaussian and complementary error function curves.

$$\operatorname{erfc} y = 1 - \operatorname{erf} y \tag{8}$$

$$\operatorname{erf} (y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-\alpha^2} d\alpha \tag{9}$$

$$x_j = 2\sqrt{Dt} \operatorname{erfc}^{-1} \left(\frac{C_B}{C_0} \right) \tag{10}$$

Example: for $C_0=5 \times 10^{18} \text{ cm}^{-3}$ and diffusion of arsenic in silicon at 1200°C , what diffusion time is necessary for the arsenic concentration to decrease by a factor of 1000 in $5 \times 10^{-5} \text{ cm}$ (0.5 microns)?

a) $C(x,t) = C_0 \operatorname{erfc} (x/\delta\sqrt{Dt})$

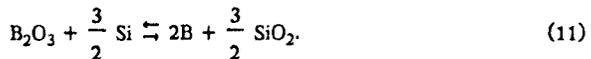
b) $\frac{C(x_j,t)}{C_0} = 10^{-3}$; $x_j/2\sqrt{Dt} = 2.32$ (from Fig. 2)

c) $t = \frac{x_j^2}{21.5D}$; $D = 2.2 \times 10^{-13} \text{ cm}^2/\text{sec}$; $t = 528.4 \text{ sec}$.

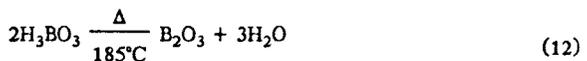
Figure 3 shows the complementary error functions plotted as a function of time on linear and semilogarithmic scales. It can be seen that the junction depth increases as the \sqrt{t} when the surface concentration is equal to a constant value.

Diffusion processes that are performed with a constant surface concentration are normally referred to as predeposition steps. Predepositions are usually done in N_2 furnace ambients with a small percentage of O_2 , and the doping species is introduced into the furnace in gaseous form. The dopant concentration in the gas N_2 stream is varied to change the surface concentration in the silicon. Typical predeposition temperatures are in the $900\text{-}1000^\circ\text{C}$ range, and times are usually 30-60 minutes. There are a wide variety of sources of dopants including liquids, solids and gases.

For the predeposition of boron the most prevalent species in the gas phase in the furnace is B_2O_3 . Once the B_2O_3 is deposited on the silicon surface there is a reaction of this oxide with the silicon to produce doping. This is shown as Equation 11.



The production of B_2O_3 may come from either one of the reactions described in Equations 12 or 13,



The source of the boron nitride in Equation 13 may be in the form of disks

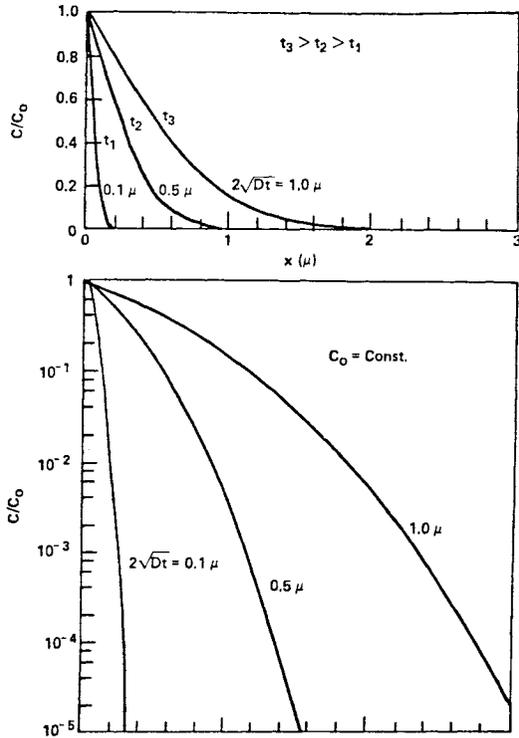


Figure 3: The complementary error function (erfc): normalized concentration vs. distance for successive times.

about the size of a silicon wafer which are placed next to the wafers in the diffusion furnace.

By varying the partial pressure of the gas phase of the dopant in the furnace, it is possible to change the concentration of impurities in the silicon. Henry's Law relates the concentration of dopants that are introduced in the furnace to the surface concentration:

$$C_0 = H p_i \tag{14}$$

C_0 = dopant surface concentration

H = Henry's constant

p_i = partial pressure of dopant gas

Figure 4 shows Henry's Law plotted. It can be seen that once the solid solubility of boron in silicon is reached, Henry's Law no longer applies. Thus, most predeposition steps operate with a high enough partial pressure in the dopant gas phase that solid solubility of the dopant is achieved in the silicon. This provides a natural control for reproducible diffusion results.

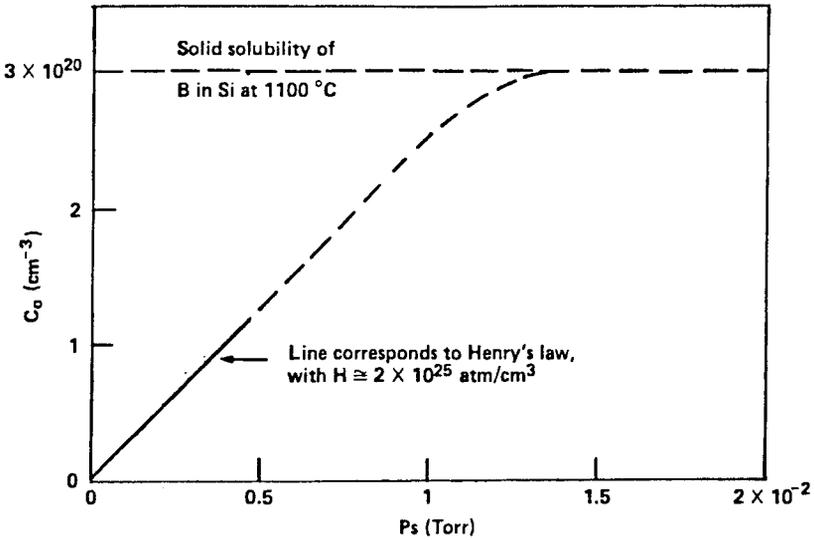
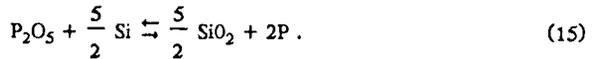


Figure 4: Surface concentrations of boron in silicon as a function of the partial pressure of B₂O₃ in the ambient at 1100°C.

For the predeposition of phosphorus the predominate species in the gas phase is P₂O₅. The doping reaction with P₂O₅ is shown in Equation 15:



Sources of P₂O₅ vapor are solid P₂O₅, red phosphorus, POCl₃, PBr₃, NH₄H₂PO₄ or PN.

For the predeposition step the goal is to deposit some number of atoms/cm² in the silicon substrate, Q(t). The way in which that number is calculated is to integrate the total concentration per cubic centimeter from 0 to ∞ as shown in Equation 16.

$$\begin{aligned} Q(t) &= \int_0^\infty C(x,t) dx \\ &= C_0 \int_0^\infty \operatorname{erfc}(x/2\sqrt{Dt}) dx \\ &= C_0 \left[\frac{2}{\sqrt{\pi}} \sqrt{Dt} \right] \end{aligned} \tag{16}$$

Once the predeposition is completed with Q atoms/cm² the next step is to redistribute the atoms to give the desired junction depth.

Redistribution or Drive-in. If Q atoms/cm² are deposited on the semiconductor surface with the following boundary conditions:²

$$C(x, t=0^-) = Q\delta(x)$$

$$C(\infty, t) = 0$$

then the distribution of impurities after diffusion for a time t is given by a Gaussian function solution to Equation 5:

$$C(x, t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(\frac{-x^2}{4Dt}\right) \tag{17}$$

This distribution is shown plotted in Figure 2 on normalized axes. The Gaussian distribution can be used to describe the impurity profile that results from a drive-in with no dopant gas in the furnace. The drive-in step is performed in several types of ambients: dry oxygen, steam, nitrogen or argon. The drive-in temperatures range from 900-1200°C.

Example: arsenic was predeposited and the resulting Q equals 1×10^{14} cm². How long would it take to drive-in the arsenic to a junction depth of 1×10^{-4} cm (one micron) in a background doping of 1×10^{15} cm⁻³? $T = 1200^\circ\text{C}$.

$$C(x, t) = 1 \times 10^{15} = \frac{1 \times 10^{14}}{\sqrt{\pi Dt}} \exp\left(\frac{-(10^{-4})^2}{4Dt}\right)$$

The solution to the equation above must be graphed and is shown where the left-hand-side of the equation intersects the righthand side of the equation in Figure 5. Note that if $x = 0$, the surface concentration C_0 decreases as $1/\sqrt{t}$ as shown in Equation 18. The junction depth can be found by taking the log of both sides of Equation 17:

$$C_0(t) = \frac{Q}{\sqrt{\pi Dt}} \tag{18}$$

$$x_j = \left[4Dt \ln \left(\frac{Q}{C_B \sqrt{\pi Dt}} \right) \right]^{1/2} \tag{19}$$

Normalized Gaussian concentration profiles versus distance are shown in Figure 6 for successive times both on linear and semilogarithmic scales.

Under certain conditions the Gaussian solution to the Second Law can be used in conjunction with short predeposition steps provided that the following condition applies:

$$\left(\frac{Dt \text{ (drive-in)}}{Dt \text{ (predep)}} \right)^{1/2} > 4$$

Otherwise, a more complex solution is necessary, and under these conditions the solution is

$$C(x, t) = \frac{2C_0}{\pi} \int_0^\mu \frac{\exp(-B(1 + \mu^2))}{1 + \mu^2} du \tag{20}$$

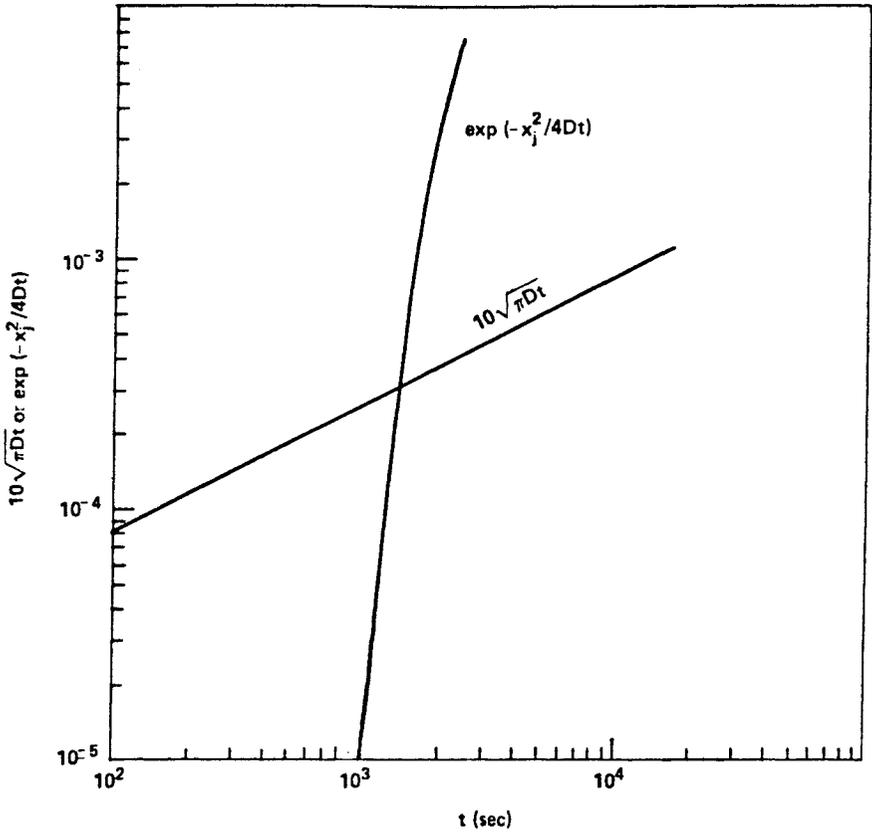


Figure 5: Graphed solution to the equation cited in the example in the text.

where

$$\mu = \left(\frac{Dt \text{ (predep)}}{Dt \text{ (drive-in)}} \right)^{1/2} ; B = \left(\frac{x}{2 [Dt \text{ (predep)} + Dt \text{ (drive-in)}]^{1/2}} \right)^2$$

The decision as to when the Gaussian solution is appropriate can be determined by plotting Equation 20 for various ratios of the \sqrt{Dt} product for the predeposition and diffusion. These results are shown in Figure 7.

Diffusion Coefficients

Diffusion coefficients are based upon the atomic behavior of the atom in a host lattice. Diffusion coefficients obey an analytical form as described in Equation 21.

$$D = D_0 \exp(-E_A/kT), \tag{21}$$

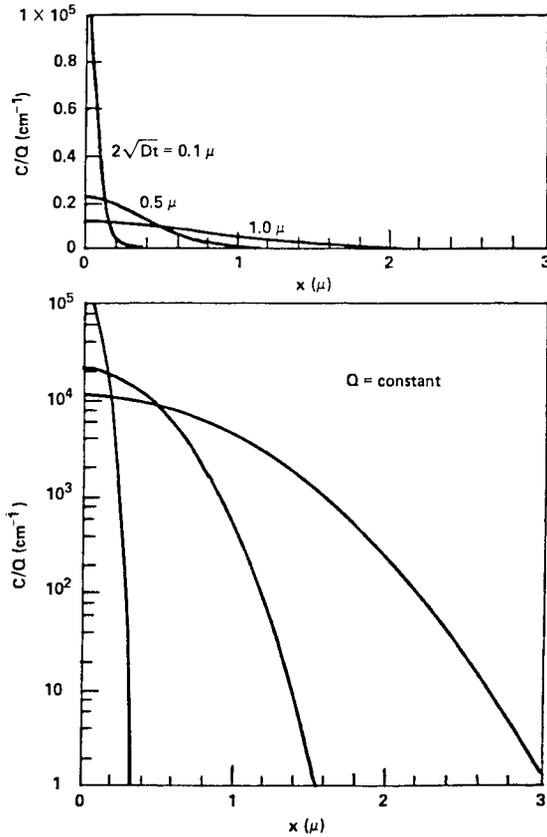


Figure 6: The Gaussian function: normalized concentration vs. distance for successive times.

where D_0 is the prefactor (cm^2/sec), E_A is the activation energy in (eV), k is the Boltzmann constant and T is the temperature ($^\circ\text{K}$). Typical diffusion coefficients of Group III and V elements in silicon are shown in Table 1. There is remarkable similarity between the diffusion coefficients of these elements in silicon. The activation energies for Group III and Group V elements are in the 3.5 to 4 eV range. A discussion of the origin of these energies and prefactors is provided in the next section.

Table 1: Diffusion Coefficients of Dopants in Silicon

Impurity	P	As	Sb	B	Al	Ga	In
D_0	10.5	0.32	5.6	10.5	8	3.6	16.5
E_A	3.69	3.56	3.95	3.69	3.47	3.5	3.9

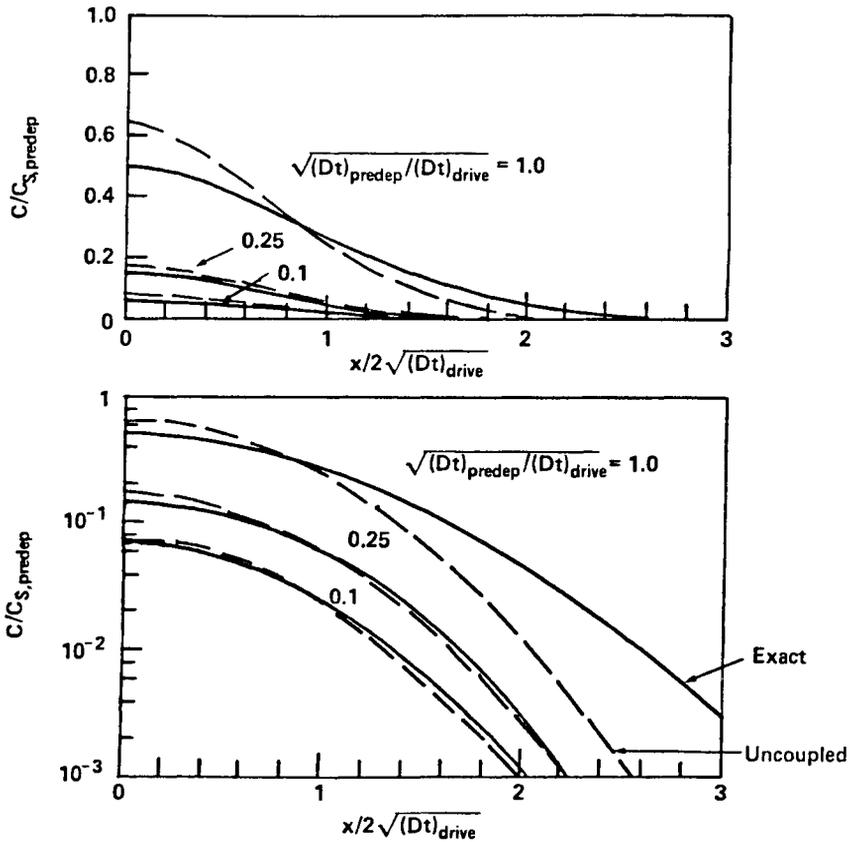


Figure 7: Comparisons between exact and uncoupled solutions to the drive-in diffusion equation for several values of Dt (drive-in)/ Dt (predep).

ATOMIC THEORY OF DIFFUSION

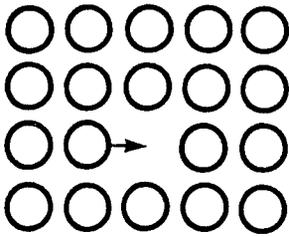
Diffusion Mechanisms

The atomistic theory of diffusion is concerned with describing how an atom gets from one part of a crystal to another. The lattice sites in a crystal are generally taken as the fixed location of the atoms making up the crystal. It is known that the atoms oscillate around these lattice sites which are their equilibrium positions. These oscillations lead to finite chances that an atom will move from its lattice site to another position in the crystal. There are several ways by which atoms can move from one site in the crystal to another. These mechanisms are

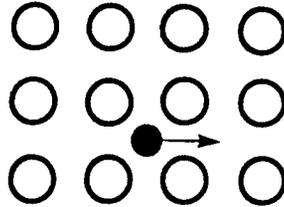
- the vacancy mechanism

- the interstitial mechanism
- the interstitialcy mechanism

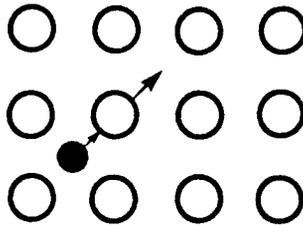
These mechanisms are illustrated in Figure 8.



(a) The vacancy diffusion mechanism.



(b) The interstitial diffusion mechanism.



(c) The interstitialcy mechanism.

Figure 8: Illustration of the dominant diffusion mechanisms in silicon.

Thermodynamic considerations require that some of the lattice sites in the crystal are vacant and that the number of vacant lattice sites generally is a function of temperature. When a lattice atom moves into an adjacent vacant site, this process is called the vacancy diffusion mechanism. In addition to occupying lattice sites, atoms can reside in the space between the lattice sites. These interstitial atoms can readily move to adjacent interstitial sites without displacing the lattice atoms. The interstitial atoms may be impurity atoms or atoms of the host lattice, but in either case they are generally present only in very dilute amounts. These atoms, however, can be highly mobile and are the dominant diffusion species in certain cases. A mechanism related to interstitial diffusion is the interstitialcy mechanism. In this process an interstitial atom moves into a lattice site by displacing the atom on that site onto an adjacent interstitial site. Although several other diffusion mechanisms may exist in semiconductors, in silicon the three dominant mechanisms are those just described.

The Flux Equation in Diffusivity

The number of atoms which cross a unit area in unit time is known as the flux. In one dimension the atoms only move to the right or left when they

change position along the x axis indicated in Figure 9. The atoms in this simple case are taken to be located in planes at x_0 and $x_0 + a_0$ as shown in the figure. The flux J is simply the concentration C times the velocity v :

$$J_x = C v \tag{22}$$

The net flux is the difference between the flux to the right and from the left:

$$J_x = \frac{1}{2} v (C_{x_0} - C_{x_0 + a_0}) \tag{23}$$

where C_{x_0} and $C_{x_0+a_0}$ are the concentrations at $x = X_0$ and $x_0 + a_0$ respectively. A factor of $\frac{1}{2}$ occurs in Equation 23 because at any one plane, half of the atoms move in the $+x$ direction and the other half moves in the $-x$ direction. When a_0 approaches 0,

$$a_0 \left[\frac{C_{x_0} - C_{x_0 + a_0}}{a_0} \right] = -a_0 \frac{dC}{dx} \tag{24}$$

and Equation 24 becomes:

$$J_x = -\frac{1}{2} v a_0 \frac{dC}{dx} . \tag{25}$$

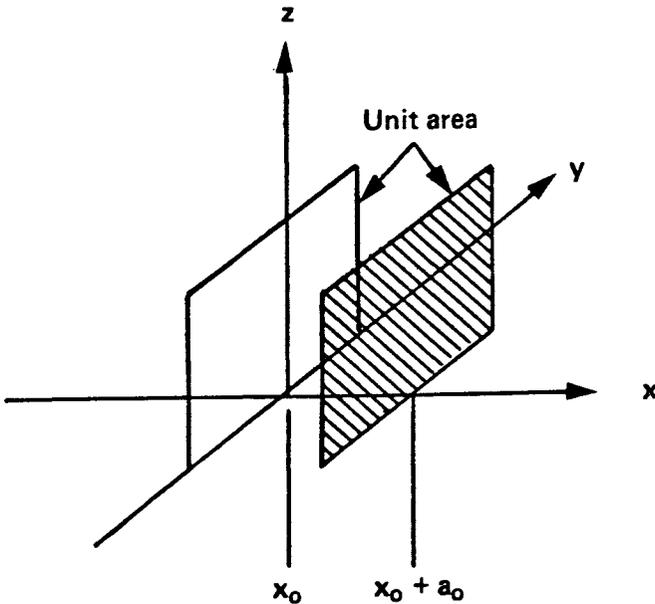


Figure 9: Flux in the x direction through the unit area in unit time. The planes of the unit area are located at $x = x_0$ and $x = x_0 + a_0$.

For motion by discrete jumps between planes a_0 apart, the velocity is the number of jumps per second, Γ , times the distance a_0 of each jump. Equation 25 may now be written as

$$J_x = -\frac{1}{2} a_0^2 \Gamma \frac{dC}{dx}, \quad (26)$$

and the quantity $(\frac{1}{2} a_0^2 \Gamma)$ is called the diffusivity or diffusion coefficient D :

$$D = \frac{1}{2} a_0^2 \Gamma \quad (27)$$

Equation 27 shows that for diffusion by a particular mechanism, calculation of the diffusivity is reduced to the calculation of the jump frequency Γ . The jump frequency by the vacancy mechanism is

$$\Gamma = X_v w \quad (28)$$

where w is the frequency at which an atom and an adjacent neighboring vacancy exchange, and X_v is the probability that the adjacent site is vacant. From statistical thermodynamics the vacancy atom fraction is given by Equation 29, where ΔS_f is the entropy of formation of the vacancy and ΔH_f is the enthalpy of vacancy formation.

$$X_v = \exp(\Delta S_f/k) \exp(-\Delta H_f/kT) \quad (29)$$

These terms are related to the Gibbs free energy change for vacancy formation through the equation

$$\Delta G_f = \Delta H_f - T\Delta S_f = -kT \ln X_v. \quad (30)$$

It is far more difficult to derive the frequency, w , from fundamental principles. Nevertheless a discussion of the physics involved in evaluation w provides useful insights into the quantities that affect diffusion.

In self-diffusion by the vacancy mechanism a lattice atom moves from a normal lattice site to a vacancy. As shown in Figure 10, the atom must move from the normal lattice site in 10a to the saddle point position in 10b to reach the vacancy at 10c. The energy at the saddle point is greater than at the equilibrium lattice sites, and atoms must be sufficiently activated in order to move to b and then c. The fraction of the lattice atoms activated to the saddle point is related to the Gibbs free energy change between positions 10a and 10b. In the same manner as for the atom fraction of vacancies, the atom fraction of activated atoms is

$$X_m = \exp(\Delta S_m/k) \exp(-\Delta H_m/kT) \quad (31)$$

where ΔS_m and ΔH_m are called the entropy and enthalpy of motion respectively.

Now, the frequency, w , at which an atom and an adjacent neighboring vacancy exchange can be written

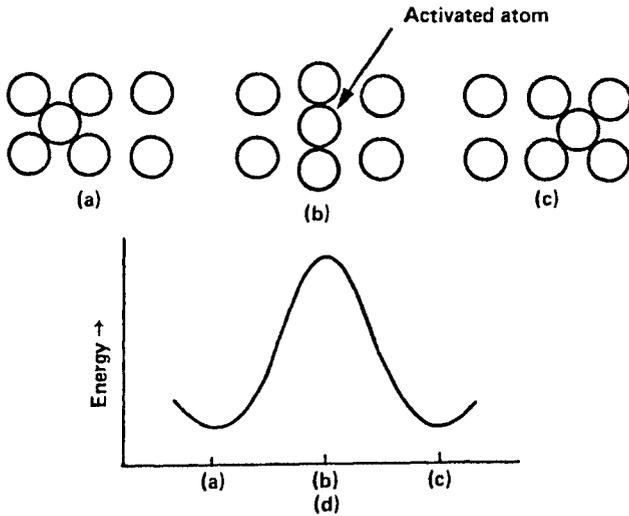


Figure 10: The sequence of (a), (b), and (c) show the movement of the atom from a normal lattice site to an adjacent vacancy. Part (d) shows the variation of free-energy as the atom moves from (a) to (c).

$$w = X_m \gamma \tag{32}$$

where the frequency, γ , is generally not known and is usually taken as the lattice vibrational frequency of an atom about its equilibrium site, which is of the order of $10^{13}/\text{sec}^{-1}$. Now, from Equations 28, 29 and 31 and 32 the jump frequency for vacancy self-diffusion is

$$\Gamma = \gamma \exp \left[\frac{(\Delta S_f + \Delta S_m)}{k} \right] \exp \left[-\frac{(\Delta H_f + \Delta H_m)}{kT} \right]. \tag{33}$$

Experimentally it is found that diffusivity is given by the Arrhenius expression

$$D = D_0 \exp (-E_A/kT) \tag{34}$$

where E_A is the activation energy. Thus, the diffusivity is

$$D = \frac{1}{2} a_0^2 \gamma \exp \left[\frac{(\Delta S_f + \Delta S_m)}{k} \right] \exp \left[-\frac{(\Delta H_f + \Delta H_m)}{kT} \right]. \tag{35}$$

Comparisons of Equations 34 and 35 gives the prefactor D_0 as

$$D_0 = \frac{1}{2} a_0^2 \gamma \exp \left[\frac{(\Delta S_f + \Delta S_m)}{k} \right] \tag{36}$$

and

$$E_A = \Delta H_f + \Delta H_m. \tag{37}$$

From the above discussion it can be seen that diffusivity is basically the product of the lattice vibration frequency, vacancy concentration and activated lattice concentration:

$$D = X_v X_m \gamma \tag{38}$$

Also the activation energy for vacancy diffusion depends upon the energy necessary to form the vacancy and to move the lattice atom into an adjacent vacancy.

Multiple Charge State Vacancy Model

From the previous discussion it can be seen that the process of diffusion depends upon the concentration of point defects in the crystal such as vacancies or self-interstitials. Therefore, if one can find ways of raising or lowering the point defect concentrations then one can effect diffusion coefficients.

For the vacancy mechanism, the single vacancy in silicon is known to exist in four charge states: V^+ , V^x , V^- and $V^{=}$, where + refers to a donor level, x a neutral species and — an acceptor level.^{3,4} The creation of a vacancy introduces a new lattice site, and thus four new valence band states in the crystal. These states are available as acceptors but are not shallow. The lattice distortion associated with the vacancy will split states from the valence and conduction bands of the surrounding atoms a few tenths of an electron volt into the forbidden gap. States split from the valence band will become donors and those from the conduction band will become acceptors. At low temperatures there should be one deep donor level, V^+ , a few tenths of an eV above the valence band edge, a single acceptor level, V^- near midgap, and a double donor level $V^{=}$ very near the conduction band edge (see Figure 11).⁵ The levels depicted in Figure 11 represent a best guess based on experiment.⁶⁻⁸

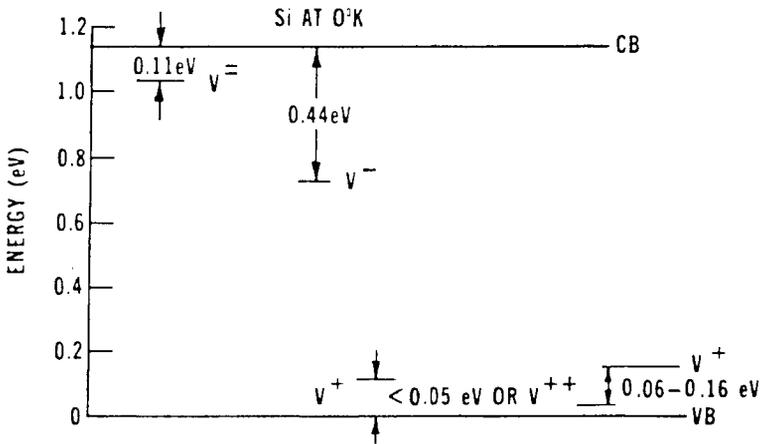


Figure 11: Estimated vacancy energy levels in the silicon band-gap at 0°K.

It has been experimentally verified that both silicon self-diffusion and the diffusion of Groups III and V impurities in silicon depend upon the Fermi-level position. The initial assumption in the vacancy diffusion model of self-diffusion is that an observed diffusivity arises from the simultaneous movement of neutral and ionized vacancies. Each charge type vacancy has a diffusivity whose value depends upon the charge state, and the relative concentrations of vacancies depend upon the Fermi-level.⁹ Calculated changes in relative concentrations of charge species versus Fermi-level, E_F , are shown in Figure 12a at $T=300^\circ\text{K}$ and in Figure 12b at $T=1400^\circ\text{K}$.⁵ Whereas at low temperature V^x will be the dominant species in intrinsic silicon, at high temperatures both V^+ and V^- would be more numerous. There is no value of E_F for which V^x dominates. Another important concept is that every time an ionized vacancy is formed the crystal must return the neutral vacancy population back to equilibrium by generating an additional vacancy. In this way, as the doping becomes more n-type or more p-type, the total vacancy concentration will increase with the increasing population of ionized vacancies. Since impurity and self-diffusion coefficients depend upon the concentration of vacancies, the diffusion coefficients will also increase with doping. Such concentration-dependent diffusion can occur when the doping level exceeds the intrinsic electron concentration, n_i at the diffusion temperature. An illustration of concentration dependent diffusion is shown in Figure 13.

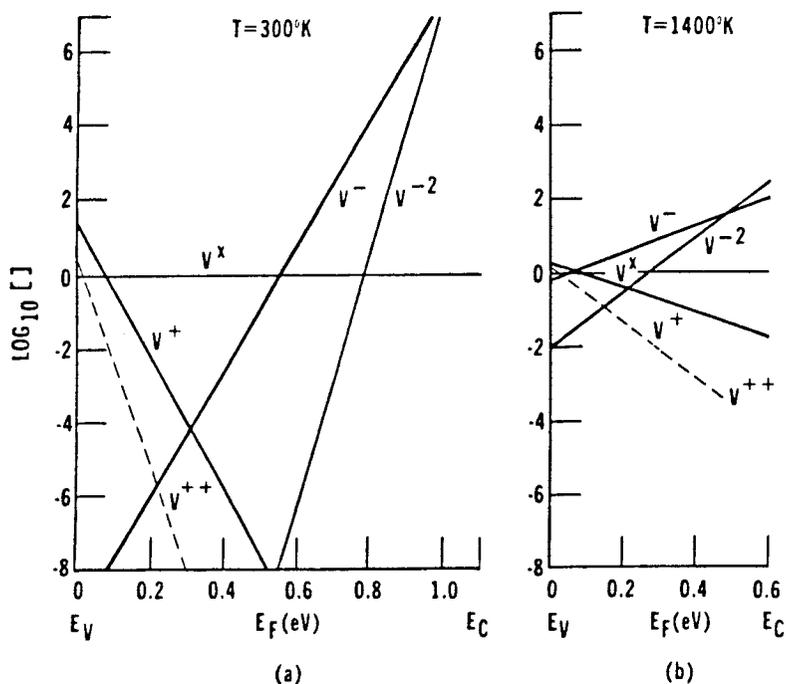


Figure 12: Calculated changes in the ratios of ionized to neutral vacancies at (a) 300°K and (b) 1400°K .

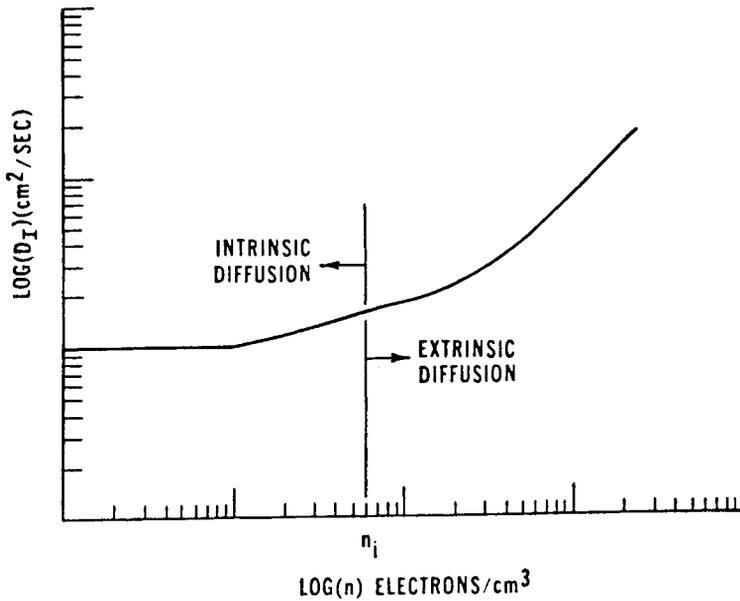


Figure 13: Donor impurity diffusion coefficient vs. electron concentration showing regions of intrinsic and extrinsic diffusion.

THE ROLE OF POINT DEFECTS IN SILICON PROCESSING

The Silicon Processing Balancing Act

Silicon oxidation and diffusion of impurities are quite related since they both occur at high temperatures and they both involve point defects such as vacancies or self-interstitials. The first level of process design involves the concept of doping and junction formation, threshold voltage control, or the gain control of the transistor. Another goal of doping is to achieve low sheet-resistance. For oxidation the primary goal is to grow controlled layers of SiO_2 . Things of concern in oxidation include the growth of stable oxides with electrical integrity, etc. If one is creating a non-planar structure it is necessary to worry about the viscous flow characteristics of the oxide and whether the viscosity is low enough to release stress. The process engineer in general spends a lot of time dealing with these first order requirements. However, the rest of the time is spent in trying to balance things that are generally not well understood. The point defect balancing act diagram is shown in Figure 14. All of the arrows in the figure indicate the directions of interactions. For example diffusion may change the concentration of point defects, and point defects themselves can affect diffusion. Oxidation produces point defects and point defects can affect oxidation. The balancing act involves point defect generation and the effect of this generation on these major processes. Diffusion may introduce strain into the lattice which can affect surface quality. Oxidation

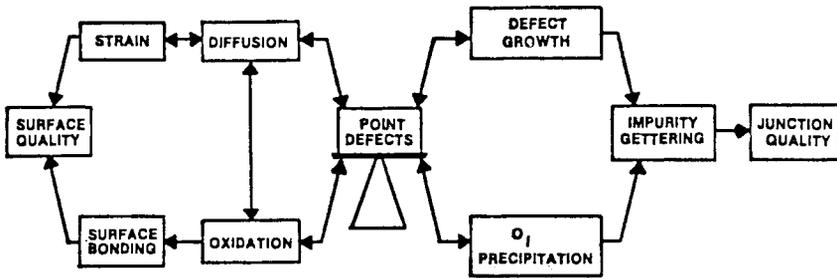


Figure 14: The point defect balancing act in silicon processing.

can influence surface bonding which also affects surface quality. As these processes produce point defects it is possible that extended structural defects can grow in the silicon. Point defects can also influence the precipitation of oxygen. Oxygen is incorporated into the crystal during crystal growth, and precipitates during subsequent heat treatments. It is known that these precipitates create good internal gettering sites for metal impurities with subsequent impact on junction quality.

Point Defects

Point defects are defined as atomic defects. There are atomic defects such as metal ions which can diffuse through the lattice as shown without involving themselves with lattice atoms or vacancies. Another type of atomic defect is the self-interstitial which in silicon is a silicon atom that is bonded in a tetrahedral interstitial site. Examples of point defects are shown in Figure 15.

One of the major controversies in solid state science currently is: what is the dominant native point defect in silicon - the monovacancy or the silicon self-interstitial? A brief review of the arguments for each species is given below

The Monovacancy

From statistical thermodynamics, it is known that if a vacancy is formed by removing an atom from the crystal and depositing it on the surface, the free energy of the crystal will decrease as the number of vacancies created increases until a minimum in this free energy occurs. Because a minimum in the free energy occurs for a certain vacancy concentration in the crystal, the vacancy is a stable point defect. Other experimental observations involving vacancies are listed below:¹⁰

1. Electron paramagnetic resonance measurements only identify the existence of vacancies or vacancy complexes in Si irradiated by electrons. The absence of Si self-interstitials has been ascribed to rapid athermal migration even at 2°K.¹¹

2. Diffusion phenomena as well as calculations of diffusion entropy and enthalpy have been successfully explained by ascribing multiple ionization levels to vacancies which are the same as those observed for the vacancy in a low-temperature irradiation experiment.¹²⁻¹⁴
3. Theoretical estimates of the heats of formation and entropies of formation of vacancies correspond well with those of the native defects observed in diffusion and quenching experiments.^{13,15-17}
4. Channeling studies of impurity-defect interactions in Si show that under helium ion bombardment the trapping efficiency of impurities for radiation-produced defects is very low near 30°K.¹⁷ Vacancies are not mobile in Si below this temperature while interstitials still are. This implies that the impurity-defect interactions involve vacancies.
5. Positron annihilation lifetime measurements which have been performed on float-zone Si at high temperatures, show that vacancy-like defects are formed.¹⁸

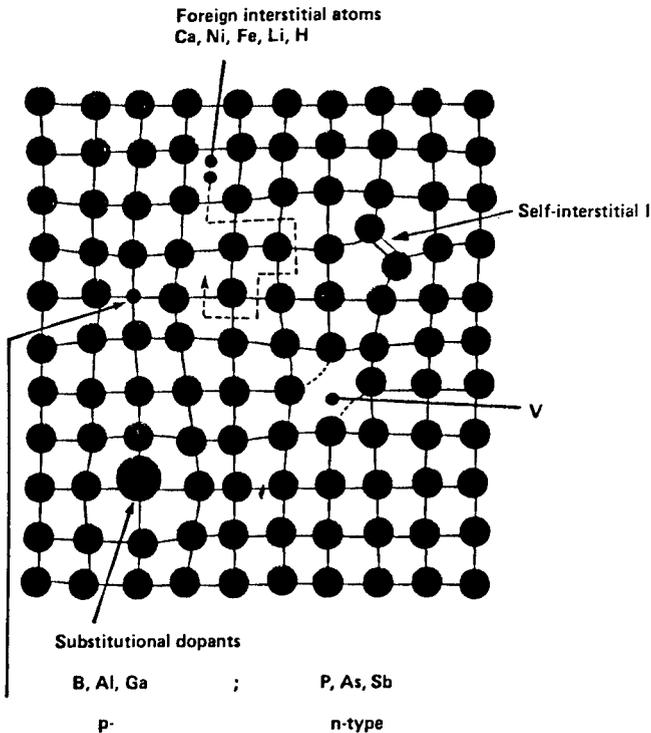


Figure 15: Examples of point defects in the silicon lattice.

The Silicon Self-Interstitial Atom

It is possible to perform a similar consistent statistical thermodynamic analysis on the existence of self-interstitials and show that they are stable point defects. Other arguments in support of the silicon self-interstitial are:

1. The great majority of dislocation loops and stacking faults in Si observed by transmission electron microscopy are judged to be of extrinsic or "interstitial" character. Though there exist four proposed mechanisms by which extrinsic type dislocations may be formed without any self-interstitials being present,¹⁰ most workers feel that self-interstitial precipitation is the dominant mechanism.
2. The picture of self-interstitials in Si developed by Seeger and Frank¹⁹ is consistent with observations indicating self-interstitial migration at low and high temperatures.
3. Evidence for the liquid drop character of B-swirl defects in Si comes from the observation that upon melting, droplets of liquid Si are formed in the interior of the solid phase.²⁰
4. In n-p structures formed by sequential diffusions of B and P, dislocation climb was observed to have occurred at the same time that the emitter-push effect was seen in the B layer.²¹ This result implies that the same point defect is responsible for both phenomena.
5. Stacking fault growth P diffusions and enhanced buried layer diffusion have been observed to occur simultaneously.²²
6. Total energy calculations show that self-interstitials form and migrate in Si with a total activation energy roughly the same as that of self-diffusion.²³

After reviewing the balance sheet of pros and cons surrounding the question of the native defect in Si, one is still left with the question: what is the native defect responsible for impurity diffusion and defect growth in Si? So far we only have clues. However, the majority opinion currently is that both types of point defects are important. Thermal equilibrium concentrations of point defects at the melting point are orders of magnitude lower in Si than in metals. Therefore, a direct determination of their nature by Simmons-Balluffi type experiments²⁴ has not been possible. The accuracy of calculated formation and migration enthalpies appears to be within ± 1 eV but do not help to distinguish whether vacancies or interstitials are dominant in diffusion. The interpretation of low temperature experiments on the migration of irradiation-induced point defects is complicated by the occurrence of radiation-induced migration of self-interstitials.^{25,26} In addition, there are indications that the structure and properties of point defects may change from low to high temperatures.²⁷ The observation of extrinsic type dislocation loops in dislocation-free, float zone Si showed that self-interstitials must have been present in appreciable concentrations at high temperature during or after crystal growth.^{28,29} However, it is unclear

whether these self-interstitials were present in thermal equilibrium or were introduced during crystal growth by non-equilibrium processes.

In view of the uncertainties regarding the native point defect in Si, it is necessary in discussions of self and dopant diffusion to take account of both types of defects.

Point Defect Models of Diffusion in Silicon

Under thermal equilibrium conditions, a Si crystal will contain a certain equilibrium concentration of vacancies, C_V , and a certain equilibrium concentration of Si self-interstitials, C_I . In diffusion models based on the vacancy, $C_V \gg C_I$ and dopant as well as self diffusion can be explained as¹³

$$D_i = D_i^x + D_i^- + D_i^= + D_i^+ \quad (39)$$

where D_i is the measured diffusivity and D_i^x , D_i^- , $D_i^=$ and D_i^+ are the intrinsic diffusivities of the species through interactions with vacancies in the neutral, single acceptor, double acceptor or donor charge states respectively. These individual contributions to the total measured diffusivity were described in a previous section.

Analogous to the vacancy model, Si self-interstitials can be assumed to be dominant such that $C_I \gg C_V$. For such a model, dopant and self-diffusion are assumed to occur via an interstitialcy mechanism.³⁰ Mobile complexes consisting of self-interstitials in various charge states and impurities are assumed to exist.

In principle, both vacancies and self-interstitials may occur simultaneously, and somewhat independently. Indeed, any relationship that may exist between C_V and C_I may be dominated by the Si surface which can act as a source or sink for either species. If a local dynamical equilibrium exists between recombination and spontaneous bulk generation, vacancies and self-interstitials would react according to



where O denotes the undisturbed lattice. The law of mass action under equilibrium for this reaction is

$$C_I C_V = C_I^0 C_V^0. \quad (41)$$

For sufficiently long times and high temperatures Equation 41 turns out to be fulfilled.^{31,32} However, it has been reported that a substantial amount of time may be required for dynamical equilibrium to occur.³² This would make vacancy/self-interstitial recombination an activated process. In addition, under point-defect injection conditions, Equation 41 may no longer be valid.

If both types of point defects are important, diffusion processes may involve both types:

$$D_i = D_i^I + D_i^V, \quad (42)$$

where D_i^I is the interstitialcy contribution and D_i^V is the vacancy contribution to the total measured diffusivity, D_i . One way in which vacancies and self-interstitials could cooperate in affecting impurity diffusion is the Watkins replacement mechanism³⁴ shown in Figure 16. Interstitial dopant impurities can be created by the exchange between a self-interstitial and a substitutional dopant atom. The newly created interstitial impurity would migrate until it finds a vacancy. Then, it is free to diffuse again as a substitutional impurity.

It is evident now that both vacancies and self-interstitials can exist in equilibrium with each other in the silicon lattice. Each species can be described by equilibrium equations of the type:

$$C_v^0 = \exp \left\{ \frac{S_f^V}{k} \right\} \exp \left\{ -\frac{\Delta H_f^V}{kT} \right\} \tag{43a}$$

$$C_i^0 = \exp \left\{ \frac{S_f^I}{k} \right\} \exp \left\{ -\frac{\Delta H_f^I}{kT} \right\} \tag{43b}$$

For silicon self-diffusion, the total diffusion coefficient could be expressed as

$$D_{Si} = f_v D_v C_v^0 + f_i D_i C_i^0 \tag{44}$$

where f_v and f_i are the fractional contributions of vacancies and self-interstitials to self-diffusion. There currently is substantial debate as to what the values of these fractional coefficients are. A diagram of the spectrum of the debate currently underway in the literature is shown in Figure 17. The concept that impurity diffusion was dominated by vacancies only was held until 1968 when Seeger and Chick³⁵ proposed that both self-interstitials and vacancies could contribute to diffusion in silicon. However, the concept of vacancies and interstitials coexisting in silicon leads to several unresolved questions such as-is there dynamic equilibrium between self-interstitials and vacancies and what is the time required to establish this dynamic equilibrium?

Experimental Observations

In order to understand whether vacancies or self-interstitials are

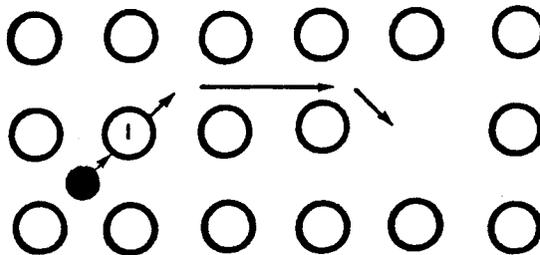


Figure 16: A schematic diagram of the Watkin’s Replacement Mechanism.

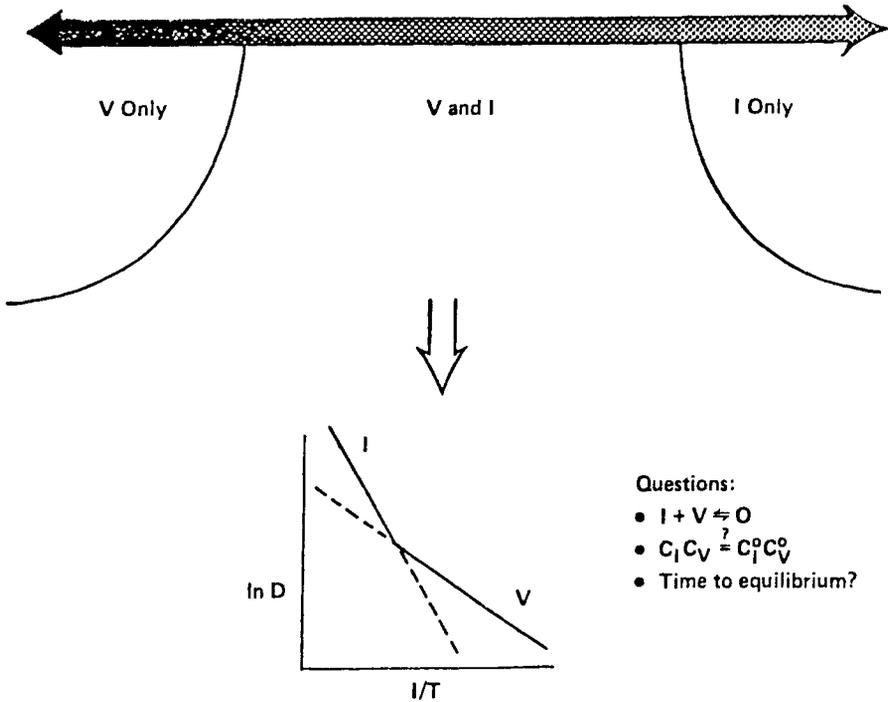


Figure 17: A diagram of the spectrum of the vacancy vs. self-interstitial debate.

involved in diffusion, there are numerous indirect observations that we must rely upon. A partial list of these types of experiments is shown in Table 2.

For example, during oxidation, enhanced diffusion of phosphorus, boron and arsenic are observed as well as retarded diffusion of antimony. However, if direct nitridization of the silicon surface occurs, the inverse effects are observed, i.e. enhanced antimony diffusion and retarded phosphorus diffusion. There also is a doping dependence associated with oxidation-enhanced diffusion. As either p or n-type doping concentration increases above n_i , the effect of oxidation-enhanced diffusion diminishes. If chlorine is introduced into the oxidizing ambient oxidation enhanced diffusion is likewise diminished.

Not only is enhanced impurity diffusion observed during oxidation but also increased stacking fault growth. A stacking fault is a plane of dislocated material that may intersect the silicon surface but which also has a bounding partial dislocation. These faults grow if sufficient numbers of self-interstitials are generated such that the concentration of self-interstitials in the lattice is higher than the concentration of self-interstitials on the bounding partial dislocation core. This process is illustrated in Figure 18. Since oxidation is a process that generates excess self-interstitials, stacking faults grow during oxidation.

Table 2

Which Model is Operating in Silicon?

For insight we need indirect observations from numerous different experiments:

- Oxidation — enhanced/retarded diffusion
 - backside oxidation
 - role of Si_3N_4 surface films
 - retarded diffusion of Sb
 - doping dependence of OED/ORD
 - effect of chlorine
 - CZ versus FZ Si

- Doping effect on OSF shrinkage
- Effect of diffusion on O precipitation
- Codiffusion studies
- TEM studies of precipitates/defects
- C_T vs. n in doped layer
- Profile Modeling
- Role of stress
- Doping dependence of oxidation

Other experiments that have been performed include irradiating uniformly doped silicon wafers with protons and observing the diffusion of the dopant after irradiation has occurred. Additional discussion of these effects will follow.

DIFFUSION IN THE PRESENCE OF EXCESS POINT DEFECTS

Oxidation-Enhanced Diffusion

As it was mentioned in the above discussion, oxidation generally enhances the diffusion of Group III and Group V elements except for antimony. These results are summarized in the Figure 19. In this figure oxidation-enhanced diffusion is generally observed by depositing a silicon nitride mask on the silicon surface which will prohibit oxidation in the regions that it covers. Then oxidation is performed in a window opened to the silicon surface, so that differential changes in junction depth can be observed. In order to explain these results, Hu³⁶ proposed a model whose essential points are:

1. Oxidation of Si at the Si/SiO_2 interface is usually incomplete to the extent that approximately 1 Si atom in 1000 is unreacted.

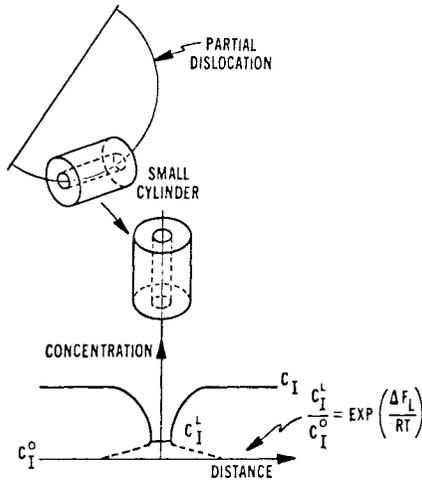


Figure 18: A model of self-interstitial diffusion from the bulk to the partial dislocation bounding a stacking fault. Under non-oxidizing conditions the concentration of self-interstitials at the fault line, C_I^L , is greater than the equilibrium bulk interstitial concentration, C_I^0 . Under oxidizing conditions, C_I is greater than C_I^L until the retrogrowth temperature is reached.

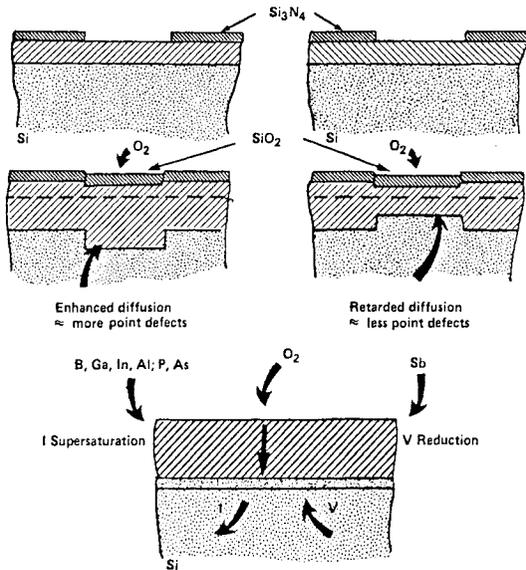


Figure 19: Experiments that illustrate oxidation-enhanced and oxidation-retarded diffusion of dopants in silicon. The supersaturation of self-interstitials associated with the oxidation process drives both effects.

2. The unreacted Si becomes mobile, severed from the lattice by the advancing Si/SiO₂ interface. These atoms enter the Si lattice interstices, causing a flux of self-interstitials away from the interface.
3. Growth of oxidation-induced stacking faults will proceed by the absorption of the generated self-interstitials. Oxidation-enhanced diffusion can occur as a result of the presence of the excess interstitials via the Watkins³⁴ replacement mechanism or by an interstitialcy process.

If the Watkins replacement mechanism is ignored, the diffusivity of an impurity atom under conditions of non-equilibrium point defect concentrations is

$$D = D_i^I(C_I/C_I^0) + D_i^V(C_V/C_V^0), \tag{45}$$

where C_I and C_V are the non-equilibrium self-interstitial and vacancy concentrations. Defining the fractional interstitialcy factor as³²

$$f_i = D_i^I/D_i^V \tag{46}$$

we can write

$$D/D_i = f_i(C_I/C_I^0) + (1-f_i)(C_V/C_V^0), \tag{47}$$

Calculations of the fractional interstitialcy components for B, P, As and Sb are shown in Table 3.^{31,37,38-40} A significant spread in the values of f_i is obtained. The value of f_i has been correlated with the amount of energy required to make a substitutional dopant atom become interstitial. Interstitial formation energies in Si are shown in Table 4. Thus, the larger the interstitial formation energy the smaller is the fractional interstitialcy component of diffusion.

It is observed that the diffusion of Sb is retarded during oxidation of the Si surface.³¹ This can be explained by assuming Sb diffuses predominantly by a vacancy mechanism, and the self-interstitials generated at the oxidizing surface combine with vacancies to reduce their concentration. Recent data from nitridation experiments suggest that f_i for P is greater than 0.7.⁴¹

Table 3: Fractional Interstitialcy Components of Diffusion Via Self-Interstitials in Silicon at 1000°–1100°C

Element	←----- f _i = D _i ^I /D _i ^V ----->				
	Fair [37]	Antoniadis [32]	Matsumoto [39]	Gosele [38]	Mathiot [40]
B	0.17	0.32	0.41	0.8-1.0	0.18
Al	0.2	--	--	0.6-0.7	--
P	0.12	0.40	0.35-0.5	0.5-1.0	0.19
As	0.09	0.43	0.45-0.75	0.2-0.5	0.16
Sb	0.13	0.15	--	0.02	--

Table 4: Estimated Interstitial Formation Energies in Silicon

Element	Interstitial Formation Energy
Si	2.2eV
Al ²⁺	2.21
B	2.26
P	2.4
As	2.5

Experiments that use the backside of the silicon wafer to inject self-interstitials and thus observe diffusion on the frontside of the wafer are illustrated in Figure 20.^{42,43} On the wafer surface, films of Si₃N₄ or Si₃N₄ on SiO₂ are deposited over previously diffused layers. On the backside a window is opened whose distance from the frontside surface can be varied by etching. It can be seen in the figure that the backside oxidation can, in fact, influence the diffusion of dopants on the frontside surface. The ratio of the junction depth under the oxidized portion to the non-oxidized portion versus distance from the backside oxidizing surface is shown in Figure 21 for boron-phosphorus and antimony. These results were obtained in float zone (FZ) silicon with no oxygen incorporated in the silicon. It can be seen that self-interstitial diffusion lengths of the order of 200-300 microns were obtained. These backside oxidation experiments show:

1. Oxidation-enhanced diffusion of B and P and oxidation-retarded diffusion of Sb involve the same point defects generated by an oxidizing Si surface.
2. The diffusion length of these point defects increases with diffusion time in FZ Si.
3. Long-range (> 100 μm) oxidation enhanced/retarded diffusion does not occur in Czochralski Si (CZ Si).
4. A Si₃N₄ layer on the Si surface is not a sink for point defects.

Doping Dependence of Oxidation-Enhanced Diffusion

Tanaguchi et al⁴⁴ found that oxidation-enhanced diffusion decreases as the concentration of the diffusing impurity increases beyond the point where concentration-dependent diffusion occurs. This effect was explained in terms of the reduction of oxidation produced self-interstitials by recombination with the increasing supply of vacancies. Fair⁴⁵ assumed that the equilibrium vacancy concentration is unaffected initially by the self-interstitials generated at the oxidizing surface. But, the quasi steady-state value of interstitial supersaturation is inversely proportional to the vacancy concentration which increases with doping above n_i. The oxidation enhanced dopant diffusivity, D_e, is then

$$D_e = D_{Si} + \Delta D_o \tag{48}$$

$$= -D_i(C_V/C_V^o) + D_i f_i (C_I/C_I^o)(C_V^o/C_V),$$

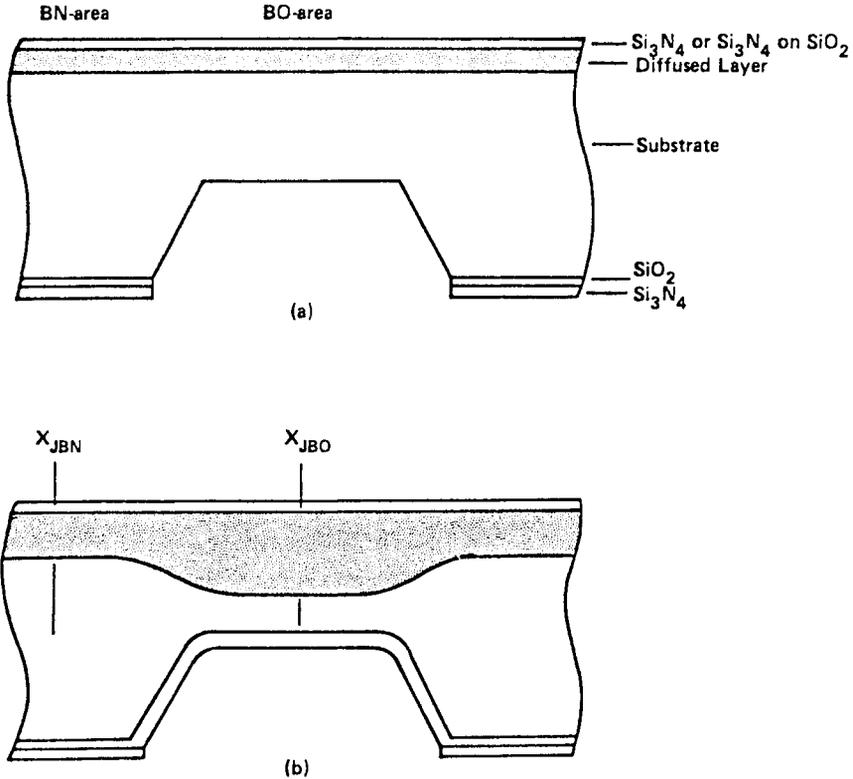


Figure 20: Experiments illustrating the use of the backside of the silicon wafer to inject self-interstitials in order to observe diffusion on the frontside of the wafer (after Mizuo).

where $(C_I/C_I^0)_i$ is the self-interstitial supersaturation under intrinsic doping conditions, and C_V/C_V^0 is the vacancy enhancement that occurs when doping exceeds n_r . This equation is divided into the contributions to substitutional impurity diffusion under non-oxidizing conditions, D_{SI} , and the enhanced contribution due to oxidation, ΔD_o . The data of Tanaguchi et al⁴⁴ are shown in Figure 22 for oxidation-enhanced diffusion of P and B versus the total number of dopant impurities per cm^2 , Q_T . The calculated values of D_{SI} and ΔD_o are shown in comparison with the data. Reasonable agreement is obtained. Thus Tanaguchi's model of self-interstitial recombination with vacancies is consistent with the high concentration diffusion models of B and P used by Fair in his calculations.

Effect of Chlorine on Oxidation-Enhanced Diffusion

If chlorine is added to oxygen in the furnace in sufficient concentrations such that stacking fault retrogrowth occurs,⁴⁶ oxidation-enhanced diffusion will become negligible.⁴⁷ This result is believed to be due to the generation

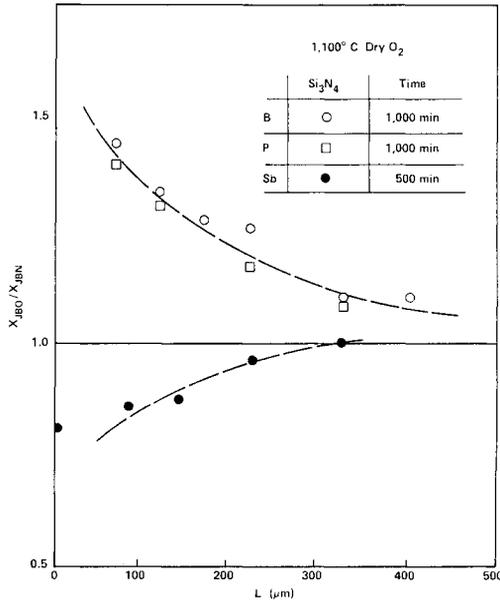


Figure 21: The ratio of the junction depth under the oxidized portion to the non-oxidized portion of the wafer vs. distance from the backside oxidizing surface.

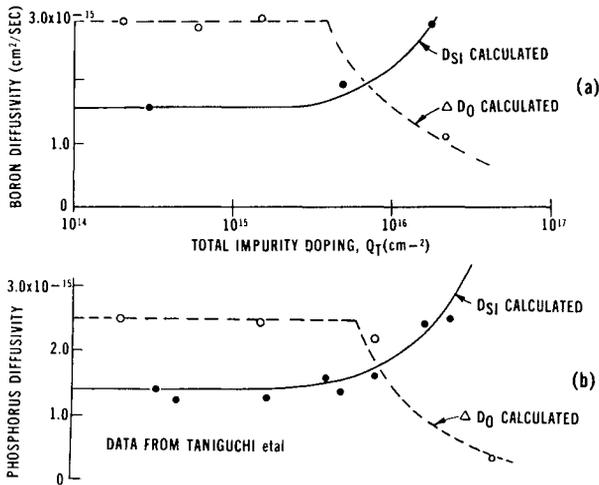


Figure 22: Measured and calculated values of boron and phosphorus diffusivity as a function of total impurity doping. Data are divided into contributions to substitutional impurity diffusion under non-oxidizing conditions, D_{S1} and the enhanced contribution due to oxidation ΔD_0 .

of vacancies at this Si/SiO₂ interface when Cl reacts with Si atoms on lattice sites to produce SiCl by the reaction



The vacancy generated is then available to recombine with a Si self-interstitial produced by oxidation:



As a result, the supersaturation of self-interstitials in the silicon surface and the bulk is reduced or eliminated, inhibiting stacking fault growth and enhanced diffusion. This effect is diagrammed in Figure 23. The effect of adding HCl to O₂ on stacking fault length after oxidation of silicon is shown in Figure 24.

CHARACTERISTICS OF SILICON SELF-DIFFUSION

In order to satisfactorily explain the self-diffusion of silicon, one must reconcile the experimental results obtained by three techniques in three temperature ranges:

1. high temperature radio-tracer measurements
2. precipitation in quenched crystals at temperatures less than 850°C

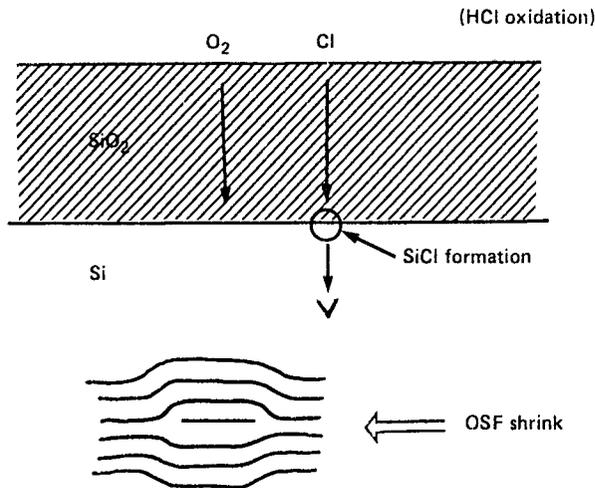


Figure 23: Diagram of SiCl formation during oxidation with the subsequent injection of vacancies. The vacancy injection reduces the concentration of self-interstitials in the bulk and causes oxidation stacking faults to shrink.

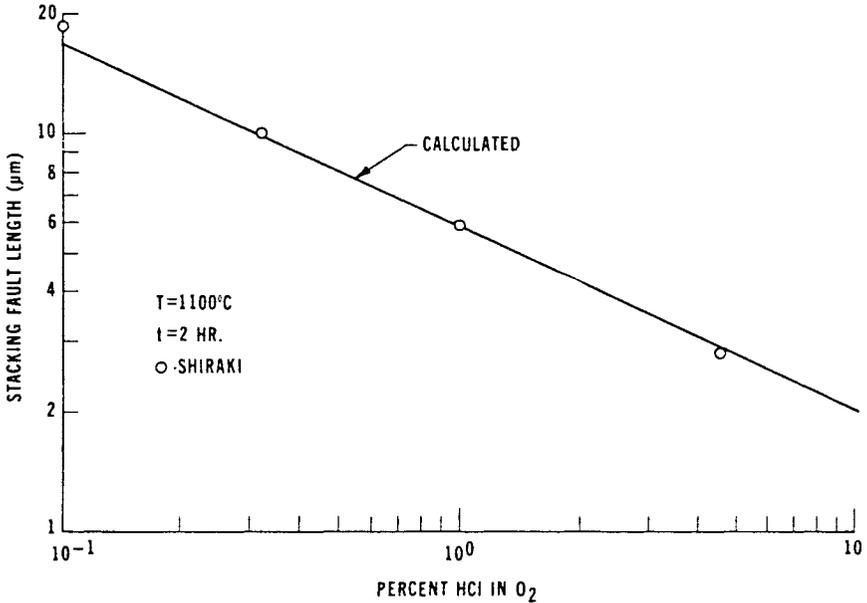


Figure 24: The effect of adding HCl to O₂ on stacking fault length after oxidation of silicon.

3. electron paramagnetic resonance measurements during the low temperature annealing (100°K) of radiation-induced defects.

Silicon self-diffusion data over the range 850 to 1380°C are shown in Figure 25.^{35,48,49} The high temperature data show an activation energy of 5.02 eV while the lower temperature show a 4.25 eV energy. Watkins and Corbett⁵⁰ reported an activation energy for self diffusion in Si of 3.9 eV. This result was obtained from low temperature annealing of E-centers, impurity-vacancy pairs at 100°K. The cause for the continual decrease in activation energy with decreasing temperature has been ascribed to different charge state vacancies dominating self-diffusion in the various temperature ranges.^{13,51,52} Thus at very low temperatures neutral vacancies may dominate self-diffusion. At high temperatures, both donor and acceptor vacancy diffusion was considered important. An alternate view was expressed by Seeger and Chik³⁵ who suggested that in Si at low temperatures, self-diffusion mainly occurs via vacancies, whereas at high temperatures it is dominated by the interstitialcy mechanism. Their observations indicate a change in the self diffusion mechanism and/or in the entropy and enthalpy of self diffusion as a function of temperature. The change in entropy with temperature can be accounted for by assuming that the form of the self-interstitial changes with temperature. For example, the entropy would increase due to a spreading out of the self-interstitial over several atomic volumes.

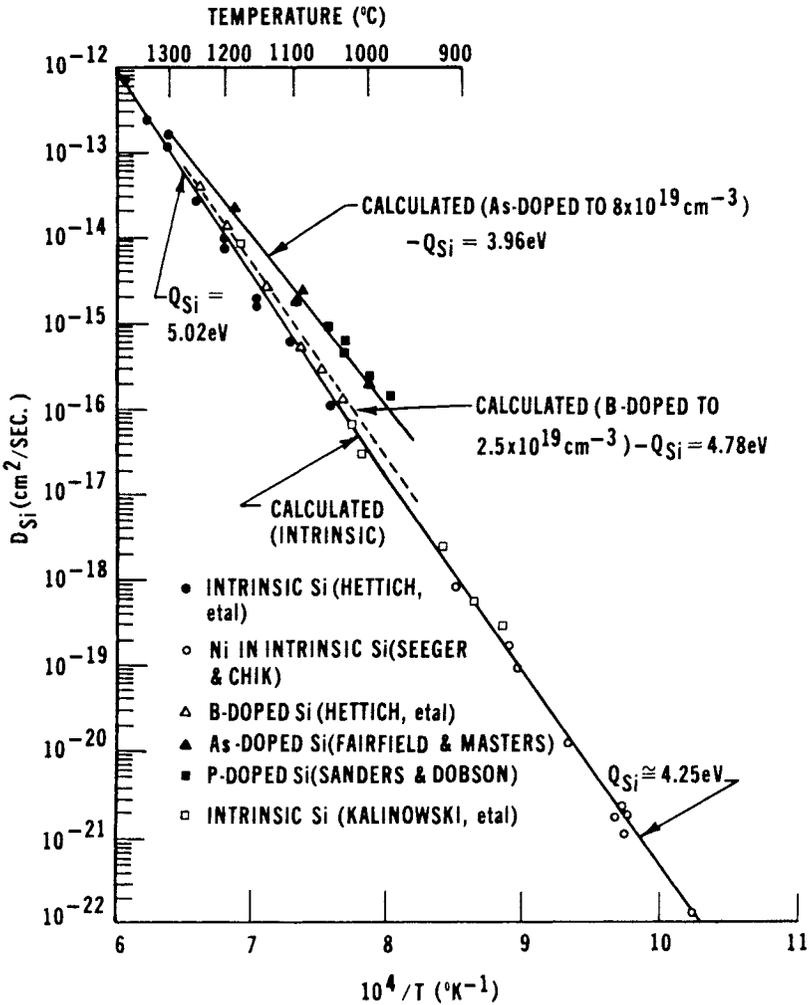


Figure 25: Self-diffusion data in intrinsic and heavily doped n and p-type silicon.

The self-diffusion data in Figure 25 shows that Si diffusion is different in heavily doped n-type or p-type silicon. In order to see how the contributions from vacancies in various ionization states can describe this effect, we will write the intrinsic self-diffusion coefficient as

$$D_{Si}^i = \frac{[V^x]}{2n_H} D_v^x + \frac{[V^-]_i}{2n_H} D_v^- + \frac{[V^=]_i}{2n_H} D_v^= + \frac{[V^+]_i}{2n_H} D_v^+ \tag{51}$$

where D_v is the vacancy diffusivity, n_H is the number of sites in the crystal, and $[V]_i$ is the intrinsic vacancy concentration. In extrinsic silicon where the

doping concentrations exceed the intrinsic electron concentration, n_p , mass action effects cause changes in ionized vacancy concentrations as shown in Figure 12. These curves obey the relations

$$\frac{[V^-]}{[V^-]_i} = \frac{n}{n_i} = \frac{[V^+]_i}{[V^+]}. \quad (52)$$

Using the relation $D_{Si} = 1/(2n_H)D_v[V]_i$, where the one-half term is the correlation factor for the diamond lattice and n_H is the number of lattice sites, the Si self diffusion coefficient becomes

$$D_{Si} = D_{Si}^x + D_{Si}^- \left(\frac{n}{n_i} \right) + D_{Si}^- \left(\frac{n}{n_i} \right)^2 + D_{Si}^+ \left(\frac{n_i}{n} \right). \quad (53)$$

The calculated curves shown in Figure 25 use Equation 53 with the following intrinsic diffusion coefficients

$$D_{Si}^x = 0.015 \exp(-3.89\text{eV}/kT), \quad (54a)$$

$$D_{Si}^- = 16 \exp(-4.54\text{eV}/kT), \quad (54b)$$

$$D_{Si}^+ = 1180 \exp(-5.09\text{eV}/kT). \quad (54c)$$

The expression for D_{Si}^- cannot be obtained by analyzing the data in Figure 25. The values of the activation energies for each term are consistent with the formation enthalpies, migration enthalpies and average free ionization energies associated with each vacancy.⁵³ Thus, at temperatures below 600°C Equation 53 predicts that Q_{Si} approaches 3.89eV as the neutral vacancy dominates self-diffusion. This agrees with the low temperature value observed by Watkins and Corbett.⁵⁰

DOPANT DIFFUSION IN SILICON

Group III and V elements as solutes in Si are unique in their ability to form strong covalent bonds with the lattice atoms. The result is they exhibit very high substitutional solubilities in Si. An important consequence of this is that it has led people to believe that these elements diffuse predominatntly via vacancies or self-interstitials. Thus, one would expect similar activation energies and pre-exponential factors for Group III or V elements in Si and for Si self-diffusion. In fact, both the pre-exponential factors and the activation energies of impurity diffusion are lower by a significant amount. To explain this phenomenon using the vacancy mechanism as an example, Hu⁵⁴ has proposed that there must exist a long-range vacancy-impurity interaction potential which would cause impurity-point defect pairing to occur. Thus, the vacancy and the impurity atom would diffuse as a pair and the additional energy required for complete dissociation would not have to be supplied. If no pairing occurred, the vacancy would have to disappear

into the lattice, and this would be the rate controlling mechanism as it is for self-diffusion. Therefore, the difference between the activation energy for self-diffusion and Group III or V impurity diffusion is less than the impurity-vacancy pair binding energy, E_b . This is illustrated in Figure 26 where a particular long-range interaction potential is assumed. The potential energy between a vacancy at a third coordination site and one infinitely removed from the impurity atom is ΔQ . Thus the activation energy for impurity diffusion, Q_i , is proposed to be

$$Q_i = Q_{Si} - \Delta Q. \tag{55}$$

Equation 55 is based upon an impurity displacement cycle which sees the vacancy first partially dissociate from the impurity atom, and go at least as far as the third coordination site to close a path around the impurity atom. If the long-range interaction potential is assumed to be Coulombic, this complements the multiple charge state vacancy model where the diffusion of donor atoms is dominated by acceptor type vacancies, etc.

Experimental evidence exists in support of coupled point defect-impurity diffusion in Si. For such a model, chemical pumping effects involving vacancies or self-interstitials would be negligibly small. Thus, for a system such as Si where the diffusivity of the solute differs considerably from that of the solvent, a vacancy or self-interstitial flux will be induced by the solute fluxes.⁵⁵ According to the chemical pump model, the influx of a fast diffusing impurity such as P or B in Si will cause an efflux of lattice vacancies or a flux of self-interstitials in the same direction as the solute atoms. However, it is observed that buried clumps of B or P approximately 5 microns from the Si surface will undergo isotropic enhanced diffusion (uniform broadening towards the surface and into the bulk) during the

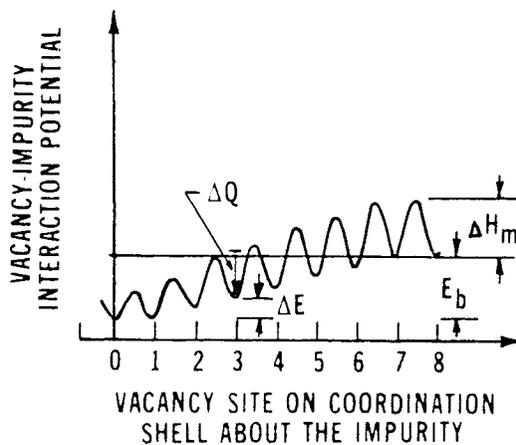


Figure 26: A schematic diagram showing a long range vacancy-impurity interaction potential which could account for the lower activation energy of impurity diffusion compared with self-diffusion in silicon.

diffusion of a high P concentration at the surface. If a chemical pumping effect were important, one would expect a buried clump of impurities to be skewed in one direction or the other depending upon whether vacancies or self-interstitials were dominant. This is not observed. This experiment is diagrammed in Figure 27.

Other types of experiments such as proton-enhanced diffusion have been used to support the notion of coupled impurity - point defects in pair diffusion. For example, consider a uniformly doped sample that is irradiated with high energy protons (see Figure 28a). The production of vacancies and self-interstitials occurs mainly in a region near the average projected range of the protons, R_p (see Figure 28b). These point defects diffuse into the bulk and towards the surface. This in turn produces a non-uniform redistribution of the dopant atoms in the Si as shown in Figure 28c.⁵⁶ The initially uniform B or P doping shows a dip centered at R_p between two smaller peaks. Thus the dopant atoms diffuse in the same direction as the diffusing point defects.

Two known mechanisms can account for these results. Either the dopant atoms respond to the chemical pumping of a self-interstitial flux, or they become tightly bound to either type of point defect and diffuse as a

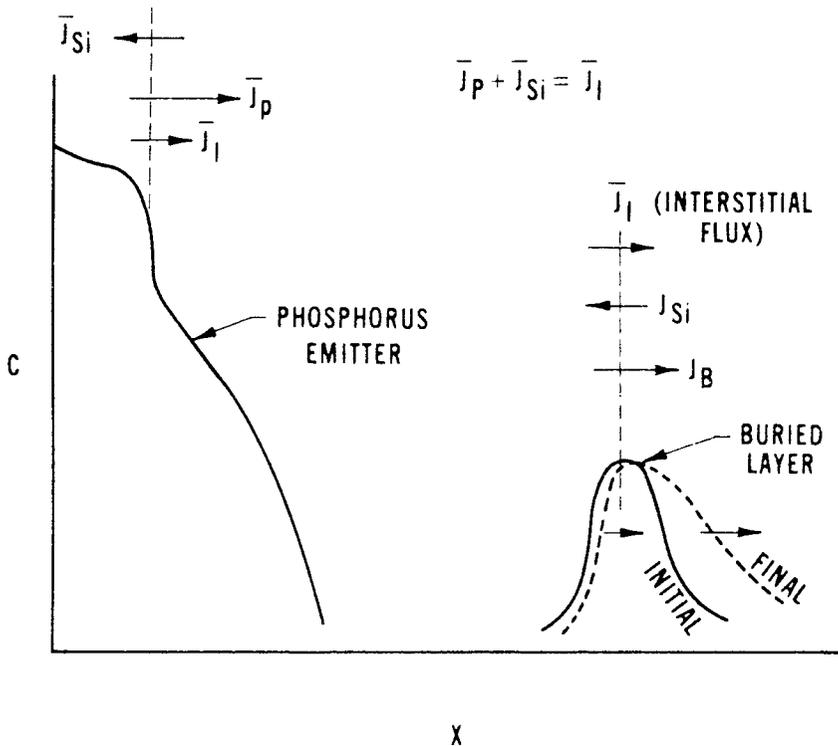


Figure 27: A schematic diagram showing the expected redistribution of a buried layer under the influence of a flux of self-interstitials chemically pumped from the surface by a phosphorus diffusion.

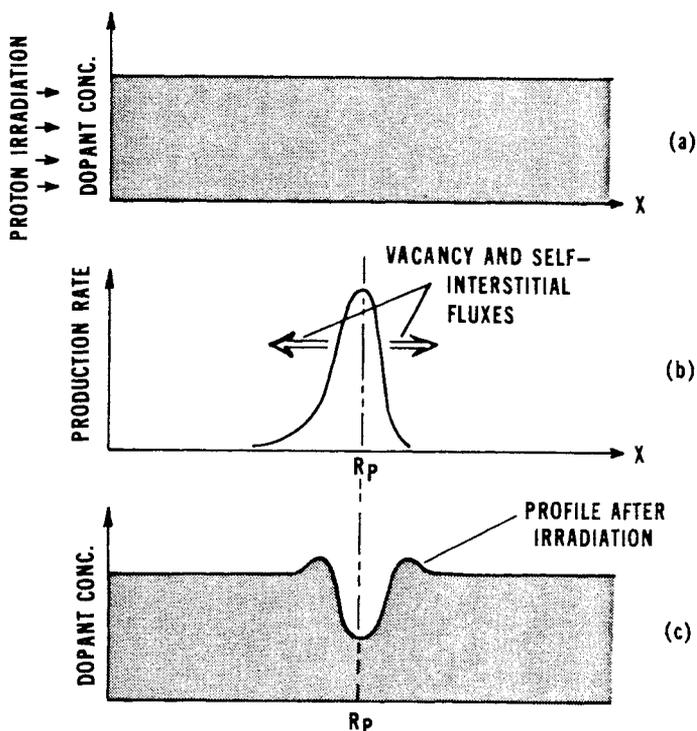


Figure 28: Proton-enhanced diffusion experiments supporting the notion of coupled impurity-point defect pair diffusion in silicon. (a) initial distribution of dopant in silicon, (b) production rate of vacancies and self-interstitials due to proton irradiation, (c) dopant concentration after irradiation.

dopant atom-point defect pair.³⁷ From our previous discussion the chemical pumping effect is unlikely. Thus the question remains: which type of point defect is more likely to pair up with the dopant atoms? The answer appears to depend in an unpredictable way on the type of dopant atom. For that reason, the following sections will describe what is known about each of the important dopants in Si.

Arsenic Diffusion Models

Arsenic is believed to diffuse primarily via a vacancy mechanism. In this section the current understanding of As diffusion via a vacancy mechanism will be reviewed. The characteristics of As diffusion as interpreted by this model are summarized in Table 5. The entropy of diffusion of the As^+V^- pair is much larger than the entropy of the As^+90V^x pair.¹³ So, in spite of the larger activation energy ($Q_{As^+V^-}=4.05\text{eV}$ and $Q_{As^+90V^x}=3.44\text{eV}$), the As^+V^- pair dominates As diffusion above 1050°C , as shown in Figure 29. Diffusion of As^+V^- pairs is less important than in the case of P. This is due to a smaller pair binding energy (1.37eV). In an oxidizing ambient little

Table 5: Characteristics of Arsenic Diffusion (Vacancy Model Interpretation)

Property	Result
$\Delta S_D^- \gg \Delta S_D^x$	As^+V^- pair diffusion dominates at $T > 1050^\circ C$
As^+V^- binding energy is ~ 0.25 eV less than P^+V^-	Fewer As^+V^- pairs than P^+V^- <ol style="list-style-type: none"> Less gettering of metal donors No emitter push above $700^\circ C$ No effect on $[O_i]$ precipitation
VAs_2 pair binding energy = 1.6 eV	<ol style="list-style-type: none"> Significant $[VAs_2]$ form at $n (= 2 \times 10^{20} \text{ cm}^{-3})$ at $1000^\circ C$ Reduces n solubility Causes retarded base diffusion in some cases
2.5 eV required to make As interstitial	Little oxidation-enhanced diffusion

enhanced diffusion of As occurs because 2.5eV (Table 5) is required to create an As interstitial atom. Thus diffusion continues to be controlled primarily by the local vacancy concentration rather than by the Si self-interstitial supersaturation created by the oxidizing surface.

It can be seen in Figure 29 that As diffusivity can be enhanced when diffusion occurs in heavily doped n-type Si. This concentration dependence is further described in Figure 30 where D_{As} versus As concentration curves are plotted. The implication of such a concentration dependence is shown in Figure 31. The As impurity profile no longer is described by an erfc solution to the continuity equation for diffusion. An approximation to this solution will be given in a later section.

Another feature of the As profile in Figure 31 is that the total As curve deviates from the measured free electron concentration curve. It has been proposed that this discrepancy is due to As clustering at high temperatures.⁵⁷ Thus, much of the study of As diffusion in Si has centered around the clustering phenomenon. The most recent contribution by Guerrero, et al⁵⁸ suggests that clustering involves As atoms along with one negative charge (electron or V^-). This result was obtained by fitting various models to room temperature free-electron concentration versus total As data such as that shown in Figure 32. The functional dependence of n on C_T (total As concentration) depends upon the source of As (implantation or diffusion) and temperature. Other authors have tried to fit similar data with calculations and have arrived at different conclusions regarding the chemical and electrical form of the As cluster.

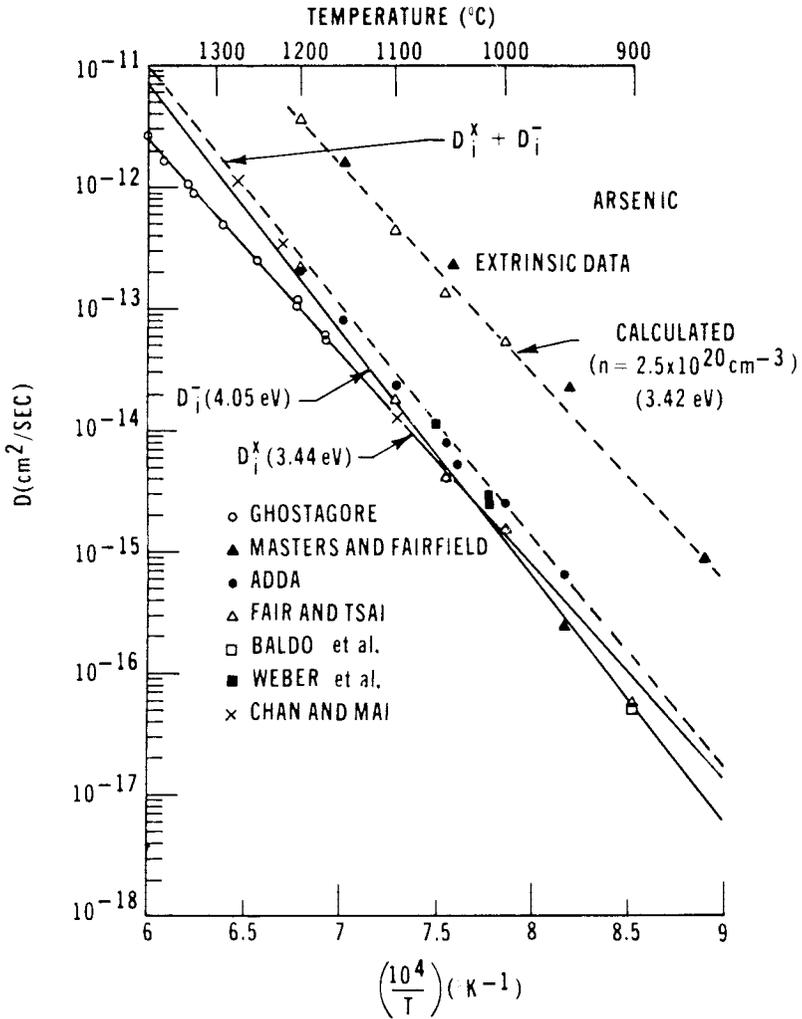


Figure 29: Arsenic diffusivities in intrinsic and extrinsic n-type silicon.

The clustering reaction can act to reduce As diffusion. Consider one model in which the As cluster is a vacancy with two As atoms— VA_{s_2} .⁵⁷ The reaction is



Then, the total As concentration is

$$C_T = C_{As} + C_{VA_s} + C_{VA_{s_2}} \tag{57}$$

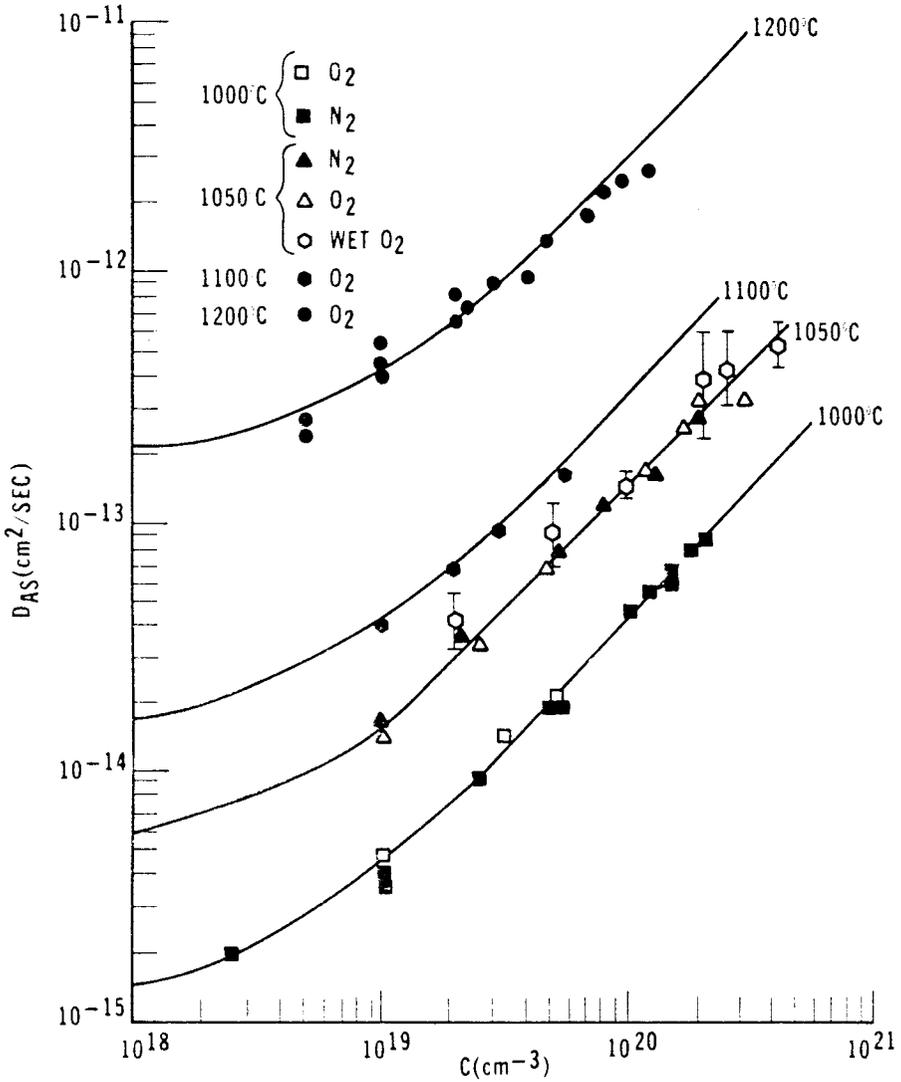


Figure 30: Arsenic diffusivity vs. total arsenic concentration. The solid curves are calculated. The data represent diffusions in various furnace ambients.

where the formation of V^-As^+ and $V=As^+$ pairs are included. By writing down the equilibrium reactions for the formation of each species in Equation 57 and setting $C_{As} = n$, then C_T becomes

$$C_T = n + K_a(T)n^2 + K_b(T)n^3 + K_c(T)n^4, \tag{58}$$

where $K(T)$'s are the equilibrium constants. Equation 58 can be used to describe the data in Figure 32.

The effective diffusion coefficient of As is defined as

$$D_{As} \equiv -J (\partial C_T / \partial x)^{-1}, \tag{59}$$

which is Fick's First Law. If J is the flux of monatomic As, C_{As} , with a linear concentration-dependent diffusivity, then

$$J = -D_i \frac{C_{As}}{n_i} \frac{\partial C_{As}}{\partial x}, \tag{60}$$

$$= -D_i \frac{C_{As}}{n_i} \frac{\partial C_{As}}{\partial C_T} \frac{\partial C_T}{\partial x}. \tag{61}$$

Substituting Equation 58 and solving for D_{As} yields

$$D_{As} = \frac{D_i (n/n_i)}{1 + \frac{4K_c(T)n^3}{1 + 3K_b(T)n^2}} \tag{62}$$

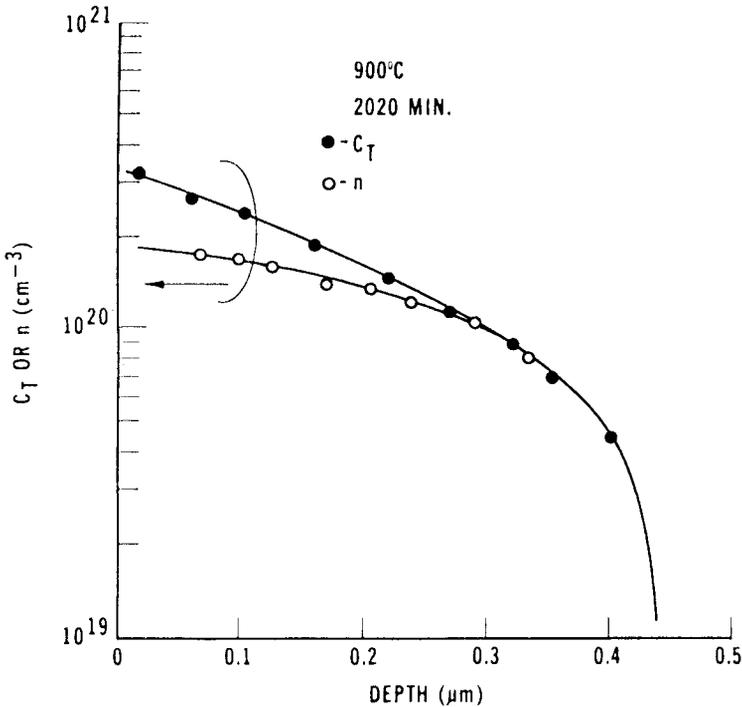


Figure 31: Total arsenic C_T and free electron profiles in silicon following an arsenic diffusion.

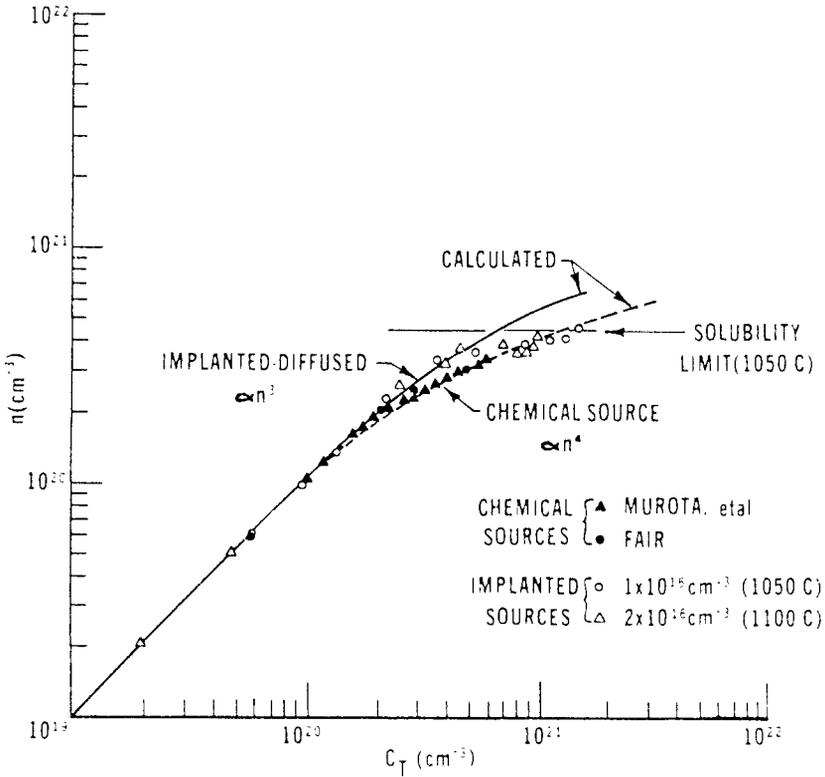


Figure 32: Electron concentration vs. total arsenic for chemical source arsenic diffusion and diffusion of an ion-implanted arsenic layer with the same integrated concentration.

D_{As} is the effective diffusivity of As in Si that results when monatomic As^+ diffuses substitutionally while inactive, stationary As complexes (VA_{s_2}) are forming to reduce the flux of diffusible ions. Equation 62 is plotted in Figure 33 and is compared with measured data of D_{As} versus C_T . The reduction in D_{As} when clustering occurs is evident.

The individual components that make up Equation 57 are shown plotted in Figure 34 at 700°C and 1000°C . At 700°C the contribution of $V=As^+$ becomes significant. This also corresponds to the temperature at which As diffusion becomes greatly enhanced, much like P diffusion at higher temperatures.

Phosphorus Diffusion Models

The characteristics of P diffusion in Si are summarized in Table 6 using the vacancy model of diffusion for interpretation. As a result of the large diffusion entropy of P with neutral vacancies, ΔS_d^* , low-concentration P diffusion is dominated by the availability of V^* vacancies.¹³ At high concentrations ($C \gg n_i$) the P^+V^- dominates diffusion. Thus, the diffusivity shows

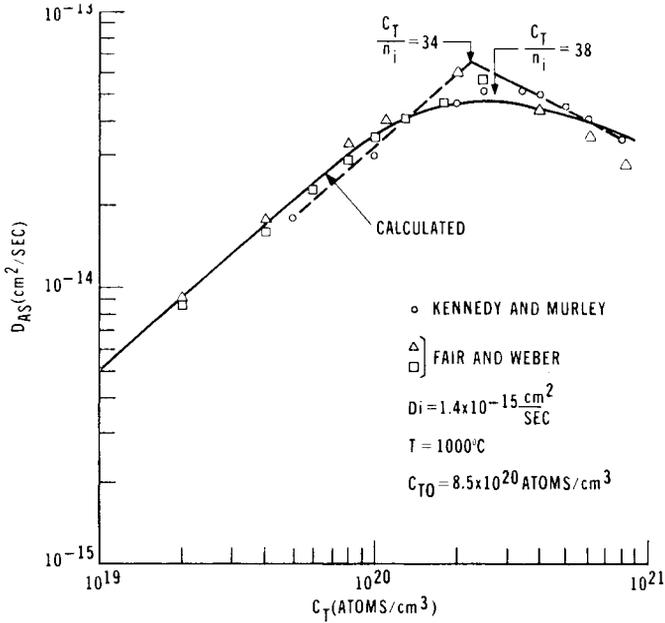


Figure 33: The effect of diffusivity of arsenic vs. total concentration for diffusions into p-type silicon at 1000°C.

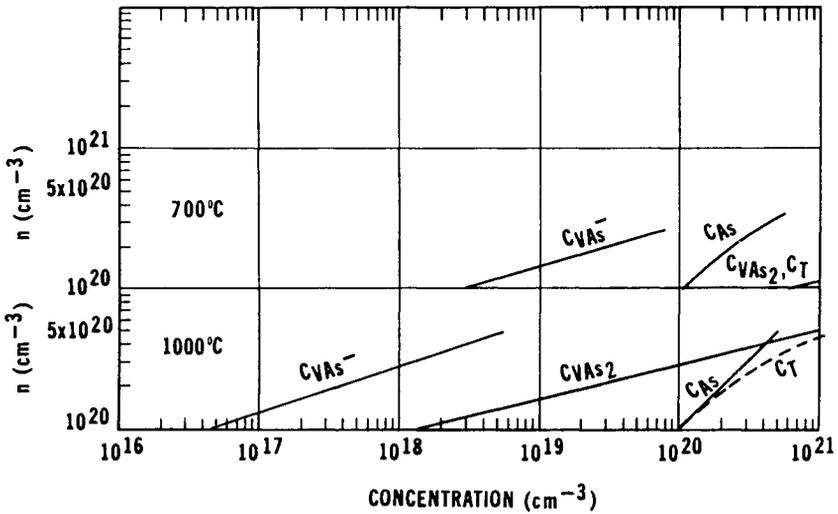


Figure 34: Arsenic and arsenic-vacancy pair concentrations vs. electron concentrations at two temperatures.

an n^2 dependence for $n \gg n_i$. The relative unimportance of P^+V^- pair migration to the total P diffusivity is due to the small size of the P atom. The small covalent radius ($r_c = 1.10 \text{ \AA}$) causes the pair migration entropy, ΔS_{PV} , to be very small.

The diffusion coefficients that have been attributed to P^+V^x , P^+V^- and $P^+V^=$ pair migration are shown in Figure 35. The Fair-Tsai¹² model uses these three diffusion coefficients to explain the unique shape of the high concentration P profile shown in Figure 36. Features of this profile include:

- a difference between total P and free carrier concentration near the Si surface
- a kink in the profile
- a tail region of enhanced diffusivity

Two more major features associated with P diffusion in Si are defect generation and supersaturation of point defects. Because the tetrahedral covalent radius of P is smaller than that of Si, contraction of the host lattice can occur at sufficiently high doping concentrations, leading to the generation of misfit dislocation arrays. The generation of excess point defects and, thus, the formation of a tail on the P profile has been proposed to be a result of the dissociation of $P^+V^=$ pairs.¹² At P concentrations above 10^{20} cm^{-3} the dominant diffusing species is thought to be the $P^+V^=$ pair. According to the Fair-Tsai theory, these pairs dissociate when n drops below 10^{20} cm^{-3} at the diffusion front. The vacancy then changes charge state ($E_c - E_f = 0.11 \text{ eV}$ is the second acceptor level of the monovacancy), and the binding energy of the $P^+V^=$ decreases. The probability for dissociation is thus enhanced. The resulting vacancy flux is shown in Figure 37.

The correctness of the Fair-Tsai theory has recently come into question as a result of experiments in which a high concentration of P was diffused

Table 6: Characteristics of Phosphorus Diffusion (Vacancy Model Interpretation)

Property	Result
Large ΔS_D^x (11.6k)	Diffusion via V^x is dominant at low concentrations
Large $P^+V^=$ pair binding energy (1.57 eV)	$P^+V^=$ pair dominates high concentration diffusion <ol style="list-style-type: none"> a. $P^+V^=$ pairs compensate monatomic P^+ b. Gettering of donor metal ions $\rightarrow P^+M^-$
$P^+V^=$ dissociation at $E_c - E_f = 0.11 \text{ eV}$	<ol style="list-style-type: none"> 1. Emitter push effect 2. Defect shrinkage near junction 3. Reduced $[O_i]$ precipitation
Small covalent radius	<ol style="list-style-type: none"> 1. Misfit strain <ol style="list-style-type: none"> a. Gettering by misfit dislocations b. Reduced diffusivity through band gap narrowing 2. $\Delta S_{P-V}^- < \Delta S_{As-V}^-$

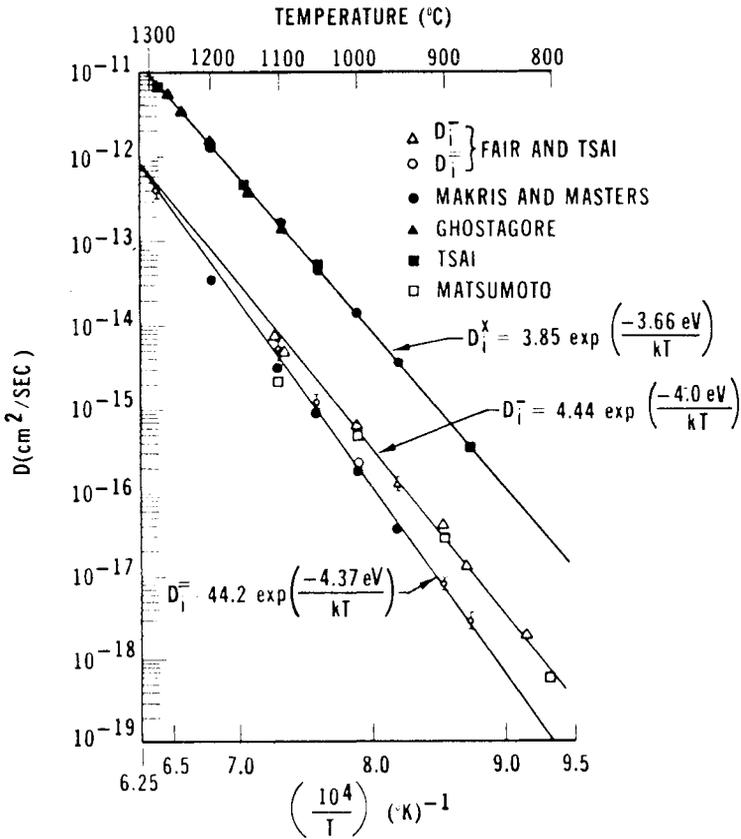


Figure 35: Phosphorus diffusivities in intrinsic silicon.

into an epitaxial layer grown over an Sb buried layer.^{59,60} The point-defects generated by the P surface diffusion retarded the diffusion of the buried Sb layer. Retarded diffusion of Sb has also been observed under an oxidizing surface in which Si self-interstitials are generated.^{31,32}

Recently these experiments were repeated in which the P diffusions were characterized and the surface concentrations were above the solid solubility limit. Over the temperature range 900-1150 $^{\circ}\text{C}$, reduced and enhanced diffusion of buried Sb or As layers respectively was observed.²² In addition, extrinsic stacking faults over the buried layers were observed to grow at the same time. Thus, the generation of self-interstitials seems to be associated with supersaturated P diffusion. Along these same lines, Nishi and Antoniadis⁶¹ have shown that oxidation-induced stacking faults (OSF) grow faster or shrink slower with increasing phosphorus/ cm^2 both within the P-doped layer and below this layer. The P was introduced from a supersaturated chemical deposition system. Plots of P diffusion junction depth and change in OSF length reduced by \sqrt{t} are shown in Figure 38 versus

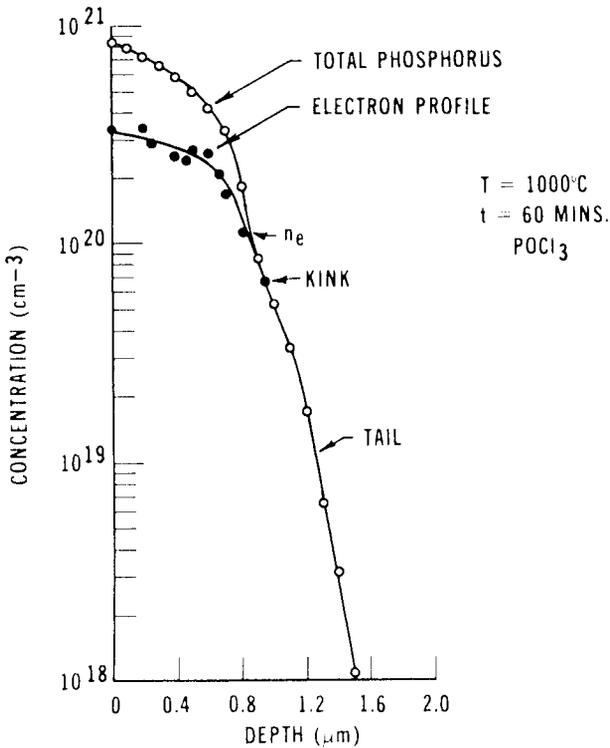


Figure 36: Total phosphorus and electron concentration profiles obtained by secondary ion mass spectrometry and differential conductivity measurements, respectively.

phosphorus dose, Q_p . These data were extracted from Nishi's work. Note that the doping conditions which create rapid growth of OSF's in the P layer did not enhance the P junction depth. This may be due to P precipitation or the recombination of generated self-interstitials with vacancies which P needs to diffuse.

Two recent experiments have yielded considerable support for an interstitialcy mechanism for P diffusion. Fahey, *et al*⁴¹ have observed that direct nitridation of Si produces a supersaturation of vacancies such that P diffusion is substantially reduced below the surface and Sb diffusion is enhanced. Estimates of the fractional interstitialcy component of P by these authors are 70-100%. Results are shown in Figure 39.

In another experiment, Nishi, *et al*⁶¹ have observed stacking fault growth beneath P diffused layers which had surface concentrations below solid solubility. This experiment provides strong evidence that the point defect injected by P diffusion is the self-interstitial. No quantitative model exists to explain this result. However, calculations of the self-interstitial supersaturation, C_i/C_i^* , caused by high concentration P diffusion are shown in Figure 40. These calculations are based upon measured stacking fault growth beneath P diffusions.

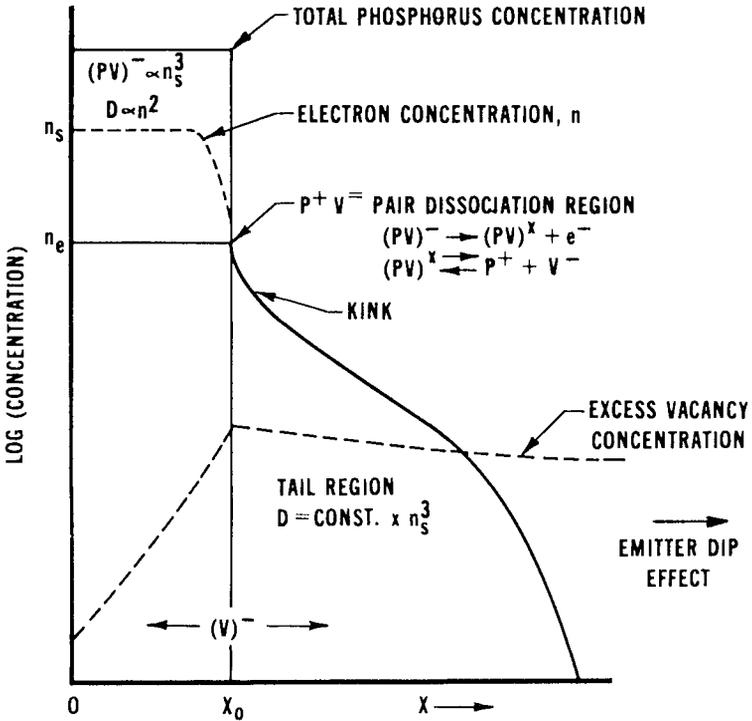


Figure 37: Idealized phosphorus profile and vacancy generation model (after Fair and Tsai). P^+V^- pairs formed in the surface region dissociate when the electron concentration drops to $n = n_e$. At this concentration the Fermi-level coincides with the second acceptor level of V^- . The freed vacancies diffuse until they recombine with P^+ atoms in the tail region.

Boron Diffusion Models

The vacancy model for B diffusion assumes that under non-oxidizing conditions, B diffuses exclusively by exchanging with V^+ vacancies.⁶² This is because the B^-V^+ migration energy is approximately 0.5eV lower than for other impurity/vacancy pairs. As a result its diffusivity is enhanced if $p > n_i$ and is reduced if $p < n_i$. A summary of the characteristics of B diffusion are shown in Table 7.

Diffusion coefficient data for B are shown in Figure 41 for intrinsic diffusion and extrinsic diffusion in both p^+ and n^+ Si.⁶² Boron is unique in that its diffusivity can be reduced by up to a factor of 10 when it diffuses in n^+ Si.⁶² Additional data showing this effect are shown in Figure 42 where D_B normalized to D_i (intrinsic diffusion coefficient) is plotted versus p/n_i . In B-doped Si compensated by donor dopants, $p/n_i \leq 1$. Calculated curves that fit these measured data show that the energy level of the V^+ vacancy is 0.05 eV above the valence band of Si.

The importance of this effect is illustrated in Figure 43. Measured B profile data are shown for a B implant into Si with an As background doping of $5 \times 10^{19} \text{ cm}^{-3}$.⁶² After annealing at 1050°C for 23 min. the B has diffused less

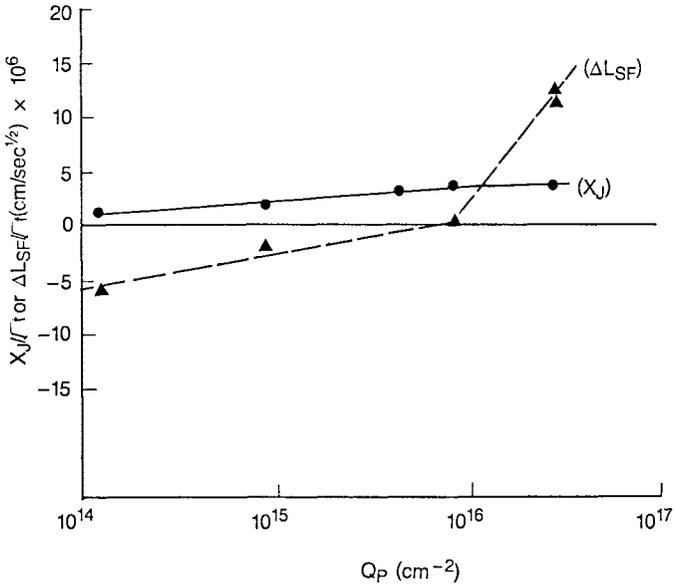


Figure 38: Phosphorus diffusion junction depth and oxidation stacking fault length change as a function of integrated phosphorus concentration (Nishi, et al).

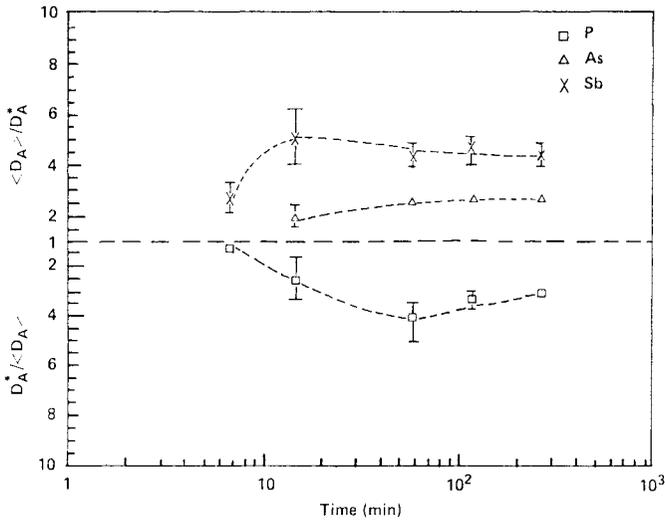


Figure 39: The effect of direct silicon nitridation on phosphorus, arsenic and antimony diffusion. The nitridation process creates a supersaturation of vacancies which substantially enhances antimony diffusion, partially enhances arsenic diffusion, and reduces phosphorus diffusion.

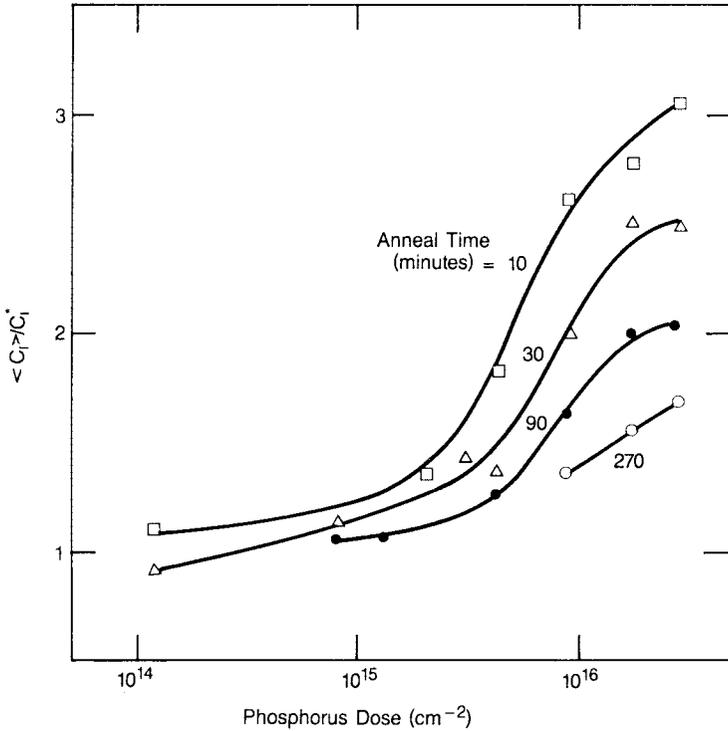


Figure 40: Calculated self-interstitial supersaturation vs. integrated phosphorus concentration (after Nishi, et al).

Table 7: Characteristics of Boron Diffusion (Vacancy Model Interpretation)

Property	Result
B^-V^+ migration energy is ~ 0.6 eV	Relatively fast diffuser dominated by $[V^+]$
Small tetrahedral covalent radius	<ol style="list-style-type: none"> 1. Small diffusion entropy 2. Misfit strain <ol style="list-style-type: none"> a. Dislocations b. Good gettering c. Reduced diffusivity
2.26 eV required to make B interstitial	Oxidation enhanced diffusion
Forms stable oxides - B_2O_3 and HBO_2	Segregation into growing SiO_2 from Si

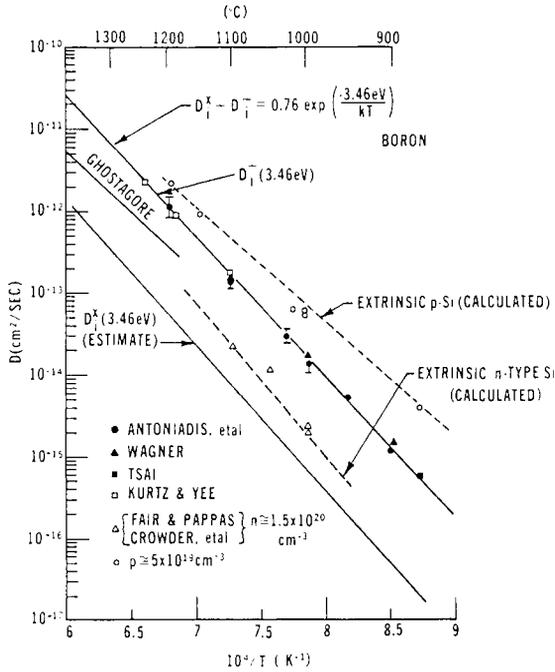


Figure 41: Boron diffusivities in intrinsic and extrinsic n and p-type silicon.

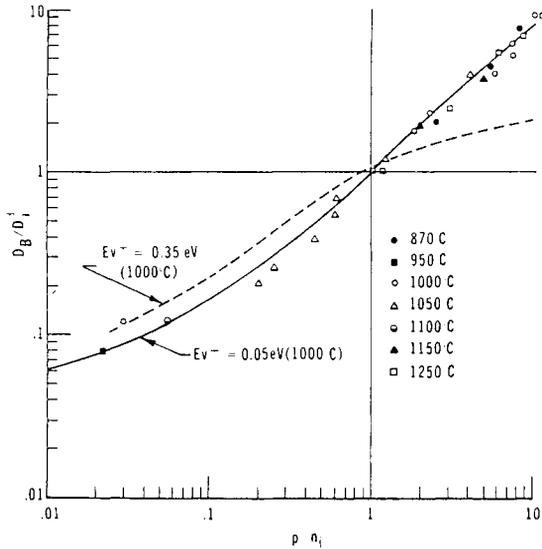


Figure 42: Normalized boron diffusivity vs. p/n_i . The solid curve is calculated with the vacancy donor level at $E_v + 0.05$ eV. The dashed curve is calculated with the vacancy donor level at $E_v + 0.35$ eV.

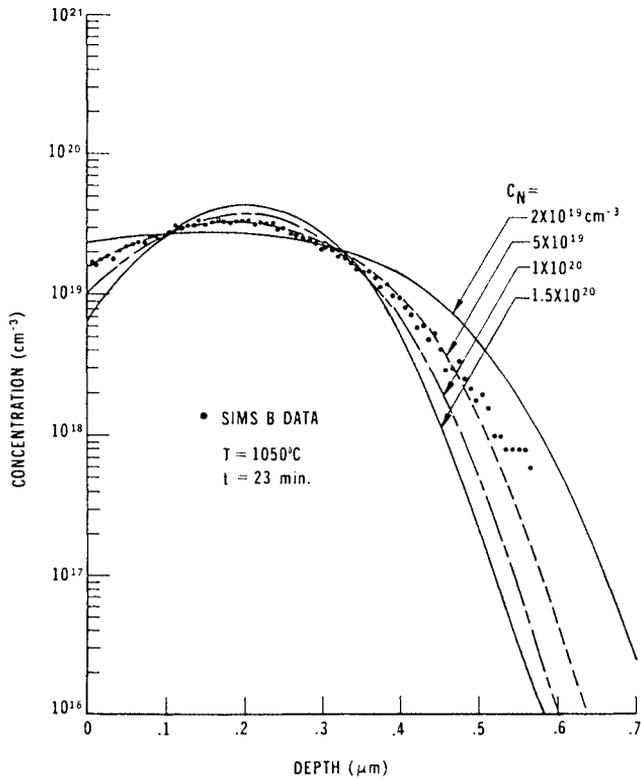


Figure 43: Calculated boron profiles as a function of n-type background doping, C_N . Experimental SIMS data of a 10^{15} boron/cm², 50 keV implant diffused at 1050°C for 23 minutes in N₂ in an n-type background doping of 5×10^{19} cm⁻³.

than if the background As were 2×10^{19} cm⁻³ and more than if higher As concentrations were present.

The results of experiments utilizing oxidation, direct nitridation of Si and the emitterpush effect in the presence of high concentration P diffusion show that B diffusion is enhanced in the presence of self-interstitial supersaturation. It is reasonable to expect, therefore, that B diffusion occurs to some degree by an interstitialcy mechanism under non-injecting conditions as well. Estimates of the fractional interstitialcy component of B diffusion range from 17%-100%.

DESIGN CONSIDERATIONS FOR IMPLANTED-DIFFUSED LAYERS

In this section useful curves and equations are developed to allow designers to estimate diffusion temperatures and times for establishing junctions from ion implanted layers in Si.

Arsenic Diffusion

For surface concentrations $\gg n_i$, the diffusivity of As is

$$D(C) = \frac{2D_1C}{n_i} \tag{63}$$

The continuity equation that must be solved with Equation 63 is

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left[D(C) \frac{\partial C}{\partial x} \right]. \tag{64}$$

No exact analytical solution exists, but an approximate solution is

$$C = C_s (1.00 - 0.87Y - 0.45Y^2) \tag{65}$$

where

$$Y = x \left(\frac{8C_s D_1}{n_i} \right)^{-1/2} \tag{66}$$

Equation 65 is plotted in Figure 44 on normalized scales and compared with data.

Equation 65 can be solved for the junction depth, x_j , when $C = 0$. Thus, for ion-implanted layers that are diffused;

$$x_j = 2 \left(\frac{Q_T D_1 t}{n_i} \right)^{1/3} \tag{67}$$

Use was made of the result obtained by integrating Equation 65 to give the “dose” of arsenic, Q_T :

$$Q_T = \int_0^{x_j} C dx = 0.53 C_s x_j \tag{68}$$

Equation 68 is shown plotted in Figure 45 and compares well with measured data from As implants that were diffused at the temperatures shown. Equation 67 is shown plotted in Figure 46 and is compared with measurements.

Example: A 3×10^{15} As/cm² implant is performed in Si at 50 keV. What is the junction depth after an 1100°C, 1 hour anneal in a N₂ furnace ambient?

Using Equation 67:

$$\begin{aligned} x_j &= 2 \left(\frac{Q_T D_1 t}{n_i} \right)^{1/3} \\ &= 2 \left(\frac{(3 \times 10^{15} \text{ cm}^{-2}) (2 \times 10^{-14} \text{ cm}^2/\text{sec.}) (3600 \text{ sec.})}{1 \times 10^{19} \text{ cm}^{-3}} \right)^{1/3} \\ &= 0.57 \times 10^{-4} \text{ cm.} \end{aligned}$$

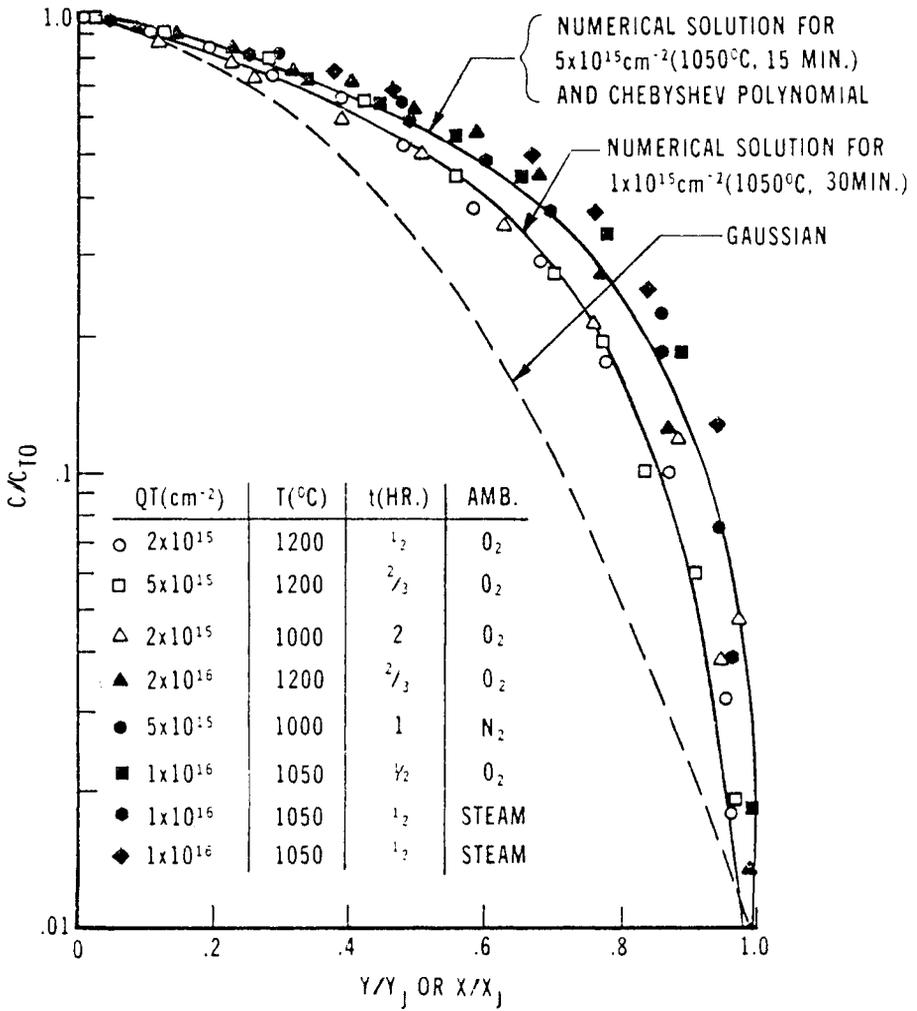


Figure 44: Normalized total arsenic profiles of implanted-diffused layers of silicon. The exact numerical solutions to the diffusion equation are compared with the polynomial approximation for two different implant doses.

The surface concentration after anneal is

$$C_s = 1.9 Q_T/x_j$$

$$\approx 1 \times 10^{20} \text{ cm}^{-3}$$

Substitution of Equation 67 into 68 and solving for C_s gives

$$C_s = 0.94 \left(\frac{Q_T^2 n_i}{D_j t} \right)^{1/3} \tag{69}$$

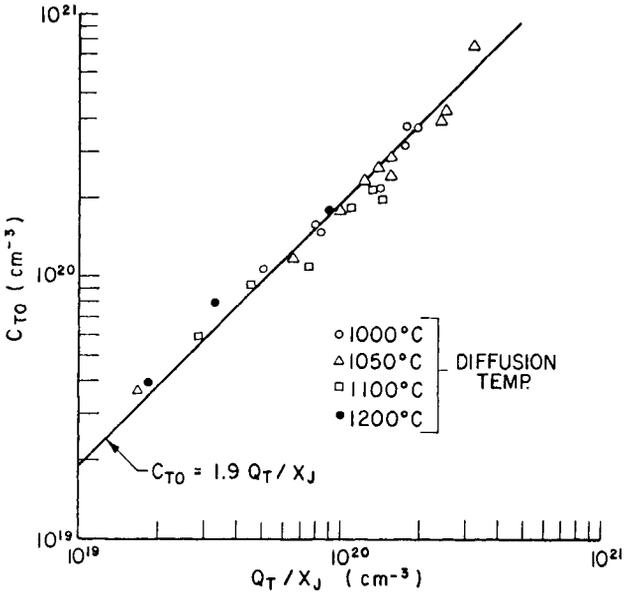


Figure 45: Total arsenic surface concentration vs. average doping.

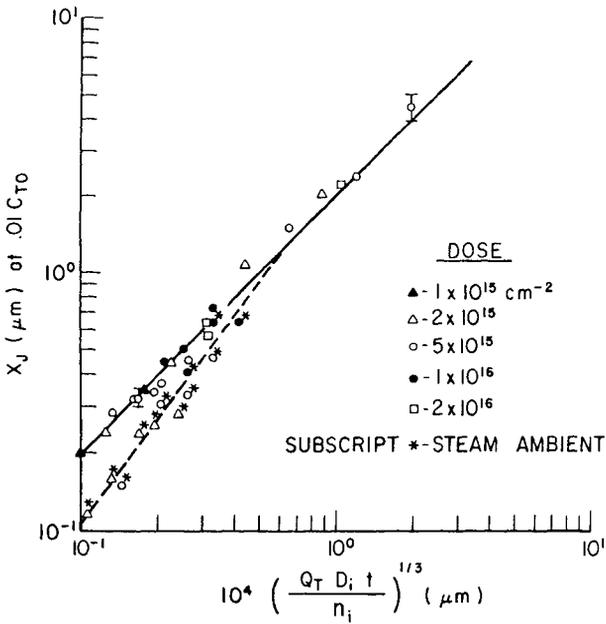


Figure 46: Time, temperature and dose dependence of the junction depth of arsenic implants after diffusion.

Thus, C_s decreases during annealing of an implanted As layer as $t^{-1/3}$.

Sheet resistance measurement can be used to characterize high dose As implanted layers after diffusion:

$$R_s = 1/(q\bar{\mu} \int_0^{x_j} C dx), \tag{70}$$

where R_s is the sheet resistance (ohms/square) and the effective bulk carrier mobility in the diffused layer. It has been shown that Equation 70 yields

$$C_s = \frac{6.26 \times 10^{15}}{(R_s x_j)^{3/2}}, \tag{71}$$

and the time dependence of R_s is

$$R_s = \frac{1.7 \times 10^{10}}{Q_T^{7/9}} \left(\frac{n_i}{D_i t^{1/9}} \right). \tag{72}$$

Equation 71 is plotted in Figure 47 and compared with measured electrically active surface concentration data. Equation 72 is a useful design equation and is shown plotted in Figure 48 for various As implant doses. The times and temperatures required to achieve a given sheet resistance for a given dose are shown. Using Figure 48 in conjunction with

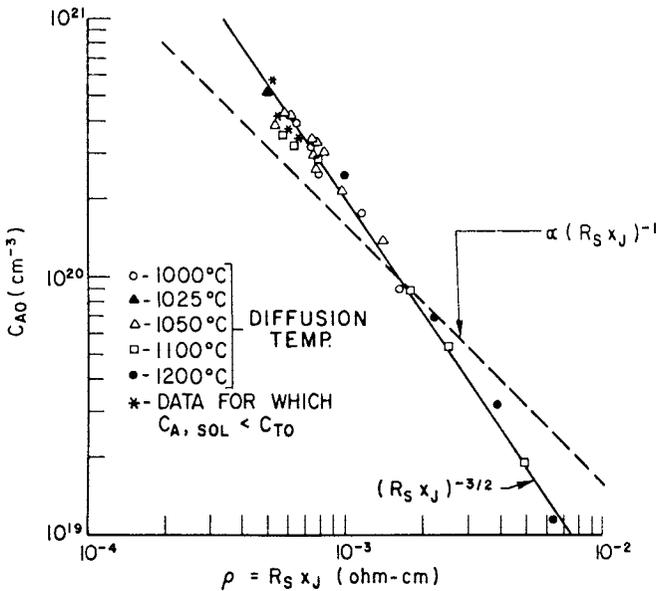


Figure 47: Electrically active arsenic surface concentration vs. average layer resistivity.

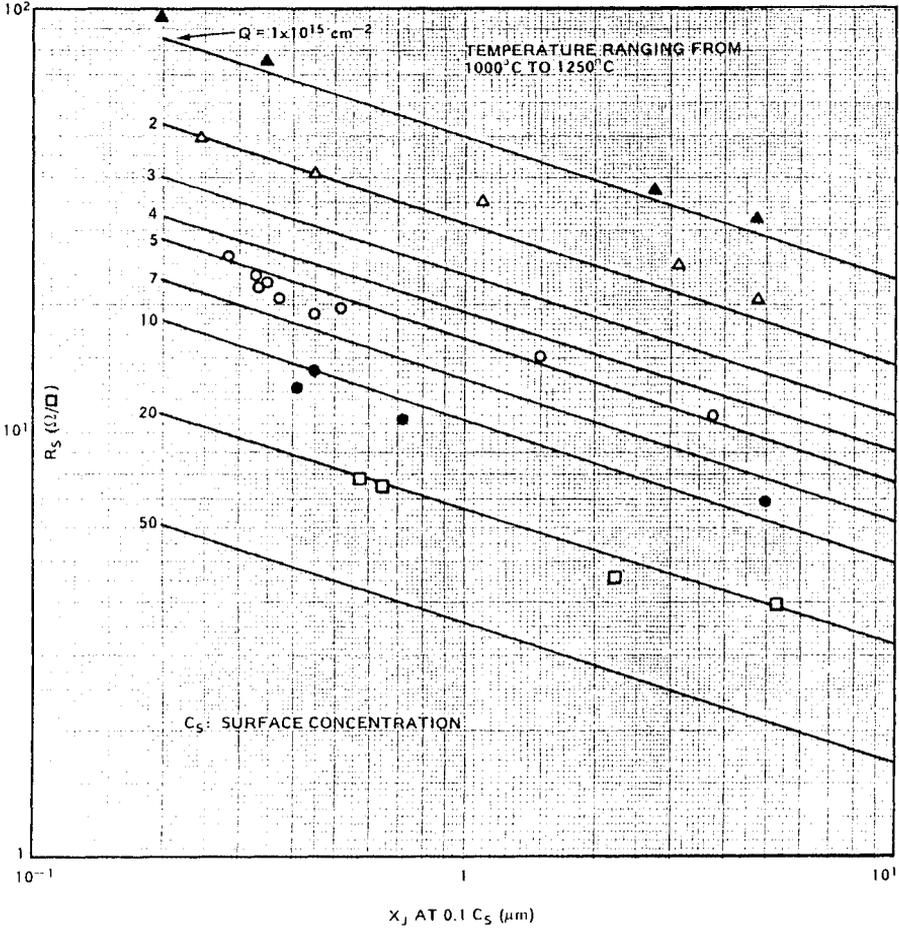


Figure 48: Sheet resistance vs. junction depth for implanted arsenic layers diffused in N_2 or O_2 .

Figure 49 allows one to design a simple one-step process. The procedure is:

1. From Figure 49 pick the As dose required to give the desired R_s and x_j values.
2. From Figure 48 pick the time and temperature combination that intersect with the R_s and dose values specified.

Phosphorus Diffusion

As pointed out earlier, high concentration P diffusion is complicated by multiple species diffusion mechanisms. Thus, no simple solution exists to the diffusion equation. However, some useful curves have been generated for designing ion implanted P layers with a single step anneal.

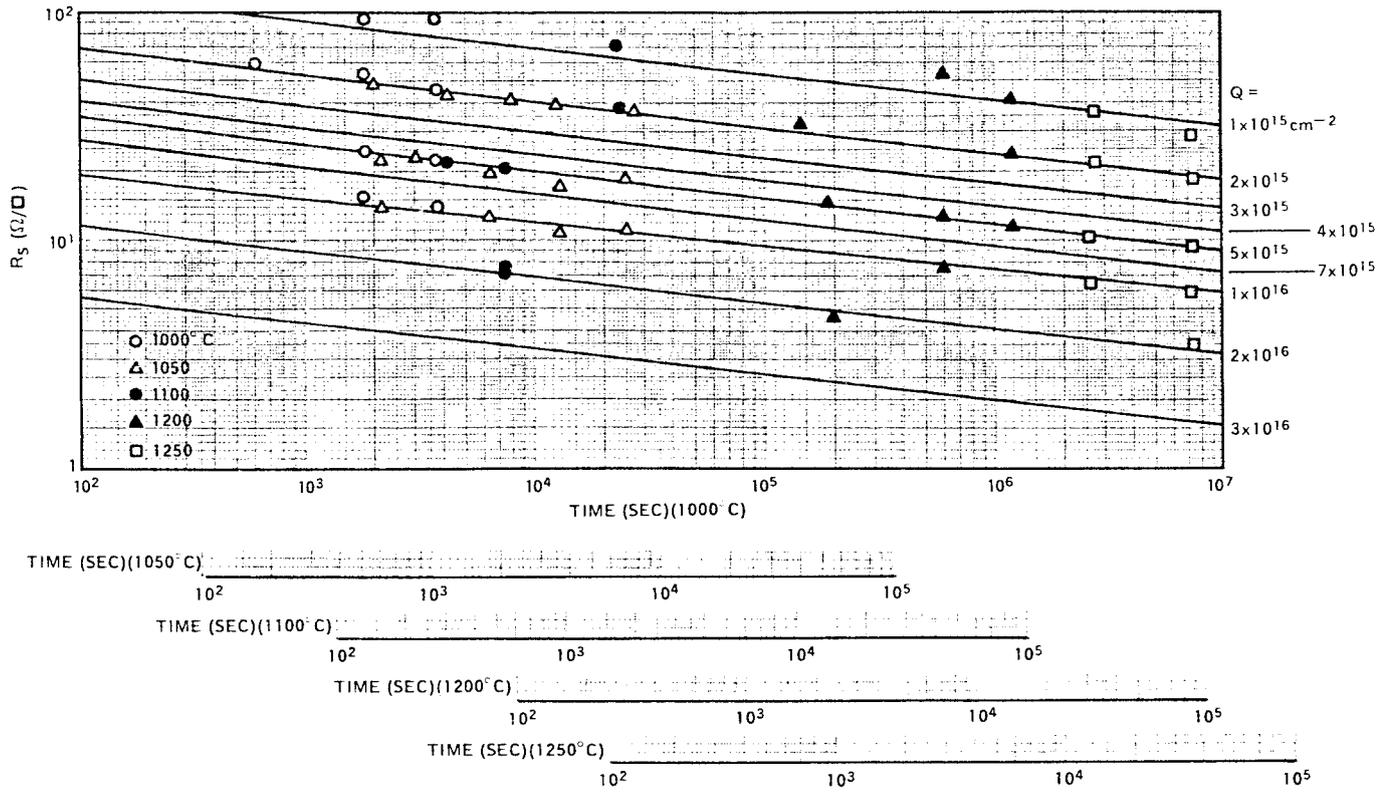


Figure 49: Time, temperature and arsenic dose dependence of sheet resistance (diffusions in N_2 or O_2).

Calculated curves of R_s versus x_j for various P implant doses are shown in Figure 50. These curves are similar to the As curves shown in Figure 49. The times and temperatures required to achieve a given sheet resistance for a given dose of P are shown in Figure 51. The design procedure is similar to the procedure described in the As section.

ION IMPLANTATION

Ion implantation is the introduction of ionized atoms into a silicon substrate with sufficient energy to penetrate beyond the surface region. This requires that the ions have energies above 3 keV if they are either boron, phosphorus or arsenic. Such an energy will allow these atoms to penetrate beyond any surface layer of native SiO_2 . Therefore, any barrier effect of the surface oxide on impurity introduction is avoided.

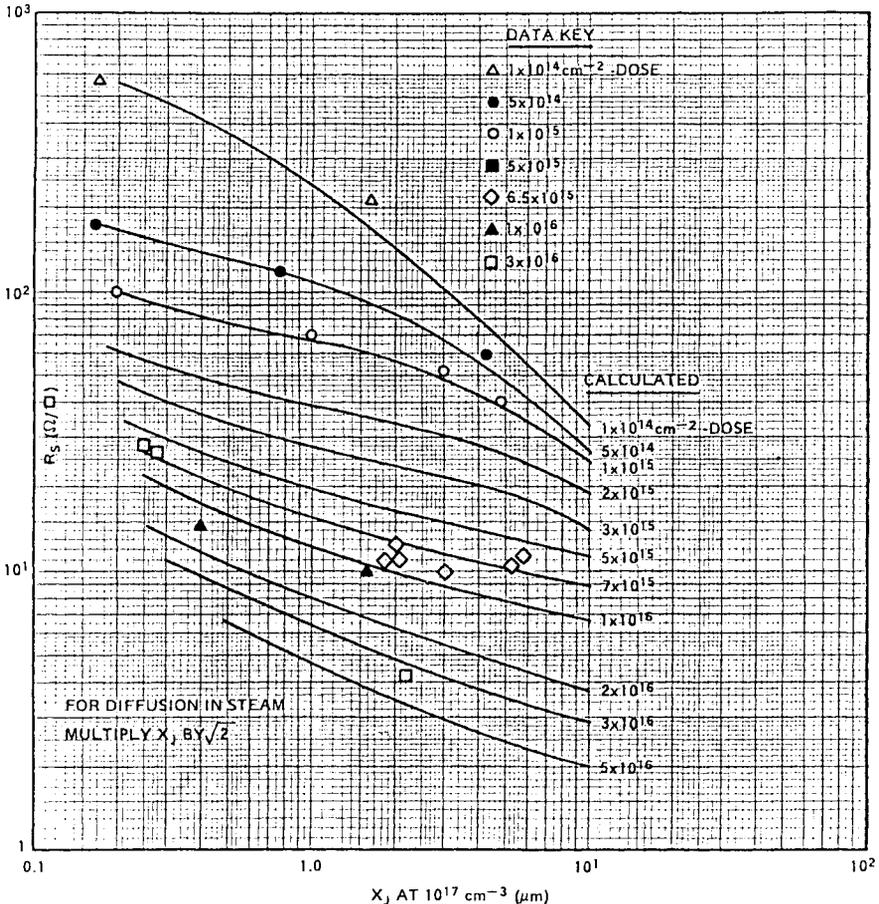


Figure 50: Sheet resistance vs. junction depth for implanted phosphorus layers diffused in N_2 or O_2 .

The advantages of using ion implantation as a predeposition technique are:

- precise control over impurity dose, depth, profile and area uniformity
- excellent reproducibility
- wide choice of masks; SiO₂, Si₃N₄, polysilicon, photoresist, etc.
- low temperature processing
- small lateral spreading of dopant (self alignment)
- vacuum cleanliness

The disadvantages of ion implantation are:

- expensive, complicated equipment
- junctions are not automatically passivated

Some typical parameters for ion implantation of dopant atoms in silicon are:

- implanted ions: phosphorus, arsenic, antimony and boron
- dose: 10¹¹/cm² to 10¹⁶/cm²
- energy: 5 keV to 2 MeV
- depth of implant: 100 angstroms to 1 micron
- reproducibility and uniformity: + or – 5%
- temperature: usually room temperature

Example: a 50 microamp beam of high energy ions implanted into a 3 inch wafer is equivalent to a dose rate of 6 × 10¹² atoms/cm² per second. The dose is calculated by the equation

$$Q_T = \frac{I \text{ (amps)}}{A \text{ (cm}^2\text{)}} q$$

For comparison, a chemical predeposition of arsenic from a doped oxide source with a surface concentration of 5 × 10²⁰/cm³ and a junction depth of 0.1 micron gives a total deposited dose from Equation 68 of

$$\begin{aligned} Q_T &= 0.53 C_s x_j \\ &= 2.6 \times 10^{15} \text{ As/cm}^2 \end{aligned}$$

Therefore in applications where low predeposition dose is needed (less than 10¹⁵/cm²), ion implantation has a natural advantage and should be used over chemical source predepositions. Ion implantation is also useful at higher doses for dopant control. Masks can be made of any convenient material used in VLSI fabrication. As a result ion implantation has become the primary doping source in the following applications:

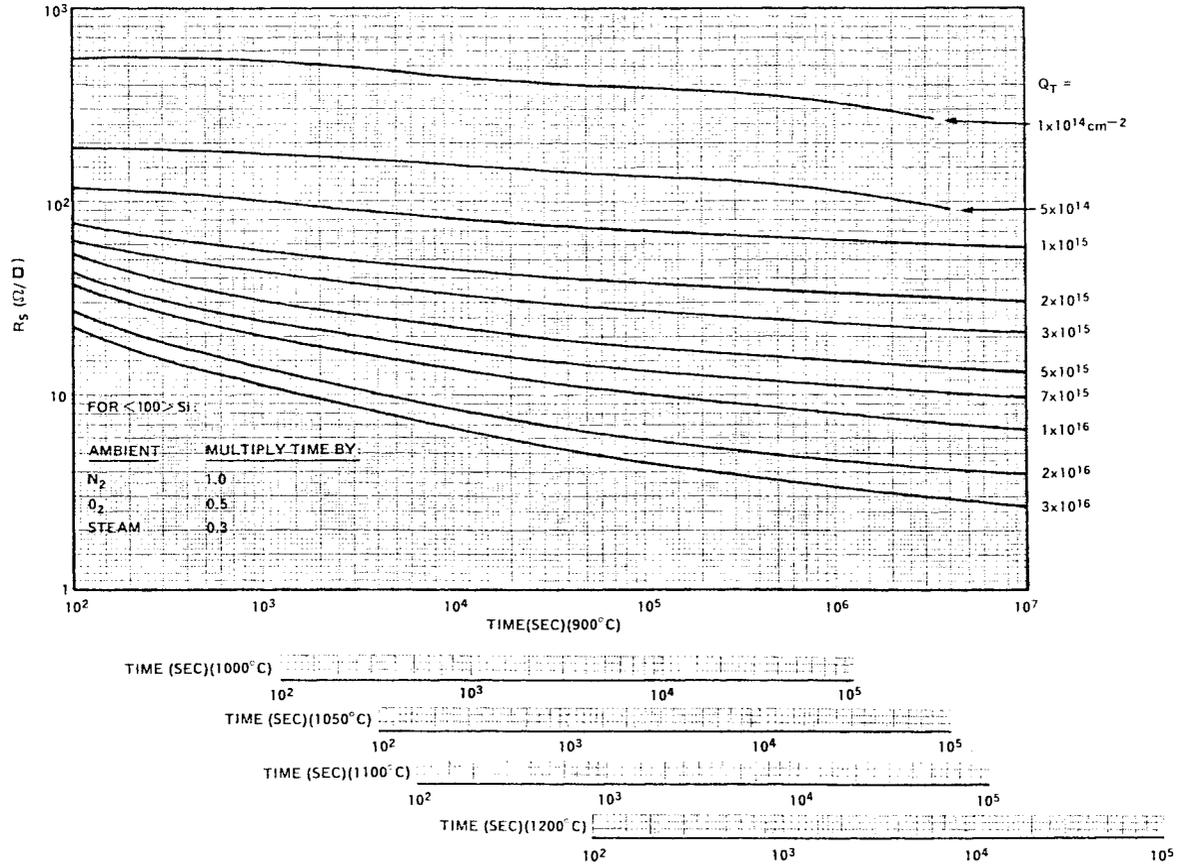


Figure 51: Sheet resistance vs. time, temperature, and dose for phosphorus-implanted, diffused layers in silicon.

- base regions in bipolar transistors for precise gain control
- high value ion-implanted resistors
- MOS threshold voltage adjustment
- p and n wells for CMOS devices
- emitters and bipolar transistors
- source/drains in self-aligned MOS

Ion Implant System

A drawing of an ion implantation system is shown in Figure 52.⁶³ The important features of this system are:

- (1) Gaseous source of vaporizable material such as BF_3 or AsH_3 at high accelerating potential. A valve controls the flow of gas to the ion source.
- (2) A power supply to energize the ion source and accelerate the ions into the mass separation mechanism.
- (3) An analyzer magnet that selects only the ion species of interest and rejects other species.
- (4) Beam sweeping electrodes with sawtooth voltages applied to raster the beam and give a uniform implantation across the wafer.
- (5) A target chamber consisting of an area defining aperture, Faraday cage for measuring current and a wafer feed mechanism.

Simple Range Theory

An understanding of the mechanisms which govern the slowing down of ions in silicon is of considerable importance in achieving controlled and reproducible dopant distributions in ion implantation. Factors such as ion energy and ion species, the crystal orientation and temperature of the implant all influence the range distributions obtained. There is, therefore, the possibility of a much greater controlled variation of the distribution of dopants introduced into the substrate than exists in the case of chemical predeposition. If the implantation behavior is well understood, as a function of ion energy and dose etc., it is feasible to program these parameters to achieve almost any desired distribution, and in practice this processing is highly automated.

Before considering the ranges of ions it is appropriate to discuss the differential functions (dE/dx), known as the stopping power or specific energy loss. Here E is the ion energy and x is the distance into the crystal, usually measured along the direction of incidence of the ions. It is customary to distinguish between two major processes of energy loss; 1) due to elastic Coulomb interactions between the screened nuclear charges of the ion and target atom and the other due to inelastic interactions of the ion with bound or free target electrons.

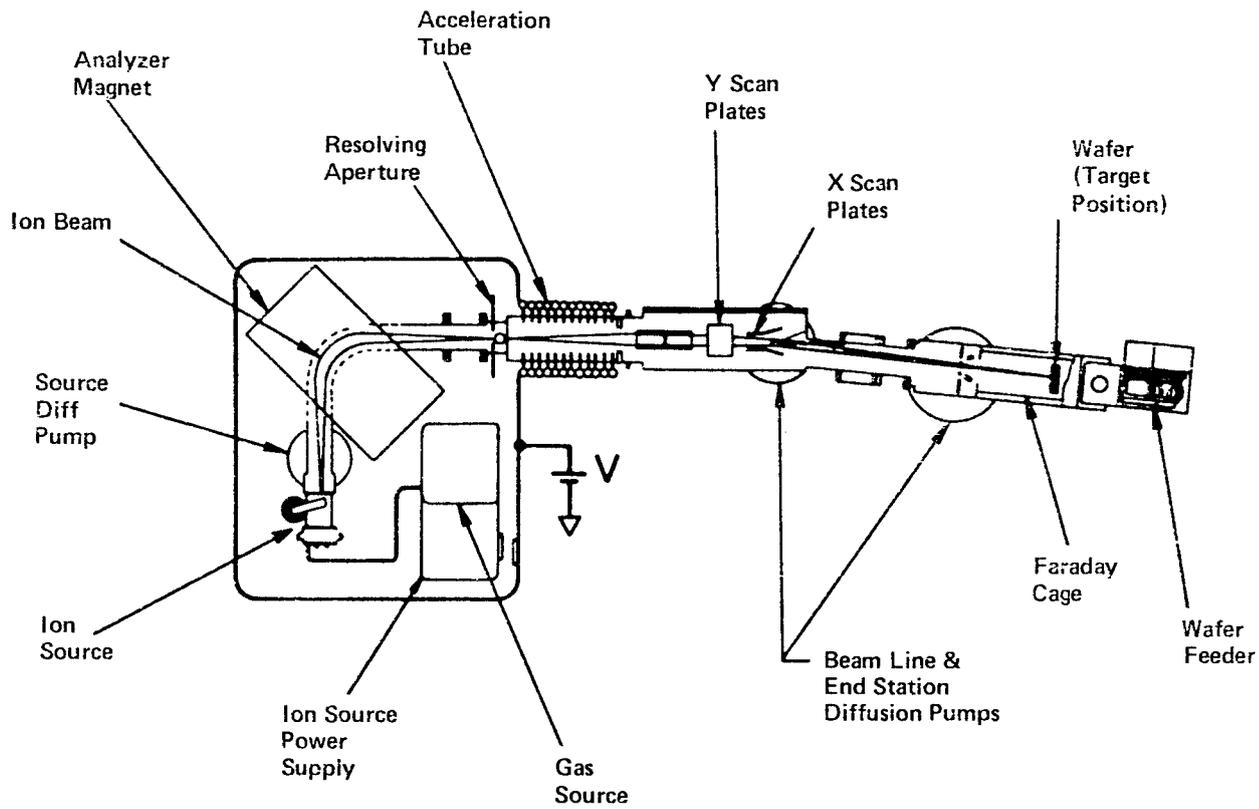


Figure 52: A schematic diagram of a typical commercial ion implant system. (After Varian-Extrion, DF-3000 brochure.)

The total specific energy loss is taken to be the sum of two separable components - nuclear and electronic:

$$-\frac{dE}{dx} = N \left[S_n(E) + S_e(E) \right] \tag{73}$$

where the units of dE/dx are eV/cm, and $S_n(E)$ is the nuclear stopping power (eV cm²), $S_e(E)$ is the electronic stopping power (eV cm²), and N is the target atom density which is $5 \times 10^{22}/\text{cm}^3$ for silicon. The variation of these two components is shown schematically in Figure 53. It can be seen that both increase with energy, reach a maximum, and then decrease again. At the lowest energies nuclear stopping dominates and is responsible for most of the angular dispersion of an ion beam. At higher energies electronic collisions are the most important, and in slowing down to rest from these energies the bulk of the particle energy is dissipated in the form of electronic, rather than nuclear motion.

Reference coordinates for the important ion implantation parameters are shown in Figure 54. The projected range, R_p , is derived from the total range of the incident ions. The total range can be calculated from Equation 73 if S_n and S_e are known. Thus, rearranging terms in Equation 73 and integrating gives Equation 74.

$$R = \int_0^R dx = \frac{1}{N} \int_0^{E_0} \frac{dE}{\left[S_n(E) + S_e(E) \right]} \tag{74}$$

Note that R will be the average total range and we would expect a distribution of R , R_p and the transfer straggle R_T . Such typical distributions are shown in Figure 55 in which the depth distribution of implanted ions in an amorphous target are shown for two cases—first relating to when the incoming ion has a mass smaller than the target, and the second—referring to the case where the incoming ion has a mass greater than the

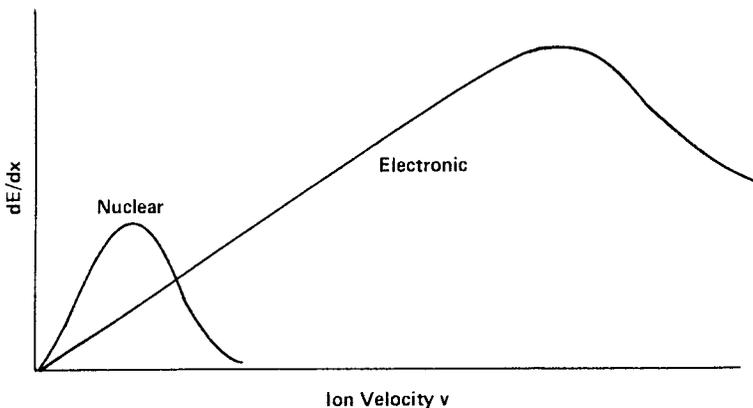


Figure 53: Behavior of the nuclear and electronic contributions to the specific energy loss dE/dx as a function of ion velocity v .

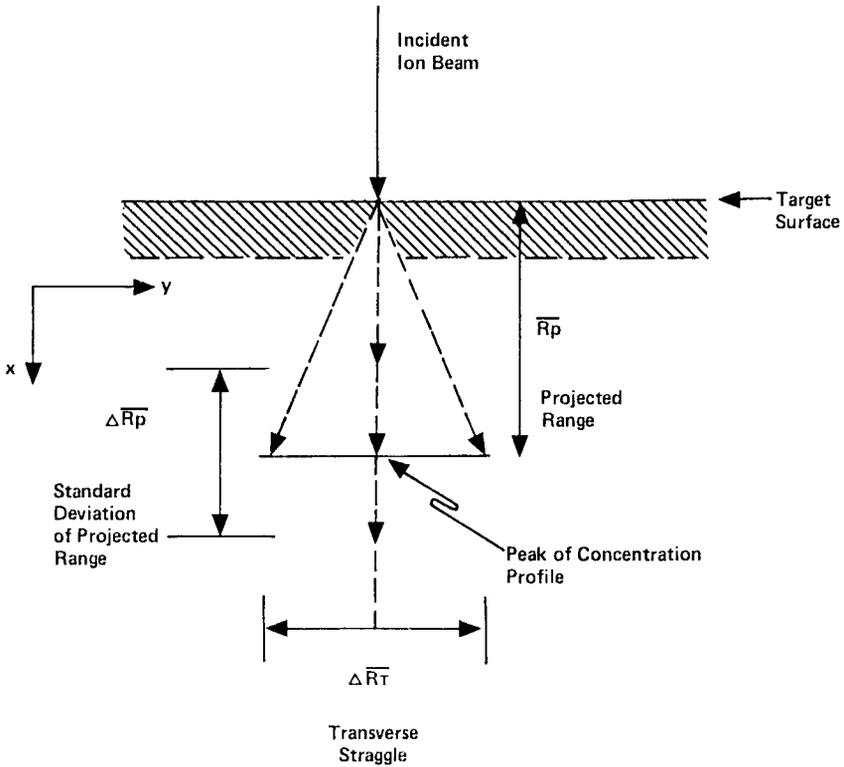


Figure 54: Reference coordinates for the important ion implantation parameters.

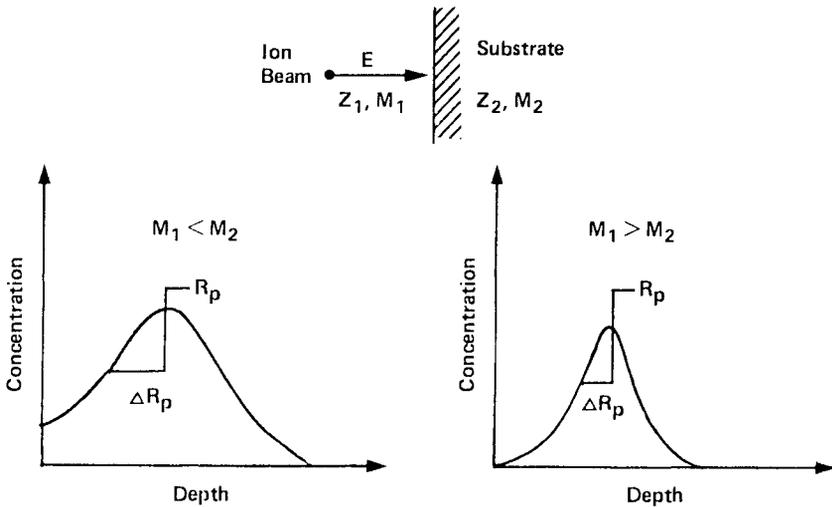


Figure 55: The depth distribution of implanted atoms in an amorphous target for the case in which the ion mass is less than or greater than the mass of the substrate atoms.

target atom mass. In general, the lighter the ion relative to the substrate the broader the distribution of implanted impurities will be. Thus:

$$\begin{array}{ccc} \text{Light Ions} & & \text{Heavy Ions} \\ \hline \frac{\Delta R_p}{R_p} \text{ is large} & & \frac{\Delta R_p}{R_p} \text{ is small} \end{array}$$

Nuclear Stopping

The specific energy loss due to collisions of the ion and a target nucleus is derived by considering these as independent elastic two body interactions. The energy transferred during a two body scattering process depends upon the interaction potential between the two particles. Assuming a Born-Mayer interaction potential it can be shown that a useful approximation for the nuclear stopping function is⁶⁴

$$S_n = 2.8 \times 10^{-15} \frac{Z_1 Z_2}{Z^{1/3}} \frac{M_1}{M_1 + M_2} \quad (\text{eV cm}^2), \quad (75)$$

where S_n is the nuclear stopping power independent of E , Z_1 is the ion atomic number, M_1 is the ion atomic mass, Z_2 is the substrate atomic number (14 for silicon), M_2 is the substrate atomic mass (28 for silicon) and

$$Z^{1/3} = \left(Z_1^{2/3} + Z_2^{2/3} \right)^{1/2}. \quad (76)$$

Electronic Stopping

When moving at velocities greater than the K shell electron velocity, an ion will have a high probability of being fully stripped of its electrons. The theory of energy loss under these circumstances derives from the work of Bore who carried out classical calculations. The electronic stopping function, therefore, is dependent upon the velocity of the atom simply expressed as⁶⁴

$$S_e(E) = C_1 v = KE^{1/2} \quad (77)$$

where v is the velocity of the ion, C_1 and K are constants. K depends on both the ion and the substrate as shown in Equation 78.

$$K = Z_1^{1/6} \frac{0.079 Z_1^{1/2} Z_2^{1/2} (M_1 + M_2)^{3/2}}{(Z_1^{2/3} + Z_2^{2/3})^{3/4} M_1^{3/2} M_2^{1/2}} \frac{C_R}{\sqrt{C_E}} \quad (78)$$

$$C_R = \frac{4\pi a^2 M_1 M_2}{(M_1 + M_2)^2} \quad (79)$$

$$C_E = \frac{4\pi \epsilon_0 a M_2}{Z_1 Z_2 q^2 (M_1 + M_2)} \quad (80)$$

and a is the bore radius. For an amorphous silicon substrate, K is approximately independent of the ion and assumes the value

$$K \approx 0.2 \times 10^{15} (\text{eV})^{1/2} \text{ cm}^2$$

Critical Energy

Note that in the above equations, S_n is independent of energy and S_e increases with energy. There does exist, therefore, a critical energy at which these two functions are equal:

$$E = E_c \text{ at } S_n = S_e(E)$$

E_c is the critical energy at which nuclear stopping and electronic stopping are equal in magnitude and can be expressed as

$$\begin{aligned} \sqrt{E_c} &= S_n/K \\ &= 14 \frac{14Z_1}{[(14)^{2/3} + Z_1^{2/3}]^{1/2}} \frac{M_1}{M_1 + 28}. \end{aligned} \tag{81}$$

The following values for E_c have been determined:

- (1) E_c is approximately equal to 10 keV for boron ($Z_1=5$ $M_1=10$)
- (2) E_c is approximately 200 keV for phosphorus ($Z_1=15$ $M_1=30$)
- (3) E_c is greater than 500 keV for arsenic and antimony.

Therefore boron tends to be stopped by electronic interactions in the typical energy ranges used in implantation and phosphorus, arsenic and antimony tend to be stopped by nuclear collisions.

If E is much, much less than E_c then,

$$\frac{dE}{dx} = NS_n. \tag{82}$$

Thus the range, R , can be expressed as

$$R \approx 0.7 \text{ \AA} \frac{Z_1^{1/3}}{Z_1 Z_2} \frac{M_1 + M_2}{M_1} E \tag{83}$$

This expression is useful for arsenic, antimony and sometimes phosphorus. If E is much, much greater than E_c then

$$\frac{dE}{dx} = NKE^{1/2}. \tag{84}$$

The useful expression for the range then becomes

$$R \approx 20\sqrt{E} \text{ (\AA)} \tag{85}$$

Projected Range

The discussion thus far has dealt with the total pathlength of the ion implanted into the substrate, called the range R . A typical ion stops at a distance normal to the surface, called the projected range, R_p . Statistically speaking other ions encounter fewer scattering events in a given distance in the target and come to rest beyond the projected range. Some ions that may have more than the average number of scattering events come to rest between the surface of the substrate and the projected range. The fluctuation or straggle in the projected range is ΔR_p . The relationship between R and R_p is shown in Figure 56. The statistical distribution of implanted impurities with R_p and ΔR_p were shown in Figure 55. The relationship that exists between R and R_p is shown in Equation 86 where $b = 1/3$ for nuclear stopping and $M_1 > M_2$ i.e. for antimony and arsenic. The value of b is smaller for electronic stopping (B,P) but $1/3$ is still a reasonable approximation to first order.

$$\frac{R}{R_p} = 1 + bM_2/M_1 \tag{86}$$

Example: for arsenic $M_1 = 75$. Therefore $R/R_p = 1.12$.

The projected range corrections for calculating R_p from R are shown in Table 8. Useful calculations of the distribution parameters and concentrations are shown in Equations 87-90.

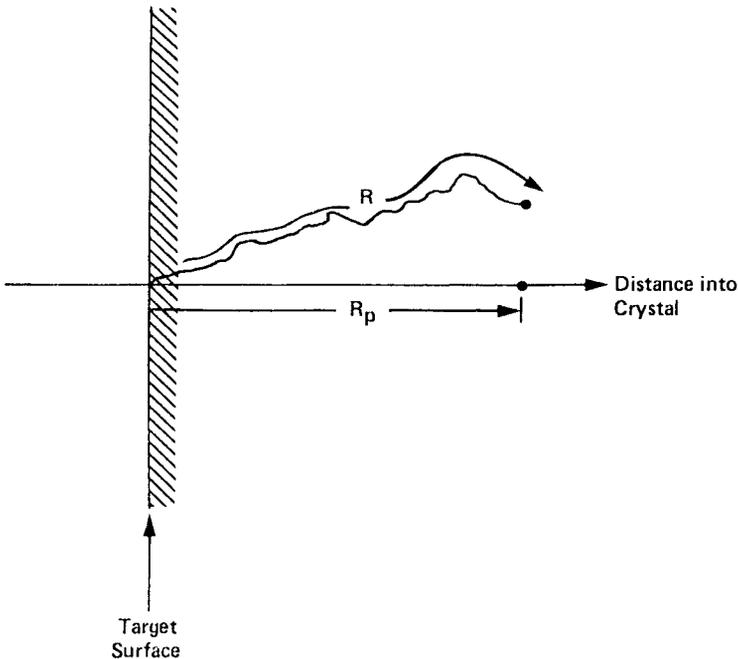


Figure 56: Relationship between total range, R , and the projected range, R_p .

Table 8: Projected Range Corrections R_p/R

Ion	Substrate	R_p/R values				Rule-of-thumb value $(1 + M_2/3M_1)^{-1}$
		20 keV	40 keV	100 keV	500 keV	
Li	Si	0.54	0.62	0.72	0.86	0.4
B		0.57	0.64	0.73	0.86	0.54
P		0.72	0.75	0.79	0.86	0.77
As		0.83	0.84	0.86	0.89	0.89
Sb		0.88	0.88	0.89	0.91	0.93

$$\Delta R_p \approx \frac{2}{3} \frac{\sqrt{M_1 M_2}}{M_1 + M_2} R_p = \frac{\text{halfwidth}@1/2 C_{MAX}}{\sqrt{2 \ln 2}} \tag{87}$$

$$C_{MAX} \approx \frac{Q_T}{2.5 \Delta R_p} \tag{88}$$

$$C(x) = C_{MAX} \exp \left[-\frac{(x - R_p)^2}{2 \Delta R_p^2} \right] \tag{89}$$

$$C(x) = \frac{1}{10} C_{MAX}@x=R_p \pm 2 \Delta R_p \tag{90}$$

$$C(x) = \frac{1}{100} C_{MAX}@x=R_p \pm 3 \Delta R_p$$

$$C(x) = 10^{-5} C_{MAX}@x=R_p \pm 4.8 \Delta R_p$$

The depth distribution or profile of stopped ions can be approximated by a symmetric Gaussian distribution function. The concentration of implanted ions as a function of position is given in Equation 89, where the maximum concentration occurs at $x = R_p$. The integral of Equation 89 is the dose, Q_T , and the maximum concentration is given by Equation 88.

Calculated curves of projected range (microns) as a function of implant energy (kiloelectron volts) shown in Figure 57 for four important silicon dopants.⁶⁵ As expected, the lighter boron atoms penetrate farther for a given energy than the heavier phosphorus, arsenic or antimony ions. Calculated curves of projected standard deviation, ΔR_p , in microns as a function of implant energy are shown for the four same dopants in Figure 58. As was stated earlier, the $\Delta R_p/R_p$ ratio is larger for light ions than for heavy ions.

Implantation Masking

The goal of masking ion implantation is to allow the doping of the substrate to occur in selected areas of the wafer. The ability of a given

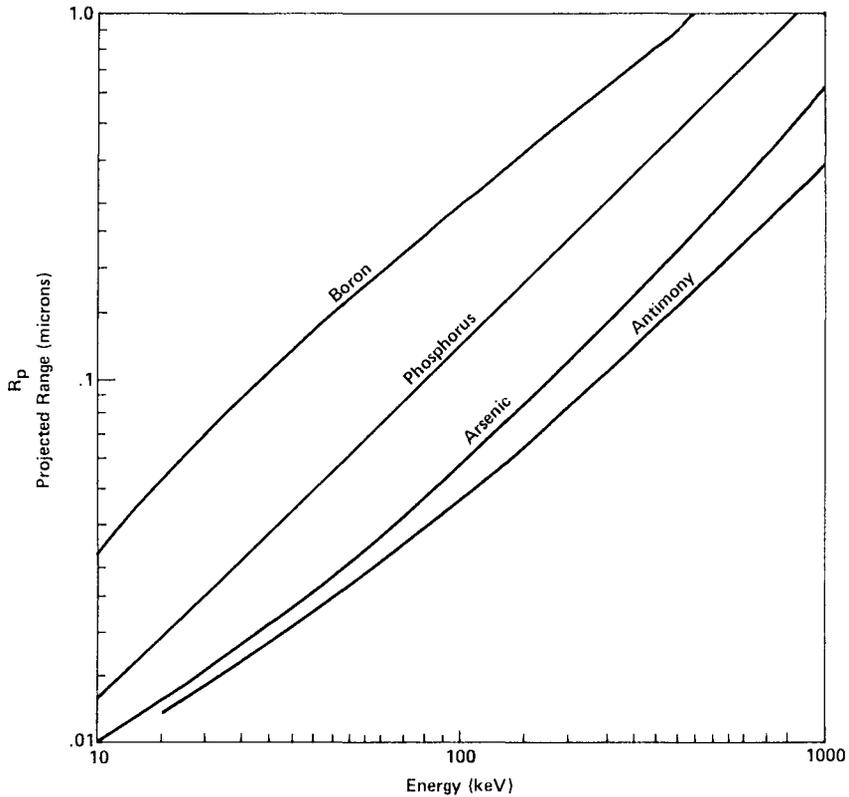


Figure 57: Calculated projected range in microns as a function of implant energy in kiloelectron volts for four important silicon dopants.

material to act as a mask depends upon the ion stopping power of the material as well as the thickness of the material. Three commonly used masks in integrated circuit technology are Si_3N_4 , SiO_2 , and photoresist. The minimum masking thickness of each of these materials is shown in Figure 59 as a function of ion energy for three dopants.⁶⁴ For example, 100 keV boron implantation requires a silicon nitride mask thickness of at least 0.4 microns, an SiO_2 thickness of at least 0.55 microns, and a minimum photoresist thickness of 0.7 microns. One problem in using photoresist as a mask is that at doses greater than $10^{15}/\text{cm}^2$, the resist mask may become difficult to etch. The chemical changes that the bombardment creates can depolymerize the resist, causing swelling and thickness changes by up to 40%.

In order to understand the sensitivity of mask thickness on the percent of ions that may penetrate that mask, reference is made to Figure 60.⁶⁶ Here, photoresist thickness is plotted as a function of ion implantation energy in a family of curves showing the percent ion penetration through the mask as a parameter. Increasing the mask thickness from 0.45 microns

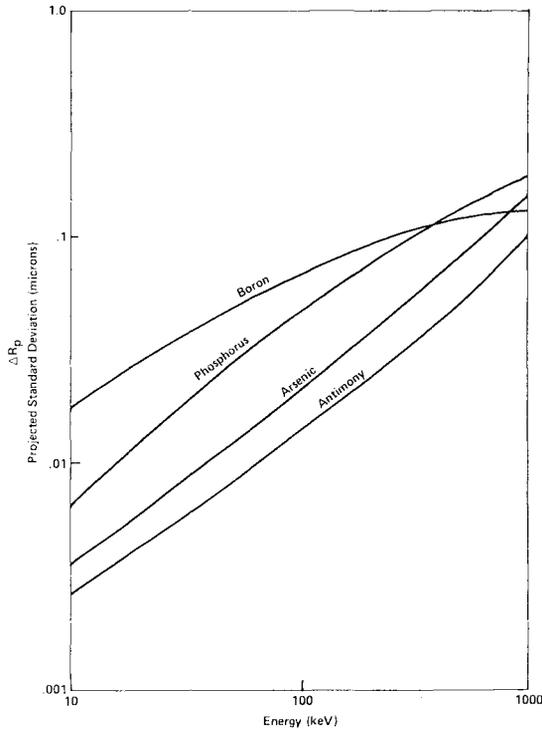


Figure 58: Calculated projected standard deviation in microns as a function of an implant energy in kiloelectron volts for four important silicon dopants.

to 0.6 microns can reduce the ion penetration percentage from 10% to 0.01% at 100 keV.

Energetic ions penetrating masks can “knock” atoms of the mask into the underlying silicon. Thus, the role of the mask in the ion implantation process is not necessarily a neutral one. This knock-on effect may nucleate damage in the silicon. An example for phosphorus into Si_3N_4 is shown in Figure 61. Concentration profiles of implanted phosphorus and recoiled nitrogen in silicon are shown as a function of the thickness of the Si_3N_4 mask. It can be seen that the measured nitrogen profiles form an exponential distribution whose surface concentration is approximately 1×10^{20} atoms/cm³. This concentration is many times above the solid solubility of nitrogen in silicon at high temperatures. Thus the excess nitrogen has the potential for nucleating damage which may grow during high temperature annealing.

Another example of knock-on damage caused by implantation through a masking layer is shown in Figure 62. This shows a schematic diagram of oxygen recoil damage caused by implantation through an SiO_2 layer.⁶⁷ Areas where recoil damage can occur are at tapers in the SiO_2 edges and where the oxide is very thin. In both cases recoiled oxygen can penetrate

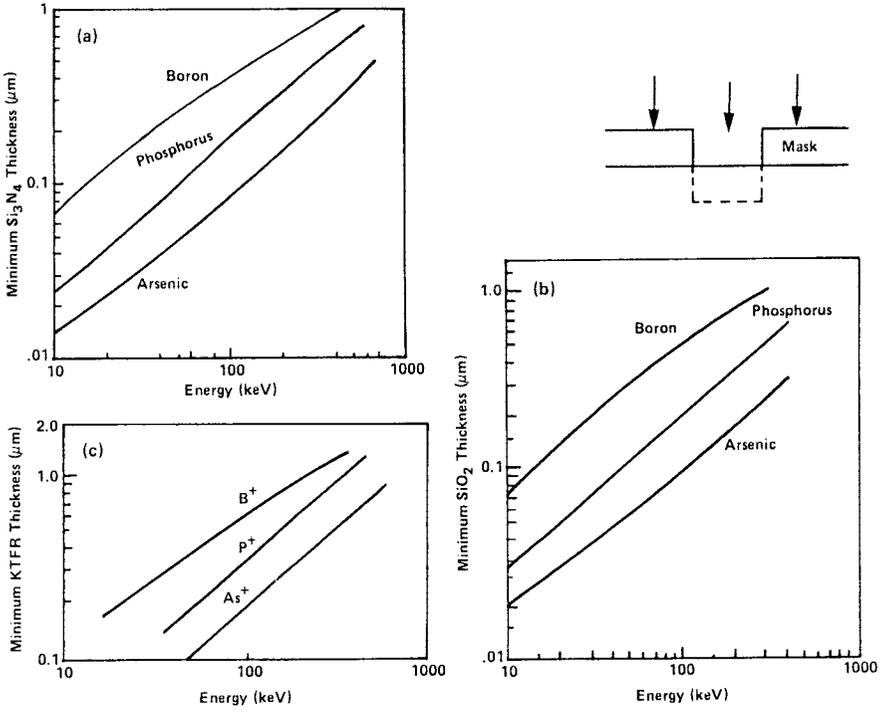


Figure 59: Minimum masking material thicknesses for the ion implantation of boron, phosphorus and arsenic.

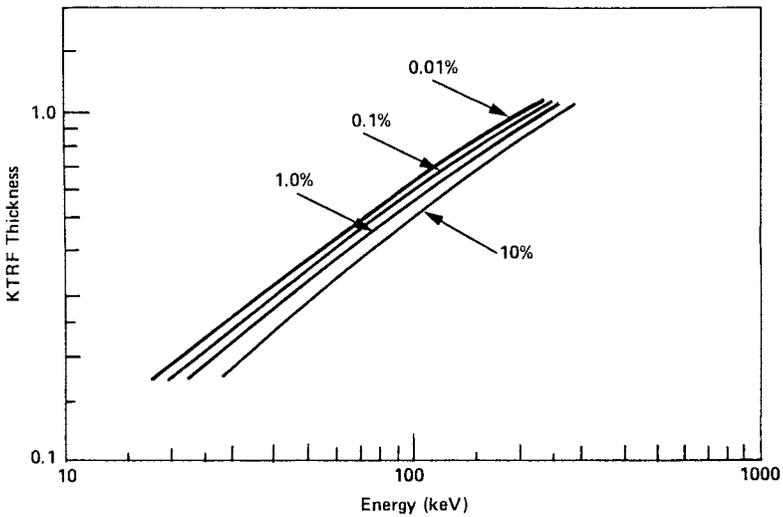


Figure 60: Percent of B^+ ions penetrating photoresist vs. energy.

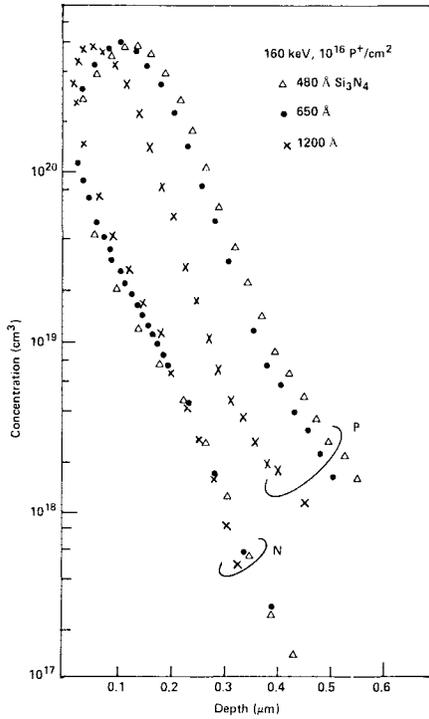


Figure 61: Concentration profiles of phosphorus and recoiled nitrogen in silicon for ion implantation through silicon nitride of 480, 650, and 1200 Å thickness at 160 keV to a dose of 1×10^{16} ions/cm².

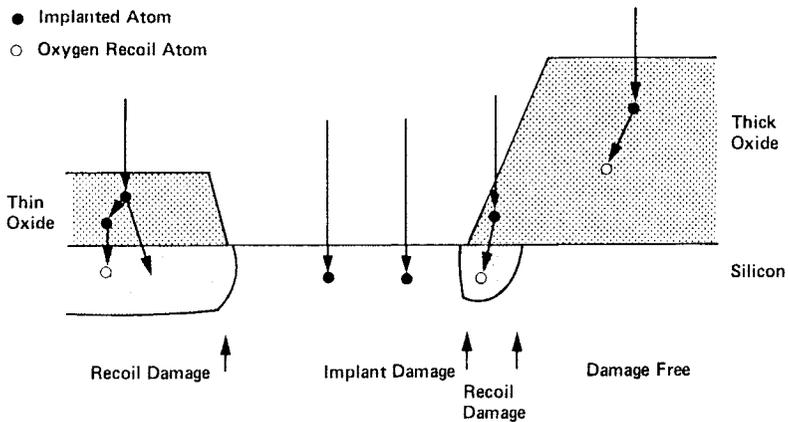


Figure 62: A schematic diagram of oxygen recoil damage caused by implantation through a silicon dioxide masking layer. Taper at SiO₂ edges can cause knock-on of oxygen atoms. Damage regions may result and any metallic precipitation in these regions will cause degradation of junctions.

into the silicon substrate. The high concentration of recoiled oxygen atoms in the substrate may cause the atoms to precipitate during high temperature processing which can nucleate the growth of extrinsic dislocations around the oxygen precipitates.⁶⁸ If the ion implanted dopant forms a junction whose space charge layer intersects this damage, leaky or shorted junctions may result.

Ion Channeling

Until about 1960 it was tacitly assumed that ion penetration in crystal-line substrates would not differ substantially from that in amorphous solids, and there was little or no experimental evidence to suggest that the regular structure of a crystal lattice was significant in atomic stopping. It is now known that low-index crystal axes or planes present large open or transparent avenues for the incident ions, and fewer atoms are exposed to the bombarding particles, resulting in the reduction of the number of atomic collisions⁶⁴. An example of this effect is shown in Figure 63 where views are presented of the $\langle 110 \rangle$ direction (open channels) vs. a random direction viewed at 10° from the $\langle 110 \rangle$ direction. Thus, if an ion beam is aligned along the $\langle 110 \rangle$ direction one would expect a much deeper

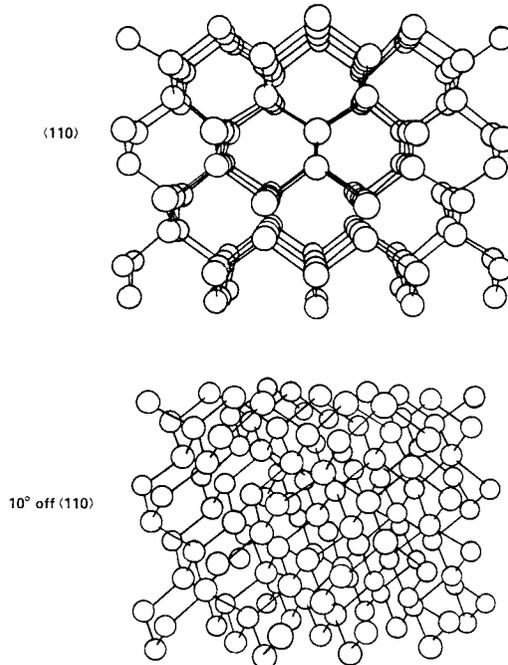


Figure 63: Atomic configuration in the diamond-type lattice (a) along the $\langle 110 \rangle$ direction, and (b) viewed along a "random" direction at 10° from the $\langle 110 \rangle$ direction.

penetration by that ion. Once the ion is inserted into a channel, the atomic potentials become operative and steer the ion towards the center of the open space or channel. The ion can be guided along the channel over considerable distances. Examples of channeling are shown in Figure 64 for potassium implanted into tungsten in the $\langle 111 \rangle$ direction and for phosphorus implanted into silicon in various orientations of the crystal off the $\langle 110 \rangle$ direction. In Figure 64a the portion of the curve labeled A represents the region of the crystal where the implanted ions strike surface atoms and no channeling occurs. Region B is where partial channeling and Region C is where full channeling occurs. Channeling is purposely avoided in most devices because it is difficult to accurately control. Thus implants are usually performed with the wafer tilted to get random "amorphous" like profiles. As an example, a 7° tilt off the $\langle 110 \rangle$ axis allows more than 99% of the ions to be stopped, as if the silicon were amorphous.

Modeling Implanted Dopants in Silicon

The Lindhard, Scharff and Schiott (LSS) range theory⁶⁵ can predict the ranges of implanted ions into solids under conditions in which the target is amorphous or oriented in a random ordered crystal lattice direction which appears amorphous to the incident ions.⁶⁹⁻⁷² Prediction of the range of ions which are either initially channeled or become channeled is difficult.⁷¹⁻⁷³ Rigorous calculations using the Boltzmann's Transport Equation and the Monte Carlo Method, which calculate individual ion trajectories, have been found to be insufficient for complex and fine structured targets.⁷⁵ In addition, implant profiles are difficult to calculate due to dynamic anneal effects and time-dependent amorphization of the substrate. As a result, in order for modeling calculations to reasonably represent implanted profiles, either the silicon substrate should be amorphous prior to implantation (especially for boron) or the modeling should be modified to use experimental data. Modeled profiles have been calculated by summing a Gaussian distribution and an exponential distribution. The exponential function was described by parameters extracted from as-implanted (unannealed) profiles. Theoretical predictions for implanted ions served as a basis for graphing the profile parameters as functions of implantation conditions. Formulae for the profile parameters were then derived from the graphs. These equations were incorporated in computer programs which calculated as-implanted concentration profiles of As ions implanted into $\langle 100 \rangle$ and $\langle 111 \rangle$ crystalline Si, for B and BF_2 ions implanted into $\langle 100 \rangle$ and $\langle 111 \rangle$ crystalline Si and for P ions implanted into $\langle 111 \rangle$ crystalline Si.

The basic shape of the desired concentration profiles for As, B and P as-implanted profiles is the sum of the two distributions:

$$C = C_{\text{peak}} \exp \left[\frac{-(x - R_p)^2}{2\Delta R_p^2} \right] + C_{\text{tail}} \exp \left[-K(x - X_{\text{tail}}) \right] \text{ cm}^{-3} \quad (91)$$

The Gaussian distribution describing the distribution of ions implanted into amorphous Si is characterized by three parameters, C_{peak} , R_p and ΔR_p . The exponential distribution is characterized by three parameters,

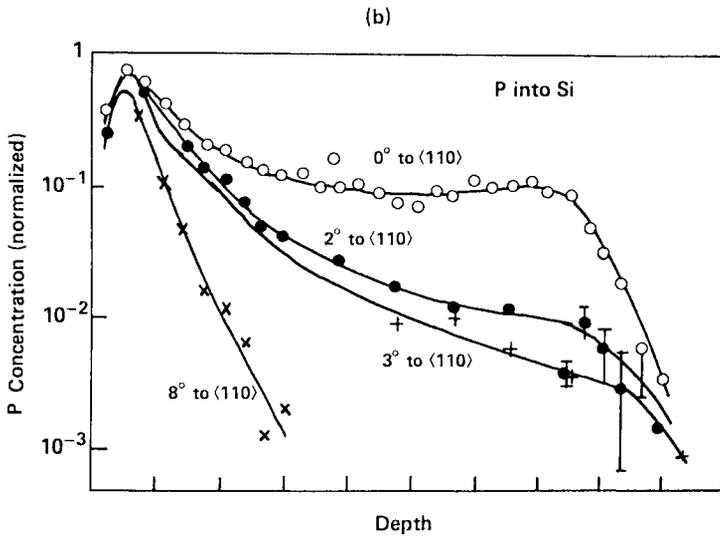
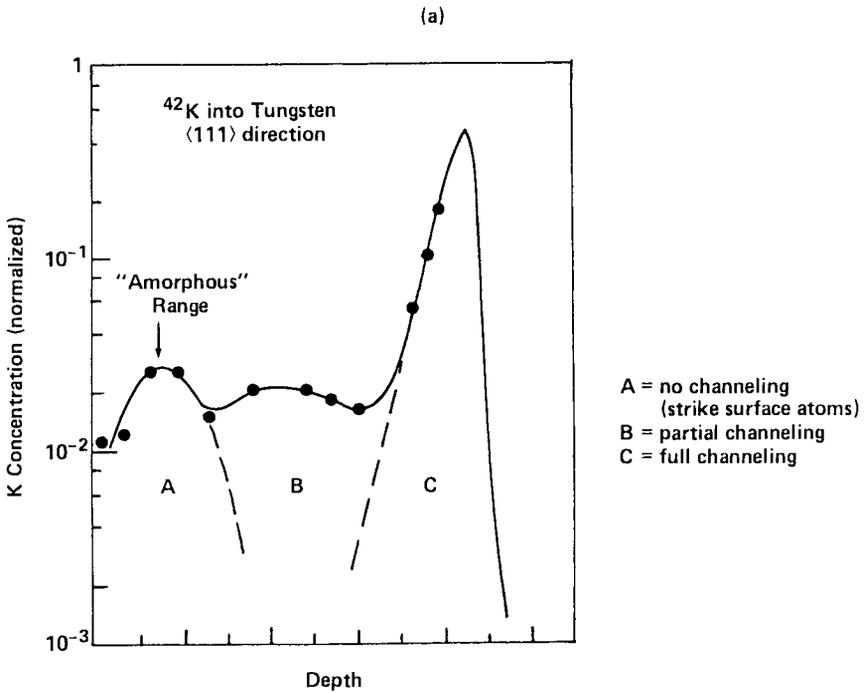


Figure 64: Examples of ion channeling for (a) potassium implanted in a tungsten in a $\langle 111 \rangle$ direction and (b) phosphorus implanted into silicon in various orientations of the crystal off the $\langle 110 \rangle$ direction.

Xtail, Ctail and K, and is assumed to describe the impurity profile of the "tail" region.^{74,76-79} Figure 65 illustrates the method by which the estimated values of key parameters which characterize the Gaussian and exponential distributions of ion implanted profiles were extracted from published figures.

The ranges of implanted ions vary sublinearly or linearly with implant energy as shown in Figure 58. Xtail is the depth at which the channeling tail is estimated to begin. Figures 66a and 66b are plots of the extracted Xtail value as a function of implant energy for B and P respectively. Empirical expressions were derived from the data and are shown drawn in the figures as Xtail.

Ctail is the parameter corresponding to the concentration of implanted ions at the beginning of the tail. The normalizing ratio Ctail/Cpeak is shown as a function of implant dose or energy in Figures 67a and 67b for B and P. This ratio is an indicator of the occurrence of channelling. As channelling occurs, Ctail approaches Cpeak. Therefore, the normalizing ratio Ctail/Cpeak will increase. The figures indicate that an increase in dose or increase in energy decreases the ratio Ctail/Cpeak. This agrees with theory since crystal damage is directly proportional to the number of ions implanted. Channeling will be minimized if the damage is sufficient for the substrate to become amorphous. Thus, more channeling is likely at lower doses and Ctail can approach Cpeak. With heavy ions, such as As, the dose to amorphize the substrate at room temperature, is approximately 2×10^{14} ions/cm². Higher doses are necessary for lighter ions such as boron and phosphorus.

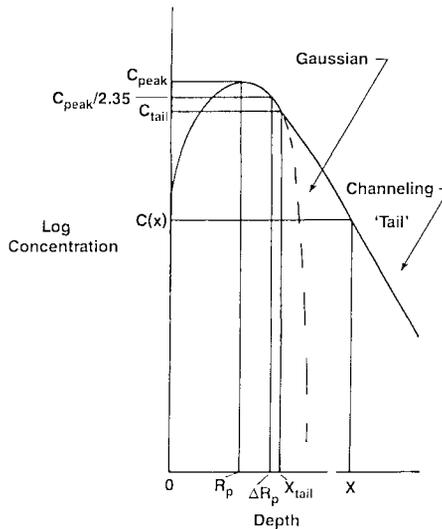


Figure 65: Illustrations of extracted parameters which characterize the distribution of an ion implanted impurity profile in crystalline silicon.

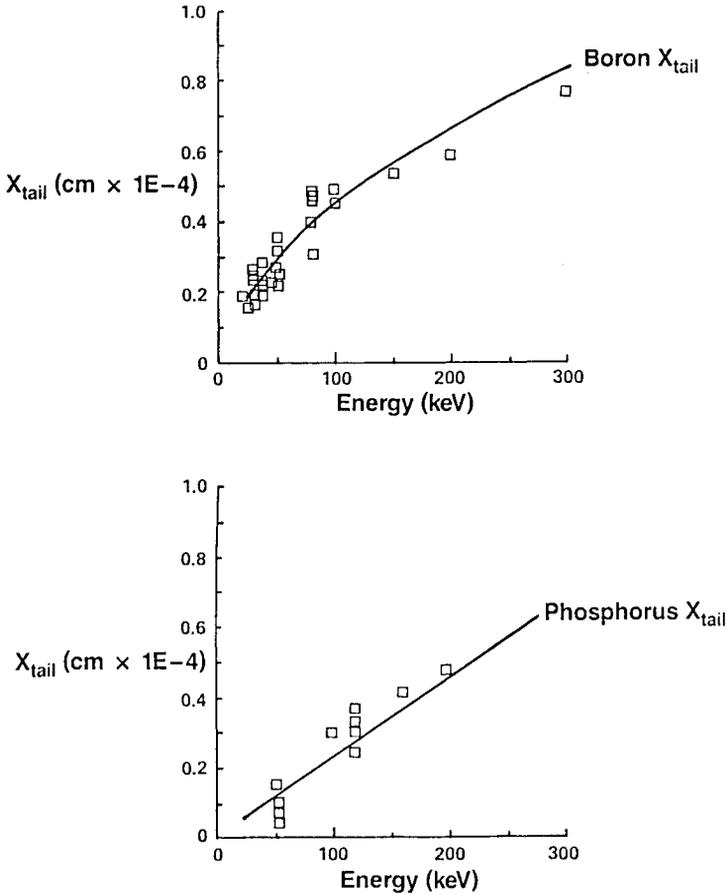


Figure 66: X_{tail} vs. energy for (a) boron implants and (b) phosphorus implants in $\langle 100 \rangle$ silicon.

The B data was not sufficiently regular to fit an equation. The average values of C_{tail}/C_{peak} are shown for the dose ranges indicated in the figure. Again, for low dose ranges, C_{tail}/C_{peak} is large indicating occurrence of ions entering open channels.

Figure 67b shows the estimated curve of C_{tail}/C_{peak} versus log dose for P for a dose \geq of $1 \times 10^{14} \text{cm}^{-2}$. Since data were unavailable for implant doses $< 1 \times 10^{14} \text{cm}^{-2}$, a constant value was used for C_{tail}/C_{peak} in the modeling programs. The data substantiate the relationship that C_{tail}/C_{peak} decreases as dose increases.

A large value of the exponential slope function, K , corresponds to a steep concentration gradient in the tail region, indicative of less channeling. Figure 68 is an illustration of the effect of an increasing slope value on the concentration profile.

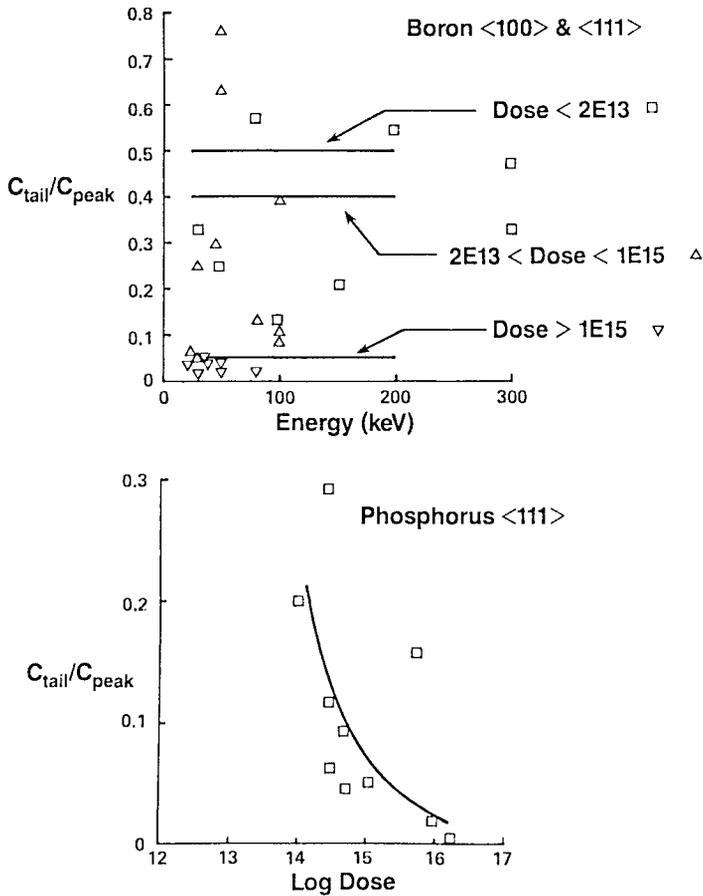


Figure 67: The normalized ratio of C_{tail}/C_{peak} as a function of implant dose or energy for (a) boron and (b) phosphorus.

The observed exponential tail is due to random scattering of ions along open channeling directions of crystalline Si substrates. For low mass ions and low energy implants, more channeling will occur. This can be understood by noting⁷⁰ the critical channeling angle, $\psi_c \propto E^{1/4}$ for low energies and at higher energies, $\psi_c \propto E^{1/2}$. Therefore, at lower energies the critical channeling angle is larger, allowing more ions to channel. K is both a function of dose and energy. For low implant doses, less crystal damage occurs allowing ions to penetrate the substrate with fewer collisions. Therefore, at low doses and at sufficient energy the implanted ions have a good chance to enter an open channel. As dose increases, the crystal becomes damaged or amorphous and less channeling occurs. K is larger for higher doses and smaller for low doses. Low mass ions, such as B, have a greater probability of entering channels since nuclear collisions are less likely to occur. Once

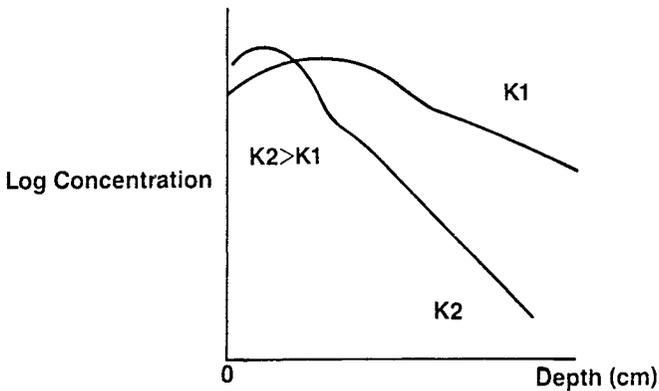


Figure 68: Illustration of the effect of an increasing exponential slope function on the tail distribution of an implanted profile.

an ion has entered a channel the penetration is deeper for ions with higher energies. This corresponds to a smaller K value as energy increases.

Figures 69a and 69b relate the slope function versus implant energy for As and B implanted ions, respectively. Figure 69a shows that for As ions implanted into $\langle 100 \rangle$ Si, the K value is more energy dependent than for As ions implanted into $\langle 111 \rangle$ Si. The critical angle for channeling of 50 keV As ions into $\langle 111 \rangle$ Si is 4.4° and ψ_c for $\langle 100 \rangle$ Si is 4.0° .⁷² The critical angle is a measure of the steering action of the atoms that comprise the walls of the channel. It is less likely for an ion in a channel in $\langle 111 \rangle$ Si to leave than an ion in a channel in $\langle 100 \rangle$. The ability of an ion to remain in the $\langle 100 \rangle$ channel at low implant energies is not as good as an ion in the $\langle 111 \rangle$ channel. Figure 69a shows calculated curves for K slope versus energy which represent the best fit to the data for As $\langle 100 \rangle$ and As $\langle 111 \rangle$ implants. In the case of the As $\langle 111 \rangle$ implants the best fit is a constant value for K .

Figure 69b is a graph of K values calculated from B profiles as a function of energy. The increase in K for B implants into $\langle 100 \rangle$ Si for doses greater than 5×10^{13} ions/cm² is a result of increased crystal damage and dechanneling with increasing dose for a given implant energy. The critical angle for channeling of a 50 keV B ions into $\langle 100 \rangle$ Si is 2.9° and ψ_c for $\langle 111 \rangle$ Si is 3.2° .⁷² It can be seen that the K value is smaller for the $\langle 111 \rangle$ Si and less dependent on implant energy than the $\langle 100 \rangle$ Si for implant dose $> 5 \times 10^{13}$.

Curve fitting model parameters extracted from the experimental data are summarized in Table 9 for B, P and As and in Table 10 for BF_2 . Based upon these parameters, representative profiles were calculated using simple computer programs which now reside in the process simulation program PREDICT. The calculated profiles using Equation 91 and a multiplicative smoothing function were compared with experimental profiles. The smoothing function is an inverse exponential function used to join the

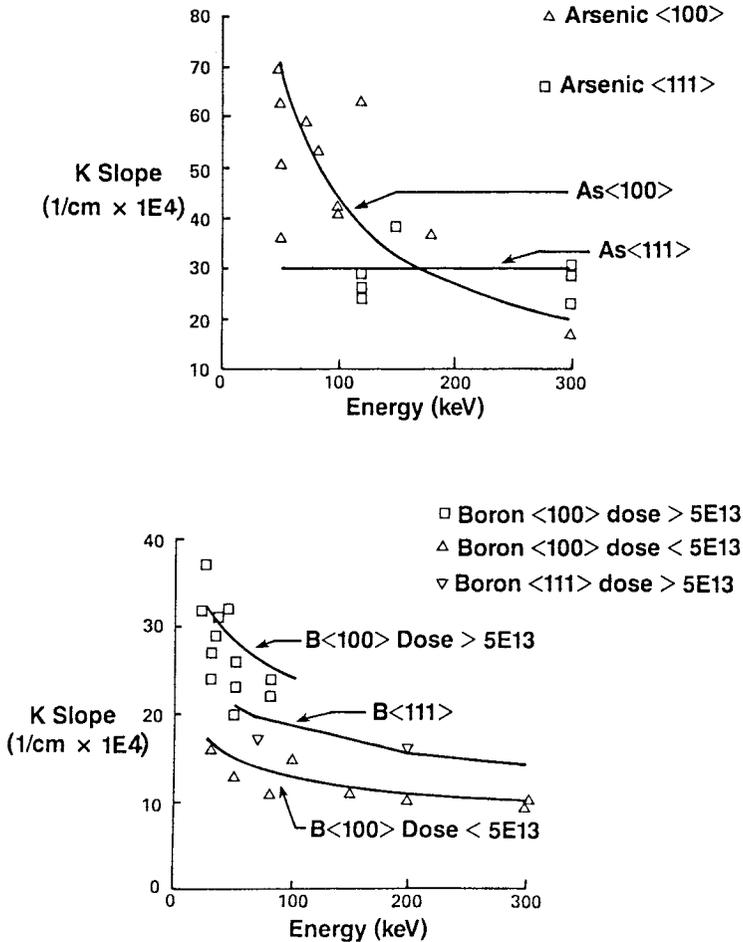


Figure 69: K vs. energy for (a) arsenic implants into <100> and <111> crystal-line silicon, and (b) boron implants into <100> and <111> crystalline silicon.

Gaussian distribution and the exponential distribution about the Xtail point without having any discontinuities in the calculated profiles.

Figure 70 shows comparisons between calculated and experimental data from published as-implanted profiles for As implanted into <100> Si for two doses. Examples of B implantation profiles into <111> Si were compared by using published data⁸⁰ in Figure 71.

Experimental errors on the model parameters may result from several factors: (1) misalignment of the target⁷³, (2) measurement variation in equipment, and (3) dose rate effects. The misalignment error is illustrated in Figure 72 using B profiles from an experiment performed by W.K. Chu, et. al.⁷³ The tilt and rotation angles of the wafer with respect to the ion beam

Table 9: Curve Fitting Model Parameters for Implanted B, P and As

Parameter	Equations		
	B	P	As
$R_p(\text{cm})$	$4.9 \times 10^{-6} E^{0.5} - 19.4 \times 10^{-6}$ (E < 100) $7.24 \times 10^{-7} E^{0.783}$ (E < 20)	$1.23 \times 10^{-7} E$	$6.7 \times 10^{-8} E$
$\Delta R_p(\text{cm})$	$6.8 \times 10^{-7} E^{0.538}$ (E < 100) $5.8 \times 10^{-7} E^{0.538}$ (E < 20)	$6.4 \times 10^{-8} E + 4 \times 10^{-7}$, <100> $5.5 \times 10^{-8} E$, <111>	$9 \times 10^{-8} E^{0.773}$
$X_{\text{tail}}(\text{cm})$	$4.2 \times 10^{-6} E^{0.5} - 7.9 \times 10^{-6}$ (E < 100) $R_p + 6.5 \times 10^{-6}$ (E < 20)	$2.25 \times 10^{-7} E + 1 \times 10^{-6}$	$1.37 \times 10^{-7} E + .03 \times 10^{-4}$, <100> $1.48 \times 10^{-7} E - .01 \times 10^{-4}$, <111>
$K(\text{cm}^{-1})$	$3.57 \times 10^5 E^{-0.222} (Q_T < 5 \times 10^{13}, <100>)$ $7.4 \times 10^5 E^{-0.222} (Q_T > 5 \times 10^{13}, <100>)$ $5 \times 10^5 E^{-0.222} (Q_T > 5 \times 10^{13}, <111>)$	$2.2 \times 10^5 (Q_T < 3 \times 10^{14}, <100>)$ $3.4 \times 10^5 (Q_T > 3 \times 10^{14}, <100>)$	$11.33 \times 10^6 E^{-0.708}$, <100> 3×10^5 , <111>

Table 10: Curve Fitting Model Parameters for Implanted BF₂

Parameter	Equation	
R _p (cm)	$5.6 \times 10^{-7} (E \times 0.22)^{0.844}$	
ΔR _p (cm)	$5.8 \times 10^{-7} (E \times 0.22)^{0.538}$	(Q _T < 2 x 10 ¹⁵)
	$4.8 \times 10^{-7} (E \times 0.22)^{0.538}$	(Q _T > 2 x 10 ¹⁵)
X _{tail} (cm)	R _p + 4.2 x 10 ⁻⁶	
K (cm ⁻¹)	$7.7 \times 10^5 (E \times 0.22)^{-0.222}$	(Q _T > 2 x 10 ¹⁵)
	$6.9 \times 10^5 (E \times 0.22)^{-0.222}$	(Q _T < 2 x 10 ¹⁵)
	$2.5 \times 10^6 (E \times 0.22)^{-0.222}$	(E < 45)

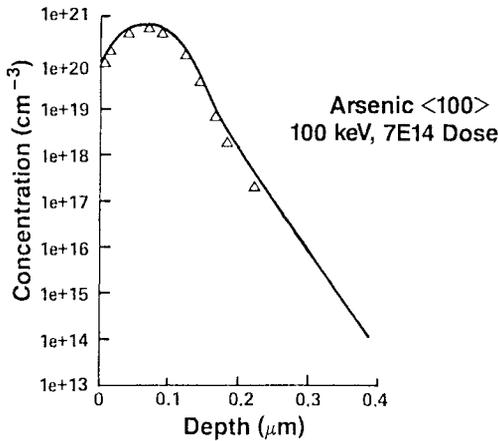
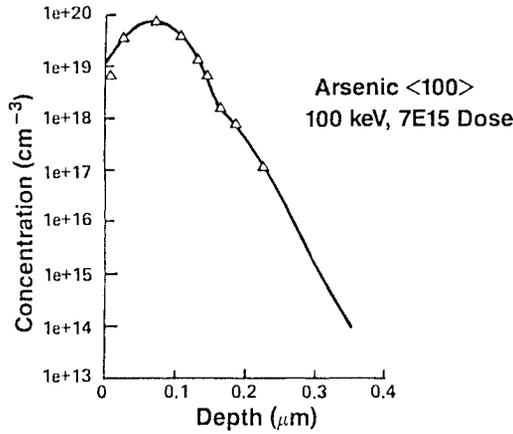


Figure 70: Examples of calculated profiles and extracted experimental data for arsenic-implanted silicon.

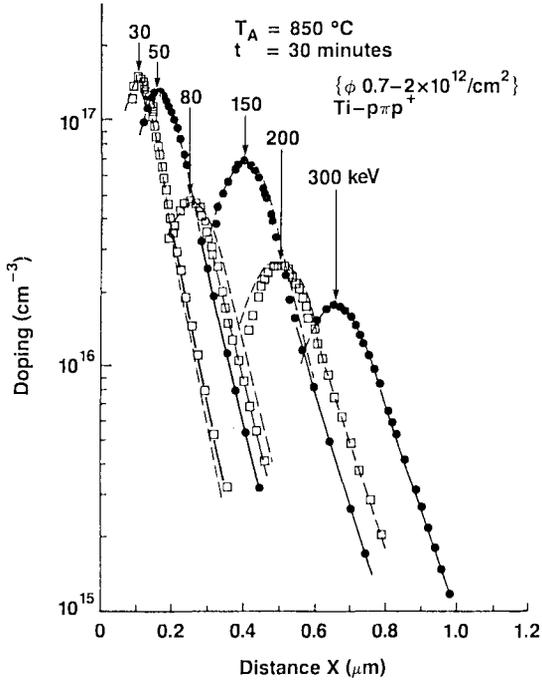


Figure 71: Example of modeled profile and extracted experimental data for boron implants into crystalline silicon.

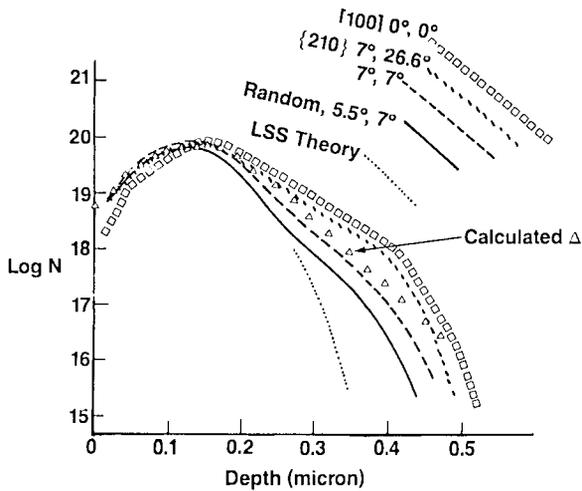


Figure 72: Example of modeled profile and extracted experimental data for boron implanted at various orientations in crystalline silicon. Boron implant was 10^{15} cm^{-2} , 45 keV.

influence the implanted impurity profiles. Using the experimental data, error measurements were calculated for the tail parameters using the profile indicating the greatest channeling, (implantation into $\langle 100 \rangle$ at 0° , and the least channeling (random orientation at 5.5° tilt and 7° rotation). The percentage difference in the X_{tail} parameter is 12%, the C_{tail} percentage difference is 92%, the C_{tail}/C_{peak} percentage difference is 89.5% and the slope percentage difference is 33.2%.

The effect of implantation beam current was not considered in formulating the models. However, this is an important factor in the channeling effect since an increase in dose rate (or beam current) will cause an increase in substrate temperature.^{80,81} An increase in substrate temperature will affect the "tail" of the concentration profile by annealing out damage occurring in the crystal. As the substrate temperature increases, the ions which have managed to enter channels can travel deeper due to the presence of fewer interstitial atoms which cause dechanneling events.⁷⁰

REFERENCES

1. P. G. Schewman, *Diffusion in Solids*, McGraw-Hill, New York, 1963.
2. J. Crank, *Mathematics of Diffusion*, Oxford University Press, London, 1956.
3. G.D. Watkins, in: *Radiation Damage in Semiconductors*, Dunod, Paris, 1964, p. 97.
4. G.D. Watkins, in: *Radiation Damage and Defects in Semiconductors*, Institute of Physics, Bristol, 1973, p. 228.
5. J.A. Van Vechten, in: *Lattice Defects in Semiconductors*, Institute of Physics, Bristol, 1975, p. 1.
6. G.D. Watkins, in: *Lattice Defects in Semiconductors*, 1974, Institute of Physics, Bristol, 1975, p. 1.
7. J.A. Naber, C.E. Mall and R.E. Leadon, in: *Radiation Damage and Defects in Semiconductors*, Institute of Physics, Bristol, 1973, p. 26.
8. L. Elstner and W. Kamprath, *Phys. Stat. Sol.*, 541 (1967)
9. D. Shaw, *Phys. Stat. Sol. (B)*, 72:11 (1975).
10. J. A. Van Vechten, *Phys. Rev.*, B17:3197 (1978).
11. C. Weigel, D. Peak, J.W. Corbett, G.D. Watkins, R.P. Messmer: *Phys. Rev. B.*, 8:2906 (1973).
12. R.B. Fair, J.C.C. Tsai: *J. Electrochem. Soc.*, 124:1107 (1977)
13. R.B. Fair: *Impurity Doping Processes in Silicon*, edited by F.F.Y. Wang, North-Holland Press, Amsterdam, 1981.
14. J.A. Van Vechten, C.D. Thurmond, *Phys. Rev.*, 14:3551 (1976).
15. J.A. Van Vechten: *Lattice Defects Semicond.*, 1975, 1:(1974).
16. J.C. Phillips, J.A. Van Vechten,; *Phys. Rev. Lett.*, 30:220 (1973).
17. M.L. Swanson, J.A. Davies, A.F. Quenneville, F.W. Saris, L.W. Wiggers, *Radiation Effects*, 35:51 (1978).
18. S. Dannefaer, G. Hogg and D. Kerr, to be published.
19. A. Seeger, W. Frank, *Conf. Ser.-Inst. Phys.*, 16:262 (1973).
20. J. Chicawa, S. Shirai, *J. Cryst. Growth*, 39:328 (1977).
21. H. Strunk, U. Gosele, B.O. Kolbesen, *Appl. Phys. Lett.*, 34:530 (1979).
22. R.B. Fair, J.C.C. Tsai, D.G. Schimmel and W. Maszara, to be published.
23. S.T. Pantelides, R. Car, P.J. Kelly, and A. Oshryama, *Amer. Phys. Soc. Meeting*, Detroit, March, 1984.
24. R.O. Simmons, R.W. Balluffi, *Phys. Rev.*, 125:862 (1962).
25. L.C. Kimberling, D. V. Lang, *Conf. Ser. - Inst. Phys.*, 23:589 (1975).

26. G.D. Watkins, J.R. Troxell, A.P. Chatterjee, *Conf. Ser. - Inst. Phys.*, 46:16 (1979).
27. A. Seeger, W. Frank, U. Gosele, *Conf. Ser. - Inst. Phys.*, 46:148 (1979).
29. P.M. Petroff, A.J. R. deKock, *Cryst. Growth*, 30:117 (1975).
30. U. Gosele, H. Strunk, *Appl. Phys.*, 20:265 (1979).
31. S. Mizuo, H. Higuchi, *Japan, pn. J. Appl. Phys.*, 20:39 (1981).
32. D.A. Antoniadis, I. Moskowitz, *J. Appl. Phys.*, 53:9214 (1982).
33. T.Y. Tan, U. Gosele, *Appl. Phys. Lett.*, 40:616 (1982).
34. G.D. Watkins, *Effects Des Rayonnements Sur Les Semiconductors*, Dunod, Paris, 1965.
35. A. Seeger, K.P. Chik, *Phys. Status Solidi*, 29:455 (1968).
36. S.M. Hu, *J. Appl. Phys.*, 45:1567 (1974).
37. R.B. Fair, *J. Appl. Phys.*, 51:5828 (1980).
38. U. Gosele, T.Y. Tan, *Defects in Semiconductors II*, edited by S. Mahajan and J.W. Corbett, North-Holland Press, Amsterdam, 1983.
39. S. Matsumoto, Y. Ishikawa, T. Niimi, *J. Appl. Phys.*, 1983, to be published.
40. D. Mathiot and J.C. Pfister, *J. Appl. Phys.*, 55:3518 (1984).
41. P. Fahey, G. Barbuscia, M. Moslehi and R.W. Dutton, *Appl. Phys. Lett.*, 46:784 (1985).
42. S. Mizuo, H. Higuchi, *Japan. pn. J. Appl. Phys.*, 20:791 (1981).
43. S. Mizuo, H. Higuchi, *Japan. pn. J. Appl. Phys.*, 21:281 (1982).
44. K. Taniguchi, K. Kurosawa, M. Kashiwagi, *J. Electrochem. Soc.* 127: 2243 (1980).
45. R.B. Fair, *J. Electrochem. Soc.* 128:1360 (1981).
46. H. Shiraki, *Japan. pn. J. Appl. Phys.* 15:1 (1976).
47. Y. Nabeta, T. Uno, S. Kubo, H. Tsukamoto, *J. Electrochem. Soc.* 123: 1416 (1976).
48. L. Kalinowski, R. Sequin, *Appl. Phys. Lett.* 35:211 (1979)
49. G. Hettich, H. Mehrer, K. Maier, *Int. Conf. Defects Radiat. Eff. Semicond.* 500 (1978).
50. G.D. Watkins, J.W. Corbett, *Phys. Rev. A*, 134:1359 (1964).
51. D. Shaw, *Phys. Status Solidi B*. 72:11 (1975).
52. R.B. Fair, *Semiconductor Silicon 1977*, edited by H.R. Huff and E. Sirtl, The Electrochem. Soc., 1978.
53. R. B. Fair, *Impurity Diffusion and Oxidation of Silicon*, edited by D. Kahng, Academic Press, New York, 1981.
54. S.M. Hu, *Phys. Status Solidi B*, 60:595 (1973).
55. F. Seitz, *Acta Crystallogr.* 3:355 (1950).
56. P. Baruch, J. Monnier, B. Blanchard, Castaing, C., *Appl. Phys. Lett.* 26:77 (1975).
57. R.B. Fair, G.R. Weber, *J. Appl. Phys.* 44:273 (1973).
58. E. Guerrero, H. Potzl, R. Tielert, M. Grassenbauer, G. Stingedden, *J. Electrochem. Soc.* 129:1826 (1982).
59. P. Fahey, R.W. Dutton and S.M. Hu, *Appl. Phys. Lett.* 44:777 (1984).
60. R.M. Harris and D.A. Antoniadis, *Appl. Phys. Lett.* 43:937 (1983).
61. K. Nishi and D.A. Antoniadis, unpublished.
62. R.B. Fair, P.N. Pappas, *J. Electrochem. Soc.* 122:1241 (1975).
63. After Varian-Extrion, DF-3000 brochure.
64. G. Dearnaley, J.H. Freeman, R.S. Nelson and J. Stephen, *Ion Implantation*, North Holland Publ. Co., Amsterdam (1973).
65. J. Lindhard, M. Scharff and H. Schiott, *Mat.-Fys. Med. Dan. Vid. Selsk* 33:1 (1963).
66. G. Baccarani and K.A. Pickar (unpublished 1970).
67. R.A. Moline, *Ion Implantation in Semiconductors* Eds. F. Chernow, J.A. Borders and D.K. Brice (Plenum Press, New York, 1976) p. 319.
68. T. Hirao, K. Inoue, Y. Yalgashi and S. Takayanagi, *Japan J. Appl. Phys.* 18, 647 (1979).
69. S.M. Sze, *Physics of Semiconductor Devices*, 2nd Edition, (Wiley), (1981).
70. G. Carter and W.A. Grant, *Ion Implantation of Semiconductors*, (Edward Arnold), (1976).

71. P.D. Townsend, J.C. Kelly and N.E. W. Hartley, *Ion Implantation Sputtering and their Application*, (Academic Press), (1976).
72. J.F. Gibbons, *Proceedings of the IEEE* 56:295 (1968).
73. W.K. Chu, et al., Semiconductor Research Annual Report, Research Triangle Park, North Carolina, (1983).
74. R.G. Wilson, *J. Appl. Phys.* 54:6879, (1983).
75. W. Maszara, C. Carter, D.K. Sadana, J. Liu, V. Ozgus, J. Wortman and G.A. Rozgonyi, 1983, Mat. Res. Soc. Symp. on Energy Beam—Solid Interactions and Transactions and Transient Thermal Processing (Eds. N. Johnson and J.C.C. Fan), (1984).
76. PREDICT, Microelectronics Center of North Carolina, Research Triangle Park, North Carolina, (1984).
77. Charles P. Ho, S.E. Hansen, Technical Report No. SEL 83-001, Stanford Electronics Laboratories, 1983.
78. H. Hofker, Philips Research Report, Supp. #8, (1975).
79. F.N. Schwettmann, *Appl. Phys. Lett.* 22:570, (1973).
80. T.E. Seidel, in *Sec. Inter. Conf. on Ion Implant.* I. Ruge and J. Graul, eds. (Springer-Verlag, Berlin), 1972.
81. D.K. Sadana, J. Washburn and C.W. Magee, *J. Appl. Phys.* 54:3479, (1983).

Microlithography for VLSI

R. Fabian Pease

*Stanford University
Stanford, California*

INTRODUCTION

The high resolution patterning of material for the fabrication of VLSI circuits is one of the key pacing technologies in the evolution of microelectronics. Feature sizes are now approaching the wavelength of light while the patterned area extends up to 1 cm^2 for a single chip and up to 150 cm^2 over a wafer. Thus, the pattern complexity is equivalent to ten thousand television images. However, these patterns are required to be defect free, with feature size control of about $0.25 \text{ }\mu\text{m}$. To make a complete circuit, up to a dozen such patterns must be overlaid with an accuracy of $0.25 \text{ }\mu\text{m}$. Projections of scaling of integrated circuit devices indicate that dimensions may shrink by another factor of four.

For most VLSI circuits making just one such pattern calls for many steps. First a mask, a pattern of chromium on a glass substrate, is formed, usually by electron beam lithography. After etching the pattern in the chromium film, the mask image is projected onto a resist-coated wafer. Following development of the resist, the pattern is transferred into the underlying circuit material. By "lithography" we mean the generation of the pattern in the resist. Transferring the pattern from the resist to the circuit material is usually done by etching which is described in other chapters. Thus, two lithographic steps are required to generate each level on the wafer, the first to generate the mask pattern and the second the wafer pattern.

Each lithographic step has several processes;

- (i) Spin coating the workpiece with a thin ($0.1\text{-}2 \text{ }\mu\text{m}$) film of resist

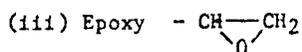
- (ii) Formation of the appropriate, high resolution, aerial image through the use of a focused electron beam in the case of mask making and an ultra-violet image of the mask for wafer exposure.
- (iii) Interaction, both physical and chemical, of the aerial image with the resist in the exposure process.
- (iv) Development of the resist pattern by selectively dissolving away the exposed (for a positive resist) or unexposed (for a negative) regions of the resist film.

In the sections below we describe these processes. We first describe mask making in considerable detail because of its relative simplicity and then go on to describe the more complex processes for wafer exposure. In the case of wafer exposure the aerial image can be formed by contact or proximity printing but these techniques have largely been supplanted by projection. A final section is a brief litany of emerging new technologies together with a list of references where the reader can pursue these topics further. The main thrust of this chapter is to present an account of today's mainstream lithographic technologies.

FORMING THE RESIST FILM

Both negative and positive electron beam resist are widely used for mask making. Because the image is built up by scanning a focused beam sequentially over each address, the process is slow. Therefore, there is considerable emphasis on resist sensitivity which is usually, unfortunately, quantified as the dose in C/cm^2 required to bring about the desired chemical change.

Negative resists are long chain polymers containing groups which, on electron bombardment, form crosslinks between adjacent chains.¹ Such groups include:



Those portions of the film that are crosslinked are insoluble in a suitably chosen developer. Typical required doses range from 10^{-7} to $10^{-5} C/cm^2$ for 10 keV electrons.

Positive electron resists are also long-chain polymers. However, positive resists contain groups which, on electron bombardment, cause chain scission and hence locally reduced molecular weight. The exposed regions are now selectively soluble in a suitably chosen developer. Examples of such materials are poly(methyl methacrylate) (PMMA)² and poly(butene-1 sulfone) (PBS)³. The chain scission process also results in gas evolution which may also promote the selective solubility of the exposed regions.

Typical required doses are somewhat higher than those for negative acting resist and range from 10^{-6} to 10^{-4} C/cm² for 10 keV electrons.

The resists as purchased are solutions of the above polymers in volatile solvents. A few ml. of the solution are dispensed onto the center of the wafer which is spun at a pre-determined rate while the solvent evaporates to leave a polymer film of predictable thickness. The theory of the spinning action is quite complicated because some material is lost through centrifugal action and the viscosity of the remaining material changes during evaporation of the solvent.⁴ Often the thickness, t , of the remaining film is given by the expression:

$$t = k_s^2 / \sqrt{\omega}$$

where k is a constant related to viscosity

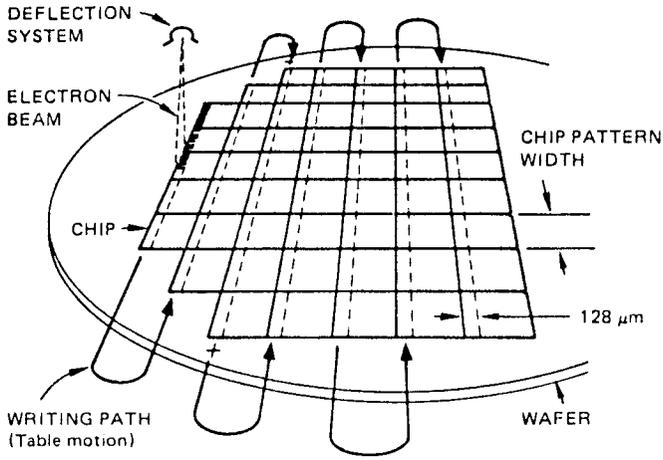
s is the solids content in the solution expressed in percent by weight.

ω is the angular velocity.

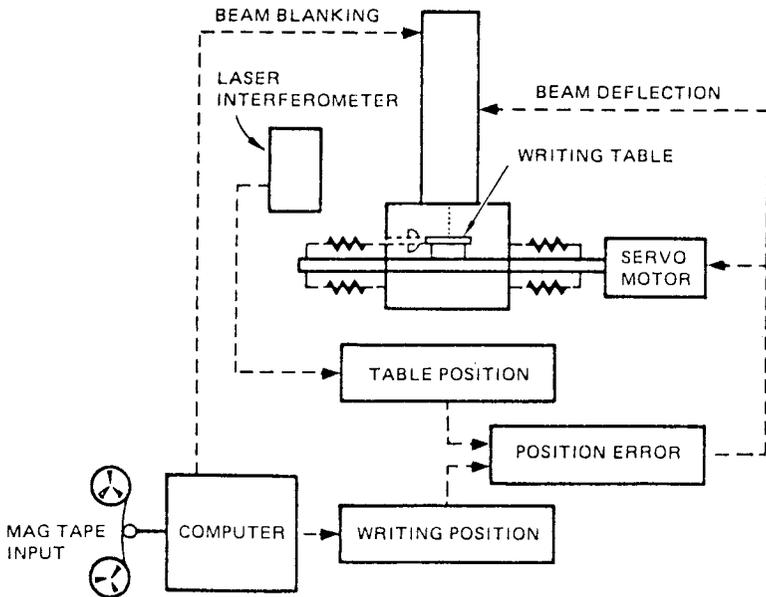
For mask making the surface of the substrate must be quite flat, much less than $0.1 \mu\text{m}$ over distances of up to 1 cm and less than $1 \mu\text{m}$ overall. The minimum thickness of the resist is set by the maximum allowable defect density, less than one per square inch. Although there exists little quantitative information on defect density versus resist thickness, the minimum value is usually $0.4 \mu\text{m}$. Unlike masks, the surface of wafers may contain step heights exceeding $1 \mu\text{m}$ due to previous patterning steps. Resist film thicknesses of up to 2 to 3 μm may be required to coat these surfaces uniformly.

GENERATION OF THE AERIAL IMAGE FOR ELECTRON BEAM MASK MAKING

The image is formed by focusing a writing electron beam onto the resist film and scanning the beam across the film in the appropriate pattern. In general, mechanical scanning of the workpiece supplements the electrical scanning of the beam. One advantage of this arrangement is that large area electrical scanning becomes less important and makes it easier to maintain the focus and positional accuracy of the deflected beam. The writing strategy and a schematic view of the most widespread commercial instrument, "EBES" developed at Bell Laboratories⁵, is shown in Figures 1 and 2. The choice of the convergence semi-angle α is important. Too large a value results in unacceptably large beam diameters due to aberrations of the final beam forming lens. Too small a value results in unacceptably low beam currents due to limits of source brightness and space charge. These limits are discussed in detail elsewhere⁶ and are summarized in Figure 3. It should be noted that if the electron energy is increased then the beam current can be increased proportionately. The specifications for the current version of EBES, Perkin-Elmer's MEBES 3, are shown in Table 1. The writing rate is 80 MHz, or 8×10^7 exposed addresses per second. A rough calculation of the exposure time indicates that at a $1/4 \mu\text{m}$ address size the rate comes to 2 cm²/minute. Also note that the optimum value of α is about 10^{-2} radians. Although beam diameters of less than 5 nm have been used



(a) WRITING STRATEGY



(b) SCHEMATIC VIEW

Figure 1: (a) Writing strategy used in Bell Laboratories Electron Beam Exposure System "EBES": a single stripe of the circuit pattern is read out repeatedly from memory then is written on to each chip site. (b) Schematic view of EBES. The feedback system employs beam deflection to compensate for table position errors (After D.R. Herriott et al.⁵).

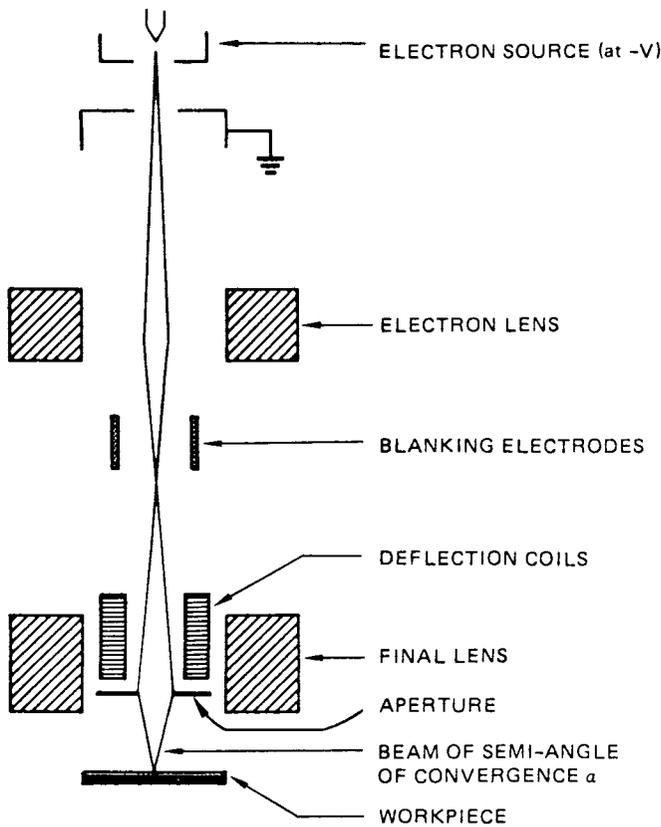


Figure 2: Schematic view of an electron beam column employing two lenses which focus a demagnified image of the source onto the workpiece. The beam can be blanked, swept away from the aperture, by energizing the blanking electrodes. The deflection coils serve to scan the beam across the workpiece while the aperture serves to control the convergence semi-angle α . Larger values of α lead to increased current but at the expense of larger aberration disks.

for specialized patterning⁸ most mask generation is carried out with beams diameters in the range of 0.1 to 0.5 μm , respective currents of 5 nA to 300 nA and an electron energy of 10 keV.

INTERACTION OF ELECTRONS WITH THE WORKPIECE

The electrons entering the resist film not only bring about the chemical effects alluded to earlier, but are also scattered by the film and substrate so that the lateral extent of the interaction can exceed the diameter of the impinging beam. This effect is known as the proximity effect and has been the subject of large numbers of papers. The effect lends itself well to

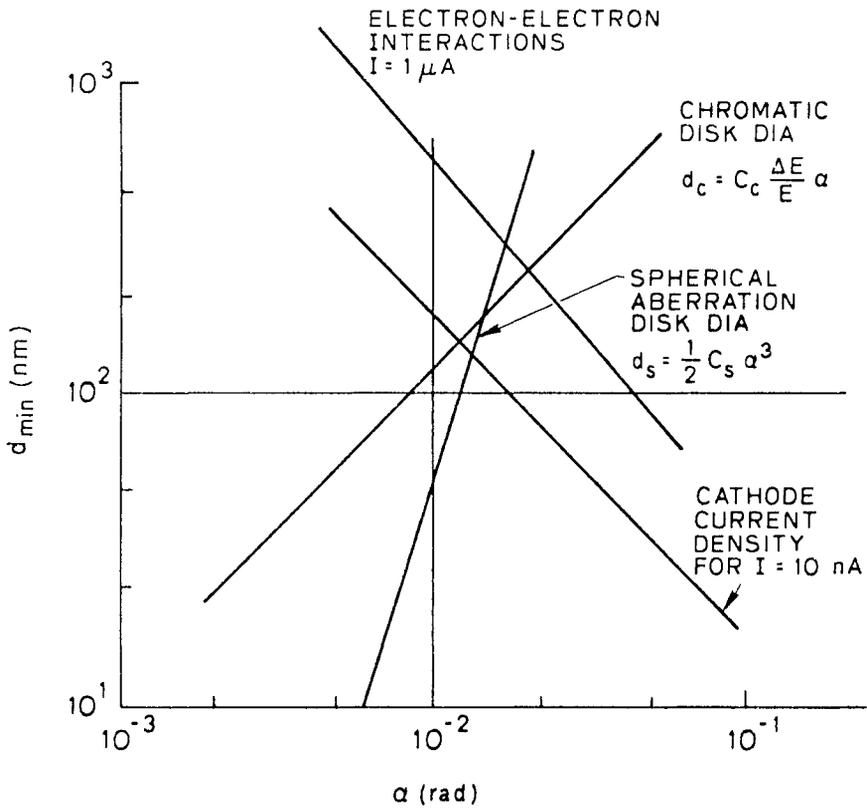


Figure 3: Contribution to focused electron beam diameter as a function of semi angle of convergence α . Often the total beam diameter is estimated by adding in quadrature the contributing diameters. The electron energy, E , is 10 keV with energy spread $\Delta E = 3$ eV, chromatic aberration coefficient 4 cm, spherical aberration coefficient $C_S = 10$ cms, cathode current density 10 A/cm², and column length 66 cms.

Table 1: Outline Specifications of a Commercially Available Electron Beam Pattern Generator "MEBES 3" Manufactured by the Perkin Elmer Corporation

Address and Beam Diameter	0.1 μ m to 1.1 μ m
Maximum Writing Area	6.1" x 6.1"
Level-to-Level overlay Position Accuracy	0.12 μ m
Electron Energy	10 keV
Writing Rates	40 MHz, 80 MHz
Required Temperature Control	$\pm 0.1^\circ$ C

modelling by a Monte-Carlo technique which has been verified by experiment.^{9,10} Monte-Carlo simulations of 10 keV and 20 keV electrons are shown in Figure 4. Qualitatively the extent of scattering of the electrons is much less for the lower energy 10 keV electrons, however, the 20 keV electrons are scattered less by the resist so that there is a sharper concentration of energy dissipation around the point of impact. Fortunately, for making photolithographic masks with minimum dimensions of 1.5 to 2 μm , the use of 10 keV electrons and 0.4 μm resist is a reasonable compromise and is a widely used combination which does not require elaborate correction for proximity effects. If either thicker resist or finer dimensions are required, then higher electron energies and some form of local dose variation for correction of proximity effects may become desirable. However, for most of today's mask making, the above combination of 10 keV electrons and a resist thickness of 0.4 μm allows us to treat the proximity effect as just another contribution to beam diameter. The magnitude of the contribution can be determined by simulation¹¹ or by experiment^{9,10} to be

$$J(r) = J(o) \exp \left(- \frac{r^2}{r_p^2} \right)$$

where J_r is the current density at radius r , and $r_p = 0.2 \mu\text{m} \pm 0.1 \mu\text{m}$ for the case depicted in Figure 4a. At higher voltages two gaussian terms are needed for an adequate description.

The chemical effect of electron bombardment has already been described as generating crosslinks between adjacent polymer chains to create an insoluble gel. We can invoke a simple model to describe quantitatively the resist's behavior in terms of the parameters, sensitivity and contrast.

The starting point of the model is to assume the following:

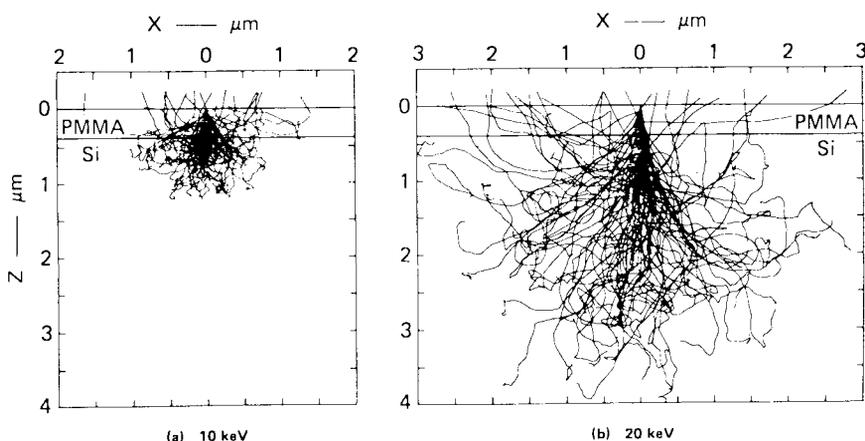


Figure 4: Trajectories of electrons scattered by a typical target comprising 0.4 μm polymeric resist on a silicon substrate. Substitution of glass as the substrate material would give rise to a slightly greater extent of scattering.¹¹

- (i) The local rate of generating crosslinks is proportional to the local power dissipated by the electron beam.
- (ii) When two or more molecules are crosslinked together they form an insoluble group.
- (iii) The number of crosslinks per unit volume obeys a Poisson distribution.
- (iv) There are many crosslinking sites per chain such that only a negligible fraction form crosslinks at the lowest exposure needed to crosslink virtually all molecules.
- (v) Associated with each chain is a volume, V_m , within which a crosslinking event results in a bond to a neighboring chain.
- (vi) The effect of the resist material on the incoming beam is unaffected by crosslinking, i.e. the distribution of power dissipation is constant during exposure.
- (vii) The energy $\Lambda(z)$ where z is the depth into the resist. We shall initially assume that Λ is constant with depth.

For very low exposures, only isolated pairs of molecules are crosslinked together. On development, these pairs per se may not be dissolved, but all the surrounding uncrosslinked material is dissolved and washed away so that no resist remains after development. At some critical dose enough chains are crosslinked together so that a skeletal gel is formed which remains after the developing solvent has leached out the uncrosslinked polymer chains. Beyond this exposure level, we assume that each crosslink that bonds a previously uncrosslinked molecule contributes that molecule to the gel. The fractional thickness, T_n , of resist remaining after development is equal to the fraction of (identically sized) molecules, each with the same volume, V_m , that have at least 1 crosslink to neighboring molecules. If N_c is the mean local concentration of crosslinks then the mean number of crosslinks per volume V_m is $N_c V_m$. The fraction of molecules with no crosslink to neighboring molecules is, from the Poisson distribution, $\exp(-V_m N_c)$. Therefore, the local fractional thickness remaining is

$$T_n = 1 - \exp(-V_m N_c)$$

for doses in excess of the critical dose and

$$N_c = \Lambda g \frac{Q}{q}$$

where Q is the local dose (in C/cm²)

g is the number of crosslinks generated per eV dissipated (the sensitivity of the resist material)

q is the electronic charge

$$T_n = 1 - \exp(-V_m \Lambda g Q/q)$$

Therefore, a plot of T_n versus $\log_{10} Q$ can be drawn (Figure 5). The general form is clear and understandable. At doses below the critical dose $T_n = 0$.

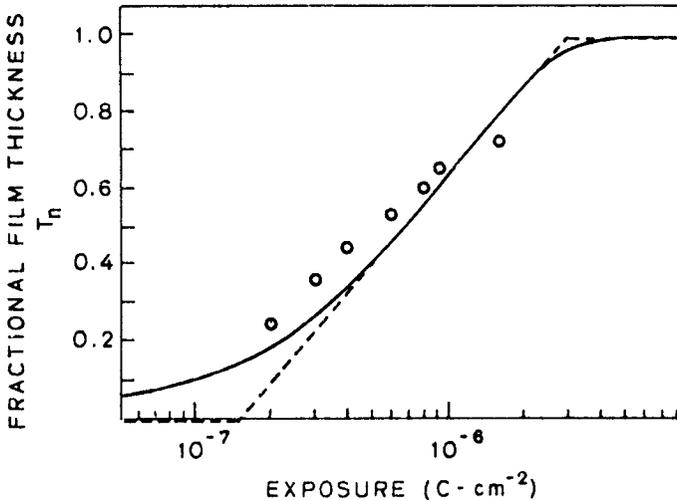


Figure 5: Fractional thickness remaining after development T_n as function of exposure for PGMA-co-EA negative electron resist (COP) for $V = 10$ kV. The solid line is the model, the circles are experimental points (Reference 12).

At increasingly higher doses, increasing numbers of molecules are cross-linked until virtually all molecules are crosslinked together so that continued exposure adds virtually nothing. However, the shape of the curve in Figure 5 can also be used to quantify two resist parameters. One is the sensitivity or dose required to bring about the required value of T_n . Thus, for $T_n = 0.5$, $-V_m \Lambda g Q/q = \ln 0.5$.

$$Q \rightarrow \frac{0.7 \times 1.6 \times 10^{-19}}{V_m \Lambda g} \quad \text{C/cm}^2$$

This looks quite reasonable since when the required dose is reduced for larger V_m , a film made of larger molecules requires a smaller concentration of crosslinks to bond them all together. Also, at large values of Λ the energy dissipated per incident electron per unit depth, and g , the number of crosslinks per eV dissipated will also tend to lower Q .

The other parameter is the contrast, or gamma, which is a measure of the minimum ratio of exposures needed to bring about a required difference in T_n (often 0 to 50%). If we approximate the curve in Figure 5 as three straight line segments (shown dotted) then the slope of the central portion is the contrast (γ) defined as $dT_n/d(\log_{10} Q)$. If this slope is equal to the maximum slope of the solid curve then we can quantify γ as

$$\left. \frac{dT_n}{d(\log_{10} Q)} \right|_{\max} = \frac{1}{e \log_{10} e} = 0.85$$

Note that this value is independent of g, Λ, V_m . According to this simple

model the contrast of negative acting, crosslinking resists is about 0.85 and independent of their constituents. As it happens a great many negative acting electron beam resist materials do have approximately this value of contrast. Of course, a high value is desirable because this means that less contrast in the aerial image is needed to bring about a given relief image in the resist. A number of materials are reported to have values of up to 1.6. There are a number of possible explanations. One is that those few materials with abnormally high values are the most attractive and hence are selected for intense study. Another is that the above simple model may not apply so well to these few materials. For example, a higher value of γ is obtained if we postulate that more than 2 molecules must be crosslinked together to form an insoluble gel. Nonetheless, the simple model above accounts for the main features of negative acting, crosslinking, electron-beam resist. It is quite straightforward to embellish the simple model. A spread in the values of molecular weight can be modeled as an equivalent spread in the values of V_m . It can be determined, as has been reported experimentally¹³, that this will give a lower value of γ than a material with a single value of V_m or molecular weight. A non-uniform value of Λ can also be accounted for by describing each elemental layer according to the above model. It so happens that for the case we are concerned with, 10 keV electrons exposing 0.4 μm thick resist the assumption that Λ be uniform throughout the resist thickness is not a bad one.

The negative resist most frequently used for mask making is poly (glycidyl/methacrylate-co-ethylacrylate) referred to colloquially as COP.¹³ It has a contrast of about 0.9, close to the value predicted by our model, MW of 180,000 and a g value of .01 crosslinks per eV. Determining V_m is tricky but if we take V_m as the mean volume occupied per molecule and the specific gravity as 1 gm/cc, then $V_m \approx 2 \times 10^{-19} \text{cc}$. From experiment and from simulation of electron scattering, a 10 keV electron, on the average, dissipates 1/3 of its energy in a resist film 0.4 μm thick;

$$\text{whence} \quad \Lambda \rightarrow \frac{10,000}{3 \times 0.4 \times 10^{-4}} \rightarrow 10^8 \text{ eV/cm/electron}$$

$$\text{whence for} \quad T_n \rightarrow 0.5 Q \rightarrow 3.6 \times 10^{-7} \text{ C/cm}^2$$

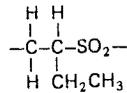
This is remarkably close to the value shown in Figure 5 although a figure of $2 \times 10^{-7} \text{ C/cm}^2$ is usually recommended for this material.

To determine the 2-dimensional relief image obtained as a result of 10 kV electron beam of known diameter exposing a 0.4 μm -thick negative resist, we can employ a crude model in which the local thickness after developing is given by the original thickness multiplied by the value of T_n corresponding to the local energy dissipated throughout the thickness of the resist. In practice neither this nor any other model is yet used because the negative crosslinking resists used swell considerably on developing. A doubling of the thickness during development has been observed. Even with a developing solvent mixture optimized to minimize swelling the shape of the image is set by the rheological behavior of the resist in the developing and rinsing solvents and in the subsequent post-development

bake. Some example of images in negative electron resist are shown in Figure 6. Thus these particular materials do not lend themselves either to useful modelling or to high resolution, steep-sided patterns. For making patterns of minimum critical dimensions of $4\ \mu\text{m}$ or greater, however, such materials are in widespread use because of their sensitivity, ease of use and adhesion.

For mask making the subsequent pattern transfer step is accomplished by: wet etching of a 60 nm thick chromium film with a buffered ceric ammonium nitrate solution. Thus good adhesion is necessary but a steep sided resist profile is not.

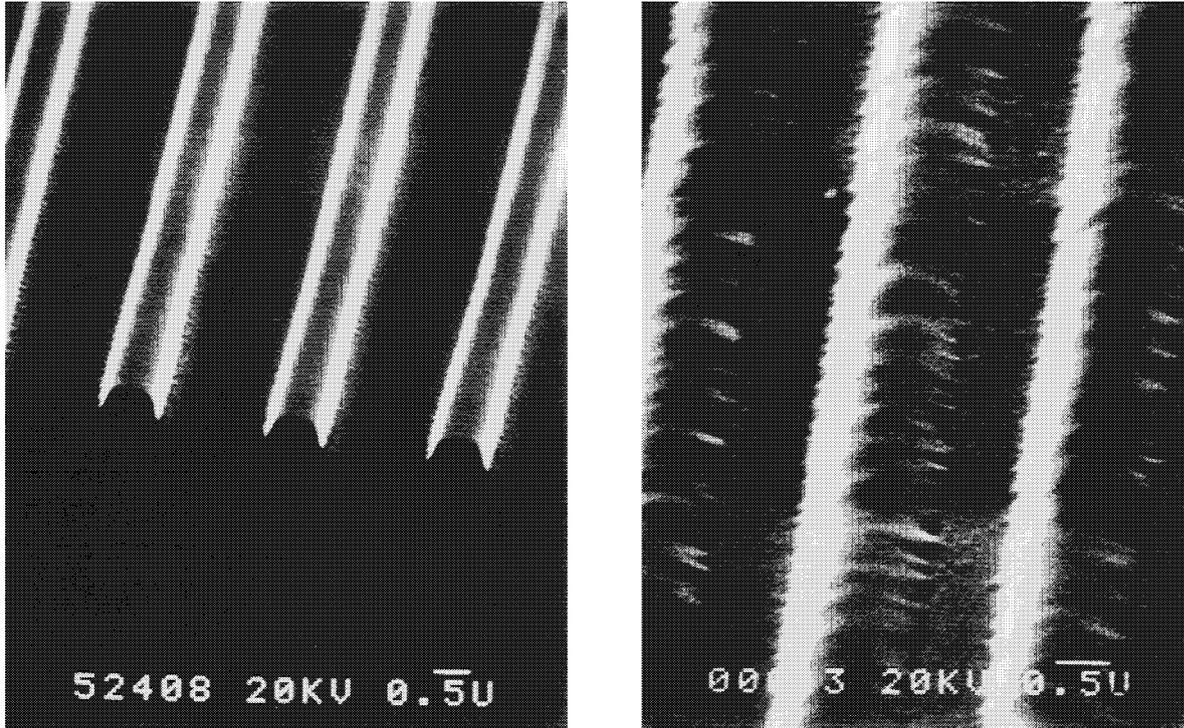
The most frequently used positive resist material for electron beam mask making is poly (butene 1-sulfone) ("PBS").³ The repeat unit of the chain is:



On irradiation by electrons, the predominant chemical reaction is chain scission accompanied by the evolution of SO_2 gas. The exposed regions can be selectively dissolved away in a solution of water and methyl iso-amyl ketone (MIAK). The curve of T_n versus dose of 10 keV electrons (Figure 7) indicates that the minimum required dose is 8×10^{-7} C/cm² and that the contrast is about 1.5. The curve shown in Figure 7 depends critically on the development process since considerable swelling occurs on development, though to lesser extent than with COP. Thus, a model of the exposure and development process that has some theoretical basis and that also reliably predicts the resulting relief image does not yet exist for this material. To compound the issue, it is not always easy to obtain good SEM micrographs of PBS patterns because the material is believed to deform as a result of the electron irradiation in the SEM. At the minimum exposure level, feature edges have very shallow slopes (Figure 8), although, with much higher exposures ($20\ \mu\text{C}/\text{cm}_2$ at 20 kV) and modified developing, crisp submicron features have been reported.¹⁴ In practice, masks with controlled feature sizes down to $1\ \mu\text{m}$ are produced using $0.4\ \mu\text{m}$ thick PBS resist and 10 keV electron exposure. The linewidths have a standard deviation of less than $0.08\ \mu\text{m}$ and defect densities are adequately low. As with COP, the subsequent pattern transfer step is wet etching of a 60 nm chromium film with a buffered solution of ceric ammonium nitrate. Examples of portions of chromium mask made in this way are shown in Figure 9.

EXPOSURE AND DEVELOPMENT OF PHOTORESIST ON SEMICONDUCTOR WAFERS

We must now expose the resist on the wafer in accordance with the mask pattern. The simplest way of accomplishing this is by contact printing or proximity printing. The former technique allows excellent resolution but is prone to defects especially when many contacts are needed to assure accurate alignment. Thus, contact printing is not in widespread use for



a

b

Figure 6: SEM photograph of PCMS negative electron resist pattern (a) exposed with a 0.5 μm diameter electron beam, (b) exposed with a 0.1 μm diameter electron beam (courtesy of H.S. Choong).

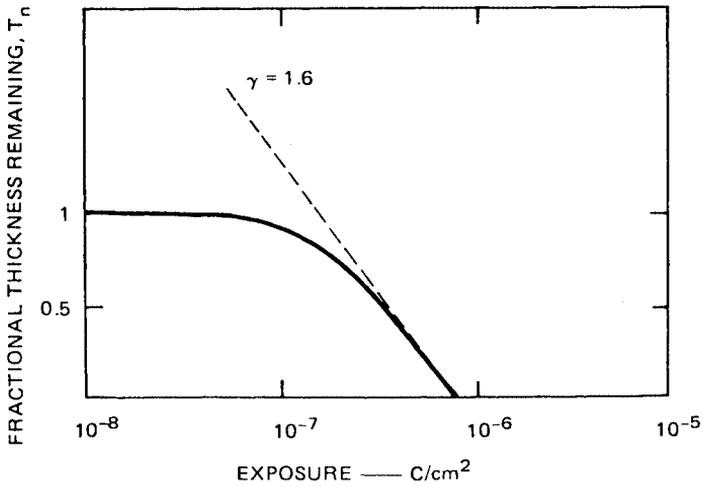


Figure 7: Fractional thickness remaining after development versus exposure for poly (butene 1-sulfone) resist exposed with 10 kV electrons (Reference 3).

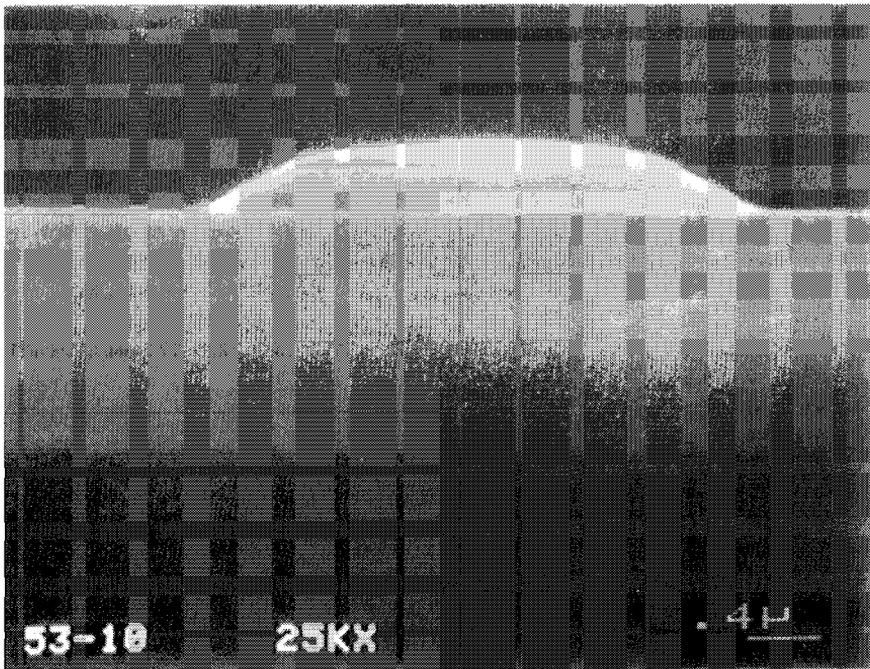
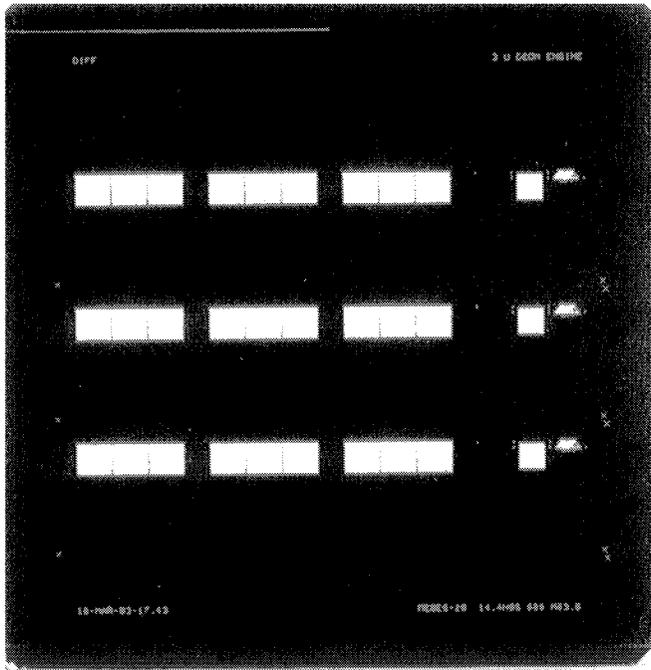
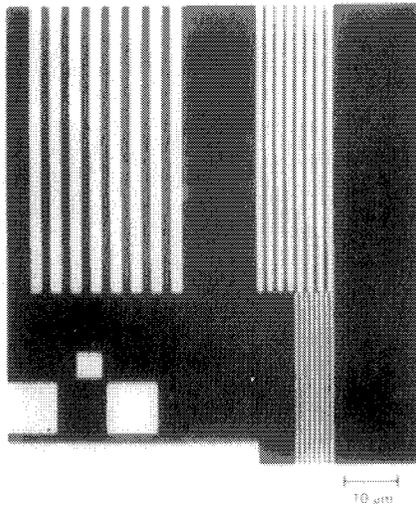


Figure 8: SEM photograph showing end view of an image formed by electron exposure of PBS resist (photograph courtesy of E. Crabb and G. Eiden).



(a) 3 x 3 INCH ARRAY



(b) RESOLUTION OBTAINED WITH PBS RESIST

Figure 9: Two views of chromium masks made with electron beam exposure of PBS resist (photographs courtesy of D.H. Dameron and J.P. Ballantyne).

VLSI. The defect problem is alleviated by proximity printing in which a gap of 15-30 μm exists between mask and wafer. Unfortunately, this gap degrades resolution so that this technique is no longer the method of choice for high density circuit manufacture. The method now adopted is to project an image of the mask onto the wafer by means of a lens or mirror system. There are two main classes of projection exposure tools also referred to as "mask aligners". The first is the scanning projection aligner in which a doubly reflecting system (Figure 10) forms an erect image over an arc-shaped field of view. The optics are such that over the field of view there are essentially no aberrations for numerical aperture values (NA) up to 3. By mechanically scanning both the mask, containing the pattern to be transferred to the wafer, and the wafer simultaneously through the illumination over the field of view an image of the complete pattern on the mask is transferred to the resist on the wafer. This scanning projection technology has been in use for about eight years and can be used for the manufacture of circuits with linewidths down to less than 2 μm . The specifications for one such system, the Perkin-Elmer Micralign 660 HT[®] are outlined in Table 2. The other form of projection aligner is usually referred to as a "stepper"

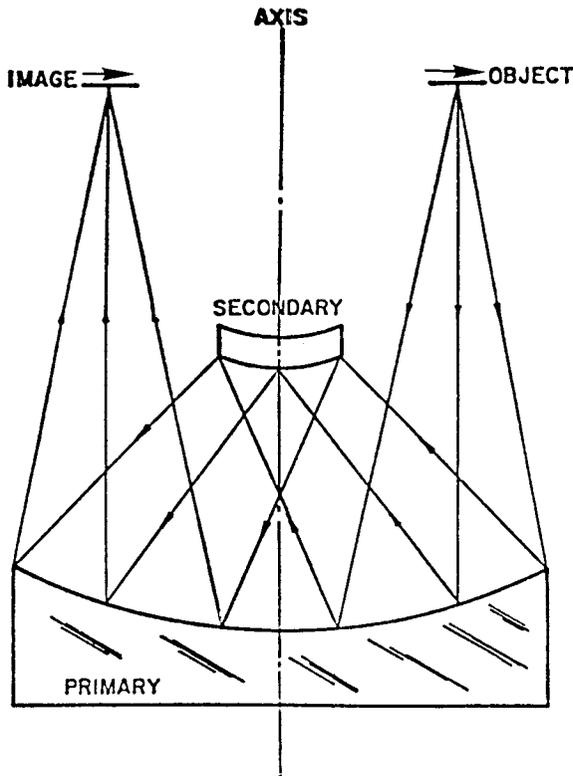


Figure 10: Schematic view of doubly reflecting optics used in scanning projection printing.

Table 2: Outline Specifications for the Perkin-Elmer "Micralign 660 HT" Scanning Projection Printer

Throughput:	100 6-inch wafers per hour
Exposing Wavelengths:	240 nm to 436 nm
Resolution:	0.9 μm to 1.25 μm lines and spaces
Auto-Alignment:	$\pm 0.25 \mu\text{m}$ (98% of population)
Magnification Compensation:	$\pm 2 \mu\text{m}$ range $\pm 0.25 \mu\text{m}$ precision
Machine-to-Machine Overlay:	± 0.45 (98% of population)
Uptime:	In excess of 90%

because the image of the mask pattern occupies only a small portion ($1 \times 1 \text{ cm}^2$) of the wafer area and the whole area is filled up by stepping the wafer mechanically and repeating the exposures. Refracting focusing is usually used for the optics. With this technique, VLSI chips with feature sizes of $1 \frac{1}{4} \mu\text{m}$ are being manufactured. The specifications of two such steppers are shown in Tables 3 and 4. With these techniques both positive and negative photoresist are used although the former is becoming the most popular.

FORMATION OF THE AERIAL IMAGE IN PROJECTION MASK ALIGNERS

In both the scanning projection and the stepping aligners the physics of image formation is the same. An outline of the basic scheme is shown in Figure 11. A mask is illuminated and the projection optics focuses an image of the mask at the wafer surface. As with electron beam lithography

Table 3: Outline Specifications of the Ultratech Model 1000 Wafer Stepper

Throughput:	24 6-inch wafers per hour
Exposing Wavelengths:	390 nm - 450 nm
Numerical Aperture:	0.315
Effective Partial Coherence:	0.45
Resolution:	$[0.8\lambda(\text{NA}) = 1.1 \mu\text{m}$ lines and spaces]
Field Size:	20 mm x 7.6 mm (max. rectangle)
Machine Alignment	$\pm 0.26 \mu\text{m}$ (2-sigma)
Total RMS Overlay Precision:	$\pm 0.35 \mu\text{m}$

Table 4: Outline Specifications of the GCA Model 6300B Wafer Stepper

Throughput:	40 150 mm wafer/hour
Alignment Precision:	$\pm 0.15 \mu\text{m}$ TIR, Global
Numerical Aperture:	0.30
Resolution:	(working) $1.1 \mu\text{m}$ $[0.8\lambda(\text{NA})]$
Field Size:	20 mm diam.
Exposing Wavelength:	436 nm (g line)
Magnification:	1/5 X

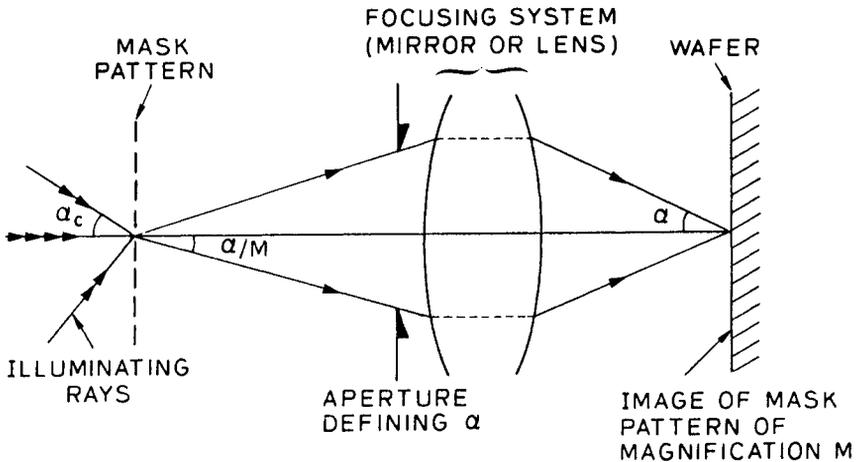


Figure 11: Basic optics required for projection printing.

the semi-angle of convergence, α , is a critical parameter although for somewhat different reasons. Because the optical engineer can arbitrarily control the shape of the refracting or reflecting surfaces, the images are virtually aberration-free over the field of view. The sharpness of the aerial image is set by diffraction, provided that the wafer surface is in the plane of best focus. When the illumination is in the form of a plane wavefront and the mask pattern is a 1-dimensional grating whose amplitude transmissivity is a sine wave, then there is only first order diffraction. If α is large enough to accept these diffracted beams then, according to the Abbe principle for microscopic image formation,¹⁵ the image is formed with full contrast at the wafer surface. The converse is true, that if α is too small to accept the diffracted beams then there is zero contrast at the wafer and the grating is

“unresolved”. The relation between α and the maximum resolved spatial frequency ν_{\max} at the wafer is

$$\nu_{\max} \rightarrow \frac{\sin \alpha}{\lambda}$$

where λ is the wavelength of the exposing radiation. The term $n \sin \alpha$, where n is the refractive index of the space between the lens and resist, is the “numerical aperture” of the lens (NA). The early projection aligners had $NA = 0.15$ but values in excess of 3 are now common. Thus with the illumination specified and with $\lambda = 404 \text{ nm}$, a frequently used wavelength available in mercury arc lamps, $\nu = \nu_{\max}$ (Figure 12, dotted line).

Illumination of the above type is referred to as “coherent” illumination. An alternative, incoherent, form of illumination is to have many wavefronts impinging on the mask over a wide range of angles $\pm\alpha_c$, greater than the angle $\pm\alpha/M$ subtended at the mask by the lens aperture, and with no systematic phase relations between them. Now it is possible for a ray whose direction is $-\alpha/M$ to give rise to a ray diffracted by $+2\alpha/M$ in that both the undiffracted and diffracted ray are accepted by the aperture and give rise to small, but non zero, contrast for spatial frequencies of $2 \nu_{\max}$. For smaller spatial frequencies a greater fraction of the incident beam will give rise to acceptable diffracted beams and so the image contrast will approach one. For this, incoherent, illumination the coherence factor can be described as $\sigma = M \sin \alpha_c / \sin \alpha$. When $\sigma = 0$ we have perfectly coherent illumination (our first case), when $\sigma > 1$ the illumination is usually regarded as incoherent, and intermediate cases are partially coherent. In Figure 12 are plotted values of image contrast as functions of normalized spatial frequency $\nu_N \equiv \nu \lambda F$ where F , the “F number”, is defined as $1/2NA$. Most mask aligners operate with $\sigma = 0.7$ because this gives the necessary high image contrast (> 0.6) for good relief images in most resist materials. Lower

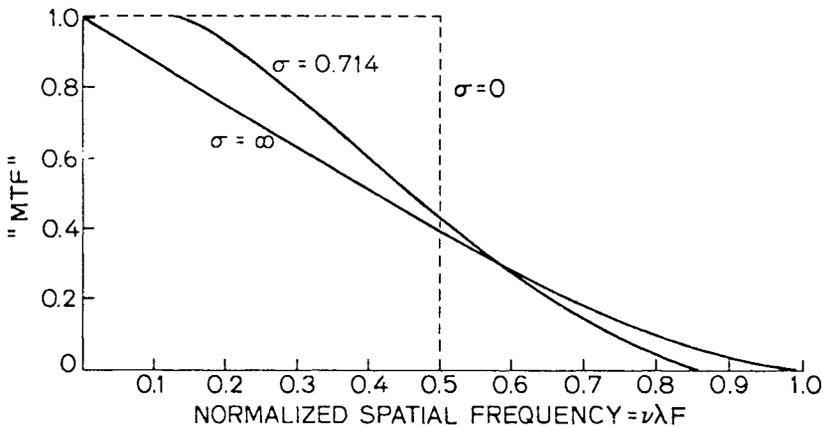


Figure 12: Contrast (“MTF”) of image of a bar pattern of equal lines and spaces as a function of normalized spatial frequency for 3 different values of coherence factor σ of the illumination (from Reference 16).

values of σ give rise to the appearance of fringes at sharp edges which is obviously undesirable.

As a digression we might point out that the human eye responds well to very low contrast, down to 0.05, images. Hence, light microscopes operate with high values of σ so that spatial frequencies close to $2\nu_{\max}$ can be resolved. Thus a mask aligner with $\sigma = 0.7$, $\lambda = 0.4 \mu\text{m}$ and $\text{NA} = 0.3$ might generate resist patterns with a maximum period of $1/\nu$ where

$$\nu = \frac{0.8 \text{ NA}}{\lambda} = 0.6 \mu\text{m}^{-1}$$

or $0.8 \mu\text{m}$ lines and spaces whereas a high resolution light microscope objective lens with $\sigma > 1$, $\lambda = 0.5 \mu\text{m}$, $\text{NA} = 0.5$ might yield a visibly resolved image corresponding to a spatial frequency given by $1.8\nu_{\max} = 1.8 \mu\text{m}^{-1}$ or $0.25 \mu\text{m}$ lines and spaces.

The reader should be cautioned against attempting to generalize the response of the projection optics to arbitrary mask patterns for intermediate values of σ . For a value of $\sigma = 0$, the system is linear with respect to amplitude and for $\sigma = \infty$ it is linear with respect to intensity. For intermediate values, the system is non-linear and extension to patterns other than a grating pattern is tricky.¹⁶ The curves in Figure 12 are for a square wave object with sharp transitions between zero intensity and full intensity. It is interesting to point out (from Figure 12) that light optics with a minimum wavelength of 190 nm, set by the transparency of the mask substrate, and $\text{NA} = 0.5$ can generate images with 10% contrast at a spatial frequency of $4 \mu\text{m}^{-1}$ or $0.125 \mu\text{m}$ lines and spaces. Such a system in practice would require resists of much higher contrast than are presently used and reflective focusing optics of very high numerical aperture. In practice contact printing has yielded lines of $0.2 \mu\text{m}$ widths.¹⁷ So, the limits to the resolution of optical lithography are not firmly set by diffraction but by increasingly difficult practical problems in pushing the diffraction limit further into the submicron regime.

In modeling the aerial image formation process, we usually assume a grating pattern and use curves similar to those of Figure 13. We are usually interested in spatial frequencies such that only the first order diffracted rays are accepted and so the aerial image is a sine wave of contrast indicated by Figure 13. We can also model the case of out-of-focus images as might occur when the wafer is not flat and there is no dynamic correction. Treatment of this case is beyond the scope of this chapter, and the reader is referred to the monograph by King.¹⁶ The most widely used modelling program SAMPLE¹⁸ requires as inputs the values of λ , widths of exposed and unexposed lines, NA and, in later versions, the extent of defocus.

INTERACTION OF ULTRA-VIOLET LIGHT WITH PHOTORESIST

The most commonly used wavelengths are 436 nm, 404.7 nm or less commonly 365 nm. All 3 lines are present in mercury arc illumination. Because the energy of the photons is insufficient to directly cause cross-linking or chain scission, a photoactive compound (PAC) is a third compon-

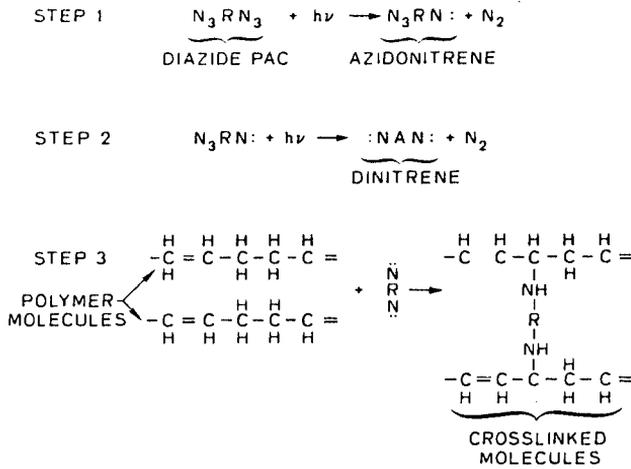


Figure 13: Outline mechanism of crosslinking in negative photoresist (after Blais, Reference 19).

ent incorporated, the other two being the base polymer and the solvent, into both negative and positive resists. In the case of negative photo-resist, the PAC is a di-azide material which decomposes and bleaches under UV irradiation to form nitrenes which bring about crosslinking of the base polymer molecules (Figure 13). Positive photo-resist (PPR) also contain a PAC with diazide groups which on UV irradiation bleaches and decomposes to render the local base polymer, a novolak resin selectively susceptible to dissolution by an aqueous alkaline solution.

We can model the exposure of both negative and positive photoresist by embellishing the simple exposure model used for the exposure of electron beam resist. The treatment follows that of Dill and co-workers.²⁰⁻²³

Initially we will treat the case where the resist is on a non-reflecting substrate. The resist material when unexposed absorbs light quite strongly. As the exposure proceeds, the PAC bleaches thus reducing the absorbance of the resist. This effect can be seen by viewing the resist in a high power light microscope in the reflection mode. After viewing the resist at 1000X magnification for several seconds, during which time the resist becomes clearer, viewing the same area at 400X magnification will reveal a small bleached area representing the area illuminated when at 1000X magnification. Thus, unlike the case of the electron beam resist, we have the complication that the local power dissipation varies both with depth, z, and with time, t. This behavior can be modelled by assigning an absorptivity $\alpha(z,t)$ to the resist material, where α contains both a labile component which decays on exposure and a non-labile component which is unaffected by exposure.

Thus, across an elemental depth, Δz , the local intensity drop is given by

$$I \rightarrow I_0 \exp(-\alpha \Delta z)$$

The labile component is modeled as $AM(z,t)$, where M is the normalized concentration of the PAC compound and therefore, $M = 1$ at $t = 0$, at the start of the exposure, and 0 at $t = \infty$. The non-labile component is a constant B . Thus

$$\Delta I \rightarrow I \exp \{ [-AM(z,t) + B]\Delta z \}$$

We model the rate of decomposition of the photoactive compound as being proportional to the local concentration and the local intensity such that $\partial M/\partial t = I(z,t)M(z,t)C$ where C is a constant corresponding to the intrinsic sensitivity of the material.

The parameters A, B , and C which describe the exposure characteristics of the resist can be determined by measuring the transmittance $T \equiv I_{(z_0)}/I_{(0)}$ of a resist film of thickness Z_0 on a transparent substrate of matching refractive index, real part. Note that $I_{(0)}$ is the intensity just inside the top surface of the resist and is related to the incident intensity, I_{inc} , by

$$I_{(0)} \rightarrow (1-R) I_{inc}$$

where R is the reflection coefficient at the top surface of the resist. Depending on the experimental arrangement a similar correction may have to be made for the emerging ray. For now, we will define $I(z_0)$ as the intensity at a depth Z_0 just inside the resist film of thickness Z_0 . Because the substrate and resist have matched refractive indices we will assume that there is no reflected wave.

Determining A and B is quite straightforward as at $t = 0$

$$T \rightarrow T_{(0)} \rightarrow \exp [- (A+B)Z_0]$$

and at $t = \infty$

$$T \rightarrow T_{(\infty)} \rightarrow \exp [-BZ_0]$$

Determining C is a bit more tricky and involves measuring the initial rate of change of T .

$$\text{In general } T \rightarrow \exp \left\{ - \int_0^{Z_0} [AM(z,t) + B] dz \right\}$$

$$\begin{aligned} \therefore \frac{dT(t)}{dt} &\rightarrow -T(t) \frac{d}{dt} \left(\int_0^{Z_0} AM(z,t) dz \right) \\ &\rightarrow -T(t)A \int_0^{Z_0} \frac{\partial M(z,t)}{\partial t} dz \end{aligned}$$

$$\begin{aligned} \text{At } t \rightarrow 0 \quad \frac{dT}{dt} &\equiv \left. \frac{dT}{dt} \right|_{t=0} \rightarrow -T_{(0)}AC \int_0^{Z_0} I(z,0)M(z,0) dz \\ &\rightarrow -T_{(0)}ACI_0 \int_0^{Z_0} \exp [-Z(A+B)] dz \end{aligned}$$

$$\therefore \left. \frac{dT}{dt} \right|_{t=0} + \frac{-T(o)ACI_o}{A+B} 1-\exp [-Z_o(A+B)]$$

$$\therefore C + \frac{A+B}{AI_o T(o)[1-T(o)]} \left. \frac{dT}{dt} \right|_{t=0}$$

Some values for A, B, C for a positive photoresist under different conditions of treatment are shown in Table 5.

Table 5: Exposure Parameters A, B, C, for AZ1350 Photoresist (Reference 21)

Exposure Wavelengths	λ	436 nm
Refractive Index (real)	n	1.68
Labile Absorptivity	A	$0.54 \mu\text{m}^{-1}$
Non-Labile Absorptivity	B	$0.03 \mu\text{m}^{-1}$
Intrinsic Sensitivity	C	$0.014 \text{ cm}^2/\text{mJ}$

Having determined the exposure parameters A, B, C, by measuring T as a function of time, we can now describe the development behavior of negative and positive photo-resists. In negative photoresist, following Blais,¹⁹ the fraction of azide groups decomposed after a given exposure is 1-M. If all the resulting nitrene groups have had time to react, then the probability of 1 azide group in a diazide group bonding to a poly (isoprene) molecule is 1-M and the probability of both azide groups bonding to polymer molecules is (1-M)². This implies that the bonding events are independent. Thus the above double event corresponds to a single crosslinking event between 2 polyisoprene molecules, assuming that the number of crosslinking events between 2 portions of the same molecule is negligible. We can complete our model using the treatment outlined for negative electron resist in which any polymer molecule crosslinked to another is assumed insoluble in the developer and vice versa. Note however, the important difference is that to achieve $T_n=1$ we must have at least as many diazide (PAC) molecules as polymer molecules. A detailed treatment is not given here but the reader can develop the model independently by determining M(z) and then equating $N_c(z) = [1-M(z)]^2 X$, the original concentration of PAC molecules. Over the last 10 years there has been a considerable decline in the popularity of negative resist. Most negative photoresist swell considerably on development, which limits their resolution and, following postbake, exhibit sloping feature edges which are

unsatisfactory when dry etching is used at the following step. The thickness versus exposure curve of a typical negative photo resist is shown in Figure 14. As with negative crosslinking electron resists, both the contrast and the required dose are quite low, about unity and 5 mJ/cm² respectively. Recently some newer materials, e.g. Selectilux[®], have been reported to exhibit better topographic characteristics. As a result, negative photoresists may regain their popularity.

To model the development response of positive photoresist we have, as yet, no satisfactory mechanism. We resort to an empirical determination of the rate of etching away of the resist material as a function solely of M for each material. The technique for measuring the etch rate, R, as a function of M is described fully in Reference 22 and is described as

$$R \propto \exp (E_1 + E_2M + E_3M^2)$$

Reliable and quick ways of determining developing rate³⁷ are important because the values obtained depend critically on many factors such as temperature of the developer and the degree of agitation. Hence, quoted values should only be used as examples. Some results are shown in Table 6.

Thus, we have characterized the interaction of ultra-violet light with photoresist in terms of the resulting normalized concentration, M, of

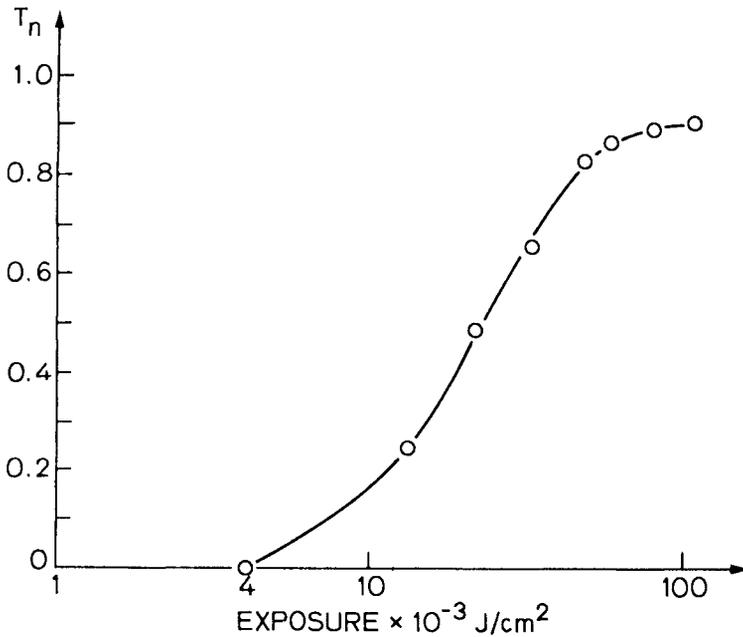


Figure 14: Fractional thickness remaining, T_n, versus exposure for Kodak KMR 752 negative photoresist. Notice that T_n never reaches 1 suggesting that the extent of crosslinking is limited by the availability of photoactive compound (after Blais, Reference 19).

Table 6: Development Parameter $E_1E_2E_3$ for AZ 1350° Photoresist (Reference 22)

Development Process 1:1 AZ Developer: Water at 22°C

$$E_1 = 5.27$$

$$E_2 = 8.19$$

$$E_3 = 12.5$$

photoactive compound. We then described how this value of M is related, in the case of negative resists, to the concentration of crosslinks and hence, to T_n . In the case of positive photoresists, the etch rate in the developer for given process conditions is treated as a unique function of M .

EXPOSURE AND DEVELOPMENT OF PHOTORESIST FILMS ON REFLECTIVE SUBSTRATES

When the resist is on a strongly reflecting substrate, such as a metal film, the amplitude of the reflected wave inside the resist can be comparable with that of the forward wave. A standing wave pattern is set up with nodes at the metal surface and at distances of integral numbers of $\lambda_r/2$ away from the surface, where $\lambda_r = \lambda/N$ where N is the real part of the refractive index of the resist. Thus, the resist will be preferentially exposed in layers separated by $\lambda_r/2$. Because there is absorption in the resist, both the forward and reflected wave decay as they proceed through the resist. So, the interference is strongest nearest the metal surface where the two waves have approximately the same amplitude. This absorption can be described by assigning to the resist material a complex refractive index $\underline{N} = n-jk$, so that k now refers to absorptivity and n is the real part of the refractive index. For most resist materials $n \gg k$ so that the $n \doteq |\underline{N}|$.

The case for a perfectly reflecting metal substrate is depicted in Figure 15. At each interface the relevant boundary conditions are that both electric fields \underline{E} and magnetic fields \underline{H} must be continuous. The amplitude reflection coefficient R can be determined from the true simultaneous equation resulting from the two boundary conditions at the top interface:

$$1 + R = E_2 (1 - e^{-j\phi})$$

$$1 - R = NE_2 (1 + e^{-j\phi})$$

where $\phi = 4\pi Z_0/\lambda (n+jk)$

Whence $R = \frac{1 - \exp(-j\phi) - n[1 + \exp(-j\phi)]}{1 - \exp(-j\phi) + n[1 + \exp(-j\phi)]}$

The power absorbed in the film is simply $1 - R^2$, the difference in power of the incident and reflected wave, outside the resist. Thus, determining R is

important in knowing how much of the incident beam's power is absorbed in the resist. Evaluating the above expression for R will indicate that as Z_0 is increased $|R|$ will be at a local maximum whenever Z_0 is an even multiple of $\lambda_R/4$ and a minimum for an odd multiple of $\lambda_R/4$, anti-reflection case. The reader may wish to check this independently for $j = 0.02$, $N \doteq 1.68$, and $\lambda = 404$ nm. The significant conclusion is that as the resist thickness changes by $\lambda_R/4$ ($\doteq 60$ nm), the power coupled into the resist can change by a factor of 2 or 3. Thus, there are two serious effects of the reflected wave; one is to cause a layered pattern to the exposure and the other is to make the total exposure a very sensitive function of resist thickness. Unfortunately, except at the first lithographic level, wafer surfaces are rarely flat and the resist thickness varies greatly at steps in the wafer surface, often resulting in serious variations in resist linewidth (Figure 16). One way to alleviate the layering effect is to bake the resist after exposure and before developing. This is thought to promote diffusion of the remaining PAC and hence homogenize the developing rate.²⁴ Another effective technique is to eliminate the reflected wave with a buffer layer of absorbing material. This is described more fully in the section on emerging new technologies.

In modeling programs such as SAMPLE,¹⁸ the case of reflecting substrates can be handled and the resulting resist feature profiles predicted quite well. The first set of input parameters includes:

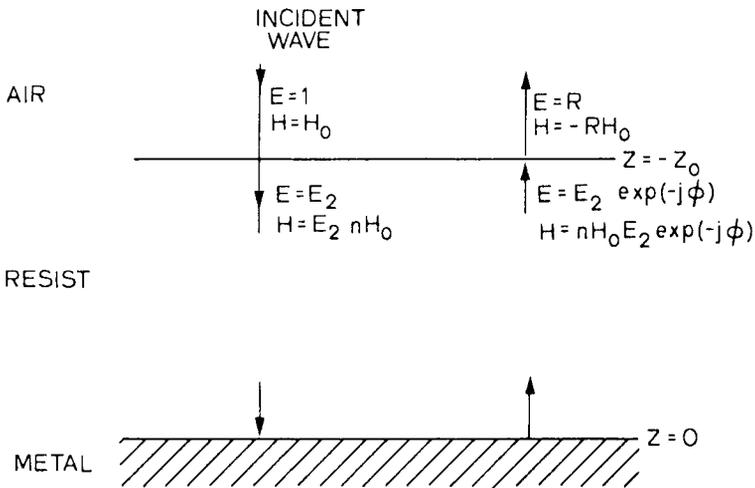


Figure 15: Schematic view of resist film on a perfectly reflecting metal film (reflection coefficient = .1). The incident wave has unit amplitude and the total amplitude reflection coefficient is R . At the air/resist interface both the electric field and the magnetic field are continuous. Because of the principle of superposition, we can treat the series of multiply reflected waves as two waves, one traveling downwards and the other upwards. The complex phase shift $\phi = 4\pi Z_0/\lambda(n+jk)$.

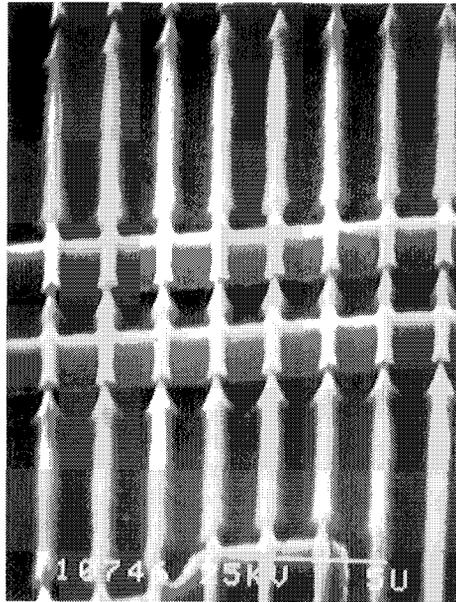


Figure 16: Linewidth variations in positive photoresist covering steps (Reference 25).

The linewidth and period of the mask pattern.

The wavelength, λ , and dose in mJ/cm^2 .

The numerical aperture NA and coherence factor σ of the aligner.

The above will give the aerial image incident on the resist. We now must supply the second set of parameters which includes:

The exposure parameters A, B, C of the resist material.

The thickness of the resist and of any underlying film (e.g. oxide).

The complex refractive index of the resist, the underlying film and of the substrate.

We can now determine $M(z)$ for each value of x where x is lateral position and Z in depth.

We now supply the parameters $E_1, E_2,$ and E_3 describing the developing action and can determine the resist profile after a succession of developing intervals. The advancing front of the etching process is controlled by the parameters $M(x,z), E_1, E_2, E_3$ and is modeled as a string of elemental

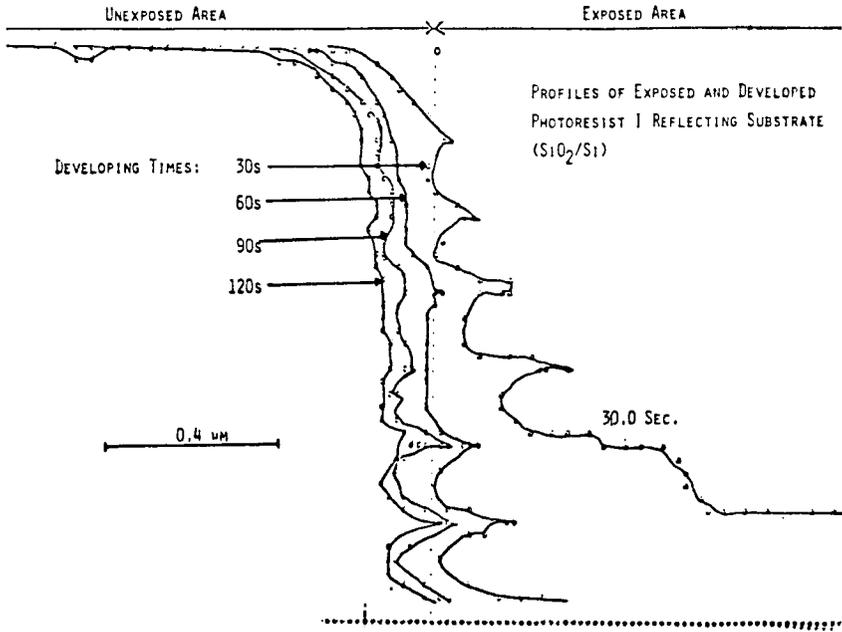


Figure 17: Examples of simulated profiles of exposed and developed positive photoresist on a reflecting substrate.

straight line segments. The point of each intersection of neighboring segments advances a distance given by equation (1) in a direction which bisects the angle between the normals of the two segments.¹⁸

Some results, obtained using an early version of SAMPLE, are shown in Figure 17. The effect of the reflecting substrate is obvious from the scalloping of the profiles resulting from the layering of the exposure maxima. A scanning electron micrograph, Figure 18, of positive photoresist profiles clearly shows the scalloping of the edges obtained by exposing on a reflecting substrate. Current versions of SAMPLE allow for multiple wavelengths of exposing radiation, non-zero defocus, and a more sophisticated description of the developing action.²⁶

EMERGING NEW TECHNOLOGIES

The research and development of microlithographic technologies is an intense area of activity. Eye catching names such as "Natural Lithography", "Brushfire Lithography" and "Vote Taking Lithography" keep emerging. For those wishing to study these developments the proceedings of the annual International Symposium on Electron Ion and Photon Beams is a good starting point.²⁷ Below we briefly review four of the most promising.

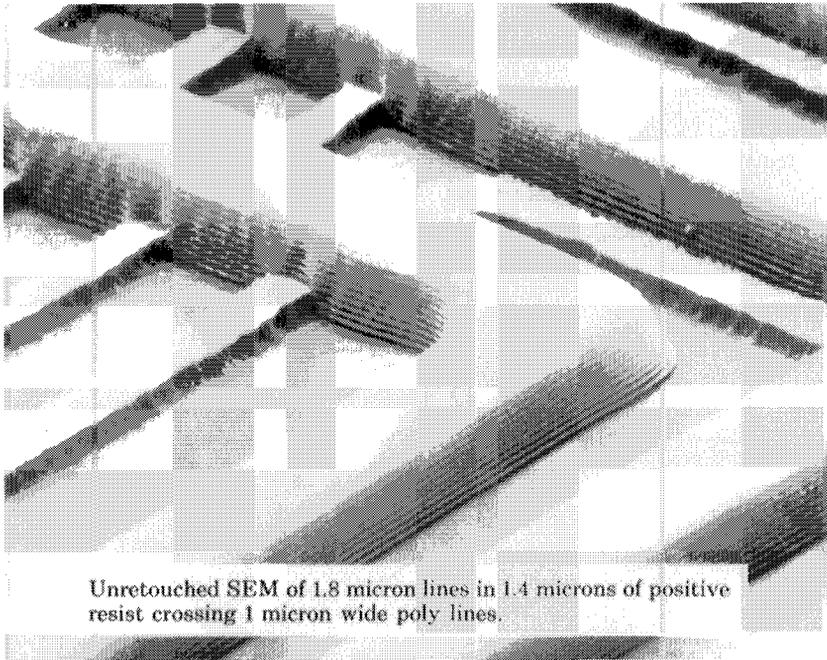


Figure 18: Scanning electron micrograph of pattern in exposed and developed positive photoresist on a reflecting substrate (courtesy of the Canon Corporation).

Multiple Level Resist²⁸

The principle of multiple level resist is illustrated in Figure 19. The first level of resist is spun on in the standard manner and may indeed be a standard resist. The action of spin coating with a thick ($1-3\ \mu\text{m}$) bottom layer is to planarise the workpiece surface. The top surface of the bottom layer is much smoother than the wafer surface. Thus, we can use a thin ($0.4\ \mu\text{m}$) top layer to adequately coat the top surface. This layer is the only one sensitive to the image of the mask pattern. Following the patterning of the top layer, the pattern is transferred to the lowest layer by any one of a variety of techniques. One popular scheme is to use a thin intermediate layer such as SiO_2 which can be easily plasma etched through the pattern in the top level. This intermediate layer then acts as a very robust mask for ion etching the much thicker first level. Some results are shown in Figure 20.

Although the multi-level resist structure was originally invented to alleviate the problems of thick, sensitive, crosslinking resists in x-ray and electron-beam lithography, it is particularly applicable to photolithography because, in addition to allowing a thinner, more uniformly thick sensitive layer, the lower layers can be rendered absorbing to eliminate the standing wave effects. Thus we can now greatly improve the linewidth control at

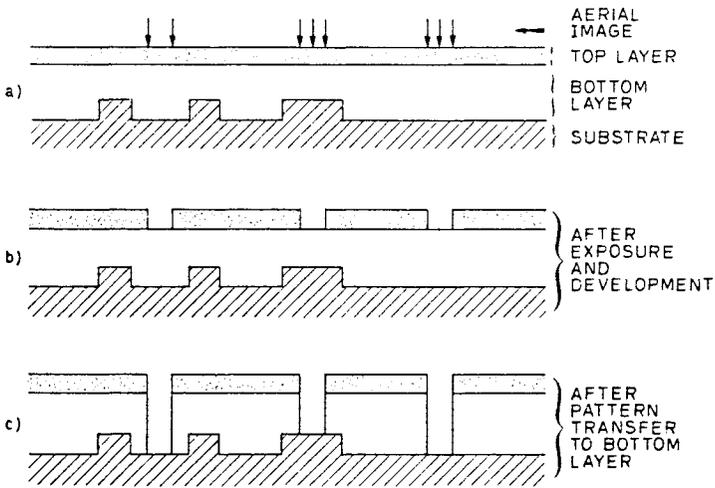


Figure 19: Principle of multi-level resist. The spin-on bottom layer is much thicker than any step height and presents a much flatter surface for the top radiation-sensitive layer. Following patterning of the top layer, by the aerial image, the bottom layer is patterned by some technique, e.g., ion beam etching, not sensitive to variations in larger thickness. Often a 3rd intermediate layer is used to provide a robust mask for etching the bottom layer.

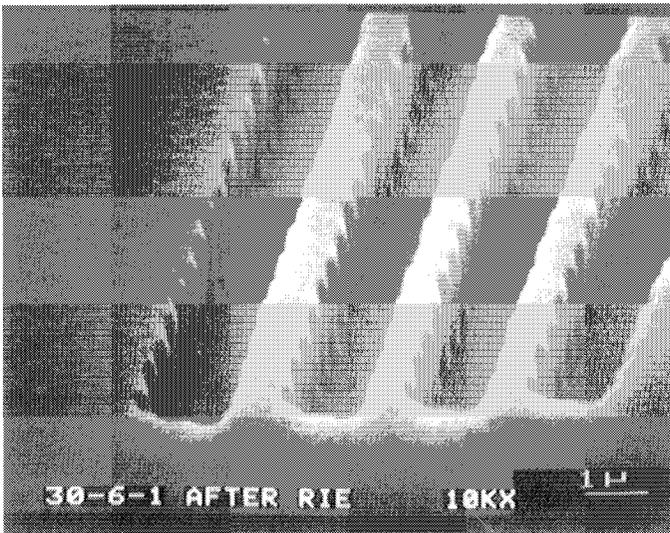


Figure 20: A tri-level resist pattern formed by electron beam lithography of the top layer (PBS resist) followed by reactive ion etching of an intermediate (800 Å silicon) layer and then oxygen ion etching of the lowest layer of overbaked (200°C for 30 min) AZ1470 resist (photograph courtesy of E. Crabbe and G. Eiden).

steps over reflecting substrates (Figure 21). The first reported manufacturing application of the multi-layer resist scheme was in the production of the CPU chip for the Hewlett Packard 9000 computer.²⁹ Stepper projection lithography was used to generate the critical levels which have a half-pitch of only $1.25\ \mu\text{m}$.

Electron-Beam Direct Write³⁰

Electron beam pattern generators have been used to generate patterns directly on resist-coated wafers since the mid 1960s. The advantages are that no mask is needed and feature sizes and overlay errors can each be less than $0.1\ \mu\text{m}$. The disadvantage is that the throughput is very low. Thus one obvious application has been in fabricating individual devices for research.

Since 1975 this technique has also been used for the production of chips that are only needed in limited quantity. The first machines used for this application could expose twenty $2\frac{1}{4}$ "-diameter wafers per hour.³¹ This throughput was achieved by projecting a shaped beam onto the wafer surface so that many address were exposed simultaneously. The drawbacks were that the minimum feature was $2.5\ \mu\text{m} \times 2.5\ \mu\text{m}$ and that linewidth control was not particularly good. In more recent machines these disadvantages have been overcome by a novel electron optical arrangement in which the shape can be varied electronically. Typically, the shape

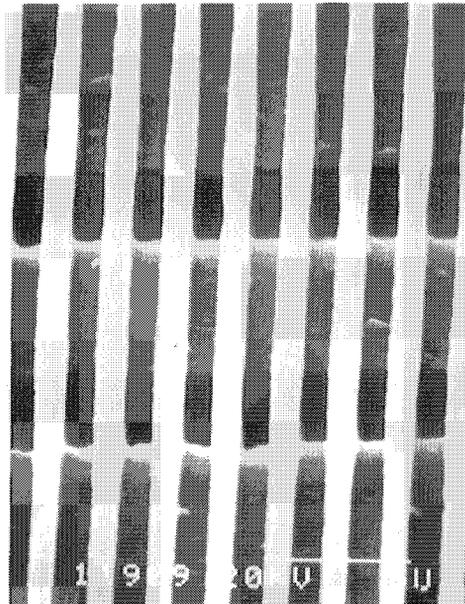


Figure 21: Optically exposed tri-level resist covering aluminum steps. Note the greatly improved linewidth control compared with that shown in Figure 16 (SEM photograph courtesy of M.M. O'Tool, E.D. Liu, M.S. Chang) (Reference 25).

is a rectangle with dimensions that range from $0.5 \times 0.4 \mu\text{m}^2$ to $2 \times 2 \mu\text{m}^2$ with changes taking place in less than $1 \mu\text{s}$.³² The specifications of a recently announced commercial instrument are shown in Table 7.

Table 7: Outline Specification of the Perkin-Elmer AEBLE - 150 Electron Beam Direct Write System

Writing Strategy:	Vector-scan, variable-shape beam
Electron Energy:	20 keV
Throughput:	up to 30 4-inch wafers per hour
Current Density:	200 A/cm ² at minimum feature size
Minimum Feature Size:	0.5 $\mu\text{m} \times 0.5 \mu\text{m}$
Overlay Accuracy:	$\pm 0.15 \mu\text{m}$ 3-sigma
Critical Dimension Accuracy:	0.08 μm

X-Ray Lithography³³

This is an approach to combine sub-optical resolution with parallel processing. The principle, outlined in Figure 22 is proximity printing with

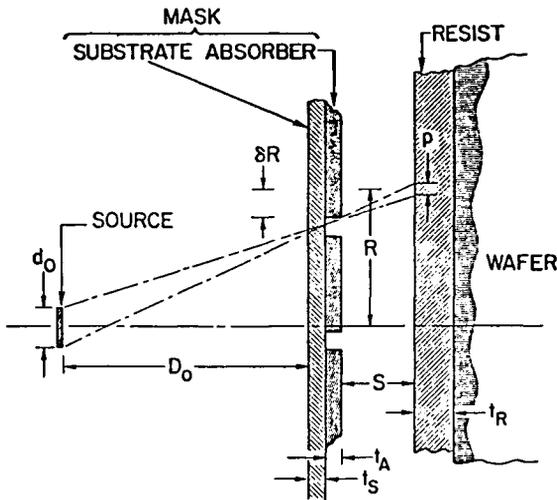
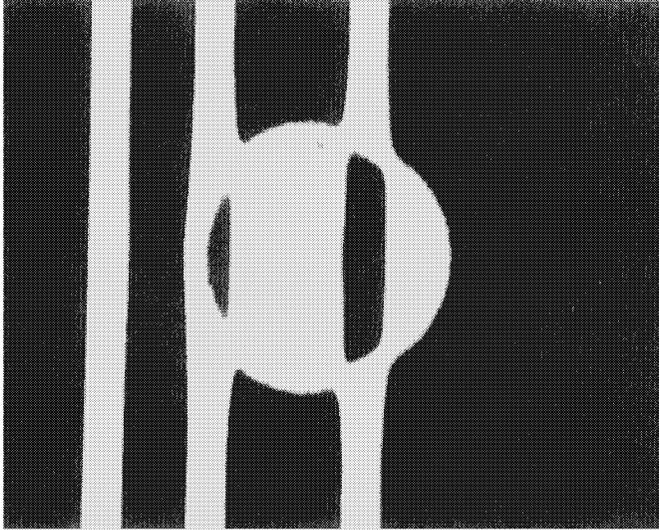
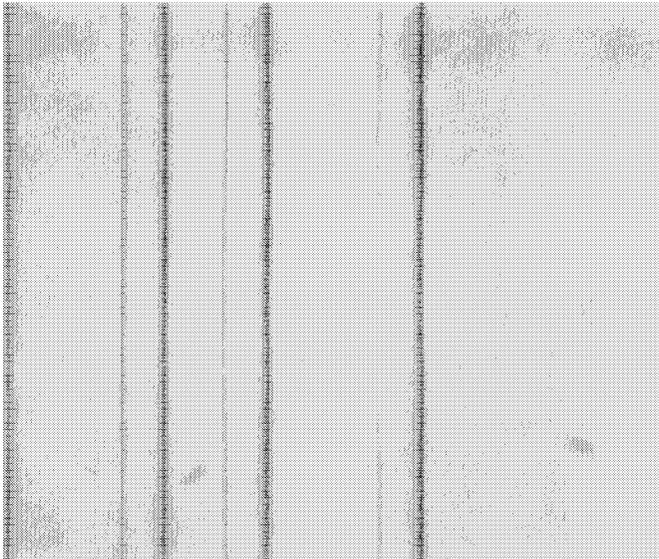


Figure 22: Principle of X-ray lithography. Because point projection is used the sharpness of feature edges is set by penumbra as well as by photoelectron generation. Varying the mask-to-wafer gap will change the magnification and can be used for correcting isotropic linear distortion but uncontrolled local variations will cause placement errors.



a



b

Figure 23: Elimination of the effect of mask defects by vote taking: (a) micrograph of a grossly defective portion of reticle mask showing both positive and negative defects, (b) micrograph of corresponding wafer resist pattern showing effective correction by using four mask fields per chip field.

the mask consisting of a thin, x-ray transparent membrane supporting a pattern of metal thick enough to absorb most of the incident x-ray beam. One of the best known of the many variations is a mask membrane of boron nitride and polyimide and an absorber pattern of $0.7\ \mu\text{m}$ —thick gold. The x-ray wavelength is $4.36\ \text{\AA}$ generated by electron bombardment of a stationary, water cooled, palladium target. The main problem is making a mask which is defect free, distortion free and exhibits less than $0.1\ \mu\text{m}$ linewidth errors in the $0.7\ \mu\text{m}$ - thick gold absorber pattern. Other problems have been in developing an attractive combination of resist and source. The use of chlorine in the resist greatly improves sensitivity to the $4.36\ \text{\AA}$ -wavelength radiation and the use of multi-level resist structures allows such a resist to exhibit good linewidth control. Exposure times of less than 1 minute per wafer have been reported for this system.³⁴ The use of novel x-ray sources such as plasma sources and synchrotron radiation is under active study. Very intense sources allow a greater choice of resist without sacrificing speed. In addition, synchrotron radiation, offers much better collimation of the beam to reduce placement error.³⁵

Vote Taking Lithography³⁶

One of the drawbacks of any form of lithography that involves masks is that defects in the masks themselves lead to degraded yield. In VTL, the effect of random defects on the masks is effectively eliminated by building up the exposure at each chip site by exposing with at least 3 identical patterns in the mask. If, for example, four mask patterns are used to build up the chip pattern then 1 random defect corresponds to a 25% perturbation of the illumination at the wafer. The trick is to obtain excellent overlay accuracy of the 4 mask patterns and to have resist with sufficiently high contrast (at least 2) that a 25% perturbation in exposure level does not result in a printed defect. Some encouraging results are shown in Figure 23, in which both negative and positive tone defects are apparent in one of the fields yet the resulting printed patterns appears adequate. The overlay accuracy obtained is better than the specification, $0.1\ \mu\text{m}$, of the stepper used (Ultratech Model 90C). This is attributed to the fact that specifications usually refer to the overlay between different machines at different times

What makes this approach to reducing the effect of defects interesting is that it is the first reported attempt to employ some form of error correction through redundancy in the mask. Hopefully, other imaginative approaches to overcoming the vexing problem of lithographic defect on the wafer will be forthcoming.

Acknowledgments

The preparation of this paper was largely supported by DARPA Contract Number MDA 903-84-K-0062. The use of material from sources other than the author's is acknowledged where appropriate.

REFERENCES

1. *Materials for Microlithography*, Ed. L.F. Thompson, C.G. Willson, and J.M.J. Frechet, 1984 American Chemical Society Symposium Series 266.
2. I. Haller, M. Hatzakis, and R. Srinivason, *IBM J. Res. Dev.* 12:251 (1968).
3. M.J. Bowden et al., in *Microcircuit Engineering*, Ed. H. Ahmed and W. C. Nixon, p. p. 239 (Cambridge University Press, 1980).
4. A.E. Emslie, F.T. Bonner, L.G. Peck, *J. Appl. Phys.* 29:858 (1958).
5. D.R. Herriott, R.J. Collier, D.S. Alles, and J.W. Stafford, *IEEE Trans. Elec. Dev.* 22:1305 (1975).
6. R.F.W. Pease, *Contemporary Physics* 22:265 (1981).
7. H.C. Pfeiffer, *Scanning Electron Microscopy 1972*, p. 113 (IIT Research Institute, 1972).
8. R.E. Howard et al., *J. Vac. Sci. Tech.* B1:1101 (1983).
9. A.A. Iranmanesh, R.F.W. Pease, C.M. Horwitz, "A Direct Experimental Determination of the Lateral Extent of Scattering of Kilovolt Electrons in a Solid Target," Stanford Electronics Labs. Technical Report (1980).
10. T.H.P. Chang, *J. Vac. Sci. Tech.* 12:1257 (1975).
11. D.F. Kyser and N.S. Viswanathan, *J. Vac. Sci. Tech.* 12:1305 (1975).
12. J.P. Ballantyne, *J. Vac. Sci. Tech.* 12:1257 (1975).
13. L.F. Thompson, J.P. Ballantyne, E.D. Feit, *J. Vac. Sci. Tech.* 12:1280 (1975).
14. D.C. Shaver, unpublished communication.
15. See any college-level textbook on optics.
16. M.C. King, *IEEE Transactions on Electron Devices*, ED-26, 711 (1979).
17. H.C. Craighead et al., *J. Vac. Sci. Tech.*, B1:1186 (1983).
18. W.G. Oldham et al., *IEEE T-ED* 26:717 (1979).
19. P.D. Blais, Proc. Kodak Microelectronics Conference 1975 p. 6.
20. F.H. Dill, *IEEE Trans. Elec. Dev.*, ED-22:440 (1975).
21. F.H. Dill et al., *ibid* p. 445.
22. K.L. Konnerth and F.H. Dill, *ibid* p. 452.
23. F.H. Dill et al., *ibid* p. 456.
24. E.J. Walker, *ibid* p. 464.
25. M.M. O'Toole et al., Proc. of SPIE Volume 275, Semiconductor Microlithography VI (1981).
26. D.J. Kim, W.G. Oldham, and A.R. Neureuther, *IEEE Trans. Elec. Dev.*, ED-31:1730 (1984).
27. Proceedings from Electron Ion and Photon Beam Symposia for 1973, 1975, 1977, 1979, 1981, 1983, 1984 are published in the following respective issues of *J. Vac. Sci. Tech.*: 10 (Nov./Dec. 1973), 12 (Nov./Dec. 1975) 15 (May/June 1978), 16 (Nov./Dec. 1979), 19 (Nov./Dec. 1981), B1 (Oct./Dec. 1983, B3 (Jan./Mar. 1985).
28. J.M. Moran and D. Maydan, *J. Vac. Sci. Tech.* 16:1620 (1979).
29. J.M. Mikkelson et al., *ISSCC 81 Technical Digest*, p. 106 (1981).
30. See, for example S. Moriya et al., *J. Vac. Sci. Tech.* B1:990 (1983).
31. H.S. Yourke and E.V. Weber, *Technical Digest IEDM* 76:431 (1976).
32. D.E. Davis et al., *J. Vac. Sci. Tech.* B1:1003 (1983).
33. D. Maydan et al., *IEEE Trans. Electron Dev.* ED-22:429 (1975).
34. G.N. Taylor, T.M. Wolf, J.M. Moran, *J. Vac. Sci. Tech.* 19:872 (1981).
35. W.D. Grobman, *J. Vac. Sci. Tech.* B1:1257 (1983).
36. C.C. Fu and D.H. Dameron, *Electron Device Letters* EDL-5:398 (1984).
37. "Resist Characterization Using a Multichannel Development Rate Monitor." Paper presented by Perkin Elmer, Applied Optics Div., 7421 Orangewood Ave., Garden Grove, CA at Sixth International Technical Conference on Photopolymers, Ellenville (1982).

Metallization for VLSI Interconnect and Packaging

Paul S. Ho

*IBM Thomas J. Watson Research Center
Yorktown Heights, NY*

INTRODUCTION

Over the past several years, a significant fraction of the resources of the semiconductor industry has been focused on the development of systems with very large scale integration (VLSI). For hardware, this consists of primarily the development of device chips and packaging architecture. Significant progress has been made in both areas with great improvement in the circuit density and performance. Advances in chip development are illustrated in Figure 1 where one can see that an exponential increase in the circuit density for memory and processor chips has been achieved with time.¹ At this time, dynamic memories of 256k bit density are available on the market and plans for manufacturing memory chips of 1M bit density have been announced by several companies in the United States and Japan. Comparable progress has been made in the development of logic chips, such as microprocessors, which have more complex structures than the memory chips. This trend, if sustainable, indicates that development of the 4 M bit memory chip is to be expected in this decade. This will require the development of ultra-large-scale integration with minimum device dimensions in the micron or even submicron range.²

Such progress is achieved largely through continuing reduction in device dimensions. In the course of this development, the metallization structures of the device and packaging have evolved into one with a high degree of complexity. It is the objective of this chapter to discuss the implications of VLSI on metallization and some of the basic problems.

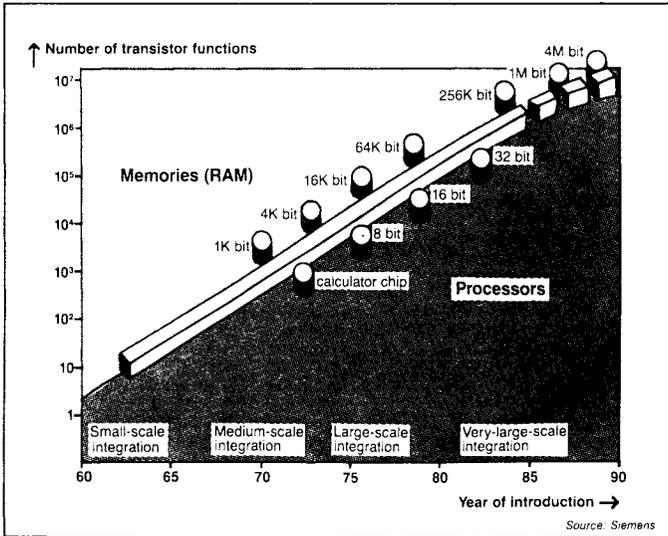


Figure 1: Progress in silicon chip technology since 1960 (Source: Siemens). Note that the degree of integration quadruples about every 3 years. While the development in microprocessors may show after the 32 bit level, the advance in dynamic memories remains constant (Reference 1).

Before discussing specific issues of metallization, it is instructive to illustrate the drastic change in the metallization structure by showing in Figure 2 two bipolar devices manufactured about a decade apart.³ The early device has only one active element in a chip with a dimension of about 2.5 mm by 2.5 mm. In contrast, the contacts in the recent chip have dimensions of about 3 micron by 3 micron. If the device density can be increased simply in accordance with the dimensions of the contacts, the number of devices for the latter can be increased about 300X300 times. Scaling of such magnitude was not accomplished because interconnections had to be provided to wire the devices. To fulfill this requirement, a very large portion of the area of this chip had to be used for the interconnect structure. As a result, this chip contains only about 1000 active elements which is about 100 times less than the density allowed by scaling the contact dimensions. Even with only 1000 elements, the overall structure required for this chip is highly complex, evolving from a simple single-level structure to one with a multilevel architecture.

In addition to device density, interconnect is important in determining the performance of the device chips. For example, by comparing the dimensions of the contact to that of the interconnect of the multilayer chip, one can see that its performance is probably not limited by the switching speed of the contacts, but rather by the time delay for signal propagation through the interconnect.

With the increased circuit density on the chip, more device functions can be assembled onto a single circuit board. While this provides a high

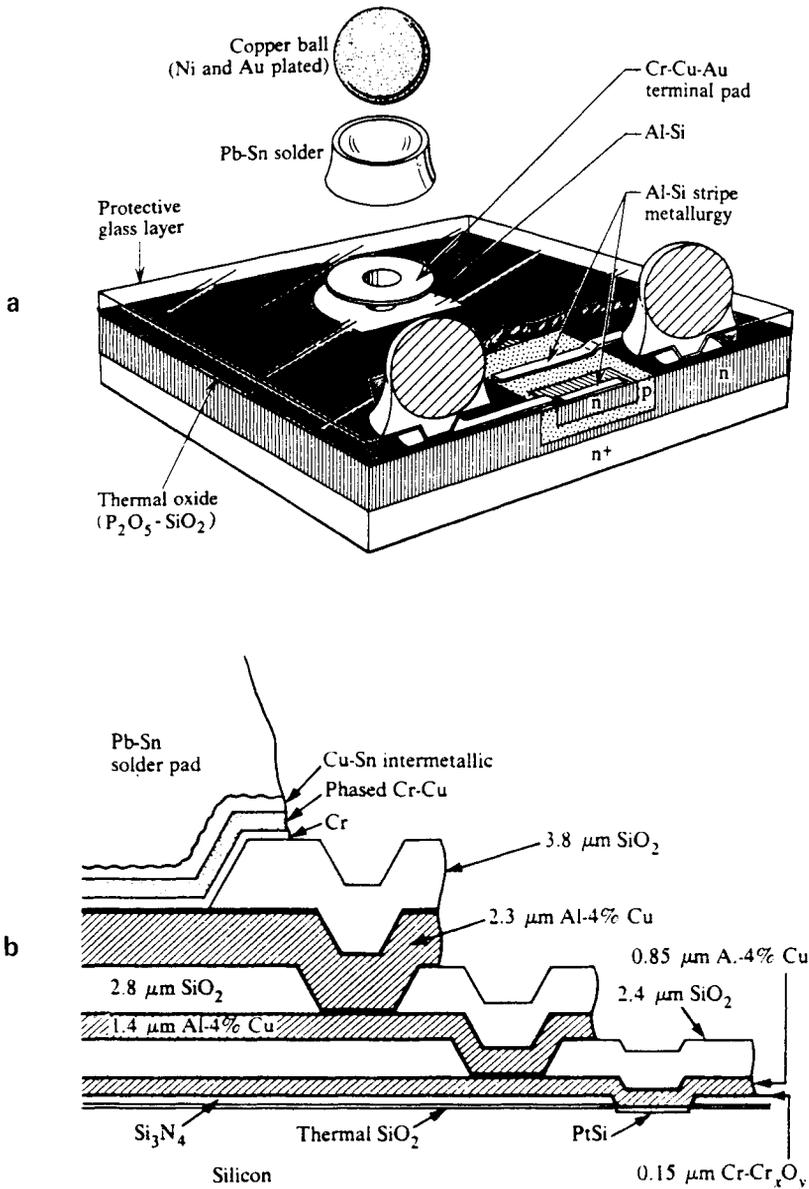


Figure 2: Schematics of interconnect metallization structure for (a) an early bipolar device and (b) a current advanced bipolar device. Note that the size of the chip in (a) containing one transistor is about 0.25 cm by 0.25 cm while the contact dimension in (b) is about 3 μm by 3 μm (Reference 3).

degree of flexibility for the functional design of the chip, to enhance the performance and the level of integration of the whole system, new packaging structures are required which can utilize the high density and performance of the device chips. This has brought forth significant improvements in the performance and level of integration of the packaging system. As indicated by the statistics in Figure 3, the wiring density in packaging has grown exponentially with time, but at a rate less steep than that of the device density.⁴ Although this has not been as well recognized in the past, it has become clear recently that packaging is an important issue, particularly for the high end computer systems where performance is the prevailing factor.

This can be illustrated by comparing the IBM 3033 and 3081 computer systems.⁵ As shown in Figure 4, the performance of the central processing unit, as measured by system cycle time, can be divided into chip and packaging portions. While the cycling time in the chip has been improved about 20%, the improvement in the packaging portion is about threefold. (The improvement in the circuit chips should not be measured by speed alone since there is significant enhancement in the circuit density which is not shown in Figure 4). The improvement is achieved primarily through a

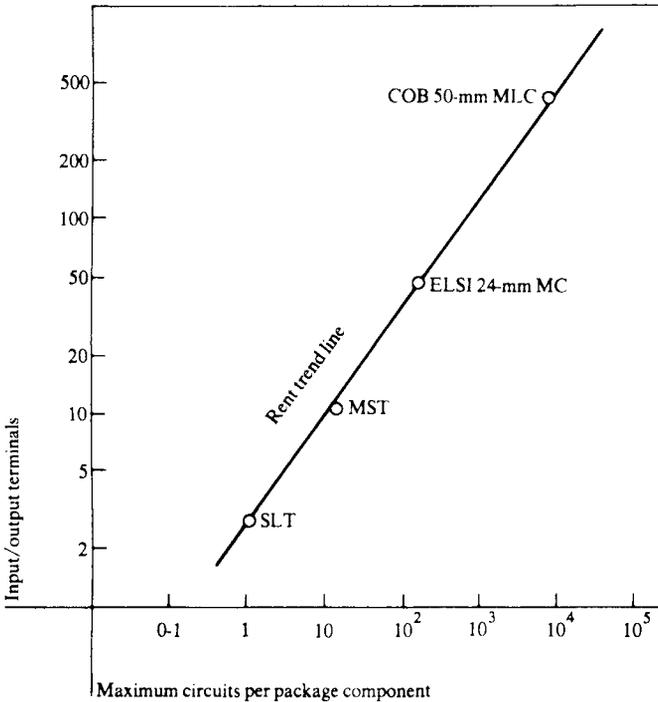


Figure 3: Progress in circuit density with time at the packaging module level. The abbreviations associated with the data points represent different versions of IBM packaging modules (Reference 4).

new design of the interconnect module and board in the packaging system. The distributions of the wiring structure at different levels of these two systems are compared in Table 1a. The new designs of the module and board provide a high level of integration which makes it possible to eliminate a large portion of the wiring cables in the 3033 system. This greatly simplifies the wiring structure and provides significant improvement in the reliability as well. It is instructive to compare the number of interconnects for the 3081 module and its equivalent in 3033 technology. As summarized in Table 1b, the high level of integration in the 3081 system makes possible a reduction of about tenfold in the number of interconnects.

Thus interconnect metallization is important in determining both the density and performance of the system. Since these are the general goals of VLSI, the optimization of the metallization structure becomes an essential part of the VLSI technology. Careful consideration should be given to not only the layout of the interconnect but also other aspects, such as material characteristics, processing reliability and manufacturing cost. As the trend continues toward device miniaturization, interconnect metallization will become even more important.

In this chapter, several basic aspects of VLSI interconnect and packaging will be discussed. It is divided into two parts: system requirements and material characteristics. The discussion of system requirements focuses on the wiring structure and the impact of device scaling. The wiring structure is a complex but fundamental issue arising from the need for interconnecting more devices when the device density increases. Device

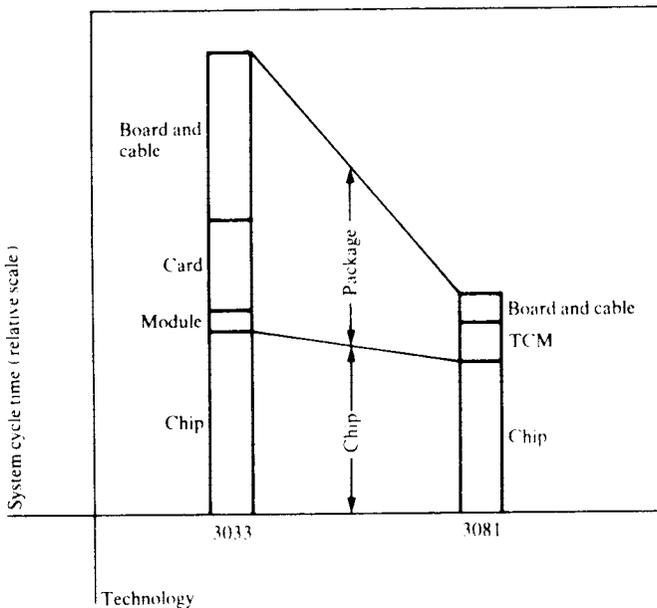


Figure 4: Comparison of system performance of the IBM 3033 and 3081 central processing units (Reference 14).

Table 1: Comparison of the IBM 3033 and 3081 Technologies

a. Percentage distribution of total wire length by packaging level (Ref. 5).

Packaging level	3033 Processor	3081 Processor
Chip	1.3	9.2
Module	—	41.9
Card	38.8	—
Board	12.9	31.1
Cable	<u>47.0</u>	<u>17.8</u>
	100%	100%

b. Average number of logic interconnections between packaging levels in two technologies (Ref. 14).

	3033 Technology equivalent	3081 Module
Chip-to-module	22,560	4368
Module-to-card	22,560	670 (no card)
Card-to-board	4,000	
Total	49,120	5038

scaling imposes certain requirements on the device structure and functions which have to be fulfilled by properly matching the materials and design of the interconnect structure. These two topics define the general requirements of interconnect metallization for VLSI. With the requirements clarified, the material characteristics of interconnect will be discussed, with emphasis placed on the thermal and electrical properties. This will be followed by discussions of two material reliability problems: contact resistance and electromigration. These problems are used to illustrate the impact of scaling on specific interconnect functions. Some of the current approaches to solve them will be indicated. It is not the intent of this article to review the different metallization schemes, nor specific device structures and processing developments. Some of these subjects are reviewed in other chapters and some have been discussed in the literature.^{6,7}

WIRING STRUCTURE

In Figure 2, one can see that the interconnect structure on a device chip serves four main functions: contact to junction and gate, interconnection between device cells, interlevel insulation, and input and output signal pads. In a computer system, interconnect structures have to be provided to connect the device chips to form a central processing unit as well as to interface with other functions, e.g. storage devices, printers and terminals. While it is beyond the scope of this chapter to discuss the overall issues of system integration, one can readily recognize that the design of the interconnect wiring structure is very complicated. The complexity originates from the combinational nature of interconnecting a very large ensemble of elements.⁸ This can be readily recognized using a simple estimate that for N elements, there are $N(N-1)/2$ possible binary connections. So, when the number of circuit elements increases beyond about 10k on a chip, the wiring becomes very complicated. Such an estimate overprojects the number of interconnects required since in practice devices are not wired randomly. Instead, the circuits are grouped in specific ways according to their functions and the placement of various types of circuits is optimized to reduce the number of interconnections. The nature of this problem has been investigated by treating it as a statistical optimization process of interconnecting a certain number of device cells subject to some overall layout of a functional block on a logic chip.^{9,10} The result follows the so-called "Rent's rule" which specifies the number of connections (or pins) P required to wire N device cells in a system by the following relationship.

$$P = aN^b \quad (1)$$

where a and b are parameters with a between 2 to 3 and b about $\frac{1}{2}$ to $\frac{2}{3}$.¹¹

This rule can be illustrated by the results obtained by Heller et al.¹⁰ Using a statistical simulation method, they calculated the wiring requirements for a logic chip as a function of the gate density. As shown in Figure 5, the number of wiring tracks required follows Rent's rule with b varying from about $\frac{1}{2}$ at low gate density to about $\frac{2}{3}$ at high gate density. This example shows the empirical nature of the Rent's rule since its parameters vary with the device density. The increase in b with gate density shows that the wiring requirements actually exceeds that predicted by Rent's rule based on low gate density.

As device density increases, there is another factor contributing to the increase in the wiring requirements. This is the increase in the average length of each connection. This altogether with the increase in the number of connections causes the total wire length to increase drastically, as seen from results obtained by Heller et al. (Figure 6). These results can be used to estimate semi-quantitatively the impact on the wiring requirements of a bipolar chip as the device density increases. For example, when the number of logic circuits increases from 100 to 1000, the number of connections required increases from 13 to 20 while the total wire length increases by a factor of about 100. Assuming the wire track width can be scaled proportionally according to the numbers of circuits, i.e. by a factor of

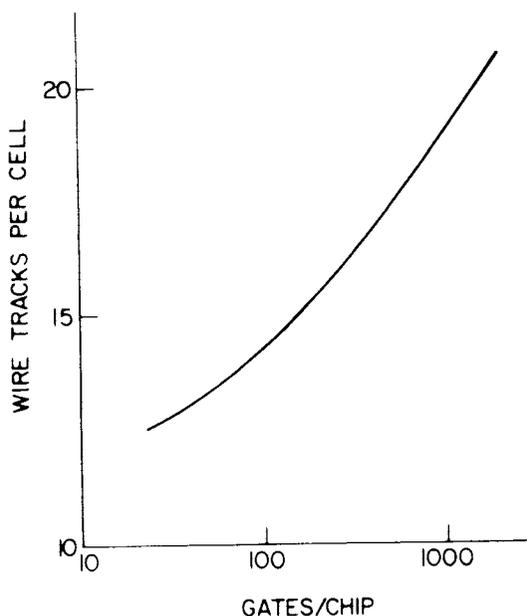


Figure 5: Results of a numerical analysis showing the "Rent's rule" for a logic chip. This rule correlates empirically the number of wire tracks required per cell to the device density on a chip (Reference 10).

$\sqrt{10}$, the total wiring area would still have to increase by a factor of $\sqrt{10}$. This indicates that as device scaling continues, the circuit layout and the chip becomes increasingly dominated by interconnect wiring. Eventually, it will become impossible to use all the circuits on a chip simply because some of them cannot be wired properly.

This can be seen from the statistics of a recent bipolar chip¹² with a surface area of 0.29 cm² and a gate count of about 1500. The total length required for wiring is 4m. With a wire channel width, line width plus line-to-line separation, of 6.5 μm , the total wiring area is 0.26 cm², which is about 90% of the surface area of the chip. Thus the structure of this chip is clearly dictated by the wiring requirements of the interconnect.

In the layout of the wiring structure, several factors are important to consider in order to optimize the device performance. First, the length of connection should be minimized. This is equivalent to minimizing the RC response time for optimum circuit performance. Second, the layout should minimize the cross talk in order to reduce the level of inductance noise coming from the random switching of individual circuits. Third, it is desirable to keep all the connections as close to the average length as possible. This reduces the random fluctuations of the switching voltage at the contact. Fourth, the placement of circuits should distribute the power dissipation evenly. This is to minimize the local heating, a problem particularly important for driver circuits. And finally, the layout should facilitate detection and correction of errors and defects.

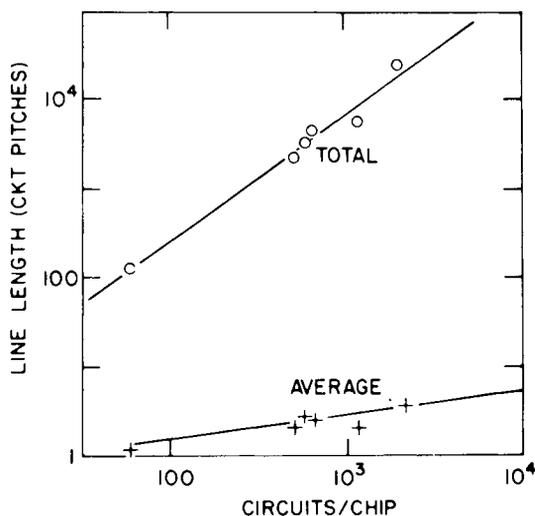


Figure 6: The average length of interconnect and the total wire length on a logic chip as a function of device density. The lengths are measured in circuit pitches, i.e. the square root of the area per circuit (Reference 12).

An effective approach in meeting these requirements is to employ multilevel interconnects. The structure of the multilevel chip in Figure 2 gives an example of a 3 level design. In this structure, all the wire tracks in one level run in one direction while those in the adjacent level run in a perpendicular direction. Such a structure provides a simple means to minimize the cross talk and the average length of the interconnects.

In the design of the multilevel structure, many material and processing requirements have to be considered in addition to wiring placement. For example, the number of levels can be reduced by packing as many lines as possible on one level. This demands high precision and good control of the patterning processes, such as lithography and metal etching. The use of multilevel interconnects can relax the wiring density requirements but causes problems in other processing areas, such as planarization and topography of interlayer contacts. The method of dealing with these problems often depends on the overall objective of the system design, e.g. performance or cost. These factors have to be optimized together with wiring placement.

The design of the wiring structure for the chip is only the first step to building the interconnections of the computer system. After the chip level, other levels of interconnections are required for assembling the chips into the central processor and for interfacing the processor storage and peripheral equipment. As shown schematically in Figure 7, the packaging system has to also provide mechanical support, power, cooling, etc. in order to assume proper function of the system.¹³ The complexity of the packaging system depends on the functional requirements of the computer. For high end systems where the design is driven by density and performance,

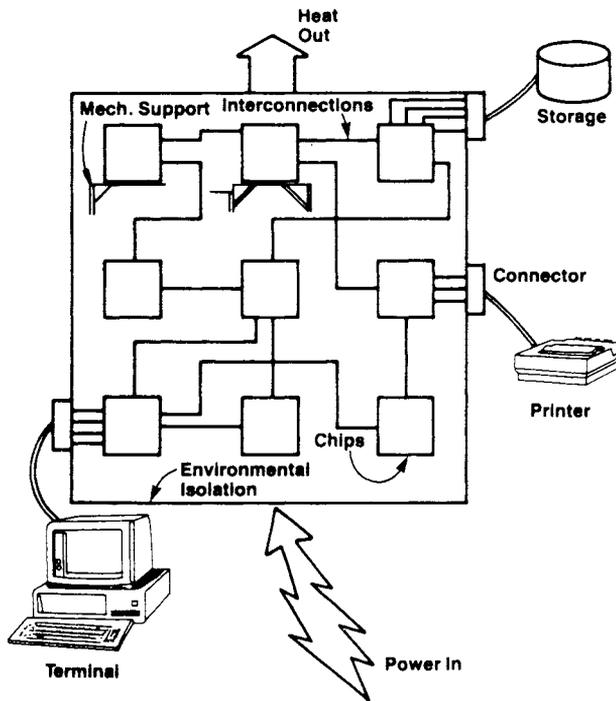


Figure 7: Schematic representation for the structure of a computer package. This structure contains thermal and mechanical supports for the chips as well as the high-speed interconnections that permit information exchange between chips and peripheral equipment (Reference 13).

packaging can become very complicated and usually requires a high level of integration and optimization. This can be illustrated by the packaging of the IBM 3081 system. This system contains two main levels of wiring structure: the module and the board. The module is designated as the thermal conduction module (TCM). Each TCM contains about 100 device chips, each of which has about 2000 circuits. It is built into a ceramic substrate of about 5.5 mm thick and 10 cm by 10 cm square (Figure 8). The wiring structure in the board is designed for each board to support 10 modules.¹⁵ It has a 20-layer structure containing about 6,000 connectives with dimensions of 60 cm by 70 cm. The wiring complexity of this board is comparable to that of the TCM.

It is interesting to estimate the wiring requirements for the module and the board on the basis of Rent's rule. Taking the parameter a to be 2.5 and b 0.6, the number of interconnects required for each chip is about 250 connections for the 2000 circuits. To support the 100 chips on one module requires 16 times the I/O connections of an individual chip. This turns out to be about 4,000 connections for each module. These requirements evolve the module into a structure with more than 30 interconnect layers

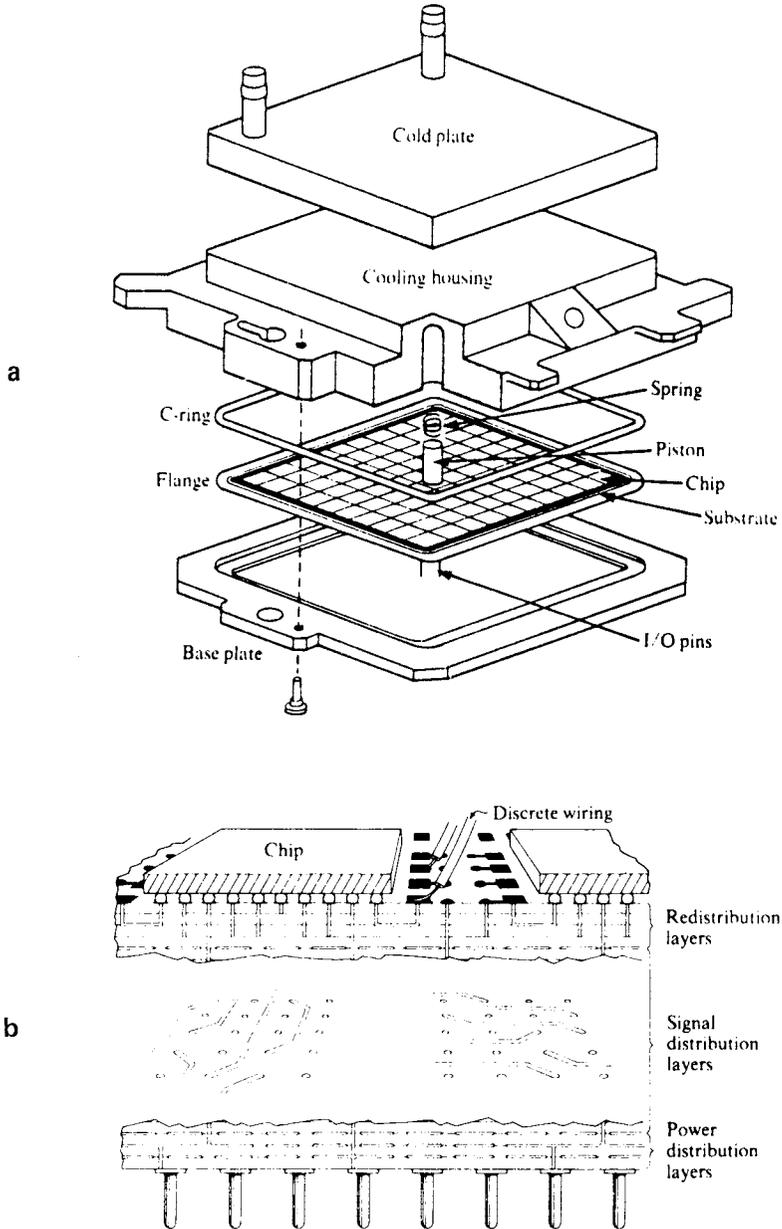


Figure 8: (a) Exploded view of the thermal conduction module assembly. (b) Schematic drawing showing the wiring structure of the multilayer ceramic substrate used in the thermal conduction module (Reference 14).

as shown in Figure 8. In order to optimize this structure, the layers are organized into three levels: the wiring redistribution layer, signal redistribution layers and power distribution layers. At the board level, to connect 10 such modules with a total of 40,000 interconnects onto one board requires about 600 connections according to the Rent's rule. Compared with Table 1 b, these estimates of 4000 interconnects per module and 600 per board are in reasonable agreement with the number of interconnections actually used.

The complexity of interconnect wiring necessitates the use of sophisticated computer-aided design techniques. It is beyond the scope of this article to review this important and fast developing area except to point out that advanced methods of statistical mechanics have recently been employed for wiring placement. For example, Kirkpatrick et al.¹⁶ employed a novel approach of simulated annealing based on the Monte Carlo technique. In such simulations, most of the interconnect requirements mentioned above can be incorporated as boundary constraints. This method can be applied to design wiring layout on a chip as well as interconnects between chips in a package module. An example of the application of this method is shown in Figure 9 where the length distributions for an interconnect structure calculated by various statistical methods are compared. The results obtained by simulated annealing has a narrower "peak" distribution, indicating the method is more efficient in comparison with other methods in placing interconnects with an average length.

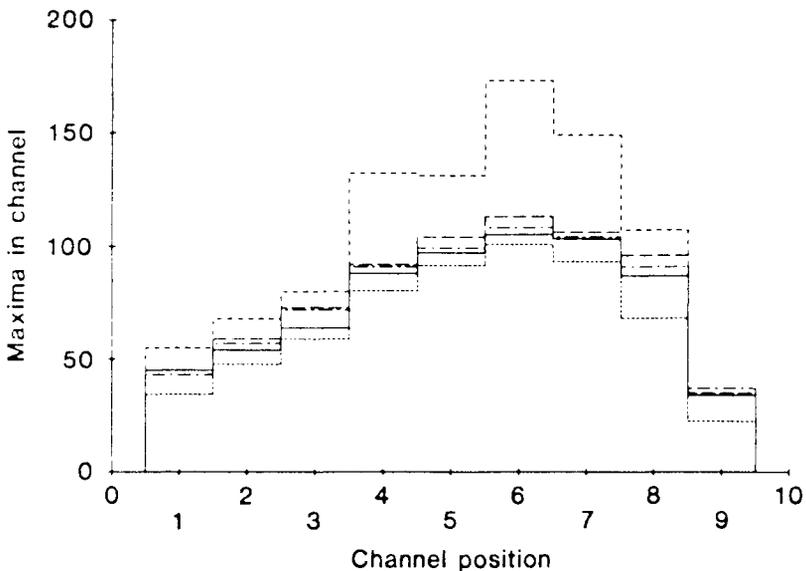


Figure 9: Histogram of the maximum wire densities within a given channel calculated using various methods of interconnect routing. The channel position provides a measure of the length of the interconnect. The top dash line represents the results obtained by the simulated annealing technique (Reference 16).

IMPACT OF DEVICE SCALING

In general, the aim of device scaling is twofold: first, to improve the device performance, particularly its speed and second, to increase the device density. These goals impose certain requirements on the function and dimension of various components in the device, which constitute the so-called scaling rules. The scaling rules differ for field effect transistors (FET) and bipolar devices because of their different operating principles. Based on a constant field approximation, Dennard et al.¹⁷ first derived a set of scaling rules for MOSFET devices by assuming all dimensions can be scaled linearly. They found that as the device dimensions were reduced by a factor of k , its performance as measured by the delay per circuit may be enhanced by the same factor if the dopant level increases by k while the voltage and current are reduced by k . On this basis, one can deduce the scaling factor for other parameters pertinent to interconnect metallization, such as the resistance, capacitance, current density, power dissipation and etc. These factors are listed in Table 2.

Table 2: Scaling Results for Some Device Parameters

Device Parameter	Scaling Factor	
	FET	Bipolar
Device dimension t_{ox} , L , W	$1/k$	$1/k$
Delay time/circuit	$1/k$	$1/k$
Doping concentration	k	k
Voltage	$1/k$	~ 1
Current	$1/k$	$1/k \sim 1$
Power density	1	$\sim k^{-k^2}$
Capacitance	$1/k$	$1/k$
Line resistance	k	k
Line response time	1	~ 1
Line current density	k	$\sim k^2$

Scaling of bipolar devices is more complicated because its vertical structure and operation principle do not allow similar linear scaling employed for FET devices. The wide spectrum of bipolar circuit designs adds additional complications to deducing a general set of scaling rules. Another parameter

relevant to the discussion of interconnect is the applied voltage. Since the switching voltage can not be less than the voltage drop across a p-n junction, which is usually about half of the bandgap of the semiconductor (e.g. about 0.6 eV for Si), scaling of the voltage is restricted. Indeed, the voltage level used in devices is usually set to be a constant, chosen as a standard by the industry. Based on constant-voltage scaling, Tang and Solomon¹⁸ employed device simulation techniques to analyze the effect of scaling on device parameters. They found that for an optimum power-delay product, the current density increases about linearly with the reduction in device dimensions. For I²L type circuits, Prince¹⁹ deduced separately a set of scaling rules based on a linear scaling of delay time. He found that the current density varies approximately with the square of the dimension. Results of these two studies lead to somewhat different scaling of device parameters which are summarized in Table II. There are certain limitations as to how far the scaling rules can be extrapolated before encountering difficulty in device processing or basic physics. For example, the minimum switching voltage in FET devices is limited by the noise level and voltage fluctuations generated by random switching and variations in line dimensions. This restricts the lower bound that the voltage level in an FET can be reduced. Usually, the gate voltage can not be substantially reduced below 0.5 Ev. This restricts the scaling of the voltage in FET devices to a level similar to that of the bipolar devices. Line thickness is another parameter that can not be reduced infinitely. Practical thickness is usually not less than about 5000 Å. There are additional limitations coming from basic device physics.²⁰ Such problems will affect the scaling rules, particularly for devices with dimensions in the submicron range.

In spite of these difficulties, the results summarized in Table II provide some general guidelines to examine the requirements imposed by device scaling on the interconnect functions. For FET devices, scaling of the device dimensions and circuit delay by k requires the doping concentration to increase by k while voltage and current decrease by k . On this basis, the response time of an isolated interconnect line will remain unchanged. (The implication of a constant line response time on the system performance and complications due to interference between conductor lines will be discussed in the next section.) In addition, the $1/k$ scaling of the voltage and current is balanced by the k^2 increase in device density. This results in a constant power density for the chip although the current density increases by k . Thus to a first order approximation, scaling of FET devices will not change the cooling requirements at the chip level in spite of the increase in the device density. It does require, however, the current-carrying capability of the line to be higher.

For bipolar devices, the fact that the voltage level cannot be scaled in accordance with the device dimension causes increases in power density and line current density exceed the scaling factor. Depending on the desired performance level, these two parameters can increase as much as k^2 , making the requirements for cooling and current-carrying capabilities significantly more severe than for FET devices. These factors plus the intrinsic complexity of the device structure make the development of an interconnect structure for bipolar devices considerably more difficult than for FETs.

To illustrate quantitatively the increasing demands due to device miniaturization, the values of several key device parameters are given in Table 3 for three stages of device development. These highlight several key areas for future concern, including the formation of shallow contacts with high dopant concentrations, the reliability of very thin gate insulators, the development of conductor lines with high electromigration resistance and the design of efficient cooling structures. While some of these problems will be discussed in the following sections, it becomes clear that scaling brings forth a set of requirements in addition to those from wiring placement for the design of VLSI interconnects. In general it is difficult to design a structure optimizing all these factors. A practical approach is to choose a certain combination of metallization and structure to optimize some designed functions of the system. For example, for bipolar logic circuits where speed and power dissipation are important, the metallization should be of low resistance and with reasonable cross-sectional area to reduce the current density and time delay. In contrast, for FET memory chips where density and reliability are important, the line dimensions should be minimized with the possibility of using thermally stable but resistive refractory metals.

Table 3: Scaling Trends in Device Dimensions

Period	Line	Junction	Oxide	Current Density	
	Width	Depth	Thickness	(10^5 A/cm^2)	
	(μm)	(μm)	\AA	FET	Bipolar
1970's	3-6	~ 1	~ 1000	0.2-0.5	0.5-1
1980-85	1.5-3	~ 0.25 - $.5$	250-500	0.5-1	2-4
Late 1980's	0.5-1.5	$\sim .1$ - $.25$	100-250	1-2	8-15

ELECTRICAL CHARACTERISTICS

When a signal pulse is transmitted through a conductor line, a certain amount of time is required for the signal level to rise from zero to its original magnitude. This rise time corresponds to the wiring delay and can be calculated by treating the conductor as a transmission line. For the usual situation of long pulses, i.e. pulse lengths longer or at least comparable to the wire lengths, the rise time equals 2.3 times the RC constant, where R and C are the resistance and capacitance of the wire respectively. The RC constant can be expressed as follows

$$R = \frac{\rho L}{W t_m},$$

$$C = \frac{\epsilon WL}{t_{ox}}$$

$$\text{and} \quad RC = \frac{\rho \cdot \epsilon \cdot L^2}{t_m \cdot t_{ox}} \quad (2)$$

where ρ is the resistivity, L the wire length, W the wire width, t_m thickness of the wire, ϵ the permittivity, and t_{ox} the thickness of the oxide. Equation 2 shows that if t_m and t_{ox} can be scaled in the same manner as L , the line delay would remain constant (see Table 2 also). Therefore, the line delay becomes an increasingly larger portion of the total circuit delay as the device dimension decreases. Eventually it can become a substantial part or even dominate the system response time for very small dimensions. In practice, linear scaling is difficult to accomplish for devices with submicron dimensions since t_m and t_{ox} are usually limited to about $0.5 \mu\text{m}$ and 100\AA respectively. In addition, L of all the interconnects does not scale uniformly as we discussed previously in the wiring placement section.

In general, the effect of the line delay will become a problem when the minimum dimension reaches below about $2 \mu\text{m}$. The impact is less for FET memory circuits than for bipolar logic circuits because of the bipolar circuits more complex wiring structure. For the 3-level bipolar chip with about $2 \mu\text{m}$ minimum dimension shown in Figure 2, the wiring delay constitutes a significant portion of the system processing time. In addition, for VLSI applications, the chip dimension is usually enlarged in order to accommodate the high device density which, when combined with a more complex wiring structure, results in broadening of the overall length distribution of the interconnects. This widens the distribution in the RC time constants with substantial increases for the portion of the wires with long lengths.

To minimize the RC delay, it is usually important to use materials with low resistivity and permittivity to build the interconnect structure. In Tables 4 and 5 are summarized some of the physical properties for the commonly used conductor and insulator materials.^{21,22} (The resistivities for thin films of these metals are not given here since they depend in general on various parameters, e.g. method of deposition and grain structure. They are about 20-40% higher for pure and large grain films of Al and noble metals but can be 2 to 3 times higher for impure refractory metal films.) Because of their excellent conductivities, it is clear why Al, Au and Cu are the most commonly used metals. However, in certain applications, because of processing requirements (e.g. the annealing temperature) or device design (e.g. the high-density self-aligned polycide gate), materials of lesser conductivities, such as refractory metals, silicides and even highly doped polycrystalline silicon are used.

Table 4: Selected Properties of Metals
(Reference 21)

Metal	Melting Point(°C)	Bulk Electrical Resistivity ($10^{-8}\Omega\text{-m}$)	Coefficient of Thermal Expansion ($10^{-6}/^{\circ}\text{C}$)	Thermal Conductivity (W/mK)
Ag	960	1.6	19.7	418.4
Al	660	2.65	23.6	220
Au	1063	2.2	14.2	297.06
Cu	1083	1.7	17.0	393.29
Pd	1552	10.8	11.0	71.13
Pt	1774	10.6	9.0	71.13
Mo	2625	5.2	5.0	146.44
W	3415	5.5	4.5	200.83
Ni	1455	6.8	13.3	92.05
Cr	1900	20	6.3	66.94

Table 5: Selected Properties of Insulators
(Reference 22)

Materials	Dielectric Constant†	Thermal Expansion ($10^{-6}/^{\circ}\text{C}$)	Strength (KPSI)	Thermal Conductivity (CGS)
96% Al_2O_3	9.3	6.4	46.0	0.06
92% Al_2O_3	8.5	6.5	48.0	0.04
Electrical Porcelain	5.5	4.4	13.0	0.004
SiO_2	3.8	0.6	8.0	0.005
Si_3N_4	6.0	3.0	85.0	0.08
Al N[8]	8.8	4.5	53.3	0.02
Glass-Ceramic [7]	7.5	4.2	42.6	0.01

† Dielectric constant is defined as the ratio of permittivity.

For the insulators, SiO_2 is the universal material used to form the dielectric layer on circuit chips. It has a low dielectric constant of 3.5 and can be produced with extremely low defect density by oxidizing the Si substrate to a thickness as small as 100-200Å. This makes it well suited for gate insulator applications although for submicron devices, there is some question regarding the integrity of SiO_2 as a gate insulator for thicknesses below 100Å.²³ For interlevel insulation, silicon oxide up to 2 μm thick produced by sputtering or evaporation (often not of the exact SiO_2 stoichiometry) is often employed. Si_3N_4 is frequently used in combination with SiO_2 in spite of its high dielectric constants. Its excellent mechanical strength makes it well suited to serve as a lithographic masking material. Examples of the $\text{Si}_3\text{N}_4/\text{SiO}_2$ combined layer can be seen in Figure 2.

For packaging applications, ceramics formed by various combinations of oxides, particularly Al oxides and Si oxides, are common materials. For example, the multilayer ceramic module shown in Figure 8 employs materials of several oxide mixtures. The ceramic materials have relatively high dielectric constants (about 7-8) and have to be processed at elevated temperatures (above 1500°C). Special, and often complex, processes have to be developed for the application of this type of material for forming multilayer structures. For example, high temperature processing necessitates the use of metals with high melting points, such as the refractory metals. This increases the response time of the system due to the high resistivity of these metals. To circumvent these difficulties, polymeric materials are being considered to replace the ceramics in chips as well as in packaging.²⁴ The main advantages of this class of materials are the low dielectric constant (about 3.5, similar to SiO_2) and the low processing temperature (usually below 400°C). High-temperature polymers such as Polyimides are used to satisfy processing requirements where thermal stability up to 400°C is required.

Some of the problems relating to wiring delays for VLSI applications have been discussed by McGreiv, ²⁵ He has considered the change of the access time for a static NMOS (N channel) RAM with decreasing design rules. His results are shown in Figure 10. The access time decreases continuously with shrinking device dimensions down to about 1.5 to 2 μm . Below that, the effect due to the RC delay of the interconnect becomes observable and its magnitude depends on the resistivity of the material used. For refractory metal gates with sheet resistivities of 1 ohm per square cm (sheet resistivity equals ρ/t_m), a decrease in the access time can still be achieved below 1.5 μm although the gain is very small. The access time is doubled what one would expect to achieve in an ideal scaling model. For resistive polysilicon gates with 20 ohm per square cm sheet resistivity, corresponding to a 1 μm thick gate with 200 ohm-cm resistivity. The access time increases with decreasing geometry due to the RC delay. The access time becomes about an order of magnitude more than the ideal case in the submicron range.

He has also considered the problem of parasitic capacitance. His results for the variation of the capacitance components of interconnect as a function of line width are shown in Figure 11. Of the three capacitance components, only the metal-to-substrate capacitance, C_{ms} , decreases

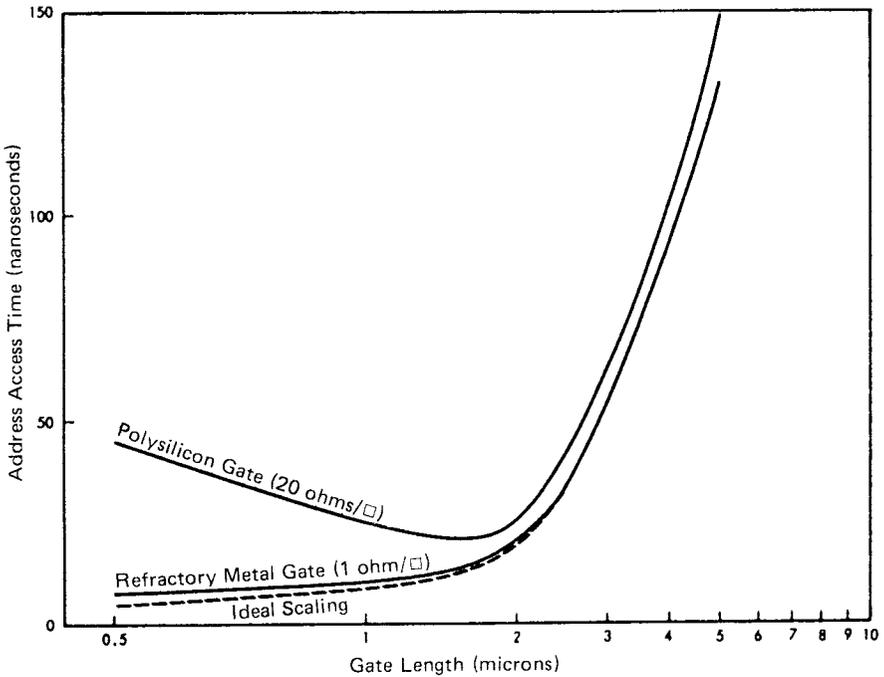


Figure 10: Variation in the access time of a 4k NMOS RAM with decreasing design rule for gate interconnects with different resistivity (Reference 25).

with scaling. The edge capacitance is relatively constant while the parasitic capacitance C_{mm} increases with decreasing line pitch. The net effect is for the total capacitance to increase in the submicron range, similar to the resistance component although the effect is less. The end result is an overall increase of the system access time.

There are other electrical problems related to wiring interconnects. The broadening of the distribution in wire length mentioned already is particularly important. The presence of a fraction of long wires will increase the overall time delay. The length variation also introduces nonuniform RC response time. This disturbs the switch synchronization of the system, which can be a particularly difficult problem for bipolar logic circuits. As line pitch decreases, inductance effect due to switching of adjacent lines increases. This generates random voltage noise which can affect device switching. Again, the problem is more serious for bipolar logic circuits. Most of these problems can be reduced by using multilayer structures with wiring running orthogonally in adjacent layers. This approach is particularly effective if combined with a statistically optimized wiring placement design, such as was mentioned previously. Multilayer structures, however, require long and complex processing steps, so the advantages will have to be balanced against cost effectiveness and the overall design objectives.

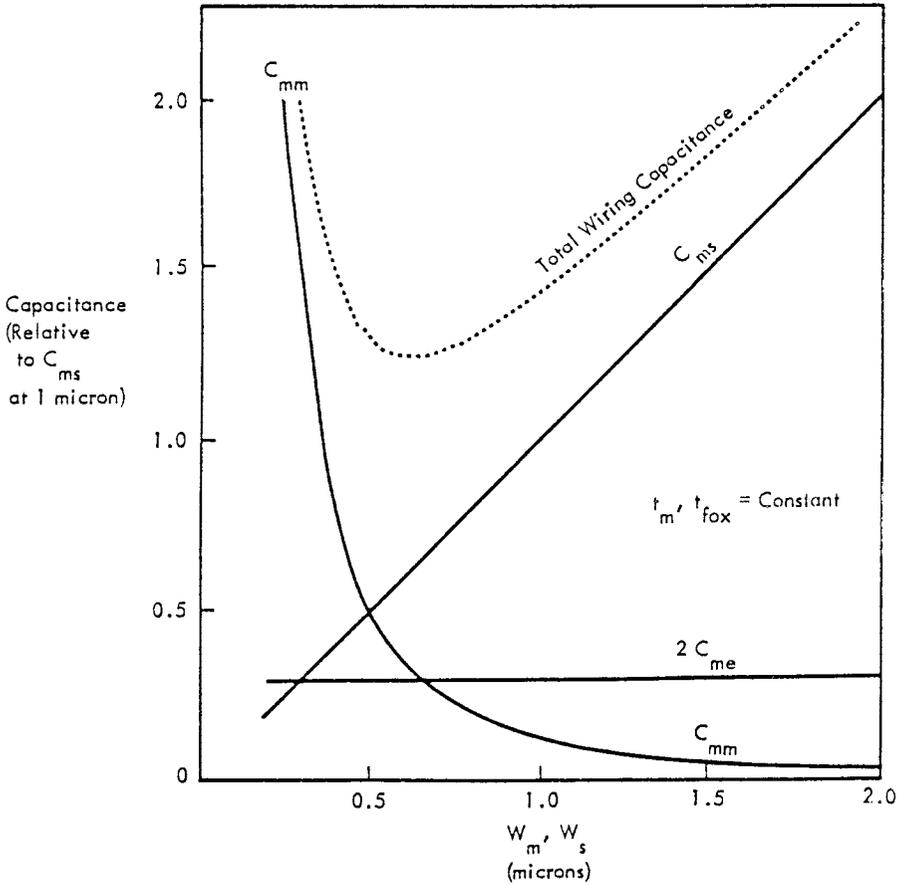
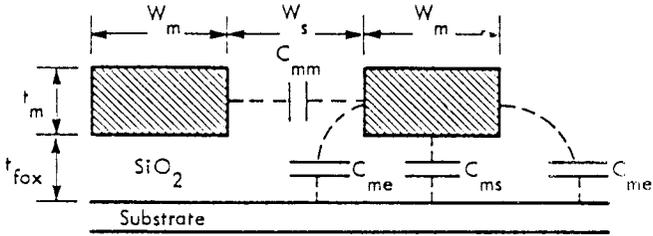


Figure 11: Variation of components of interconnect capacitance with design rules. The various components are defined in the upper figure (Reference 25).

MATERIAL REACTION

In the multilayer structure, various materials are integrated to serve the designed function of the device. Interfaces are formed with different combinations of materials, such as metal-metal, metal-insulator, metal-semiconductor and semiconductor-insulator interfaces. In addition to the interfaces, there are structural defects in thin films, such as grain boundaries and dislocations. The close proximity of layers, the different material types and the presence of structural defects are all factors contributing to mass transport in the device structure. During processing or operation of the device, atomic transport is further enhanced by the elevated temperature as well as by the external driving forces, such as the applied voltage or the current density. The resultant material reactions often change the device characteristics, giving rise to reliability problems.

The concern for device reliability will become more important as device dimensions are reduced and more complex multilayer structures are used. For example, a reduction in the vertical dimension of interlevel thickness will increase interfacial reaction which can result in junction penetration as well as a change in the dopant profile at the contact. Discussion in this section focuses on the characteristics of materials reactions in multilayer structures, emphasizing the roles of atomic mobility and driving force as well as damage formation induced by flux divergence.

In a multicomponent solid, the atomic flux can be expressed generally as

$$J_i = c_i \frac{D_i}{kT} F_i \quad (3)$$

where c_i is the mobile concentration of the i th element characterized by a diffusivity D_i at temperature T . The driving force F_i is derived from the gradient of the chemical potential μ_i as

$$F_i = -\nabla \mu_i \quad (4)$$

The chemical potential μ_i can be expressed as

$$\mu_i = kT \ln c_i + \mu_i(c_i) + \sigma_i v_i + q_i \phi_i + \dots$$

where μ_i is written as originating from the concentration c_i , the internal chemical free energy $\mu_i(c_i)$, and other external contributions, such as those from the stress σ_i and electric potential ϕ_i ; v_i and q_i are the atomic volume and charge respectively. Consequently,

$$F_i = \frac{kT}{c_i} \nabla c_i - \nabla \{\mu_i(c_i)\} - \sigma_i \epsilon_i - q_i E_i \quad (5)$$

where ϵ_i and E_i represent the deformation strain and the electric field respectively.

In this form, the driving force in a multicomponent system can originate

from three types of sources. The first source is related to the concentration gradient which is generally recognized as the diffusion term. The second force comes from the internal chemical energy gradient representing the driving force associated with the change in the chemical form of the i th element, e.g. the compound phase or the composition in a concentrated alloy. The last type relates to external constraints such as an applied stress σ_i or an electric field E_p .

Combining Equations 3 and 5, it is clear that the existence of an atomic flux requires not only that the atoms move, i.e., $D_i \neq 0$, but also that there is a nonvanishing driving force. Thus the role of the driving force is as important as the diffusivity although the present discussion will address mainly the diffusivity issue since the variety of driving forces makes a systematic discussion of the role of the forces difficult. (One exception is the later discussion on electromigration.)

The existence of an atomic flux by itself is not sufficient to cause damage formation. In order for that to occur, a local depletion or accumulation of materials is required. This condition can be expressed by the flux continuity equation as

$$\frac{dc_i}{dt} = -\nabla \cdot J_i + \frac{c_i - c_i^0}{\tau} \quad (6)$$

where the rate of accumulation or depletion equals the negative flux divergence plus the rate of dissipating the excess concentration $c_i - c_i^0$ from equilibrium (τ is the time constant for the process). In a bulk solid, the flux divergence is usually generated by gradients of macroscopic parameters, such as temperature and stress. In a multilayered device structure, it can originate from two other types of inhomogeneities. One is associated with the interfaces where two different materials join, and the other is due to the presence of microstructural defects, e.g. grain boundaries and dislocations. These two types of structural inhomogeneities provide high diffusivity paths for mass transport which often give rise to flux divergence. Grain boundary triple points or abrupt changes in grain size are well-known examples of structural inhomogeneities. (See, for example, the discussion in Reference 26.) For multilayered devices, such defects are usually more important than the macroscopic ones since their effect on atomic transport is significantly higher. Structural imperfections will become even more important as further device scaling produces steeper gradients in the structure as well as closer proximity of the interfaces.

The characteristics of diffusion in thin films have been reviewed by Balluffi and Blakely²⁷ and Gupta and Ho²⁸. The diffusivities via various defect structures can be summarized as a function of a homologous temperature T/T_m , where T_m is the absolute melting point, as shown in Figure 12²⁹. Using this data, the relative contributions of different diffusion processes in thin films can be estimated. For metallic films in the device operating temperature range of 0.3-0.6 T/T_m , diffusion is dominated by grain boundaries and dislocations instead of lattice defects (Figure 13). For nominal films with grain size L in the 1 μm range, i.e. $\log(1/L) \approx 4$, grain boundaries usually dominate.

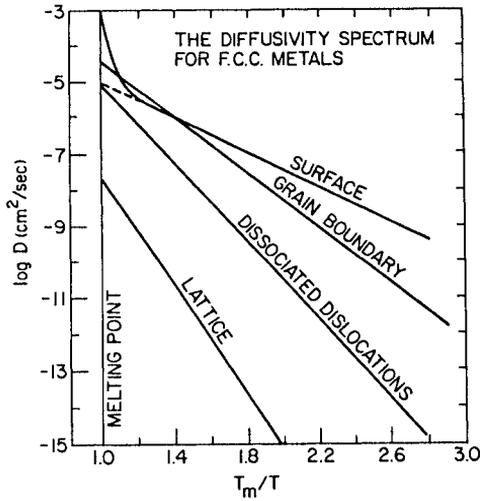


Figure 12: Summary of diffusivities via various types of structural defects. The temperature scale is normalized to the absolute melting temperature T_m (Reference 29).

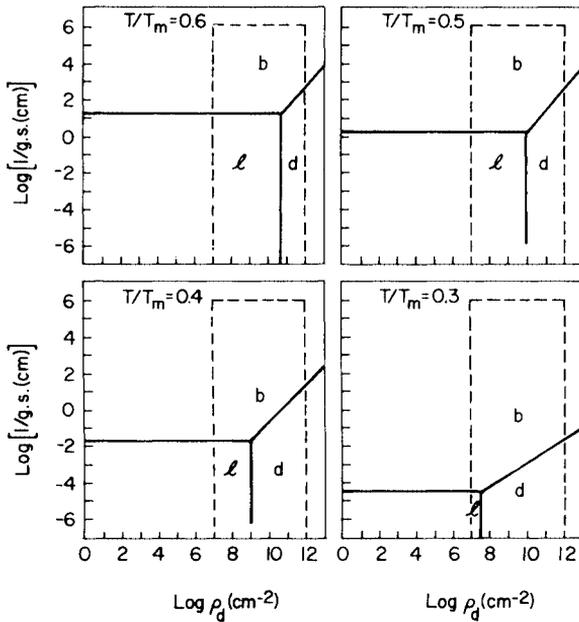


Figure 13: Regimes of grain size (g.s.) and dislocation density ρ_d over which lattice diffusion (l), grain boundary diffusion (b), or dislocation diffusion (d) is dominant. The calculation is based on steady-state diffusion through a thin film specimen of an fcc metal as a function of the homologous temperature (T/T_m) (Reference 27).

The amount of grain boundary diffusion can be estimated from the diffusion distance $(D_b t)^{1/2}$, where D_b is the grain boundary diffusivity and t is the time. Known values of D_b have been summarized for metal films in Figure 14.³⁰ Between $0.3T/T_m$ and $0.6T/T_m$, D_b is approximately 10^{-15} – 10^{-10} cm^2s^{-1} , taking the thickness of the grain boundary δ to be 10^{-7} cm. Accordingly, for a period of 10^3 sec., the diffusion distance in the boundary is in the range of .01 to 1 μm . At higher temperatures, such as those used in device processing, about 400°C , the diffusion distance can be one to several orders of magnitude higher. Comparing these diffusion distances with the dimensions of submicron device structures, the diffusion distance as estimated shows that atomic transport via structural imperfections in about 10 min. is sufficient to cause material reaction throughout the device structure. Thus it is clear that the high rate of diffusion is a basic reliability problem in VLSI technology.

This has led to the use of diffusion barriers.³¹ The general approaches include the use of single-crystal or epitaxial films, the use of amorphous films, stuffing the boundaries with impurities or second phase particles, and by imposing a barrier with a high activation energy for diffusion, e.g. nitrides, carbides or oxides. Except for the last approach which is close to a true diffusion barrier, the others often achieve the results by changing the nature of the structural defects instead of reducing the intrinsic rate of the diffusion process. The use of refractory metallic compound, e.g. TiN, will be described later as an example of a barrier for reducing material reaction at junction contacts.

The correlation between diffusivities and the absolute melting point shown in Figure 12 provides a useful calibration for the relative thermal stabilities of various materials used in interconnect structures. For example, comparing Al (T_m , 933°K) and refractory metals (e.g., T_m of Mo, 2983°K), equivalent stability at $0.5 T_m$ corresponds to 200°C for Al, but about 1200°C for Mo. For device structures requiring high-temperature reliability, the refractory metals are significantly superior than Al and are often used in spite of their high resistivity.

METALLIZATION RELIABILITY

Junction and Gate Contacts

The scaling results in Table 2 indicate the impact on junction properties is to increase the contact resistance while reducing the junction depth. This problem affects both Schottky diodes and ohmic contacts although these two types of junctions have different reliability requirements because of their specific circuit functions. The problem of contact resistance has been investigated by several groups³² and the results are shown in Figure 15. The variation of the specific contact resistance R_C with a dopant concentration, N_D , is comprised of two regions: the charge transport is mainly controlled by thermionic emission for N_D less than 10^{19} per cm^3 , but by tunneling for N_D more than 10^{19} per cm^3 . To reduce the contact resistance, it is essential to operate in the tunneling region which requires high dopant concentrations.

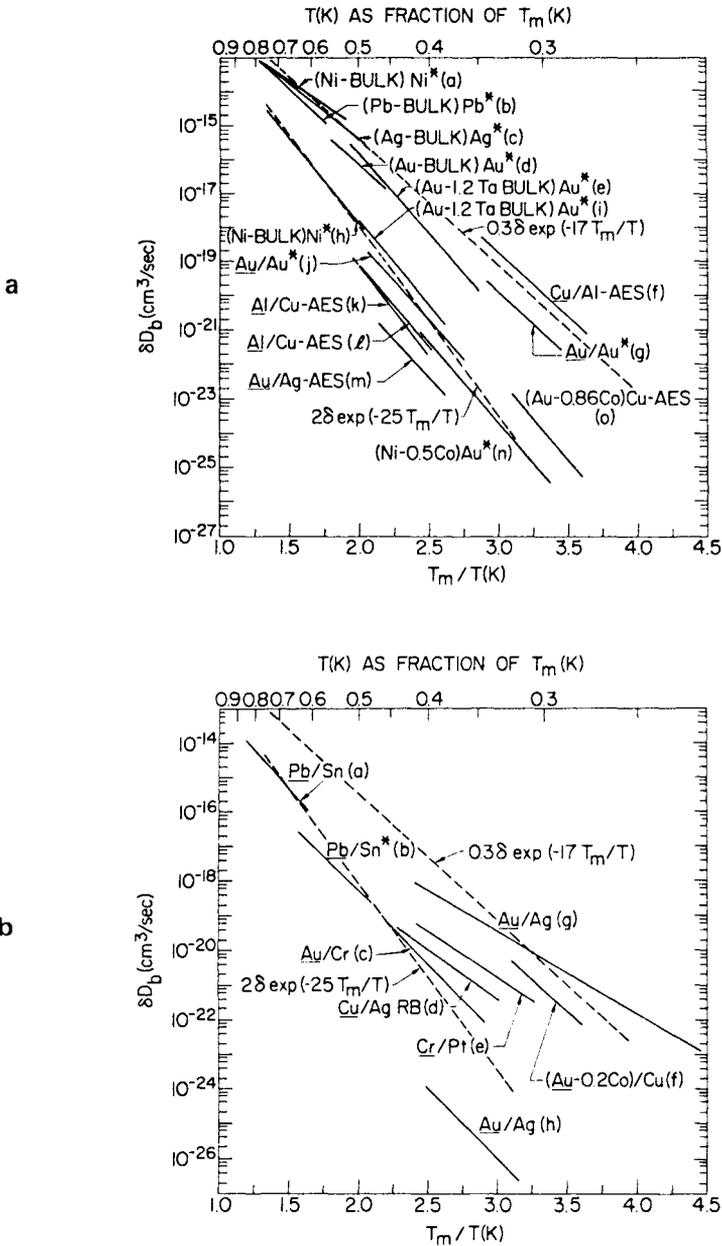


Figure 14: Plot of the grain boundary diffusivity against the reciprocal normalized temperature T_m/T for (a) data obtained by sectioning techniques in thin films and some bulk materials and (b) data obtained by permeation techniques in thin films (Reference 30).

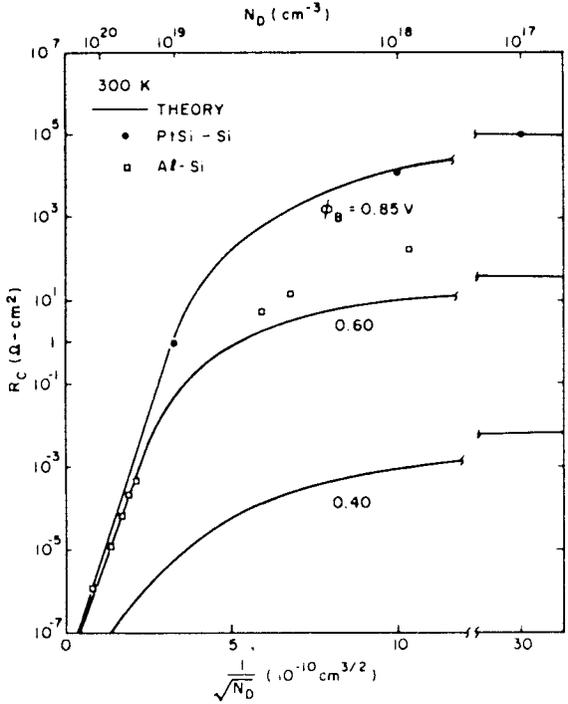


Figure 15: Theoretical and experimental values of specific contact resistance. Note that for doping level exceeding 10^{19} per cm^3 , R_c is dominated by the tunneling process while for doping level of 10^{17} per cm^3 , thermionic emission dominates and R_c becomes constant (Reference 32).

To see the effect of scaling, consider a typical contact of PtSi on n-type Si. Even for a high dopant concentration of 10^{19} cm^{-3} where conduction is dominated by tunneling, the specific contact resistivity is about $10^{-6} \Omega\text{-cm}^2$. For a contact of $3 \mu\text{m} \times 3 \mu\text{m}$ dimension, the contact resistance is about 10Ω . It will increase to 100Ω when the contact size is reduced to $1 \mu\text{m} \times 1 \mu\text{m}$. An increase in resistance of this magnitude can degrade the switching time and reduce the voltage swing to affect the normal device operation. Since the increase is nonlinear with scaling (see Table 2), the problem becomes even more significant for submicron junctions. The usual remedy for the problem is to increase the dopant concentration to ensure conduction by the tunneling mechanism. For example, when the doping of the PtSi contact is increased to 10^{20} cm^{-3} , the resistivity is reduced by about a factor of about 10^5 . Contact junctions with such high dopant concentrations have many material problems, such as the stability of the dopant profile during processing and interfacial resistance due to contamination. Thus one can readily understand the extensive recent interest in studying the formation of shallow contacts with high dopant concentrations.

The concern for the selection of a metallization scheme extends

beyond the initial formation of shallow contacts. Particularly important in this regard are the thermal stability of the contact and the retention of shallow junction profiles during processing. It has been recognized that the usual method of using Al to form junction contacts has severe limitations for VLSI applications. The difficulty arises from Al penetration of the junction as a result of the formation of Al spikes caused by Si dissolution into Al³³. This is a particularly significant problem for shallow contacts since it induces leakage current which, in severe cases, can short the junction electrically. The incorporation of Si into Al to alleviate junction penetration proves to be ineffective for shallow contacts since the precipitation of Al doped Si at the contact interface gives rise to high contact resistance.³⁴

These problems led to the use of silicides to prevent Al penetration of junction contacts. Silicides are compound phases usually formed by reacting a deposited layer of transition metal with Si.³⁵ Silicides have two important advantages for forming junction contacts. First, the formation of a silicide establishes a silicide-silicon interface below the original metal-silicon interface, thus producing a junction interface free of processing defects and contamination. Second, the amount of silicon consumed can be limited by restricting the annealing time in the formation of silicide contacts to shallow implanted junctions.

Silicides have several other material characteristics useful for integrated circuit applications.^{36,37} First, some silicides have electrical conductivities exceeding heavily doped polysilicon by several orders of magnitude. The resistivities of silicides formed on polysilicon³⁸ are summarized in Table 6. Second, it can be oxidized. And third, different metals form silicide Schottky diodes with barrier height varying from 0.5 to 0.9 eV. The combination of these material characteristics extends the usage of silicides beyond contact metallurgy into other VLSI applications. Notably among these are the use of silicides for interconnecting lines, the formation of low and high barrier diodes in Schottky transistor logic circuits and in combination with polysilicon for self-aligned bipolar contacts and MOS gates.³⁹

The use of silicides is not free of metallurgical problems. Particularly relevant are the thermal degradation of the silicide contact^{40,41} and reliability problems associated with stress generated due to silicide formation.⁴² Thermal degradation is a particular concern for the Al/silicide/Si structure when subjected to post-metallization annealing. This step is often required for removing process-induced damage in the insulator or during deposition of passivation layers. Depending on the process, the temperature can exceed 500°C for an hour or longer. The basic problem which has been recognized for some time⁴⁰ is caused by the phase instability of silicide in contact with Al.⁴¹ The dissociation of silicide is driven by the formation of an Al-transition metal compound with a higher thermodynamic stability than the silicide. This precipitates the dissociated Si (p-doped by Al) on the silicon substrate, causing changes in the junction characteristics. The sequence of steps in the degradation of the Al/Pd₂Si/Si junction is illustrated schematically in Figure 16. This mechanism of junction degradation is similar to the degradation of Al(Si)-Si contacts. Thus, it appears that the silicide serves only as a "sacrificial" barrier to delay the reaction between

Table 6: Resistivities of Various Silicides Formed on Polysilicon (Reference 38)

Silicide	Method of Formation	Sintering Temperature (°C)	Resultant Resistivity ($\mu\Omega\text{-cm}$)
TiSi ₂	Metal on Polysilicon	900	13-16
	Cosputtered Alloy		25
ZrSi ₂	Metal on Polysilicon		35-40
HfSi ₂	"		45-50
VSi ₂	"		50-55
NbSi ₂	"		50
TaSi ₂	"	900, 1000	60-70, 35-40
	Cosputtered Alloy	900 or 1000	50-55
CrSi ₂	Metal on Polysilicon	700	-600
MoSi ₂	Cosputtered Alloy	1000	-100
WSi ₂	"	1000	-70
FeSi	Metal on Polysilicon	500	150-200
FeSi ₂	"	700	>1000
CoSi ₂	Cosputtered Alloy		25
NiSi ₂	Metal on Polysilicon	900	-50
	Cosputtered Alloy		-50-60
PtSi	Metal on Polysilicon	800	28-35

Al and Si, leaving the nature of junction degradation unchanged. The thermal stability of the Al/silicide/Si system has been reviewed recently by Wittmer.⁴³

Ironically, this necessitates the use of another barrier to protect the silicide layer from reacting with Al. Two approaches have proven to be effective. One is to incorporate a thermally stable intermediate layer, such as TiW⁴⁴ or TiN⁴⁵ while the other is to use phase separation to form the barrier layer during silicide formation.⁴⁶ For example, if a PdW mixed layer is used to form the contact, Pd₂Si will form at the contact upon annealing due to its faster reaction kinetics and W will separate out to form a protective layer. In addition to being a single-step process, the consumption

SCHEMATIC PRESENTATION OF
Al/Pd₂Si/Si REACTION

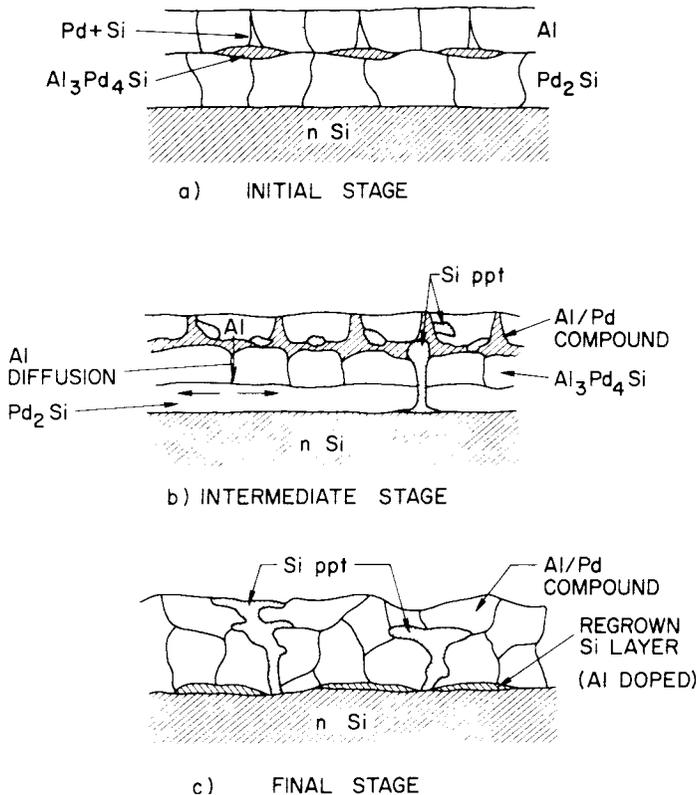
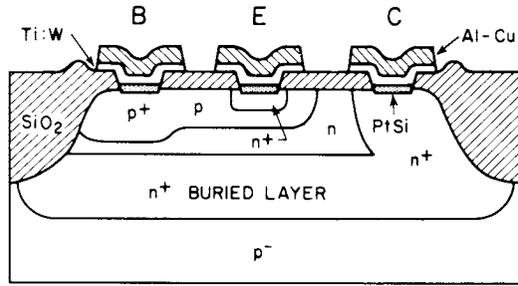


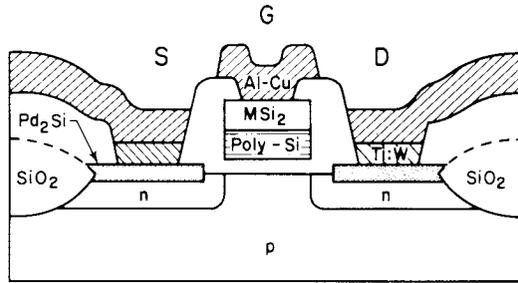
Figure 16: Schematic presentation of the different reaction stages in Al/Pd₂Si/Si junctions. Note that Si precipitate can reach the contact interface before the Pd₂Si layer is completely consumed by passing through pinholes existing in the Pd₂Si layer as shown in b) (Reference 41).

of silicon can be limited by reducing the concentration of the reacting element in the mixture. A structure making use of a TiW barrier in a bipolar device with a PtSi contact is shown in Figure 17a.

The problems associated with gate metallization systems differ from those of contacts because the metal is not in contact with Si. Since device density is important for MOS devices, more emphasis is placed on the patternability and geometrical aspects of the structure, such as edge coverage. There is a need, in addition, to reduce the resistance of the polysilicon gate line to improve the switching time. For this purpose, the approach of using a silicide-silicon combined layer, polycide, has been



(a)



(b)

Figure 17: Schematic cross sections of silicide contacts to (a) bipolar and (b) MOS devices. Note also in (b) the combined use of silicide and polysilicon for gate metallurgy (from Reference 43).

widely accepted.⁴⁷ One can compare the resistivities of silicides in Table VI to those of metals in Table 4. In Figure 17 b, an example of a polycide gate used in a recent MOS device is shown.

Electromigration

Electromigration describes the movement of atoms in a metallic conductor induced by the passage of a direct current. Its magnitude is determined by the atomic diffusivity and the current density. Electromigration induced damage in the form of opens or shorts in the interconnect lines is a result of a local divergence of the mass flux. This divergence can be generated by various types of inhomogeneities, such as those from grain size variation local heating and stress gradients.⁴⁸

With regard to electromigration, the main impact of scaling is to increase the current requirements of interconnecting lines. This problem has two basic aspects, one from the increase in the current density and the other from the reduction in device dimensions.⁴⁹ From Table 2, the current density is seen to increase linearly with the scaling of MOS devices and

more than linearly for bipolar devices. This increases the driving force for electromigration as well as the Joule heating generated in the conductor. With the heating increase as $j^2\rho$, the effect can be significantly higher than the driving force coming from the linear increase in j . The combination of these factors can raise the conductor temperature giving rise to higher atomic diffusivity and electromigration flux. This problem, together with the increased power density, necessitates an improvement in heat dissipation during device operation. The combined effect of these factors will inevitably cause the electromigration flux to increase beyond that caused by the increase in current density alone.

The increase in the current density can be estimated based on the trend in device dimensions. For present devices with dimensions of about $3\ \mu\text{m}$ or larger, the current density can reach 2×10^4 and $5 \times 10^4\ \text{A}/\text{cm}^2$ for MOS and bipolar devices, respectively. For the next generation of devices with minimum line dimensions in the range of 1.5 to $3\ \mu\text{m}$, j increases to about $5\text{-}10 \times 10^4$ for MOS and $2\text{-}4 \times 10^5$ for bipolar devices. This trend continues and can result in j exceeding 10^5 and 10^6 respectively, for submicron MOS and bipolar devices. This is an increase of $10^2\text{-}10^3$ times in the current density. Since an isolated metal wire can carry only about $10^4\ \text{A}/\text{cm}^2$ before melting, the Joule heating generated in a line carrying $10^5\ \text{A}/\text{cm}^2$ must be almost completely removed through the substrate and/or the passivating overlayer. When the current reaches $\sim 10^6\ \text{A}/\text{cm}^2$, any imperfection in the substrate, such as processing defects or interfacial barriers, can cause thermal runaway to destroy the line. Even without such a catastrophe, the heating effects of such high current densities will increase the rate of electromigration, resulting in a significant reduction in the lifetime. In practice, this is reflected by an increase in the exponent n in the lifetime equation of $t_{50} = A j^{-n} \exp(\Delta H/RT)$ outside the normal range between 1 and 2. This effect adds considerable difficulty in extrapolating the lifetime under operating conditions for submicron lines from results obtained in accelerated stress tests.⁵⁰

The other electromigration problem due to size reduction is geometry-related and caused by scaling into the submicron range. For metal films, it is generally observed that the grain size is about the same as the film thickness. For a $1\ \mu$ thick film, the common thickness of interconnecting line, there will be only a few grains spanning across a $1\text{-}3\ \mu\text{m}$ wide line. At the device operating temperature, the electromigration flux is confined to grain boundaries. With a small number of the grains across the line, each individual divergent site in the grain structure becomes potentially more damaging since a line can fail without requiring a statistical linkage of several divergent sites, as would be the case of a line many grains across. This shortens the conductor lifetime while increasing the randomness of the failure statistics, i.e. increasing the statistical deviation σ in the lifetime. Both trends have been observed in lifetests⁵¹ as well as in computer simulation for linewidths down to about $2\ \mu\text{m}$ ⁵². This effect is significant since the extrapolated lifetime for device operation can be significantly reduced by an increase in σ .

Another effect which results from the increase in the grain size-to-line width ratio is a decrease in the role of the grain boundary in mass transport.

Particularly for multilayer structures, scaling in the line width causes a reduction in the grain boundary area per unit line length relative to surface and interface areas. This increases the relative contributions of surfaces and interfaces⁵³ to mass transport. Since these structures have diffusivities generally exceeding that of grain boundaries, the total electromigration rate will be increased accordingly. Even though electromigration at surfaces and interfaces has seldomly been studied, the effect can be estimated from the respective diffusivities. As seen from Figure 12, at about $0.5T_m$, the surface diffusivity is about $10\times$ that of the grain boundary diffusivity. Thus, the total electromigration rate can be substantially increased in the multilayer structure if the surface and interface contribute.

In addition to the two aspects of electromigration discussed already, one new area of concern emerges, namely the behavior of device contacts and step coverage. In certain applications, e.g. the emitter contact in bipolar devices, performance enhancement in scaling requires as much dimension reduction as possible. Consequently, the emitter contact can be subject to current densities exceeding 10^5 A/cm². This problem is complicated by two factors. First, the contacts are often formed using a combination of materials such as silicides and polysilicon. And second, current crowding usually occurs when current flow converges vertically into a contact. The combination of these factors will change the nature of the divergent site, Joule heating and the local electromigration flux. Consequently, the formation of electromigration damage at these structural elements can be basically different from that of a metallic conductor line. Very few studies on this aspect of electromigration have been reported, particularly for small geometries of interest to VLSI applications. Although a recent investigation⁵⁴ of Al/Si contacts reported severe effects due to current crowding and Joule heating.

Studies on electromigration and life tests have been reviewed.^{48,50} While most of these investigations have focused on Al-based metallurgy for line widths more than $2\mu\text{m}$, two systems have been developed for fine lines below $2\mu\text{m}$. One is based on a grain structure modification of the AlCu metallurgy.⁵⁵ Using a low Cu concentration of 0.5wt.% and suitable annealing conditions, Al lines with "bamboo" like grain structure can be formed. With the grain boundary placed normal to the current flow, grain boundary electromigration is greatly reduced. Lifetime of such lines at $5\mu\text{m}$ was found to improve by an order of magnitude over the same material without specifically oriented grains. The lifetime showed a significant upward trend with the decrease in linewidth below $2\mu\text{m}$ although the corresponding σ seemed to increase. The observed increase in the lifetime is opposite to other studies^{51,52} and was attributed to an improvement in the uniformity of grain structures in the line. The other approach taken was to incorporate an Al-transition metal intermetallic sandwich layer in the AlCu structure.⁵⁶ The intermetallic compound was formed by reacting Al with a thin layer of transition metal, e.g. Cr and Ti. The intermetallic layer was found to improve the AlCu grain texture and also serve as a barrier for void linkage to the top and bottom AlCu layer. Between $1\text{-}2\mu\text{m}$, the lifetime was found to improve 50-100 times in comparison to AlCu while σ remained almost unchanged.

SUMMARY

In this chapter, the fundamental aspects of metallization schemes for VLSI interconnects were discussed by considering system requirements and material characteristics. To assess system requirements, the increase in wiring complexity was first explored as a statistical optimization problem.

With device dimensions approaching $1\ \mu\text{m}$, wiring placement becomes a dominant factor in the design of the chip layout and packaging structure. For the design of VLSI interconnects, device scaling brings forth a set of requirements in addition to those from wiring placements. The impact was assessed by considering the scaling rules for FET and bipolar devices, emphasizing the increasing demands on the functional requirements of the interconnect metallization. The discussions on material characteristics focused on the electrical and thermal properties with emphasis placed on the basic parameters including electrical resistivity, dielectric constant and diffusivity.

This was followed by a discussion of two important reliability problems in metallization, junction contacts and electromigration. The projection based on the scaling rules makes it clear that as the device dimension approaches about $1\ \mu\text{m}$, there will be serious materials problems for junction contacts and electromigration. There are two aspects of the problem. The first is material-related because of the use of new materials and the other geometry-related. Some approaches used to deal with these two problems have been described.

Compared with other factors which can potentially limit VLSI development, such as those imposed by the basic physics of the devices, interconnect metallization seems to be of immediate concern and will become increasingly important. The future development of VLSI may well depend on how successfully these problems can be overcome.

REFERENCES

1. *Electronic News*, July 1, 1985.
2. R.H. Dennard, *J. Vac. Sci. Technol.* 19:537 (1981).
3. P.A. Totta and R.P. Sopher, *IBM Res. & Develop.* 13:226 (1969); L.J. Fried, J. Havas, J.S. Lechaton, J.S. Logan and G. Paal and P.A. Totta, *IBM J. Res. and Develop.* 26:362 (1982).
4. D.P. Seraphim and I. Feinberg, *IBM J. Res. and Develop.* 25:617 (1981).
5. M.S. Pittler, D.M. Powers and D.L. Schnabel, *IBM J. Res. and Develop.* 26:2 (1982).
6. A.K. Sinha, *Thin Solid Films* 90:271 (1982).
7. P.B. Ghate, *Thin Solid Films* 93:371 (1982).
8. R.W. Keyes, *Int. J. Theoretical Phys.* 21:263 (1982).
9. M. Hanan and J.M. Kurtzberg, *SIAM Review*, 14:324 (1972).
10. H.R. Heller, W.F. Mikhalil and W.E. Donath, Proc. 14th Design Automation Conf., New Orleans, La. (1977). p. 32.
11. B.S. Landman and R.L. Russo, *IEEE Trans. Comput.* C-20:1469 (1971).
12. R.W. Keyes, Proc. *IEEE* 69:267 (1981).

13. W.D. Grobman, *J. Vac. Sci. Technol.* A3:725 (1985).
14. A.J. Blodgett and D.R. Barbour, *IBM J. Res. and Develop.* 26:30 (1982).
15. D.P. Seraphim, *IBM J. Res. and Develop.* 26:37 (1982).
16. S. Kirkpatrick, C.D. Gelatt, Jr. and M.P. Vecchi, *Science* 220:671 (1983).
17. R.H. Dennard, F.H. Gaensslen, H.N. Yu, V.L. Rideout, E. Bassous and A.R. LeBlanc, *IEEE J. Solid State Circuits* SC-9:256 (1974).
18. D.D. Tang and P.M. Solomon, *IEEE J. Solid State Circuits* SC-14, 679 (1979).
19. J.L. Prince, in *Very Large Scale Integration: Fundamental and Applications*, ed. by D.F. Barbe, Springer-Verlag (1982). Chp. 2.
20. See, for example, B. Hoeneisen and C.A. Mead, *Solid State Electronics*, 15:819 (1972); *ibid* 15:981 (1972).
21. *Metals Handbook*, 8th Ed. Vol. 1, Am. Soc. for Metals (1961).
22. B. Schwartz in *Electronics Packaging Materials Science*, p. 49, (E.A. Giess, K.N. Tu and D.R. Uhlmann eds.) Materials Res. Soc., Pittsburgh (1985).
23. D.J. McGreivy in *VLSI Technologies* (D.J. McGreivy and K.A. Pickar eds.) p. 211-224, IEEE Computer Society Press, Los Angeles (1982).
24. D. P. Seraphim, L.C. Lee, B.K. Appelt, and L.L. Marsh in *Electronics Packaging Materials Science* p. 21, (E.A. Giess, K.N. Tu and D.R. Uhlmann eds.)
25. D.J. McGreivy in *VLSI Technologies* (D.J. McGreivy and K.A. Pickar eds.) p. 185-197, IEEE Computer Society Press, Los Angeles (1982).
26. F.M. d'Heurle and R. Rosenberg, *Phys. Thin Films*, 7:257 (1973).
27. R.W. Balluffi and J.M. Blakely, *Thin Solid Films*, 25:363 (1975).
28. D. Gupta and P.S. Ho, *Thin Solid Films*, 72:399 (1980).
29. N.A. Gjostein in *Diffusion*, p. 241, Am. Society for Metals, Metals Park, Ohio (1973).
30. D. Gupta, D.R. Campbell and P.S. Ho in *Thin Films: Interdiffusion and Reactions* (J.M. Poate, K.N. Tu and J.W. Mayer eds.) p. 161, Wiley, New York (1978).
31. P.S. Ho, *Thin Solid Films* 96:301 (1982); R.S. Nowicki and M.A. Nicolet, *Thin Solid Films* 96:317 (1982).
32. C.Y. Chang, Y.K. Fang and S.M. Sze, *Solid State Electronics*, 14:541 (1971); A.Y.C. Yu, *Solid State Electronics* 13:239 (1970).
33. T.M. Reith and J.D. Schick, *Appl. Phys. Lett.* 25:524 (1974).
34. H.C. Card in *Metal-Semiconductor Contacts*, Inst. of Physics Conf. Series No. 22, Manchester, England (1974).
35. K.N. Tu and J.W. Mayer, in *Thin Films: Interdiffusion and Reactions*, ed. by J.M. Poate, K.N. Tu and J.W. Mayer, Wiley-Interscience (1978). Chp. 10.
36. B.L. Crowder and S. Zirinsky, *IEEE J. Solid State Circuits* 14:291 (1979).
37. S.P. Murarka, *J. Vac. Sci. Technol.* 17, 775 (1980).
38. S.P. Murarka, D.B. Fraser, A.K. Sinha, and H.J. Levinstein, *IEEE Trans. Electron Dev.* ED-27:1409 (1980).
39. C.M. Osburn, M.Y. Tsai, S. Roberts, C.J. Lucchese and C.Y. Ting in *VLSI Science and Technology/1982*, ed. C.J. Dell'Oca and W.M. Bullis, The Electrochemical Society, Inc. (1982), p. 213.
40. C.J. Kircher, *J. Appl. Phys.* 47:6394 (1976); H. Grinolds and G.Y. Robinson, *J. Vac. Sci. Technol.* 14:75 (1977).
41. U. Koster, P.S. Ho and J.E. Lewis, *J. Appl. Phys.* 53:7436 (1982); *ibid* 53:7445 (1982).
42. F.M. d'Heurle in *VLSI Science and Technology/1982* ed. C.J. Dell'Oca and W.M. Bullis, The Electrochemical Society Inc. (1982), p. 194.
43. M. Wittmer, *J. Vac. Sci. Technol.* A2:273 (1984); *Thin Solid Films* 107:99 (1983).
44. J.A. Cunningham, P.R. Fuller and C.T. Haywood, *IEEE Trans. Reliab.* 19:182 (1970).
45. C.Y. Ting and M. Wittmer, *Thin Solid Films* 96:327 (1982).
46. K.N. Tu, *J. Vac. Sci. Technol.* 19:777 (1981).

47. C.Y. Ting, S.S. Iyer, C.M. Osburn, G.J. Hu and A.M. Schweighart in *VLSI Science and Technology/1982*, ed. C.J. Dell'Oca and W.M. Bullis, The Electrochemical Society, 1982, p. 224.
48. F.M. d'Heurle and P.S. Ho in *Thin Films—Interdiffusion and Reactions*, ed. by J.M. Poate, K.N. Tu and J.W. Mayer, Wiley Interscience (1978), Chp. 8.
49. P.S. Ho, IEEE Proc. of 20th Sym. Reliab. Phys. San Diego, Ca. (1982), p. 288.
50. P.B. Ghate, IEEE Proc. of 20th Symp. Reliab. Phys. San Diego, Ca. (1982), p.292.
51. G.A. Scoggin, B.N. Agarwala, P. Peressini and A. Browillard, Proc. 13th. IEEE Symp. Reliab. Phys. (1975), p. 155.
52. J.M. Schoen, *J. Appl. Phys.* 51:513 (1980); K. Nikama, Proc. 19th IEEE Symp. Reliab. Phys. (1981), p. 175.
53. H.L. Huang, *J. Vac. Sci. Technol.* A3:705 (1985).
54. S. Vaidya and A.K. Sinha, IEEE Proc. 20th Sym. Reliab. Phys. San Diego, Ca. (1982), p. 50.
55. S. Vaidya and A.K. Sinha, *Thin Solid Films* 75:253 (1981).
56. J.K. Howard, J.F. White and P.S. Ho, *J. Appl. Phys.* 49:4083 (1978).

Characterization of Semiconductor Materials

Gary E. McGuire

*Tektronix, Incorporated
Beaverton, Oregon*

INTRODUCTION

Characterization of semiconductor materials frequently conveys the image of analyzing the single crystalline substrate. However, semiconductor materials include a broad range of high purity gases, solvents, metals, organics, dielectrics and single crystalline substrates. The analysis of these materials requires an extensive array of analytical tools in order to fully characterize them. Even a brief introduction to this array of tools is beyond the scope of this chapter. Instead, a short description of some of the more popular analytical techniques will be given with the major emphasis being placed on those techniques which provide physical and chemical information.

The ultimate goal of semiconductor device production is to produce components with the appropriate electrical properties. Due to the reduced size of the electrically active region of VLSI devices, the number of impurity and dopant atoms allowed is extremely small, quite often below the detection limit of many analytical techniques. Collectively, these elements produce electronic states within the semiconductor band gap which impact device performance, yield and reliability. The interaction of chemical impurities with physical defects can result in precipitate formation and stacking fault generation, both of which influence the mechanical and electrical properties of the substrate.

The presence, location and role of dopants and adventitious impurities are highly dynamic. The substitutional or interstitial locations for impurities are only two of the many possibilities that arise when impurities interact with defects in the crystal lattice. Device processing, especially at elevated

temperatures, constantly changes the microstructure of the device which can induce both desirable and undesirable electrical properties depending on the location in the wafer.

As a result, characterization strategies must consider the nature of the material, the impurity level and distribution, microstructure and electrical properties. Since each analytical tool provides only a small segment of the necessary information, a series of techniques may be required. This chapter illustrates where some of the more popular analytical tools may be applied and the type of information that may be obtained from them.

SURFACE ANALYSIS TECHNIQUES

Auger Electron Spectroscopy

Figure 1 shows an energy level diagram which depicts the Auger Electron (AE) process. Incident photons, electrons or ions with sufficient energy will create a core hole through the excitation of an ionizing photoelectron.¹ The atom, left in an excited state, de-excites through the emission of x-rays or Auger electrons which are characteristic of the energy levels

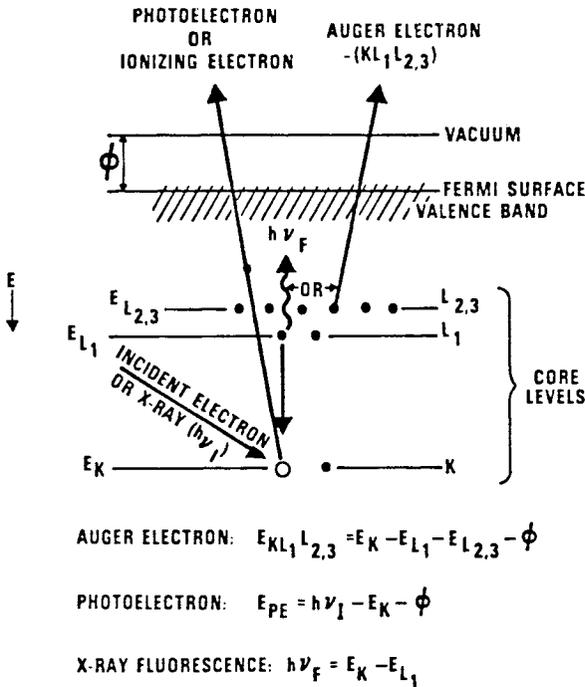


Figure 1: Energy level diagram describing the process for the emission of Auger and photoelectrons.

analyzer. In most modern spectrometers a LaB_6 filament is used to provide high brightness. Spatial resolution of from 25-50 nm may be achieved by sacrificing beam current and sensitivity. A secondary electron detector is incorporated to facilitate locating the electron beam on the sample area of interest.

The focal point of the electron optics and the electron spectrometer are designed to be identical.³ When a sample is positioned at the focal point, the electron beam irradiates the surface, giving rise to the Auger electrons which pass through the acceptance slits into the spectrometer. A negative potential applied to the outer cylinder of the spectrometer deflects the electrons through the exit slit onto the electron multiplier. By sweeping the voltage on the outer cylinder, the electron energy spectrum may be scanned.

The shallow escape depth or inelastic mean free path of electrons as a function of energy is the factor which gives AES its surface sensitivity, Figure 3.⁴ In the range of interest, from 0-2000 eV, the inelastic mean free path (IMFP) is only a few monolayers. Although many studies have been conducted to more accurately determine the IMFP, there is still a large uncertainty in this function.

Characteristic Auger transitions may be observed for all elements with three electrons or more. As a result AES is often used to survey the surface composition of materials. For example, Figure 4 shows an AES spectrum of the surface of a Si wafer coated with an Al layer that is doped with approximately 4% atomic Cu, after etching in a CCl_4 plasma. The plasma etch removes the Al but leaves residual Cu since it does not have the same volatility as Al. The Cu rich residue is only a few monolayers thick and as a result can only be detected by surface analysis techniques such as AES.

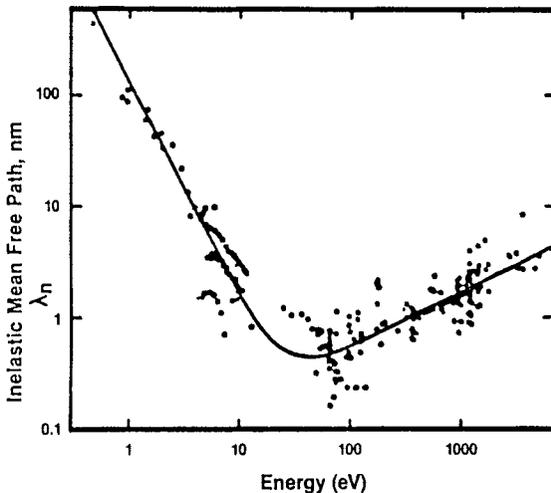


Figure 3: Plot of electron inelastic mean free path versus energy which illustrates the shallow sampling depth of the electron spectroscopies.

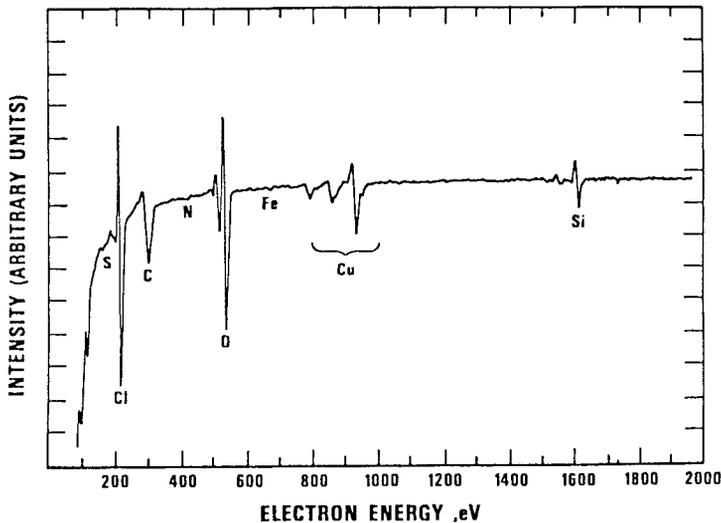


Figure 4: Auger spectrum of residue left after plasma etching a copper doped aluminum layer in CCl_4 .

By combining the surface sensitivity with ion sputtering, depth profiles of the elemental composition may be generated.⁵ This is accomplished through the use of the ion gun shown in Figure 2. The ion gun bombards the surface with a flux of inert ions in the 2-5 keV range, removing controlled amounts of material due to the transfer of momentum from the impinging ions to the surface atoms. By monitoring the Auger signal intensity of selected elements as a function of sputtering time a plot can be generated which represents the concentration as a function of depth. Figure 5 shows the in-depth profile of a multilayered sample of chemically vapor deposited Si_xN_y on thermally grown SiO_2 on a Si substrate.⁶ For amorphous materials, the profile of the boundary between succeeding layers can be quite sharp. Artifacts of the sputtering process may arise due to a variety of factors, however, the information gained as a result of an in-depth profile with a resolution of 20-50Å usually outweighs the disadvantages.

Auger spectra usually contain features which are characteristic of the surface chemistry of the system under investigation as a result of the participation of the valence band electrons in the Auger process. These features have been studied for many systems and may be used in combination with the elemental composition as a means to identify the chemical oxidation state. Figure 6 illustrates the change in the $\text{Ga L}_{3,4,5}\text{M}_{4,5}$ Auger electron kinetic energy and line shape in two chemically different environments.⁷ The Ga Auger peak is shifted by 4.9 eV for the oxide formed on GaAs by anodic oxidation relative to the energy observed for the underlying GaAs substrate. A similar shift of 5.8 eV is observed for the $\text{As L}_{3,4,5}\text{M}_{4,5}$ Auger transition for As in the anodic oxide relative to As in GaAs. Chemical shifts of this magnitude have been observed for most elements. Since the spectral features are complex and the magnitude of the chemical

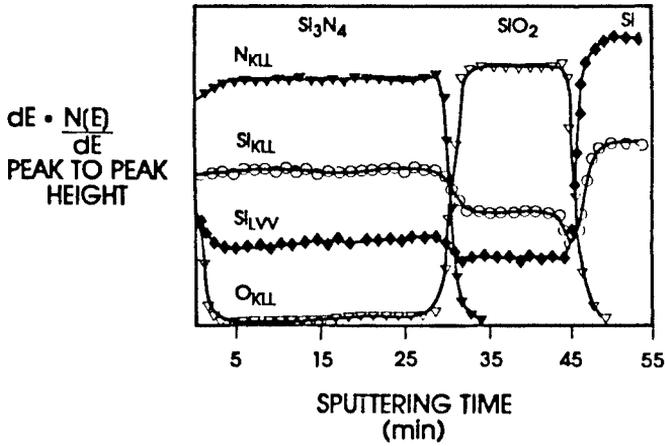


Figure 5: Auger depth profile of chemically vapor deposited Si₃N₄ on thermally grown SiO₂ over Si.

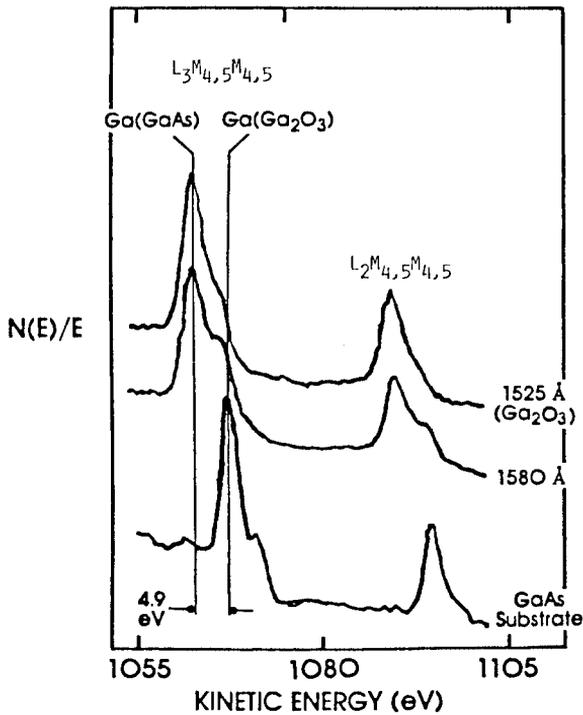


Figure 6: Gallium Auger lineshapes for anodic oxide grown on GaAs.

shift relatively small it is not simple to determine the composition of multicomponent systems.

Since the primary electron beam can be focused to a small spot and rastered scanned over the surface, elemental distribution maps may be obtained of the surface composition. This is accomplished by fixing the pass energy of the spectrometer so that only one transition is being monitored. If more than one element is of interest sequential elemental distribution maps may be generated as shown in Figure 7. One typically looks for inter-relationships in the maps as an indication of compound formation, corrosion, etc.⁸ When the Auger transitions exhibit features which are indicative of certain oxidation states these may be mapped out as well. Elemental or chemical oxidation state distribution mapping may be combined with ion sputtering to generate a three dimensional picture of the sample.

AES has been a powerful tool in the investigation of a wide range of materials problems in the semiconductor industry. One application for which AES is particularly well suited is the study of diffusion in thin films. Figure 8 illustrates schematically the three diffusion processes encountered in thin polycrystalline films. Grain boundary diffusion transports material from the underlying substrate through a thin polycrystalline film more rapidly than bulk diffusion. Rapid surface diffusion then distributes the material across the specimen. Surface analytical techniques like AES have been used extensively in the investigation of diffusion in thin metal couples used in the semiconductor industry. This may be accomplished by monitoring the surface composition while annealing the sample in-situ to determine the arrival time and increase in concentration of the diffusing species. By annealing at different temperatures, one can generate a series of curves of surface concentration versus annealing time from which an Arrhenius type plot may be generated.⁹ An alternate approach which has been used is to anneal a series of samples at different times at a fixed temperature then profile through the film to determine the extent of interdiffusion.¹⁰

AES has a detection limit of approximately 0.1% atomic or 10^{18} atoms/cm³ in Si with a sensitivity variation of 50-100 across the Periodic Table. Several handbooks of Auger data provide relative sensitivity factors for the elements which may be used to quantify experimental results.¹¹ There remains extensive work that needs to be done, however, before AES is truly a quantitative technique. In addition, even though AES has relatively poor sensitivity, it is one of the more popular analysis techniques for the evaluation of surfaces and interfaces.

AES may be utilized on a wide variety of materials, but due to the use of an electron beam for excitation it does have limitations. The electron beam and ion beam used for sputtering may induce sample decomposition. This problem is accentuated by the high current densities that occur with small probe diameters. Insulating materials may be difficult to evaluate due to sample charging effects. The imbalance of currents from the primary beam, the secondary electrons and the sample result in a surface potential which distorts the Auger electron energy.

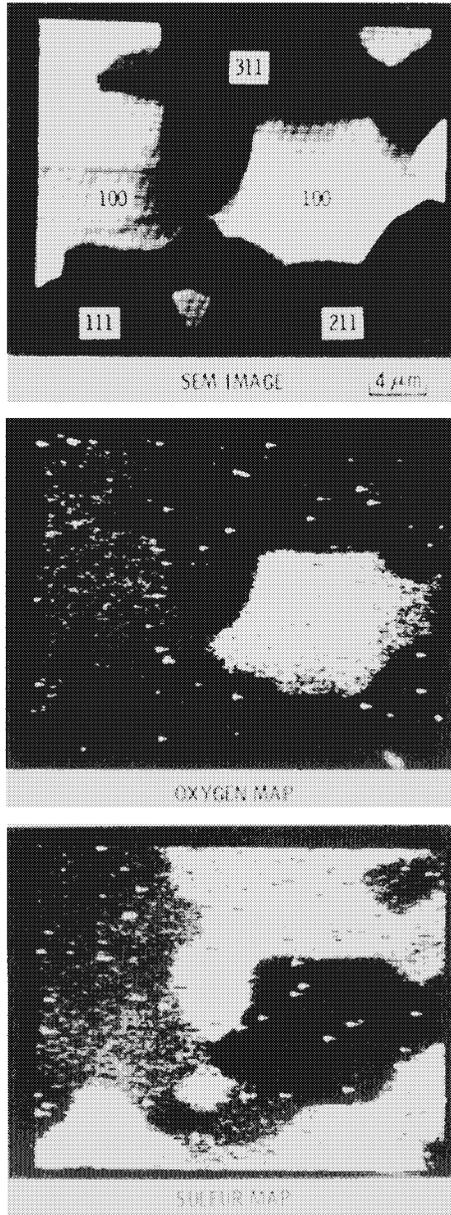


Figure 7: SEM image and O and S scanning Auger maps of zone refined Fe foil. The numbers indicate the approximate orientation of the surface normal of the various grains.

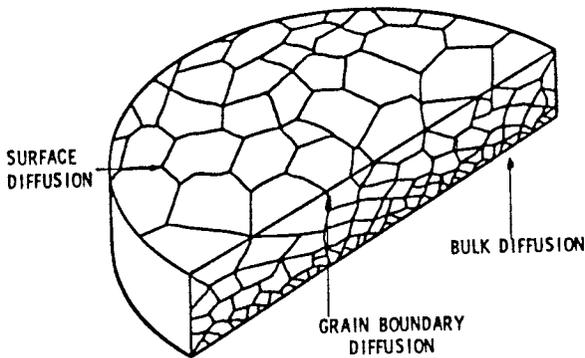


Figure 8: Schematic diagram of the diffusion paths in thin polycrystalline materials.

Photoelectron Spectroscopy

Photoelectron spectroscopy is a technique which has many similarities to AES. The same energy level diagram used to describe the Auger process may be used to describe the photoelectron process.¹² Excitation of the ionizing photoelectron may be accomplished through the use of a variety of energetic photons or charged particles. The primary focus in this text will be on monochromatic x-ray excitation of photoelectrons (XPS). Use of a monochromatic excitation source is essential to this spectroscopy, since the photoelectron's kinetic energy is directly dependent on the energy of the excitation source. By knowing the energy of the x-ray with a high degree of accuracy and measuring the kinetic energy (KE) of the emitted photoelectron from the relationship:

$$BE = h\nu - KE + \phi \quad (2)$$

one can determine the binding energy (BE) of any electron energy level less than the photon energy.

A variety of electrostatic electron energy analyzers have been produced commercially. One of the most popular is the cylindrical mirror analyzer, Figure 9, similar to that used for AES. A two stage, two cylindrical mirror analyzers in tandem, device is employed to enhance the energy resolution. An x-ray source, either an Al or Mg anode, mounted in proximity to the sample is used for excitation. The x-rays flood a broad area of the sample since they, unlike the electron source in AES, can not be easily focused. The acceptance angle of the spectrometer determines the area of analysis, which is typically a few millimeter diameter circle. With adjustable aperture slits the sampled area may be reduced to a few hundred micrometers.

Figure 10 shows a schematic diagram of another common variety of XPS spectrometer. It employs an x-ray monochromator to enhance the x-ray line width, eliminate satellite x-ray lines and focus the x-rays. The x-

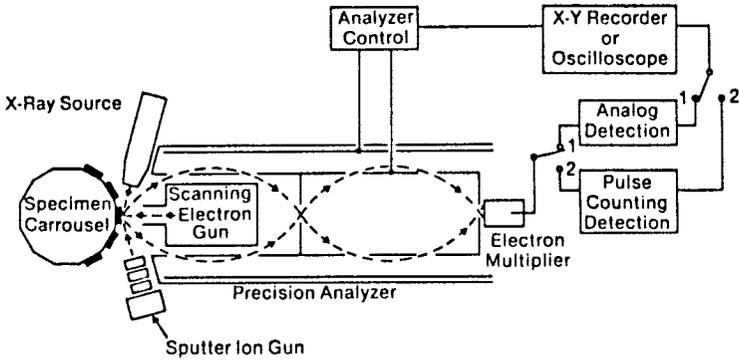


Figure 9: Schematic diagram of a two stage cylindrical mirror analyzer and x-ray source used for XPS.

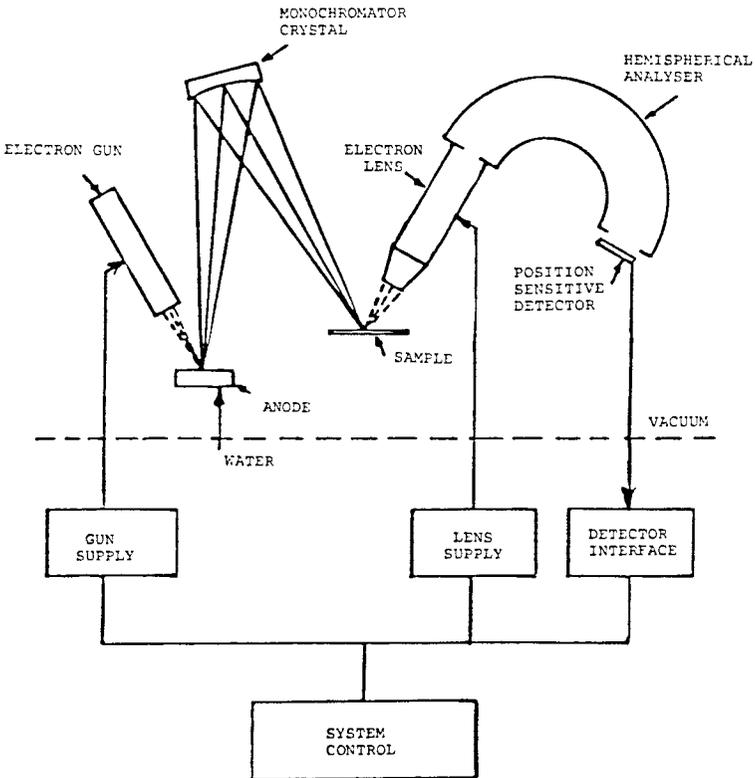


Figure 10: Schematic diagram of a XPS system utilizing a bent quartz crystal x-ray monochromator in conjunction with an electrostatic lens and hemispherical analyzer.

rays from an Al anode are allowed to diffract off of a bent quartz crystal before interacting with the sample. The natural Al x-ray linewidth is approximately 0.9 eV while that of the monochromized source is approximately 0.4 eV.¹³ The focusing properties of the monochromator produce a spot size of approximately 150 micrometers.¹⁴ Due to the loss in x-ray intensity in going through the monochromator, most spectrometers of this type employ an electrostatic lens to increase the collection efficiency of photoelectrons going into the hemispherical analyzer and a position sensitive, multiple array, detector to enhance the count rate.

Photoelectron spectrometers employ ion guns for in-depth profiling as in AES. Since the area of analysis is much larger than in AES, the ion beam is defocused in order to generate a uniform ion flux. This reduces the ion etch rate but does not prevent one from monitoring signal intensity as a function of ion sputtering time. In addition, many XPS systems have both x-ray and electron beam sources for combined multi-technique analysis by XPS and AES.

Figure 11 shows the Ag3d photoelectron spectrum. The trace illustrates the relative simplicity of the spectra. The spectral features are Gaussian-like sitting on a low background. The spin-orbit splitting of the energy levels, in this case the $3d_{5/2}$ and $3d_{3/2}$, is well characterized and easy to recognize due to the predictable intensity ratios. The spectra are usually plotted in the $N(E)$ versus BE format even though the energy analysis is of $N(E)$ versus KE. Each element exhibits a unique set of photoelectron (PE) transitions corresponding to its atomic energy levels. The PE transitions are a function of atomic number so that the energy levels of adjacent elements in the Periodic Table are all shifted to higher binding energy.

As suggested by this unique set of binding energies, XPS is a good elemental surface analysis technique. Figure 12 shows a spectrum from a

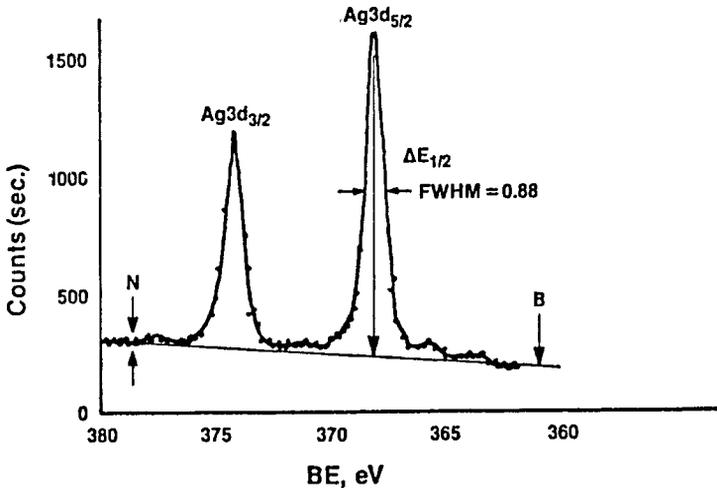


Figure 11: XPS spectrum of the Ag3d transition showing the spin-orbit splitting into the $3d_{5/2}$ and $3d_{3/2}$ components.

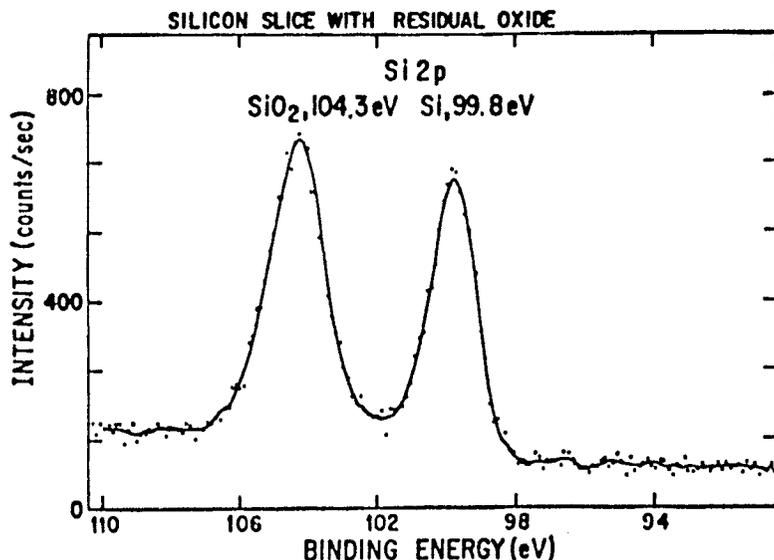


Figure 12: XPS spectrum of a Si surface with the native oxide showing the chemical shift in the Si2p transition.

clean Si wafer to illustrate this point. Two Si2p transitions are observed, one for elemental Si and one for the native oxide formed on the wafer as a result of air exposure. One can get a feel for the surface sensitivity of XPS since the native oxide thickness is typically less than 30Å. The surface sensitivity, as in AES, is controlled by the inelastic mean free path of the electron as illustrated in Figure 3, rather than the path length of the x-rays used for excitation.

The ability to distinguish different oxidation states, as in the case of Si and SiO₂, has been one of the recognized strengths of XPS. These chemical shifts in the core level binding energies are due to changes in the valence electron density due to compound formation. Although, the chemical shifts may be as large as 10-12 eV, there is frequent overlap for many compounds as illustrated in Table 1.¹⁵ However, there are many sources of chemical information in the spectra. The sources include first the identification of the elements present and their relative concentrations, then the chemical shift of the cation to determine its approximate oxidation state and finally the chemical shift of the anion to determine its oxidation state. Table 2 illustrates the magnitude of the chemical shifts observed for anionic species X⁻², XO₃⁻² and XO₄⁻² when X is S, Se and Te.¹⁶ Combination of this information gives a detailed picture of the chemistry of the surface under investigation in many cases. There are other spectral features which provide additional chemical information but which are too detailed for the scope of this text.

Figure 13 illustrates the use of XPS in the investigation of the anodization of GaAs.⁷ Both the As3d and Ga3d transitions may be observed. By combining XPS analysis with ion sputtering the composition of the ano-

Table 1: Chemical Shifts in the Cr2p_{3/2} Transition of Chromium Compounds

Chromium, Cr Atomic Number **24**

COMPOUND	2p _{3/2} BINDING ENERGY, eV						
	573				578		583
Cr							
CrB ₂							
Cr ₂ S ₃							
CrN							
Cr(CO) ₆							
K ₃ Cr(CN) ₆							
CrO ₂							
Cr ₂ O ₃							
CrI ₃							
CrBr ₃							
CrCl ₃							
CrO ₃							
K ₂ CrO ₄							
K ₂ Cr ₂ O ₇							
CrF ₃							

Table 2: XPS Chemical Shifts

Compound	Oxidation State	ΔE Relative to Elemental State		
		S	Se	Te
X ⁻²	-2	-1.4	-0.8	-0.6
X ⁰	0	0	0	0
XO ₃ ⁻²	+4	3.6	3.7	2.9
XO ₄ ⁻²	+6	5.5	4.2	3.6

W. E. Swartz, K. J. Wynne and D. M. Hercules, ANAL. CHEM., 43 1884 (1971)

dized layer is obtained at various depths. The outer surface appears to be predominately Ga₂O₃ as determined by the chemical shift in the Ga3d peak and the absence of As. In the bulk of the oxide, the composition is a mixture of As₂O₃ and Ga₂O₃ as determined from the chemical shifts in both the As3d and Ga3d transitions. At the oxide-GaAs substrate interface there is a transition region where both the As3d and Ga3d exhibit two

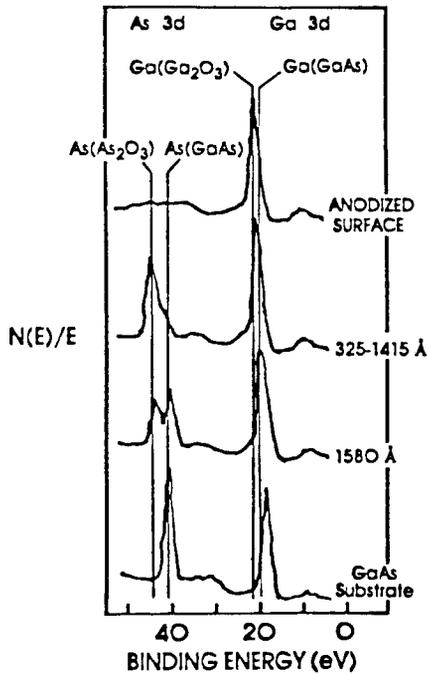


Figure 13: XPS spectrum of anodized and annealed GaAs showing the chemically shifted Ga3d and As3d transitions at various depths in the oxide and at the GaAs substrate.

peaks, one for the oxide and one for GaAs. The oxidation of many compound semiconductors has been studied by XPS. The composition of the oxide, especially as a function of depth, has been found to be strongly dependent on the method and conditions of formation.

Investigation of metallization schemes used to contact semiconductor devices is another key area where XPS has been applied. As an example Figure 14 shows the Pt4f and Si2p spectra obtained from the silicides which are formed when Pt is used to contact Si.¹⁷ Two different silicides may be formed depending on the annealing conditions used in the process. The Pt $4f_{7/2}$ and $4f_{5/2}$ doublet exhibits a chemical shift of less than 1 eV between the Pt_2Si and PtSi phases which may be formed. This chemical shift is easily detectable. The corresponding Si2p transitions have essentially the same binding energy. The chemical information obtained from XPS compliments that obtained from a variety of other thin film and surface analysis techniques in the investigation of a variety of contact materials.

Since the photoelectron spectra of many elements exhibit only small chemical shifts for a series of compounds in which the electronegativity varies over a wide range, it is frequently necessary to examine the other features of the spectrum. One of these features which frequently exhibits useful chemical information even when the photoelectron spectra do not,

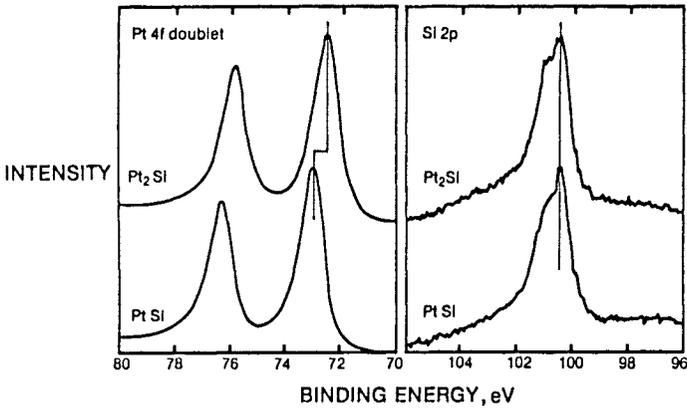


Figure 14: XPS spectrum of the Pt4f_{7/2} and Pt4f_{5/2} doublet and Si2p transition of the two silicide phases, Pt₂Si and PtSi, of platinum.

is the corresponding Auger transitions. Since Auger emission is a multi-step process in which two electrons are emitted, the Auger and photoelectron, the electron shells surrounding the atom have more time to undergo relaxation. This almost always results in larger chemical shifts for the Auger transitions.

Table 3 compares the binding energies of photoelectron and Auger transitions for a series of metals and their oxides.¹⁵ In all cases, the Auger transition exhibits a factor of two larger chemical shift over that of the photoelectron transition. However, it is important to remember that the Auger spectra are more complex and exhibit broader line widths. In addition, Auger transitions are not always available due to the limited energy range of excitation of the x-ray source.

The detection limit for XPS is approximately 0.5% atomic or 5×10^{18} atoms/cm³ in Si with a sensitivity variation of 10^2 . Data sets are available which provide relative sensitivity factors based on peak area.¹⁸ The sensitivity factors are specific to the instrument design, yet, once established the results are more easily quantified than AES data.

The x-rays used for excitation do less damage to the surface than charged particle excitation sources. There is evidence for x-ray induced desorption even though only to a minor extent. Sample charging is minor since only the secondary electron and sample return currents must be balanced.

XPS has traditionally been more of a research tool in the semiconductor industry but is becoming more of a routine analytical technique as more people realize the extensive chemical information that may be obtained from the spectra.

Secondary Ion Mass Spectroscopy

Secondary Ion Mass Spectroscopy (SIMS) is the mass analysis of secondary ions generated by ion sputtering. As illustrated in Figure 15,

Table 3: Chemical Shifts in X-Ray Excited Auger Spectra

Compound	Binding Energy, eV (Al Radiation)				
	2p	3d	KLL	LMM	MNN
Mg	49.8		300.8		
MgO	51.2		307.2		
Al	72.8		93.5		
Al ₂ O ₃	75.4		100.2		
Zn		9.9		494.0	
ZnO		10.7		498.6	
Ge		29.4		341.5	
GeO ₂		33.2		349.3	
Ag		374.0		1128.2	
Ag ₂ SO ₄		374.2		1132.2	
Sn		492.9		1048.3	
SnO ₂		494.6		1053.4	

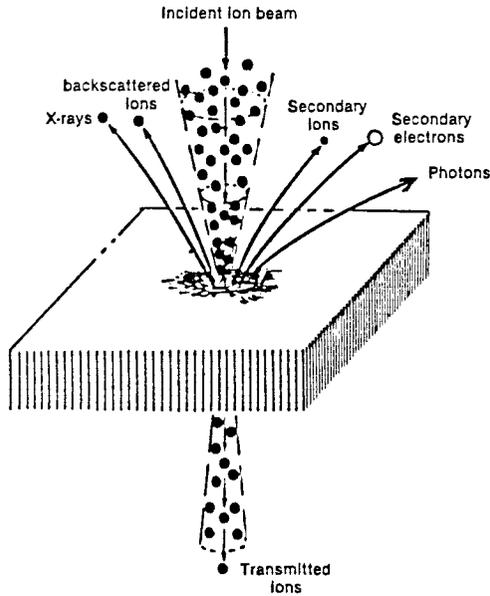


Figure 15: Schematic diagram of secondary particles and radiation generated by an incident ion beam.

bombarding the surface of a solid with an energetic ion beam generates a variety of secondary transitions, including electrons, photons and ions. Detection of any of these secondary events could serve as the basis for any analytical probe. However, SIMS equipment is optimized for the detection of positive and secondary ions.

There are several major design features which must be taken into consideration in selection and application of SIMS as an analytical tool. One of these is the use of SIMS as an ion imaging tool, ion microscope, or as an ion microprobe. The ion microscope utilizes an ion lens to image the ions removed from an area of the sample surface. It often has many of the other features of an ion microprobe and, as a result, is positioned at the top end of the line in performance and price. Figure 16 is a schematic of an ion microprobe which contains the essential features of a SIMS system. These include the ion source and secondary ion mass analyzer. The ion source shown here is a duoplasmatron source which has the capability of generating ion beams from a gas source of 1-2 micrometers spot size at up to 20 KV potential. Other ion sources include the hot filament activated gas sources which achieve 100 micrometer spot sizes at potentials of typically 5 KV maximum and liquid metal ion sources which achieve spot sizes of 2000A using field emission of low melting point metals.

More expensive instrument designs have a primary ion mass analyzer to separate the positive and negative ions and the neutrals which are produced in the source. Since the ions are charged a condenser lens may be used to focus the ions while charged deflection plates are used to position the beam or raster it over the sample surface.

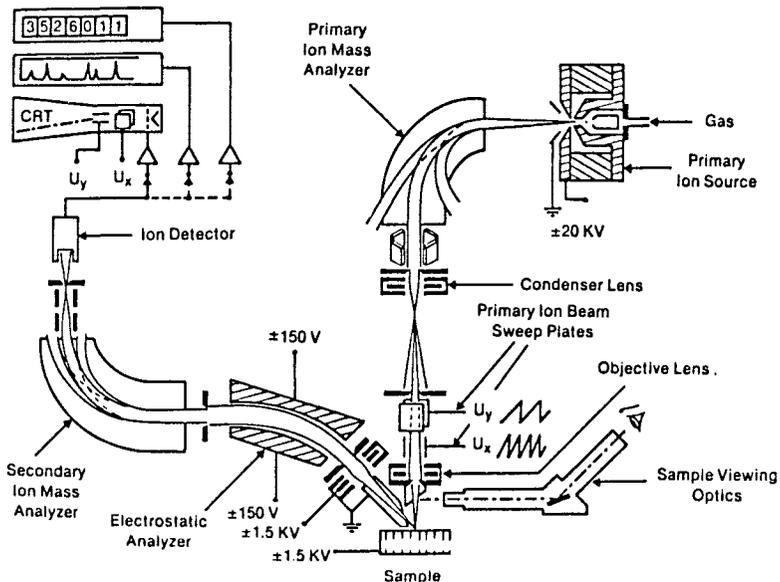


Figure 16: Schematic diagram of a secondary ion probe mass spectrometer.

Applying a potential to the sample increases the secondary ion collection efficiency due to their radial distribution. Simple electrostatic analysis is often used prior to the mass analyzer in order to select a narrow energy distribution of the secondary ions. This allows improved performance of the mass spectrometer. The mass spectrometer can be a quadrupole analyzer for low to intermediate charge to mass ratio resolution or a magnetic sector for high resolution.

Since SIMS is basically a depth profiling technique in that generation of secondary ions requires removal of material, it is necessary to generate the ions from a uniform depth. Since ion beams have Gaussian shapes, it is necessary to raster the beam over an area slightly larger than the area of analysis in order to maintain a flat sampling area.¹⁹ Otherwise, ions coming from the sidewalls of the crater would result in a signal coming from a region of shallow depth, often referred to as the memory effect. This is avoided by gating the detection system such that the signal is only accepted when the primary ion beam is away from the crater wall.

The key design consideration for SIMS instrumentation is the efficient generation of secondary ions. One aspect of this is the selection of the primary beam energy. As can be seen in Figure 17, the sputtering yield, atoms removed per incident ion, is dependent on the incident ion energy. SIMS ion sources are usually designed to operate in the 5 to 20 KV energy region. There is no benefit in going to higher accelerating voltages since the sputtering yield is flat or decreases above 20 KV.

Only a few per cent of the atoms removed by sputtering are ionized. The remainder come off as neutral atoms or atom clusters. Proper selection of the primary ion can enhance the ion yield. Electropositive primary ions enhance the yield of negative secondary ions while electronegative primary ions enhance the yield of positive secondary ions. On this basis, the favored primary ions are O^- and Cs^+ , although there are other factors which may dictate the use of other primary ions.

The secondary ion yield is a function of the electronegativity of the elements. When a negative primary ion beam is utilized the relative positive ion yield will be greatest for those elements with the lowest

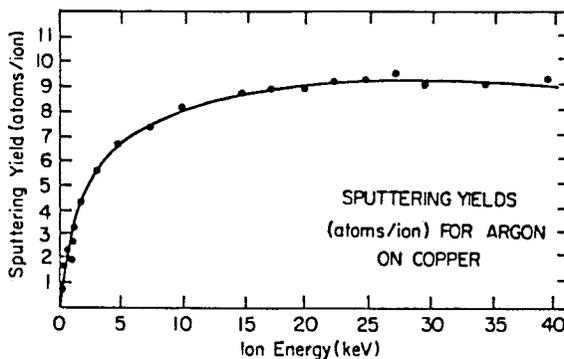


Figure 17: Plot of the sputtering yield versus ion energy for argon on copper.

electronegativity. Figure 18 illustrates the variation in secondary ion intensity as a function of atomic number, which correlates with electronegativity.²⁰ Conversely, when a positive primary ion beam is utilized the relative negative ion yield will be greatest for those elements with the highest electronegativity.

The SIMS spectrum is a plot of the secondary ion intensity versus the mass to charge ratio. As can be seen from Figure 19, the spectrum from even high purity elements like Si can be very complex. The spectrum consists of ionized atoms and isotopes, ionic complexes, and multiply ionized species. The ion intensity is usually plotted on a log scale due to the dynamic range, 10^6 , of the data.

Figure 20 illustrates the range of sensitivities²¹ available for P implanted in Si, from 10^{17} to 10^{21} atoms per cm^3 . This is a factor of 10-100 more sensitive than AES or XPS. However, SIMS data is difficult to quantify because the ion yield is matrix dependent. As a result, other techniques such as neutron activation analysis are used to normalize the SIMS data. The depth resolution of SIMS is approximately 20Å, the depth from which secondary ions are generated. As illustrated here the surface sensitivity in combination with the inherent sputtering can be utilized to generate in-depth profiles of 0.5 micrometers or more with high sensitivity. The depth scale, however, is not known accurately since the sputter rate depends on a wide range of variables. The sputter depth is frequently determined through the use of mechanical stylus techniques, ellipsometry or interferometric techniques.

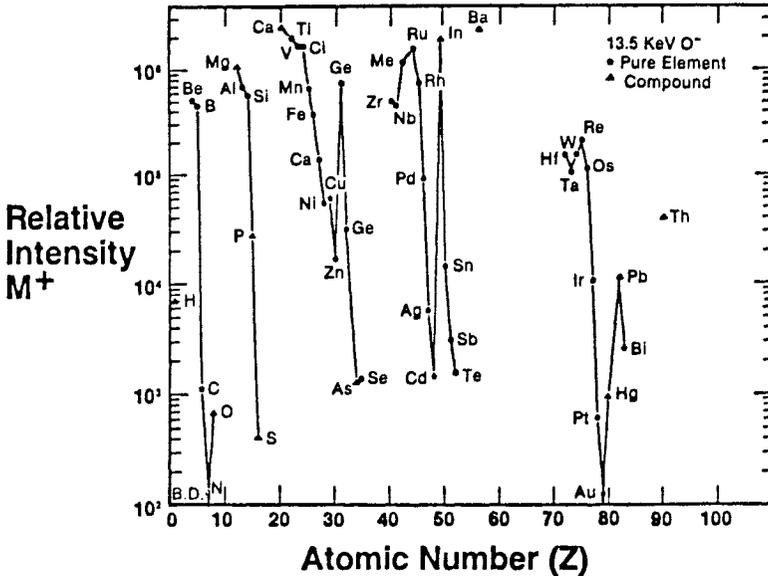


Figure 18: Plot of the relative positive secondary ion yield versus atomic number for 13.5 keV oxygen ions.

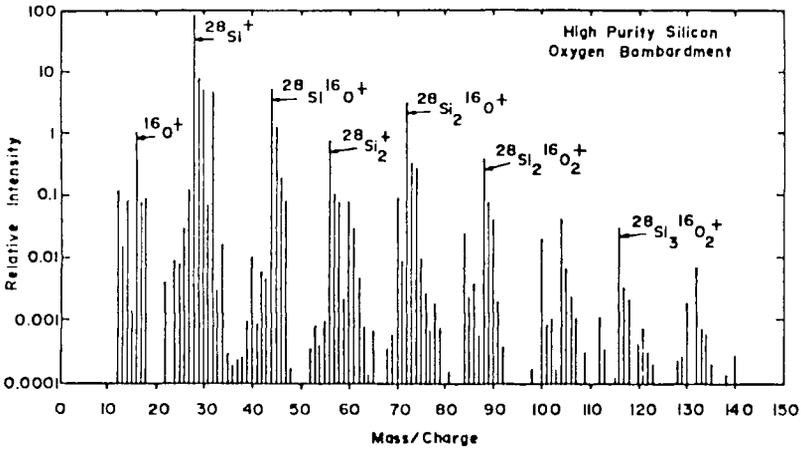


Figure 19: Plot of the relative secondary ion intensity versus the mass-to-charge ratio resulting from oxygen bombardment of high purity silicon.

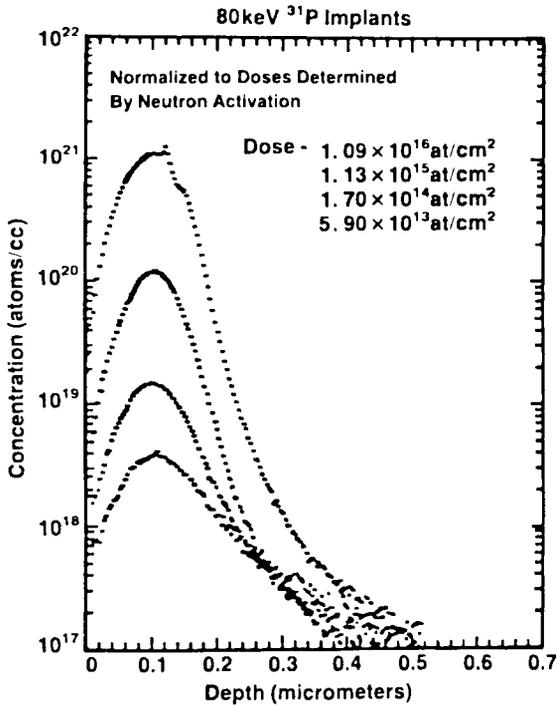


Figure 20: Depth profiles of phosphorus implanted into silicon at 80 keV showing the concentration versus depth.

Even with these limitations SIMS is a powerful tool for the analysis of dopants in semiconductor materials. Figure 21 illustrates an example where SIMS has been used successfully. Boron profiles in Si are shown before and after laser annealing.²² The dopant is redistributed as a result of laser annealing. The maximum B concentration decreases while the tail of the profile diffuses to a greater depth. It is interesting to note that the redistribution of B is concentration dependent. The tail of the implant profile is redistributed to a much greater depth after laser annealing for the high dose implant than the low dose implant.

SIMS is one of a few analytical tools capable of distinguishing isotopes. This has resulted in some well designed experiments that take advantage of this feature. For example, Coleman et al.²³ utilized SIMS to investigate

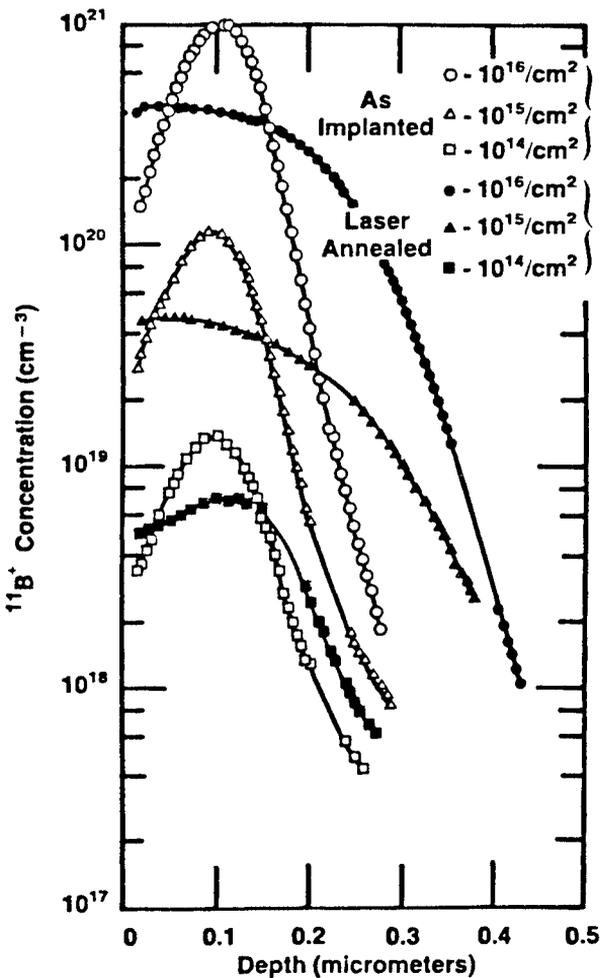


Figure 21: Depth profiles of boron implanted into silicon at different doses showing the concentration versus depth before and after laser annealing.

the anodic oxidation of GaAs. By using isotopically labeled H_2O , they were able to distinguish the mechanism by which anodic oxidation proceeds. Figure 22 illustrates the depth profiles that were obtained from a GaAs (001) wafer anodized in H_2O^{18} then H_2O^{16} and another sample anodized in the opposite sequence. From this study the authors were able to show that oxygen is incorporated into the growing oxide at the oxide-electrolyte interface as opposed to the oxide-semiconductor interface. Mass transport occurs through the interstices of the growing oxide.

The most frequent application of SIMS in the semiconductor industry is in the analysis of dopants in single crystalline substrates. However, SIMS is widely used in the analysis of a broad range of materials but with some limitations. Easily ionized elements such as Na, Li or Cl may migrate during ion bombardment. The charge that builds up on insulators during ion bombardment may reach sufficient field strength to cause charge induced ion migration. This effect may be minimized by lowering the sample temperature or by neutralizing the surface charge. This may be accomplished by a number of techniques including positioning a hot filament in proximity to the sample or exposing the surface with a scanning electron beam. The degree of success in neutralizing the surface charge, however, greatly influences the secondary yield. The ratio of neutral species versus secondary ions increases as a result of ion neutralization during sputtering.

The many factors that influence ion yield make quantification of the data difficult and limit the range of applications of the technique. By adding

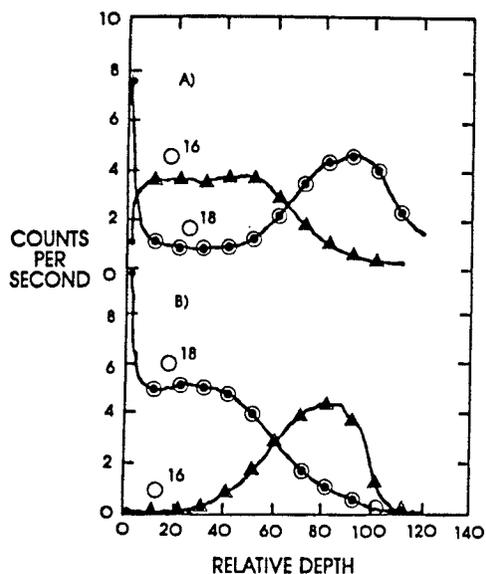


Figure 22: Depth profile of the isotopic distribution of oxygen in an anodic oxide grown on GaAs(001): curve (A) is for GaAs(001) anodized in H_2O^{18} then H_2O^{16} while curve (B) is for anodization first in H_2O^{16} then H_2O^{18} .

a photon source with sufficient energy to ionize the material removed in the sputtering process to the SIMS system, the ion yield can be increased to approximately 100%. Through the use of the additional photon source, the matrix dependence of the ion yield can be minimized. The data obtained in this manner is much easier to quantify. There is the added benefit of improved sensitivity since only ions are detected.

An example of this approach, which is frequently referred to as Sputter Assisted Laser Ionization (SALI) is shown in Figure 23. Curve (a) shows the SIMS spectrum of a GaAs wafer using static Ar⁺ bombardment, while curve (b) shows a comparable SALI spectrum with the introduction of a 248 nm laser.²⁴ The SIMS spectrum exhibits a significant difference in the Ga⁺ and As⁺ intensity. With the addition of the ionizing laser, the ion yields of Ga and As are more nearly equal and the ion yields of trace elements in the GaAs are enhanced.

Resonance Ionization Spectroscopy

Resonance Ionization Spectroscopy (RIS) is a technique which is similar to SALI. The material to be analyzed is removed from the surface of

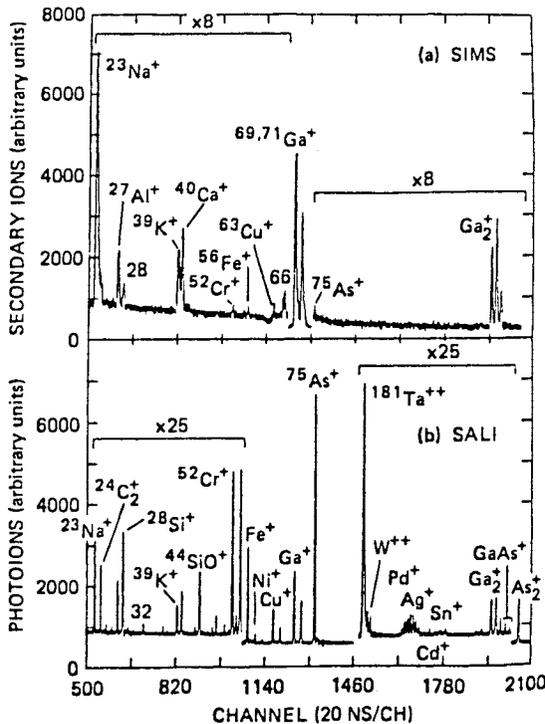


Figure 23: The positive secondary ion SIMS spectrum in trace (a) is from a GaAs wafer under static Ar⁺ bombardment. A SALI spectrum is shown in trace (b) for the same surface under static Ar⁺ bombardment, plus an accompanying 248 nm laser for ionization.

the sample by ion sputtering. A laser beam is positioned above the sample so that it intersects with the vapor cloud of atoms as shown in Figure 24. The laser is tuned to the frequency necessary to ionize the atom of interest. The ions are extracted from the sample region, then energy analyzed before passing into a mass spectrometer. The spectrum like SIMS is a plot of the intensity versus the mass-to-charge ratio.²⁵

RIS reportedly is capable of achieving sensitivities reaching 1 part in 10^{12} and a selectivity that eliminates ambiguity in the interpretation of the results.²⁶ Several factors contribute to the enhanced sensitivity of RIS over similar techniques like SIMS. RIS has a lower background. The ions that are generated during the sputtering process are extracted before the laser pulse ionizes the remaining neutral species. As a result, only ions generated in the laser pulse pass into the mass spectrometer. Additionally, only selected ions are generated in the laser pulse. Figure 25 shows the five basic schemes used in the resonance ionization process.²⁶ Typically, a tunable dye laser is adjusted so that it emits precisely the correct wavelength to excite an electron in an atom from its original state to a higher state. Occasionally, a second photon from a second laser is used to excite the atom further to an even higher state. After excitation, a second or third photon is used to interact with the excited atom, causing the electron to be released from the atom. Thereby, producing a positive ion and a free electron. The key to RIS is choosing resonantly excited states that can be easily excited and that have large photoionization cross-sections, so that they can be ionized with high efficiency. It is possible to saturate all of these processes with commercially available lasers, so that an atom in the initial state will be excited through the resonant intermediate states and into the ionization continuum with unit probability during a single laser pulse.

RIS is extremely selective, in that, only atoms of a given element are ionized. The intermediate excited states through which the ionization proceeds may be chosen such that they are uniquely characteristic of that

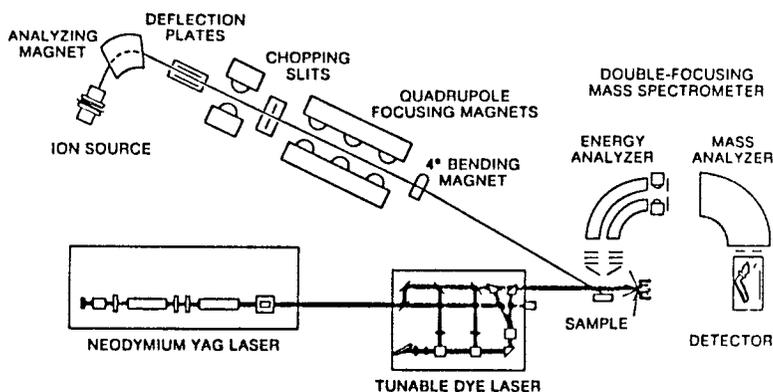


Figure 24: Schematic diagram of sputter initiated resonance ionization spectrometer.

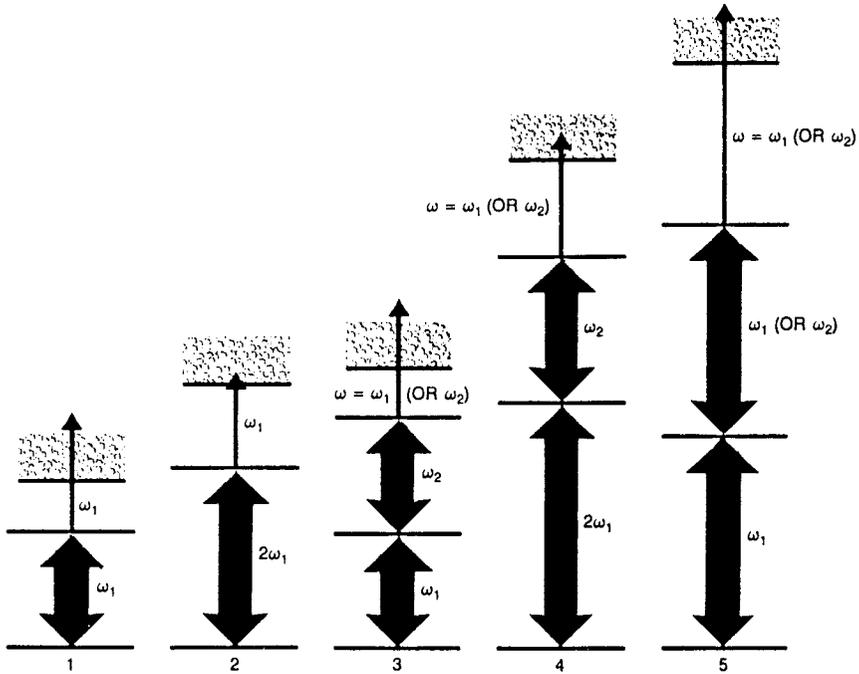


Figure 25: Schematic diagram of the five basic schemes for RIS.

element. As a result, RIS provides sensitive analysis of solid samples for all the elements except He and Ne. Sensitivities of down to 10^{10} atoms/cm³ have been reported for Na in Si.²⁷

Figure 26 shows a plot of the concentration of B in Si as measured by RIS versus the B concentration determined by electrical resistivity.²⁵ There is generally good agreement between the values down to the part-per-billion level. The RIS values for the samples lowest in B lie above the least squares line fitting the values for the samples with the higher B concentration. This is attributed to contamination resulting from sputtering of the stainless steel slits, a common problem for RIS in the detection of extremely low concentration levels.

Several approaches have been taken to remove material from the sample surface to be introduced into the laser beam where resonance ionization occurs. When an ion beam is utilized, many of the analytical capabilities are similar to SIMS. The sampling depth of approximately 20Å is determined by the sputtering process. The depth that is probed depends on how long one wants to continue the analysis. Due to the generation of a crater during the sputtering process, the ion beam must be rastered and the signal gated such that only material from the flat portion of the crater, uniform depth, is analyzed.

Since RIS is the newest of the surface analysis techniques, there is yet much to be learned about its sensitivity and sensitivity variation. It shows

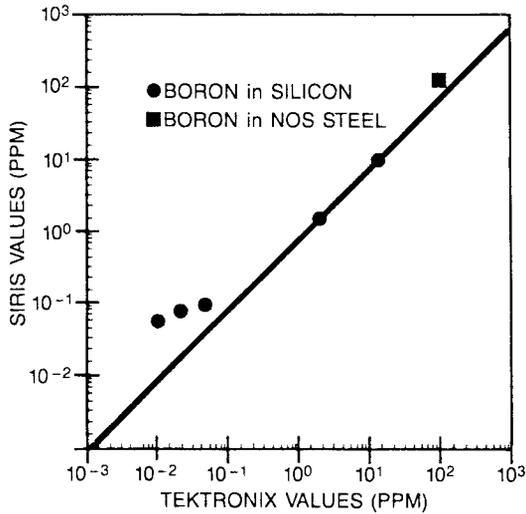


Figure 26: Correlation plot showing the boron concentration in silicon measured by RIS versus the value determined by electrical measurements.

promise of being one of the most sensitive analytical techniques available for solids analysis. Commercially produced instruments are now available which should accelerate the learning process.

Rutherford Backscattering Spectroscopy

Rutherford Backscattering Spectroscopy (RBS) is the energy analysis of ions that are backscattered from a surface. Typically ions with low mass, such as H⁺ or He⁺, are accelerated toward a target at a potential of 0.5 - 2.0 MeV. As shown in Figure 27, the target, M₂, recoils while the primary ion, M₁, scatters at an energy E₁ at an angle θ. The scattering energy, E₁, is easily calculated from the relationship,

$$E_1 = K(M_1, M_2, \theta)E_0$$

where

$$K = \left[\frac{M_1 \cos\theta + \sqrt{M_2^2 - M_1^2 \sin^2\theta}}{M_1 + M_2} \right]^2 \tag{3}$$

The scattering cross-section is a smoothly varying function of the target atomic mass as shown in Figure 28. From this curve it is obvious that the scattering efficiency is very low for those elements with low atomic mass. In addition, it is difficult to distinguish elements which have similar masses when the elements have high atomic masses.²⁸

Figure 29 shows the equipment necessary to perform RBS. The accelerator must be capable of generating MeV ions from the light elements.

Modern instruments make use of compact tandetron accelerators which allow the construction of RBS systems which are not significantly larger than other surface analysis equipment. The analysis of the backscattered ions may be achieved through the use of an electrostatic analyzer or a solid state detector. The solid state detector, the preferred detection system, is positioned in front of the target at an angle of approximately 30° from the

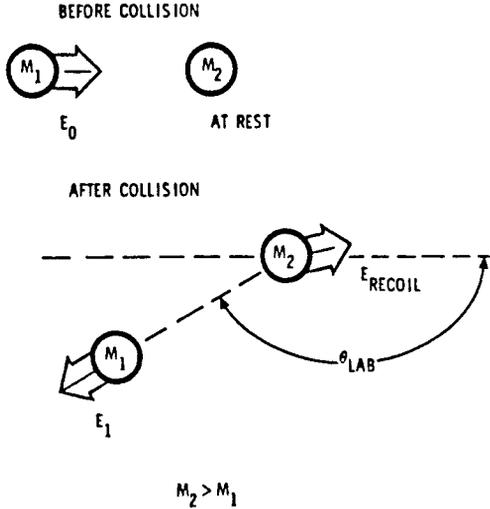


Figure 27: Schematic diagram illustrating a Rutherford backscattering collision.

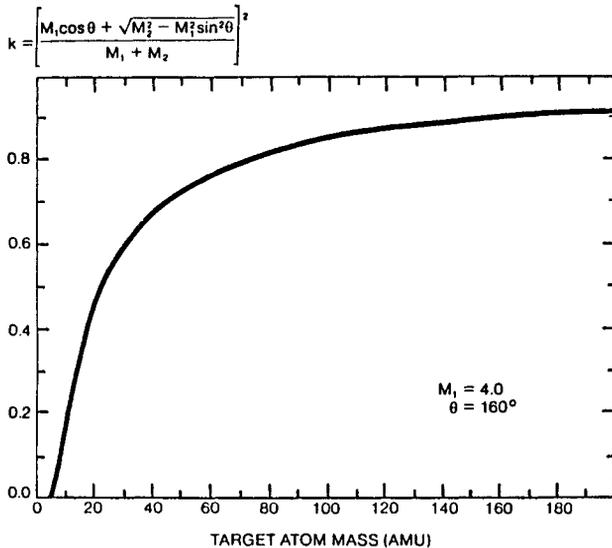


Figure 28: Plot of the Rutherford backscattering cross-section versus target atomic mass.

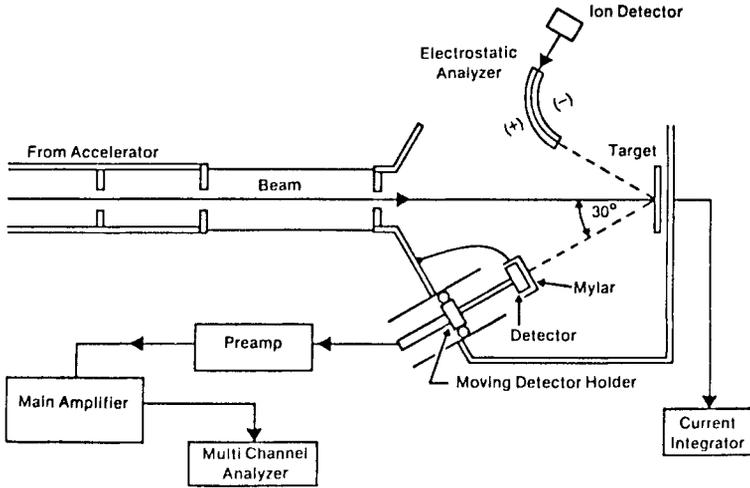


Figure 29: Schematic diagram of a Rutherford backscattering spectrometer.

incoming primary beam. A thin mylar sheet is placed in front of the detector to attenuate low energy secondary ions and secondary electrons.

The spectra are plots of scattered ion intensity versus energy. An RBS spectrum is the sum of a family of curves from each atomic mass in the target. As shown in Figure 30, KE_0 represents inelastic scattering from the front surface of the target.²⁸ At a depth X , the primary ion loses additional energy through electron scattering both going into and escaping from the solid. Since Rutherford scattering occurs at all depths, a curve is generated which is the sum of all these events. Each atomic mass in the target generates a separate curve based on its scattering cross-section.²⁹

Figure 31 illustrates the application of RBS in the analysis of the silicides formed during the interaction of Ni and Si.³⁰ The dashed line represents the as-deposited Ni on Si case, where the Ni and Si exhibit distinct scattering energies. Upon heating at 300°C for 90 minutes, Ni₂Si forms which is represented by the open circles. The curve for Ni has decreased in intensity and broadened while the silicide portion of the Si curve has moved toward the Ni. Additional heating results in a further decrease in intensity and broadening of the Ni curve and an increase in the Si curve for the silicide. Since the scattering cross-sections for Si and Ni are known, the stoichiometry for the different phases of silicide can be calculated without the use of standards. There are many examples like this in the literature where a heavy metal in a matrix of a low atomic mass element lends itself to RBS analysis.

Since RBS is essentially a non-destructive quantitative analysis technique, it is frequently used to calibrate other surface analysis techniques. It, however, has a limited range of sensitivity of about 10^{18} atoms/cm³ in a Si matrix.³¹ This sensitivity is adequate for calibrating XPS and Auger samples but not for many SIMS samples.

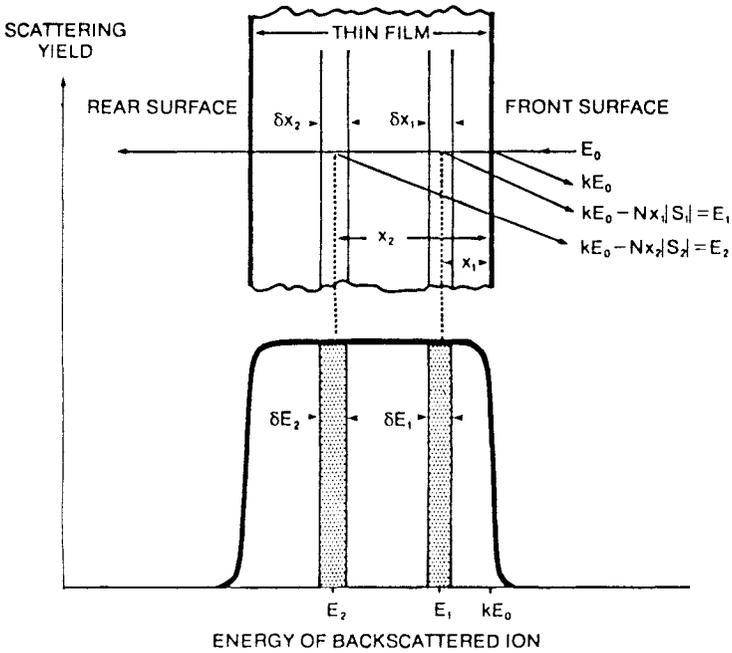


Figure 30: A plot of the Rutherford backscattering yield versus the energy of the backscattered ion with an accompanying illustration showing the scattering location in the sampled depth.

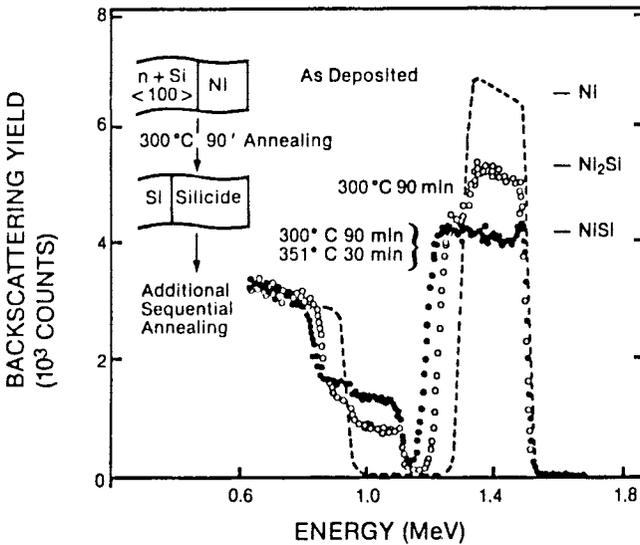


Figure 31: RBS spectra of the phases of nickel silicide formed following the deposition and annealing of nickel on silicon.

Of the surface analysis techniques, RBS is unique in its ability to distinguish whether a dopant occupies a substitutional or interstitial site in a crystalline lattice. When the primary ion beam is oriented along the crystalline planes, the ions penetrate long distances into the crystal along the open channels. Scattering occurs at crystal imperfections and interstitial impurity sites. Figure 32 compares the RBS spectra from Si implanted Si (100) samples which are positioned such that random scattering and channeling occur.³² The virgin sample exhibits minimal scattering except in the random orientation, indicating the quality of the crystal. After implantation, the crystal has undergone extensive damage which is evident in the increased scattering along the channeling direction. Subsequent heating at 550°C and 850°C anneals out much of the damage, however, the crystal quality of the virgin sample is not recovered.

The same equipment used to do RBS can be used for nuclear reaction analysis (NRA).²⁸ Instead of Rutherford scattering the primary ion must penetrate the nucleus of the target atom and induce a nuclear reaction as depicted in Figure 33. The nuclear reaction cross-section as a function of incident ion energy must be known in order to select an energy which will result in adequate yield. The energies required for NRA are frequently higher than those used for RBS.

Table 4 lists some useful nuclear reactions. NRA compliments RBS in that, many of the useful nuclear reactions are for low atomic number elements for which RBS has low sensitivity. Since the nuclear reaction

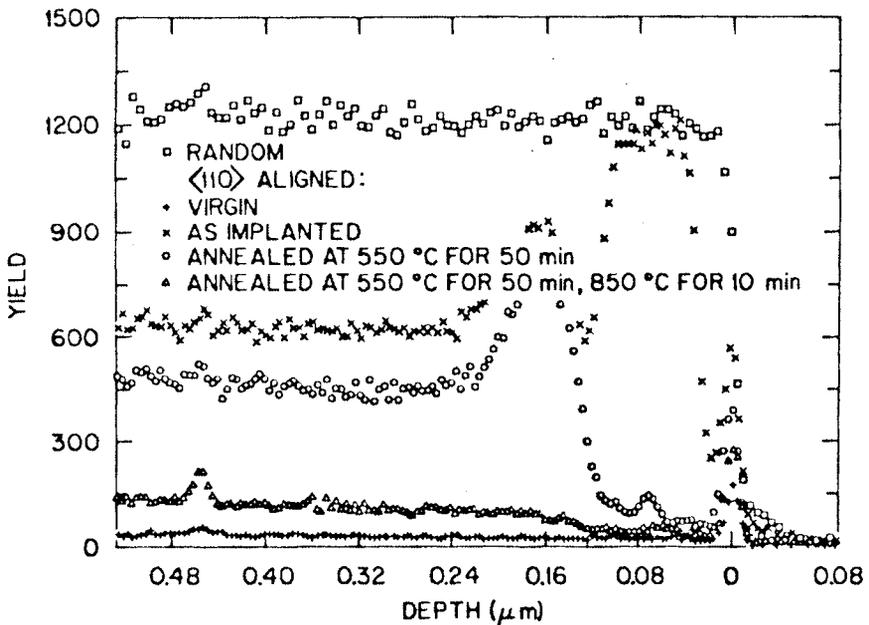


Figure 32: RBS spectra for Si(100) in the random and <110> aligned directions before and after 80 keV ³⁰Si⁺ implant and subsequent anneal at 550° and 850°C.

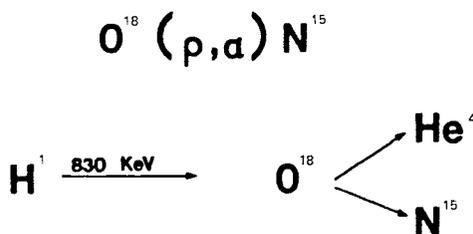


Figure 33: Schematic diagram illustrating an ion induced nuclear reaction.

Table 4: Useful Nuclear Reactions

Nucleus	Reaction
${}^2\text{H}$	${}^2\text{H}({}^3\text{He}, \text{p}){}^4\text{He}$
${}^3\text{He}$	${}^3\text{He}(\text{d}, \text{p}){}^4\text{He}$
${}^6\text{Li}$	${}^6\text{Li}(\text{d}, \alpha){}^4\text{He}$
${}^7\text{Li}$	${}^7\text{Li}(\text{p}, \alpha){}^4\text{He}$
${}^9\text{Be}$	${}^9\text{Be}(\text{d}, \alpha){}^7\text{Li}$
${}^{11}\text{B}$	${}^{11}\text{B}(\text{p}, \alpha){}^8\text{Be}$
${}^{12}\text{C}$	${}^{12}\text{C}(\text{d}, \text{p}){}^{13}\text{C}$
${}^{13}\text{C}$	${}^{13}\text{C}(\text{d}, \text{p}){}^{14}\text{C}$
${}^{14}\text{N}$	${}^{14}\text{N}(\text{d}, \alpha){}^{12}\text{C}$
${}^{15}\text{N}$	${}^{15}\text{N}(\text{p}, \alpha){}^{12}\text{C}$
${}^{16}\text{O}$	${}^{16}\text{O}(\text{d}, \text{p}){}^{17}\text{O}$
${}^{18}\text{O}$	${}^{18}\text{O}(\text{p}, \alpha){}^{15}\text{N}$
${}^{19}\text{F}$	${}^{19}\text{F}(\text{p}, \alpha){}^{18}\text{O}$
${}^{27}\text{Al}$	${}^{27}\text{Al}(\text{p}, \gamma){}^{28}\text{Si}$

cross-sections are well known, NRA like RBS is quantitative without the use of standards. This is especially beneficial for elements like H which are difficult to detect and quantify by other analytical techniques.

Summary

More attention has been given to the surface analysis techniques in this chapter than will be given to the imaging and bulk analysis techniques. This is the area of expertise of the author and one which has received increasing attention as device dimensions have decreased. Table 5 is provided as a summary of the characteristic features of these techniques that were discussed in this section.

IMAGING ANALYSIS TECHNIQUES

Scanning Electron Microscopy

Scanning electron microscopy is surface imaging of solids using

Table 5: Summary of Surface Analytical Techniques: Capabilities and Limitations

CHARACTERISTIC	AES	XPS	SIMS	RIS	RBS	NRA
EXCITATION SOURCE	e ⁻	X-RAY	ION	ION	ION	ION
DETECTED EMISSION	e ⁻	e ⁻	ION	ION	ION	ION
ELEMENTAL DETECTION	Z > 3	Z > 1	Z > 1	Z > 3	Z > 2	Z > 1
ELEMENTAL IDENTIFICATION	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	GOOD	GOOD
SENSITIVITY VARIATION	10 ²	10 ²	10 ⁴		10 ³	10 ³
DETECTION LIMITS	0.1%	0.5%	10 ⁻⁴ %	10 ⁻⁷ %	10 ⁻³ %	10 ⁻³ %
CHEMICAL INFORMATION	YES	YES	NO	NO	NO	NO
LATERAL RESOLUTION	50 NM	150 μM	1 μM	1 μM	1 MM	1 MM
DEPTH RESOLUTION	5A	5A	20A	20A	100A	100A
DEPTH PROBED		← SPUTTER DEPTH →			10 ⁴ A	10 ⁴ A
DEPTH ANALYSIS		← DESTRUCTIVE →			NON-DESTRUCTIVE	NON-DESTRUCTIVE
BEAM INDUCED DECOMPOSITION	HIGH	LOW	LOW	LOW	LOW	LOW
SAMPLE CHARGING	YES	MINOR	YES	YES	NO	NO
STANDARDS REQUIRED	YES	YES	YES	YES	NO	NO
MATRIX EFFECTS	MINOR	MINOR	MAJOR	MINOR	NO	NO
SPECIAL FEATURE	SPATIAL RESOLUTION	CHEMICAL INFORMATION	HIGH SENSITIVITY	HIGH SENSITIVITY	NON-DESTRUCTIVE DEPTH ANALYSIS	LIGHT ELEMENT ANALYSIS

electron beam generated secondary electrons. The equipment utilized frequently has additional analytical tools that take advantage of the other types of radiation generated by electron beam excitation. Figure 34 illustrates the electron beam interaction with a solid. The primary beam may be focused to a spot 50A in diameter.³³ Upon interacting with the solid, secondary electrons are generated which are utilized to image the surface. As the high energy electrons penetrate the solid they undergo scattering which increases the interaction volume. Some of the primary electrons will be backscattered toward the surface from an area much larger than the region exposed to the primary beam. The energetic primary electrons will ionize the atoms in the solid producing x-rays which are characteristic of the elements that are present. With a suitable set of detectors, all of these events may be monitored.

Secondary electrons are low energy even though the primary electron beam is several keV or higher. Figure 35 shows the average energy distribution of secondary electrons from metals.³⁴ The peak in the distribution is below 5 eV. In order to efficiently collect the secondary electrons, a high potential bias is applied to a scintillator tube which is positioned in proximity to the sample. The signal is converted to light and fed out through a light pipe to a photomultiplier tube as shown in Figure 36. Images are generated by synchronizing the raster of the electron beam with the output of the photomultiplier tube.

Backscattered electrons have the same energy as the primary electrons. The backscattering coefficient, like that of ions in RBS, is a well known, smoothly varying function of atomic number as shown in Figure 37.

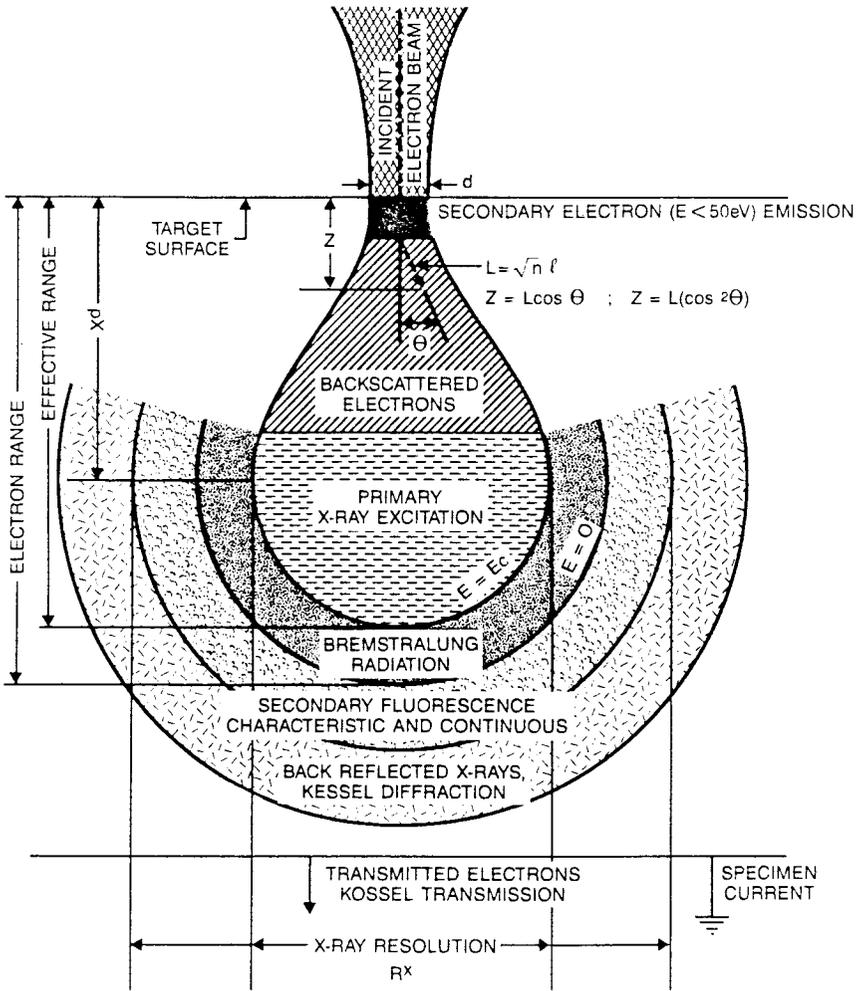


Figure 34: Diagram illustrating the interaction of the primary electron beam with a solid surface in the production of secondary and backscattered electrons, x-rays and other secondary radiation.

Since the backscattering yield varies more than the secondary electron yield across the Periodic Table, backscattered electrons will yield better image contrast in many situations.³⁵ The information depth for backscattered electrons as a result of the energy dependence of the escape depth is 10^2 greater than secondary electrons.³⁴

One of the most common analytical attachments to the SEM is the energy dispersive x-ray spectrometer (EDX). The high energy primary electron beam excites x-rays which are characteristic of the elements which are present in the solid to a depth of 0.5 micrometers or greater

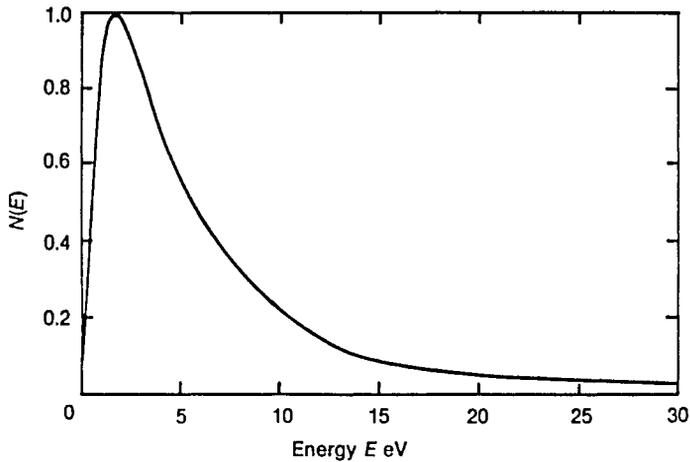


Figure 35: Plot of the average intensity of secondary electrons from metals as a function of energy.

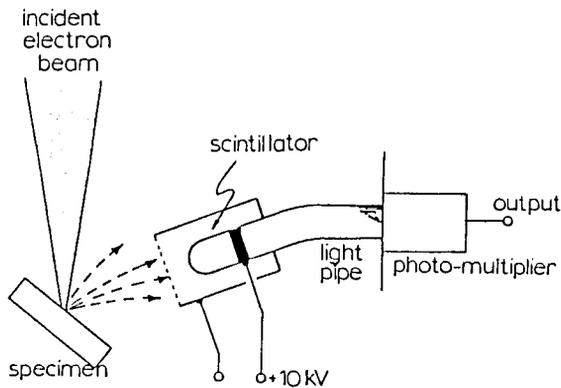


Figure 36: Schematic diagram of a scintillator tube used for the detection of secondary electrons.

depending on the accelerating potential.³⁶ The emitted x-rays are detected by a solid state detector which is positioned in the vicinity of the sample, Figure 38. The detector is a Li doped Si crystal which is biased at high voltage. X-rays interacting with the detector create electron-hole pairs which are swept through the detector due to the high voltage bias. The charge pulse is converted to a voltage pulse by a charge-sensitive preamplifier. The voltage distribution can be displayed on a cathode ray tube or an x-y recorder. The useful energy range for EDS systems is from 1.0 to 20 keV which limits the analysis to light element with $Z > 9$. It is possible to operate without the Be window in front of the detector in an

ultra-high vacuum system. This permits the analysis of the lighter elements down to C.

The major disadvantage of EDX is its ability to operate in the pulse counting mode and detect simultaneously the characteristic x-rays for all elements above F in the Periodic Table. A full spectrum may be obtained in

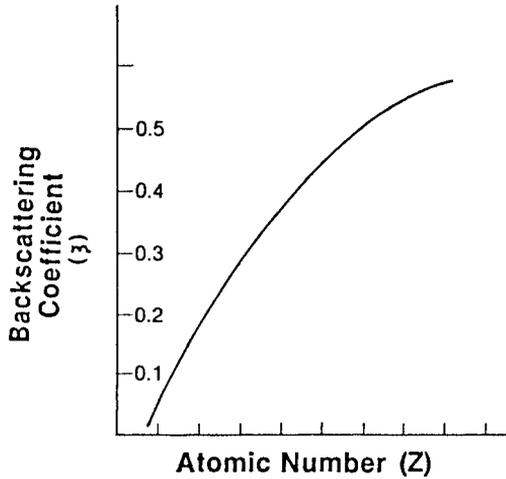


Figure 37: Plot of the electron backscattering coefficient versus atomic number.

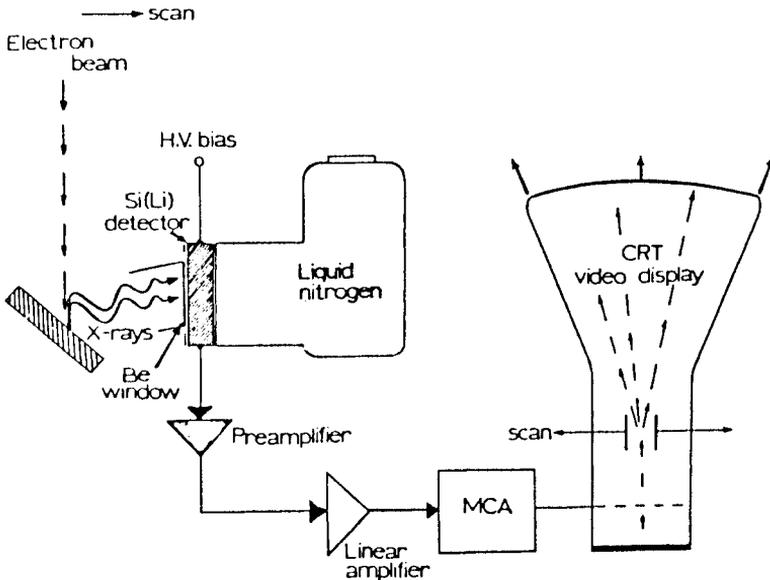


Figure 38: Diagram illustrating the detection of electron beam excited x-rays with a solid state lithium doped silicon detector.

a much shorter time with the EDX analyzer than with the wavelength dispersive x-ray (WDX) analyzer but at the expense of energy resolution. The EDX analyzer has a resolution of approximately 150 eV, whereas, the WDX analyzer has a resolution of 5 eV. The Li doped Si detector also requires liquid nitrogen cooling to keep the Li from diffusing and rapidly degrading the performance.³³

A schematic diagram of the wavelength dispersive detector is shown in Figure 39. The electron beam excited x-rays interact with a diffraction crystal which disperses the x-rays. As the crystal is rotated, the different wavelength x-rays are allowed to enter the detector. A variety of crystals are used in order to optimize the energy resolution and collection efficiency of the broad range of x-ray energies for elements $Z > 6$. The detection system may be used to generate a spectrum of x-ray intensity versus wavelength from which the characteristic x-ray lines may be identified. It may also be operated at a fixed wavelength, so that the detector output represents an intensity map of the sample surface for one characteristic x-ray. The most commonly used detector for the WDX spectrometer is a gas flow proportional counter. When an x-ray enters the tube through a thin window on the side and is absorbed by an atom of the gas it causes a photoelectron to be ejected which then loses its energy by ionizing other gas atoms. The electrons are then attracted to a central wire which gives rise to a charge pulse.

For bulk samples more than a few micrometers thick, spatial resolution for elemental analysis does not improve for probes much less than 1

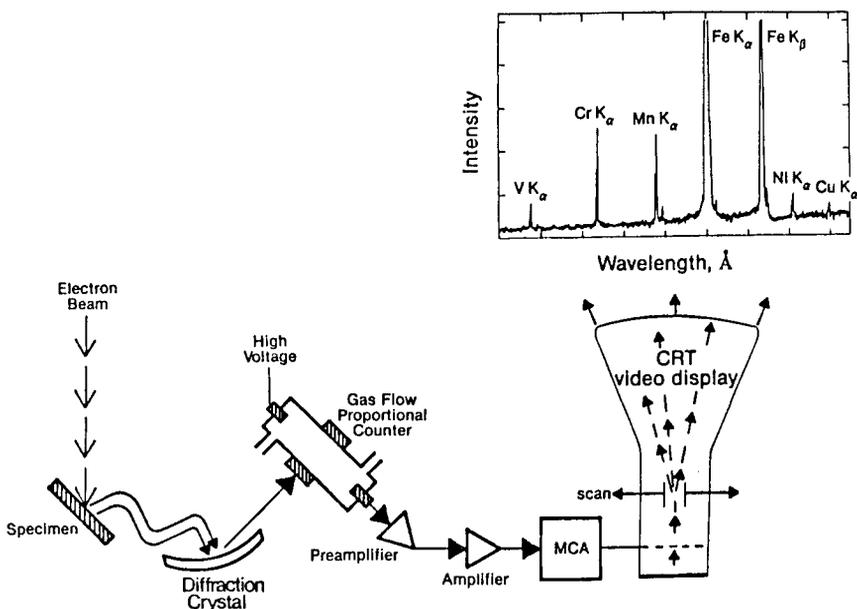


Figure 39: Diagram illustrating the detection of electron beam excited x-rays with a wavelength dispersive detector.

micrometer in diameter since the volume of x-ray production is determined by electron beam scattering. Spatial resolution equal to or smaller than the probe diameter can be obtained for thin foils.

One of the most rapidly growing applications of the SEM is voltage contrast. As IC geometries continue to shrink, it becomes increasingly difficult to test device performance with physical probes. Since the electron beam in the SEM may be focused to 50Å, it can easily be localized on any feature of modern IC structures. By introducing an electron spectrometer in the sample chamber to energy analyze the secondary electrons, the SEM becomes a voltage probe of the surface. The voltage applied to conductive leads alters the secondary electron energy in the local area. The electron spectrometer detects these differences. Many different types of spectrometer have been applied to voltage contrast, however, one of the most common is the retarding field spectrometer. As shown in Figure 40, the spectrometer is mounted in the path of the electron beam. A small aperture allows the electron beam to pass through to the sample which is by necessity an operating device. Voltage contrast is used to test for conductive lead continuity, propagation delays and etc., where circuit designs or process parameters are being verified.³⁷

Scanning Transmission Electron Microscopy

The scanning transmission electron microscope (STEM) utilizes an electron beam much like the SEM but at higher accelerating potential. A

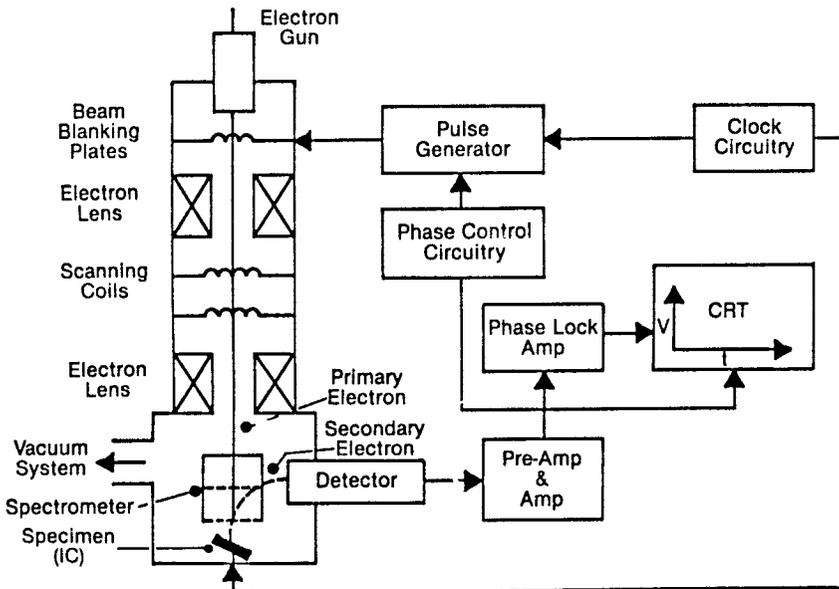


Figure 40: Schematic diagram of electron beam optical column equipped with a retarding field electron spectrometer for voltage contrast measurements.

higher accelerating potential is utilized since only electrons that are transmitted through thinned samples are imaged. The accelerating potential depends on the sample thickness and atomic mass but is typically 100-200 kV. When the directly transmitted beam is imaged this is usually referred to as the bright field image. By utilizing the diffracted beam for crystalline materials, a change in contrast is observed in what is referred to as a darkfield image.

As shown in Figure 41, the STEM can accommodate a variety of analytical attachments. Like the SEM, the STEM can be utilized for secondary electron and backscattered electron imaging and x-ray analysis from the front side of the sample and lattice imaging, microdiffraction and electron energy loss from the transmitted beam.³⁸

The STEM is capable of imaging features as small as 2Å since the sample is thinned. Electron scattering which occurs in the bulk of the sample is minimized since the primary beam has a short path length through the sample. Samples are thinned to approximately 2000Å through a combination of chemical and/or mechanical polishing and ion milling. Ion milling is continued until a photodetector mounted behind the specimen shuts the system down due to the light emitted from the ion gun filament. The specimen is mounted on a wire grid for ease of handling since the sample is prepared in a separate vacuum chamber.

Samples may be thinned vertically or horizontally. Figure 42 shows a micrograph of the cross-section of an epitaxial Si layer grown on an Si substrate. Stacking faults occur in the epitaxial Si layer as a result of impurities, possibly native oxide, that remain on the single crystalline substrate after cleaning.³⁹ Cross-sections like this may be utilized to evaluate various epitaxial growth techniques, implantation damage, deposited films and contact formation.

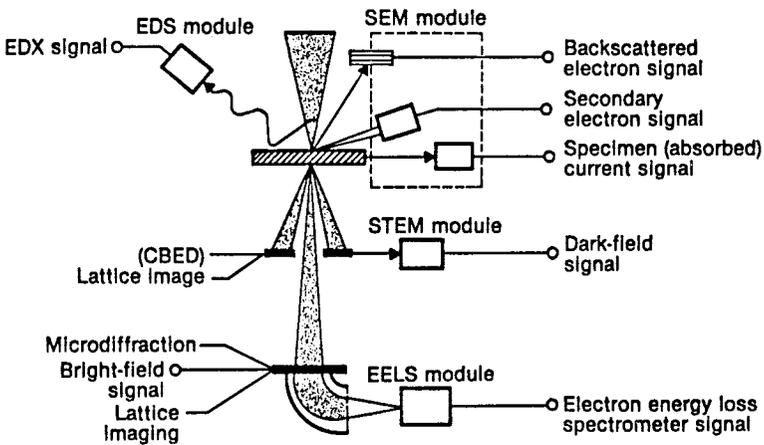


Figure 41: Diagram of the analytical attachments and modes of operation of a STEM for evaluating thinned specimens.

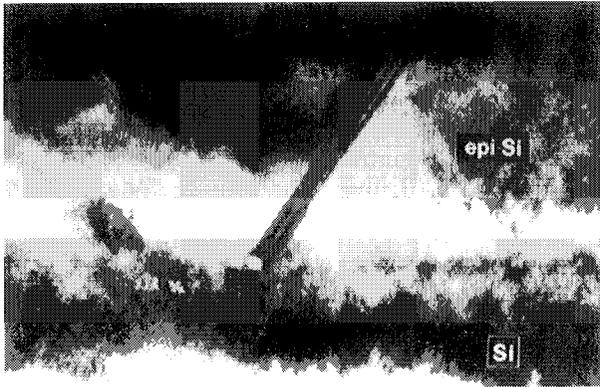


Figure 42: Cross-sectional TEM micrograph of an epitaxial silicon layer grown on a silicon substrate showing the presence of stacking faults caused by remnants of the native oxide.

Since the metallization systems used in device structures are polycrystalline, they may be evaluated by a combination of microdiffraction patterns and transmission electron micrographs. Figure 43 illustrates the nature of the electron diffraction pattern that may be observed.³⁸ Single crystalline samples produce ordered electron diffraction patterns which depend on the crystal structure of the system being studied. From the pattern it is possible to deduce the indices of the crystal plane giving rise to the diffraction spots. It is possible to experimentally determine the crystallographic direction of a dislocation or the plane of a stacking fault. As the sample becomes more disordered, the ordered diffraction pattern is accompanied by diffuse rings until only the diffuse rings appear for randomly oriented samples. Figure 44 illustrates how the electron diffraction patterns correlate with the TEM micrographs of thin films with an initial composition of $\text{Si/Ti} = 4.5$. The sample heated to 650°C shows very fine grain size and a diffuse diffraction pattern. Comparable samples heated to 850°C and 1050°C show grain growth and an increase in order in the diffraction pattern.⁴⁰

High temperature heating cycles employed in device processing frequency lead to precipitation in the lattice of single crystalline substrates. Depending on the nature of the precipitate and its location, it may have a beneficial or harmful role in device processing. Figure 45 shows the consequences of a precipitate.⁴¹ The precipitate grows, usually during thermal cycling, causing strain on the lattice. If the strain is sufficiently large, a stacking fault will propagate through the crystal lattice. The presence of precipitates may be minimized by lowering the device processing temperatures or by eliminating the elements which lead to precipitation. These frequently are transition metals which form complexes with the O present in Czochralski grown Si.

The composition of these precipitates may be evaluated by x-ray

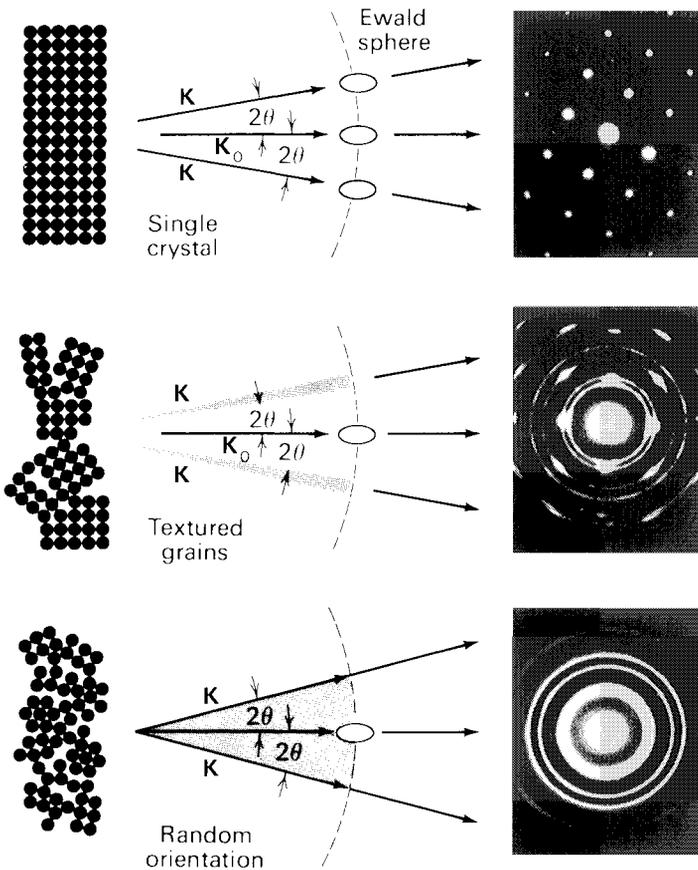


Figure 43: Diagram of the electron scattering that occurs from single crystalline, polycrystalline and randomly oriented films.

analysis or electron energy loss spectroscopy (EELS). A schematic illustration of an EELS apparatus is shown in Figure 46. After passing through the specimen, the electron energy is analyzed utilizing a magnetic sector electron spectrometer. A magnetic instrument is utilized since it is the only type of electron spectrometer with the resolving power to handle the high electron energies necessary for STEM analysis. The high energy primary electrons lose energy passing through the sample due to ionization of the energy levels of atoms present. This results in the loss peaks at discrete energy levels. EELS is best suited to the analysis of light elements, however, the analysis may be localized to the precipitate region.

Because of the perfection of silicon crystals STEM is of limited use until the crystal slices are heat treated as in diffusion or oxidation. However,

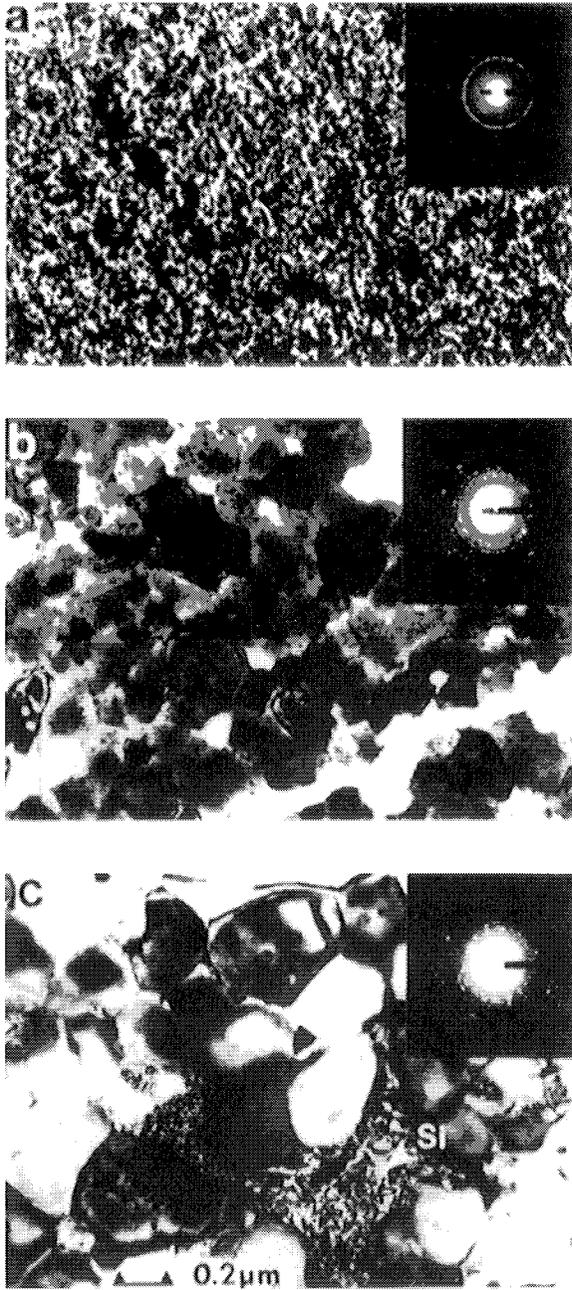


Figure 44: TEM micrograph and diffraction patterns of samples with initial composition $\text{Si/Ti} = 4.5$ reacted at (a) 650°C ; (b) 850°C and (c) 1050°C .

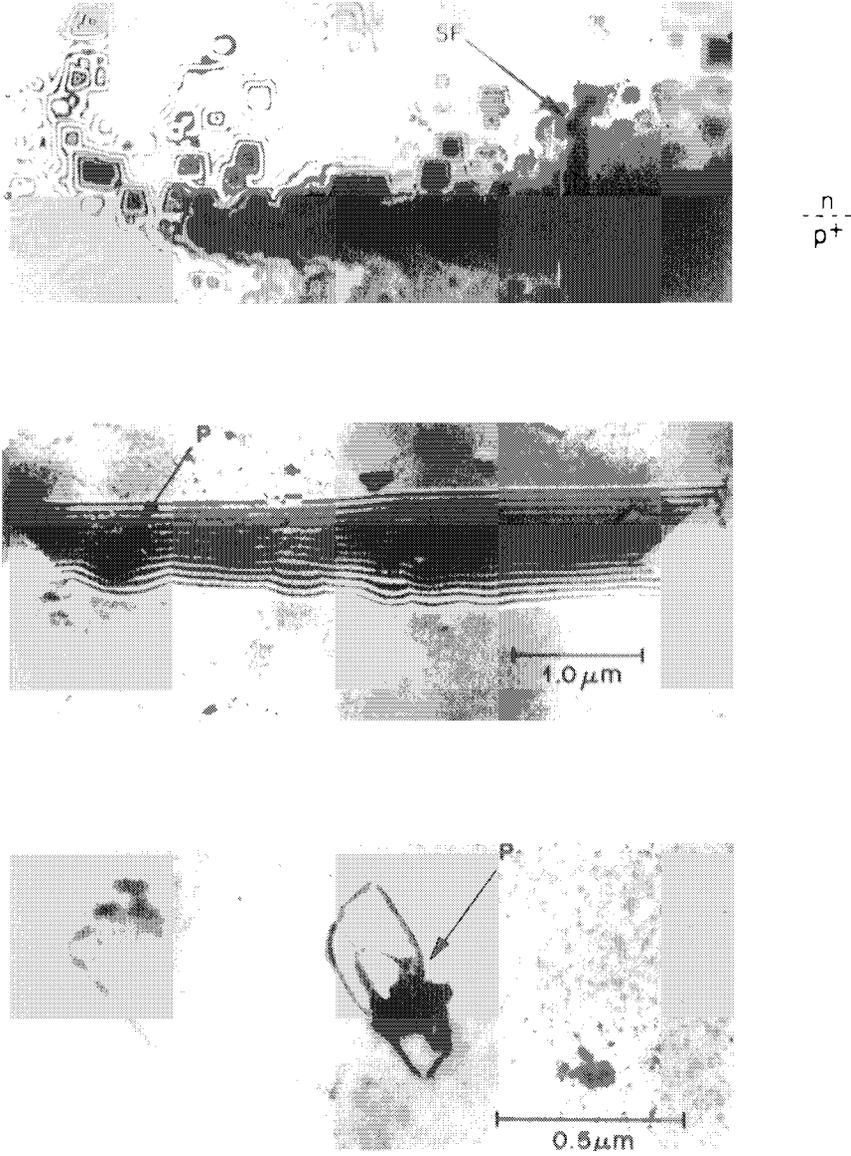


Figure 45: TEM micrographs of a decorated stacking fault with a precipitate found at a pn-junction with a leakage problem.

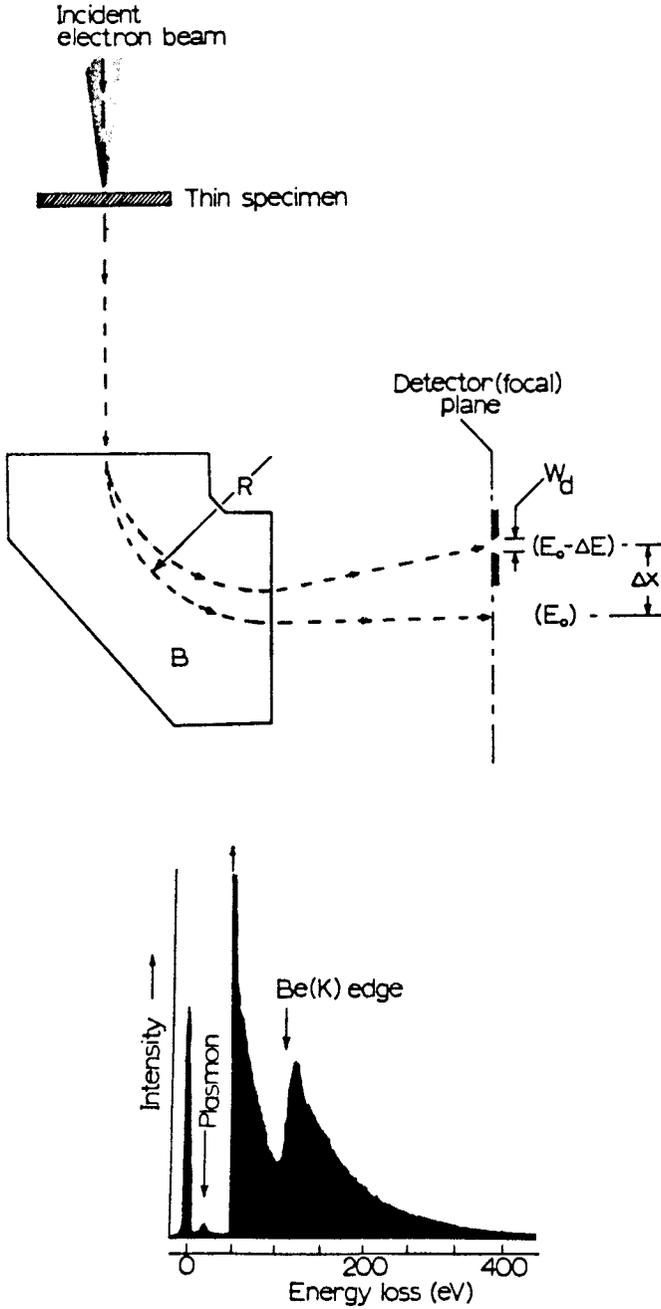


Figure 46: Schematic diagram of a magnetic sector electron energy loss spectrometer.

once defects, such as precipitates or stacking faults, are introduced then STEM is the best technique to understand the materials science involved. Defect analysis is accomplished through a combination of imaging, crystallographic and chemical analysis. The major weaknesses of STEM are the limited area of analysis, a few hundred micrometers, and the extremely time consuming sample preparation, 0.5-2 days. As a result, STEM is restricted in the range of applications where it may be applied on a routine basis.

X-ray Topography

X-ray topography (XRT) is a technique which provides a photographic image of the distribution of defects within a crystal. Topographs are recorded from either transmitted or reflected x-rays.⁴² Contrast on the photographic recording medium occurs from either orientation contrast, where a portion of the crystal is misaligned, or extinction contrast, where the lattice around a defect is distorted.

X-ray topography can image an entire wafer but at a resolution of not greater than approximately 1 micrometer under carefully controlled conditions. More typical values are in the 10-20 micrometer range. As illustrated in Figure 47, the x-ray source is positioned at a distance sufficiently far from the sample so that the x-rays appear to be collimated. The objective is to image x-rays that are diffracted by the single crystalline sample at a fixed Bragg angle. The sample is moved between a set of defining slits at the appropriate angle in parallel with x-ray sensitive film. When the proper Bragg angles are chosen the diffracted x-rays expose the film. If precipitates, stacking faults, or other crystal imperfections are present dark images will appear on the film due to the local change in the diffraction angle.

XRT is a non-destructive technique which may be utilized for process evaluation following a series of steps. Figure 48 shows a topograph which

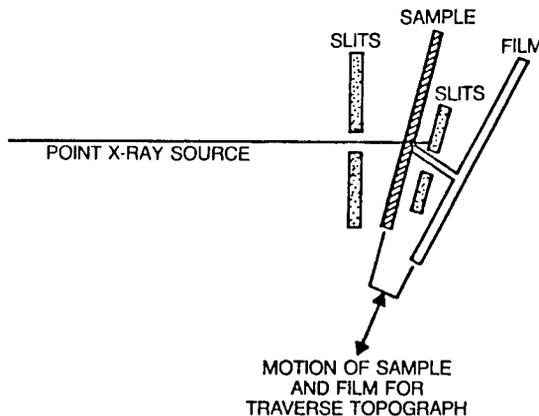


Figure 47: Schematic diagram of the equipment necessary to accomplish x-ray topography.

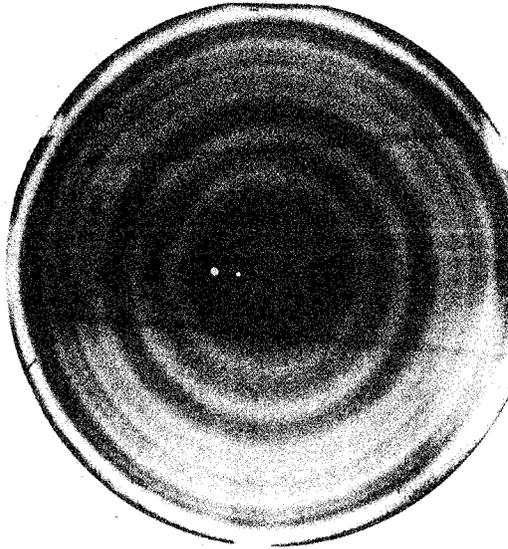


Figure 48: X-ray topograph of Czochralski grown silicon wafer showing radial distribution of oxygen precipitates.

was taken from a Czochralski (CZ) grown Si wafer. The CZ growth technique utilizes a quartz liner which gradually dissolves in the Si melt. The dissolved O precipitates following high temperature processing steps producing the swirl pattern observed in the XRT topograph. XRT may be utilized to monitor the crystal growth process, not realtime but after the wafers are processed.

XRT is used to identify or study process steps which introduce damage to the substrate. The x-ray topograph in Figure 49 shows stacking faults which extend from the identifying label generated on an Si wafer using laser writing. By analogy, one can imagine using XRT to study processes which generate damage to getter impurities or high temperature processes which invariably nucleate precipitates. Because the images of the defects are 100-1000 times larger than those obtained with electron microscopy, interpretation is difficult. XRT images only the strain not the defect causing the strain.

BULK ANALYSIS TECHNIQUES

Fourier Transform Infrared Spectroscopy

Infrared absorption spectroscopy has been utilized for many years in the evaluation of semiconductor materials. With the advent of Fourier transform infrared (FTIR) spectroscopy, which has better sensitivity than older grating instruments, infrared spectroscopy has seen a resurgence in

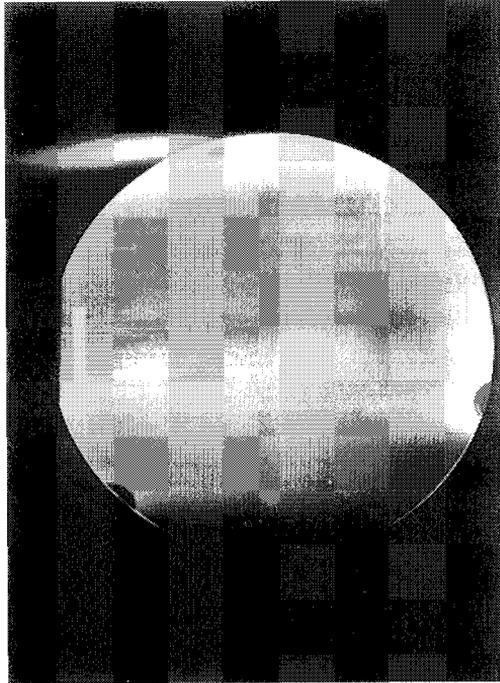


Figure 49: X-ray topograph of silicon wafer following laser scribing showing the presence of stacking faults.

interest. One version of FTIR spectroscopy is shown schematically in Figure 50. Radiation from the IR source passes through a beam splitter onto a fixed and movable mirror. The IR radiation is recombined in an interference pattern which is determined by the position of the movable mirror. Some of the IR radiation passing through the sample is absorbed before striking the detector. The resolution and accuracy of FTIR is strongly dependent on the repeatability and accuracy of positioning the mirror.

FTIR is not only a sensitive bulk analysis technique but also capable of determining the net and total impurity concentration. The energy level diagrams provided in Figure 51, provide some insight into the IR absorption process.⁴³ At room temperature, shallow donor and acceptor impurity levels will be ionized. By lowering the sample temperature to liquid He temperature, the ionized states will be depopulated. Electrons are excited to bound states when IR radiation is absorbed with no energy greater than or equal to the bandgap. The IR radiation absorbed is proportional to the net impurity concentration. If the sample is simultaneously irradiated with a second photon source with energy greater than the bandgap, the IR radiation absorbed is proportional to the total impurity concentration.

Table 6 lists the strongest IR absorption lines and sensitivity for

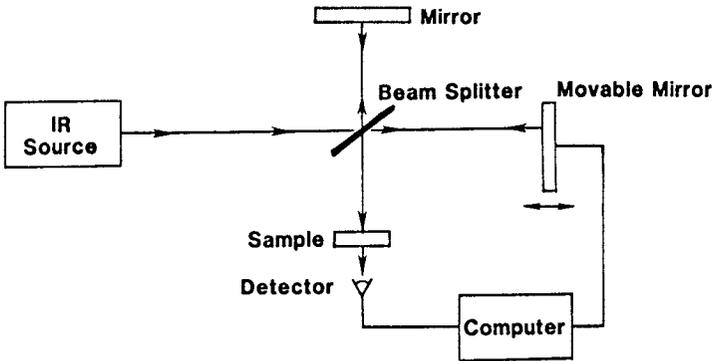


Figure 50: Schematic diagram of a Fourier Transform infrared spectrometer.

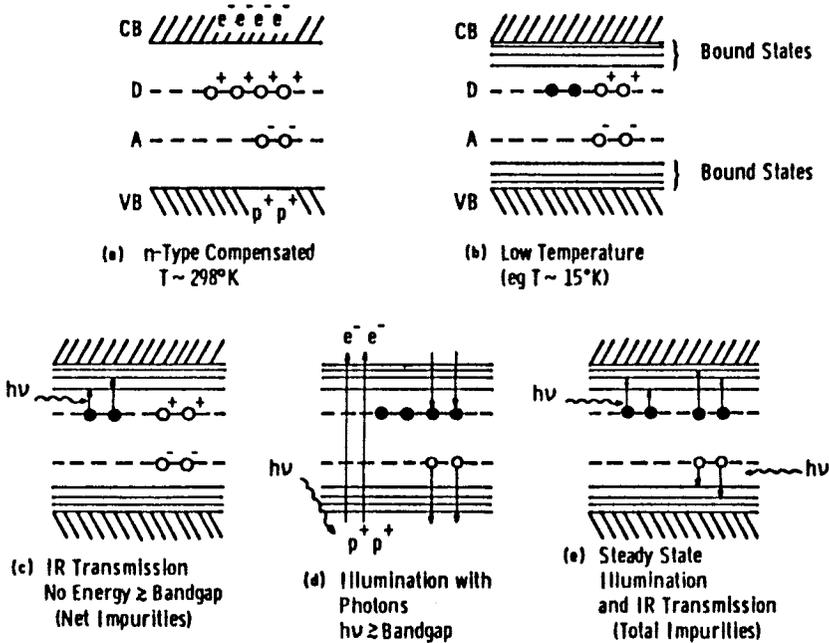


Figure 51: Energy level diagram illustrating the population of donor and acceptor levels in the band gap of silicon at room temperature and 15°K.

impurities commonly found in Si.⁴³ The stated sensitivities for all the impurities except C and O are for a sample temperature of 12°K. Since C and O are not electrically active in Si, there is no significant gain in sensitivity obtained by lowering the sample temperature. The FTIR sensitivity is enhanced by using a thick specimen, 5 mm, since it is proportional

to optical path length. Thinner specimens may be used with a proportionate drop in sensitivity. Quantitative analysis is based upon the measured percent transmission at the characteristic wavelength compared to standards.

Even though there is not a significant gain in sensitivity for C and O at lower sample temperatures, the IR absorption bands are sharper as shown for O in Figure 52. The sharper bands make it easier to distinguish O atoms

Table 6: FTIR Absorption Line Frequencies and Sensitivities

Impurity	Strongest Absorption Line (cm ⁻¹)	Sensitivity (Atoms/cm ³)
Phosphorus	315.9	1.5 x 10 ¹¹
Arsenic	382.2	2.7 x 10 ¹¹
Antimony	293.6	3.1 x 10 ¹¹
Boron	319.7	3.1 x 10 ¹¹
Aluminum	471.7	2.1 x 10 ¹²
Gallium	548.0	2.7 x 10 ¹²
Indium	1175.9	1.6 x 10 ¹³
Carbon	605.0	2.5 x 10 ¹⁶
Oxygen	1136.0	2.5 x 10 ¹⁵

Sample Thickness - 5 mm

Sample Temperature - 12° K except Carbon and Oxygen which were at room temperature

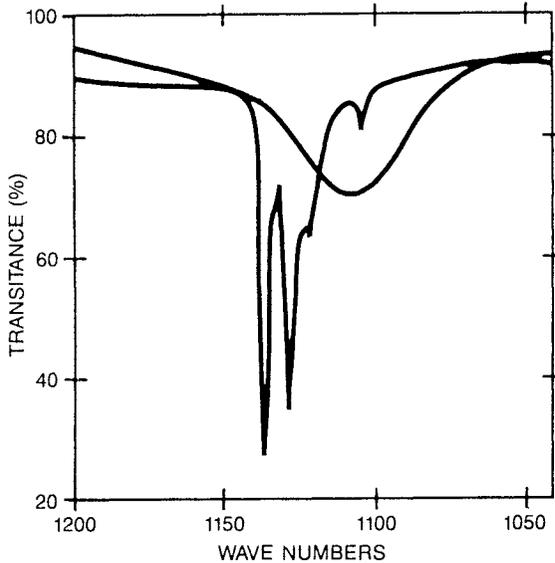


Figure 52: Infrared absorption spectrum of oxygen in silicon at 300° and 55°K.

that occupy different lattice sites. The use of FTIR to study the interstitial and substitutional O concentrations as a function of the Si substrate thermal history has been a major tool in the development of process parameters.⁴⁴

FTIR is a powerful tool for the investigation of compound semiconductor materials, in addition to Si. The spectra, however, tend to be more complicated since the impurity atom can occupy multiple sites. Figure 53 illustrates this point with the example of GaAs doped with ^{28}Si and ^{30}Si . Both the ^{28}Si and ^{30}Si can occupy Ga and As sites resulting in four possible IR peaks.⁴⁵

FTIR has also been used in the determination of Si epitaxial layer thicknesses in the range from 1 micrometer to 200 micrometers. The epitaxial layer thickness is proportional to the number of interference fringe spacings which occur when IR radiation is reflected off of the

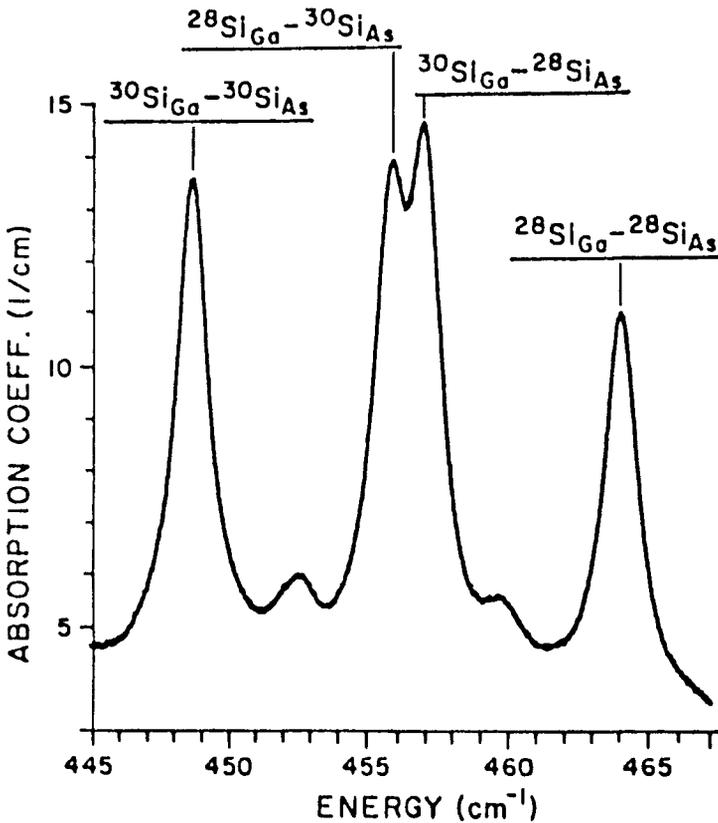


Figure 53: Infrared absorption spectra at liquid nitrogen temperature for GaAs sample doped with both ^{30}Si and ^{28}Si showing the isotopic combinations for the pair absorption.

interface between an epitaxial layer and a Si substrate with an approximate 10^2 higher dopant density. The relationship is defined as follows:

$$T = \frac{M}{2N(\nu_2 - \nu_1)} \quad (4)$$

where T is the epitaxial layer thickness, M the number of interference fringe spacings, N the refractive index and ν is the wavelength in cm^{-1} . Since FTIR analysis is non-destructive, it has become a routine analysis tool for epitaxial layer thicknesses.

Deep Level Transient Spectroscopy

The study of the electrical properties of matter by capacitance measurements is a well established technique. Many semiconductors do not readily lend themselves to such methods since they are too highly conductive except in the form of reverse biased p-n junctions. The reverse bias creates a depletion layer which is analogous to a parallel plate capacitor. The basis of the technique is the measurement of the junction capacitance C and slope $\Delta C/\Delta V$ as the reverse bias is increased. In the most usual form of measurement ΔC and ΔV come from point measurements of C versus V . The capacitance may also be measured at some fixed voltage by superimposing a small oscillating voltage on the applied bias. The depletion width remains approximately constant and the junction capacitance, $C = \Delta Q/\Delta V$, is determined by measuring the current induced by the oscillating voltage. When deep levels are present the oscillating voltage will uncover charge associated with deep levels.⁴⁶

Deep states are those which are positioned deep in the band gap as contrasted to shallow states which are located near the band edges. Deep defect states are referred to as traps, recombination centers, generation centers, deep levels and deep impurities. These terms often have precise meanings but a particular defect state may be both a trap and a recombination center depending on such variables as doping and temperature.

Because the rates of defect state capture and emission processes are temperature dependent, variation of the junction temperature, while monitoring the junction capacitance at constant frequency, can often give the same information as frequency response. However, temperature variation also effects the equilibrium Fermi level, E_F . As the temperature is increased E_F may cross a defect state resulting in a capacitance step. The emission rate and Fermi level effects are mutually exclusive. The temperature dependent measurement is, therefore, inherently ambiguous which complementary frequency response measurements may resolve.⁴⁷

Electronic transmissions within the space charge region consist entirely of emission processes. Because states within the space charge region have no possibility of being filled by capture processes, emission processes can only be observed following the forced introduction of carriers which are captured by defect states. Initially the junction is reverse biased to establish space charge conditions. States in the space charge region are empty because no mobile carriers are available for capture. A bias pulse

toward zero bias will momentarily collapse the space charge region, making majority carriers available for capture. When the pulse is turned off and V_R re-established, the junction capacitance is reduced because compensating majority carrier charge has been trapped in the space charge region. This charge can subsequently be excited and swept from the space charge region by the applied junction potential.

Alternatively, a pulse into forward bias can be used to inject a minority carrier population. However, any injection pulse using either optical or voltage excitation introduces both majority and minority carriers. The relative ratio of minority to majority carriers may be varied by varying the magnitude of the injected current.

Defect states which are filled during the excitation pulse will return to their initial state if sufficient excitation energy, thermal or optical, is present. Such processes can be followed as junction capacitance transients. The strength of the method is that each defect may be studied independently through its unique activation energy for carrier emission, cross-section for carrier capture and cross-section for optical excitation.

The capture of carriers normally proceeds exponentially during an injection or zero bias pulse. By measuring the capacitance as a function of pulse duration, it is easy to obtain the capture rate from the slope of $\log [C_\infty - C_{(T)}]$ vs T where T is the duration of the pulse. For majority carrier capture, the cross-section is obtained from the expression

$$C = \sigma \langle V_N \rangle n \quad (5)$$

where, for example, n is the concentration of electrons and $\langle V_N \rangle$ is the average thermal velocity of electrons.

The main advantage of Deep Level Transient Spectroscopy (DLTS) is its high sensitivity to electrically active traps.⁴⁸ The expression

$$N_T = \frac{\Delta C}{2C} \cdot N_D \quad (6)$$

where N_D is the background doping concentration, provides an approximation to the trap concentration. For 4 ohm/cm material, N_D is approximately 10^{15} . A change in capacitance of 0.1 fF may be determined for a C of 10pF with modern instrumentation. From the above expression a sensitivity of the order of $10^{10}/\text{cm}^3$ may be obtained by DLTS for electrically active traps. DLTS has the disadvantages that all traps are not electrically active or necessarily unambiguous. In addition, it requires device fabrication, is sensitive to doping concentration and has poor spatial resolution. Yet, DLTS is one of the most sensitive tools available for the evaluation of semiconductor materials.

Photoluminescence Spectroscopy

Photoluminescence (PL) Spectroscopy is a measure of the intensity of radiation versus wavelength emitted as a result of radiative recombination of electron-hole pairs or excitons excited from their thermal equilibrium states by optical excitation.⁴⁹ An electron-hole pair excited from the

ground state can recombine radiatively through various kinds of recombination processes, as shown schematically in Figure 54. The most simple recombination process is band-to-band recombination where a free electron excited in the conduction band recombines radiatively with a free hole excited in the valence band. Impurities which introduce traps, donor or acceptor levels, in the band gap provide alternate paths for de-excitation. When an electron or hole is captured by a trap center and then the trapped carrier recombines radiatively with the remaining carrier, this is called band-to-impurity recombination. When both the excited electron and hole are captured by different trap centers and then the trapped electron and hole recombine radiatively, this is known as donor-acceptor pair recombination.

At low temperatures, a generated electron-hole pair becomes an exciton. An exciton is a complex with an electron and a hole bound together by Coulomb attraction which can move freely as a quasiparticle in a semiconductor crystal. These free excitons decay to the ground state through free-exciton (FE) recombination accompanied by luminescence. Impurity-exciton complexes are formed when free excitons are bound to impurity centers. Bound excitons (BE) radiatively decay at just below the free-exciton energy.

Analysis of PL spectra provides a large amount of information about semiconductors and the impurities and defects that are present in them. The most effective application of PL is the identification of shallow impurities. This is accomplished by measuring the characteristic positions of the BE luminescence lines at low temperature. The spectral positions will differ depending on the impurity, while the intensity is related to the concentration. PL has been used in the analysis of elements such as B, P, Al, As, and N in Si in the concentration range of $10^{11} - 10^{15}$ atoms/cm³. It has also been used to study C, Si, Mn, Mg, Te, etc., in GaAs down to 10^{13} atoms/cm³.⁵⁰

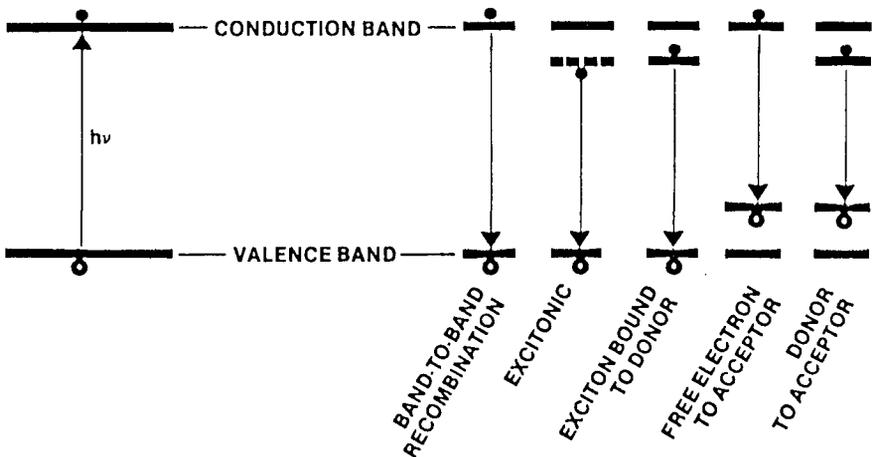


Figure 54: Diagram of the possible photoluminescence transitions.

The PL intensity is not directly related to shallow-impurity concentration because of competing non-radiative decay processes for the BE. The intensity also depends on the excitation level. It has been found empirically that good correlation can be obtained between impurity levels determined by electrical measurements and the intensity ratio of the BE to FE when recorded at moderate excitation levels. Measurement of the intensity ratios minimizes the influence of variables dependent on the crystal growth and process conditions. Tajima has obtained the calibration curves shown in Figure 55 for B and P in float-zone refined Si.⁵¹ The concentration range between 10^{11} to 10^{15} atoms/cm³ represents the practical range over which PL may be applied to Si.

Room temperature PL due to band-to-band recombination can be utilized to characterize thermally induced defects in Si. Some thermally induced defects in Si act as non-radiative recombination centers which trap excess carriers. The presence of such non-radiative recombination

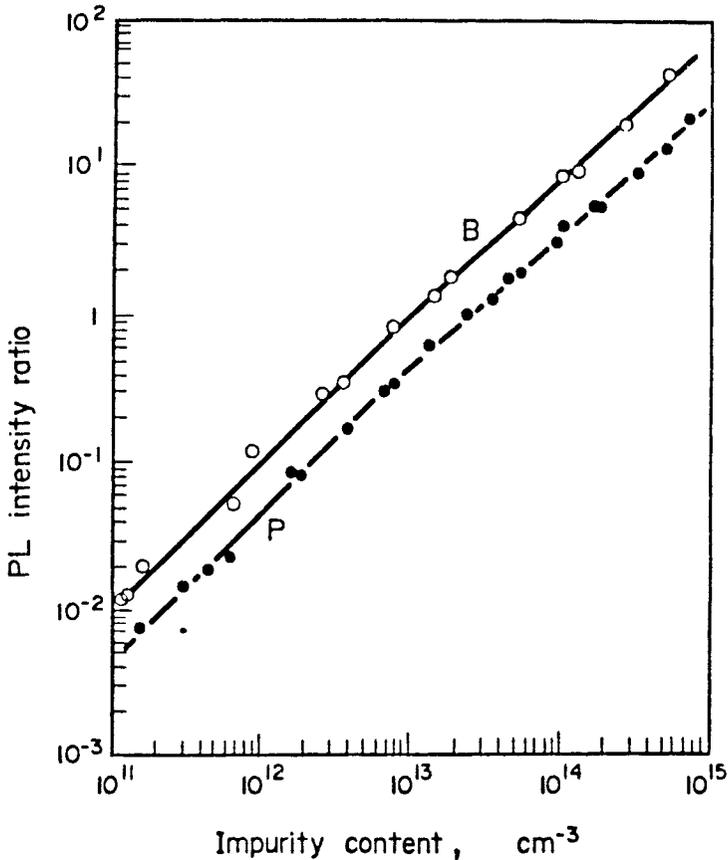


Figure 55: Calibration curves for the P donor and B acceptor concentrations in Si from analysis of the BE and FE photoluminescence intensity ratios.

centers leads to a reduction in the PL intensity. Figure 56 shows the dependence of the PL intensity on annealing temperatures for samples held at liquid helium and room temperature.⁵² The structure and intensity of the excitonic peak at low temperature changes little as a result of annealing. The intensity of the broad band-to-band recombination peak decreases rapidly with increasing annealing temperature. This rapid decrease in intensity is attributed to the increase in the thermally induced non-radiative recombination centers in the crystal. It is believed that the thermally induced defects are induced by oxygen precipitates since a strong correlation was found between the etch-pit density and the PL intensity, Figure 57.⁵²

PL is a non-destructive technique which requires minimal sample preparation. It is restricted to analysis of single crystalline wafers or epitaxial layers. The sampling depth is approximately 3 micrometers, the optical attenuation length. Through the use of laser excitation spatial

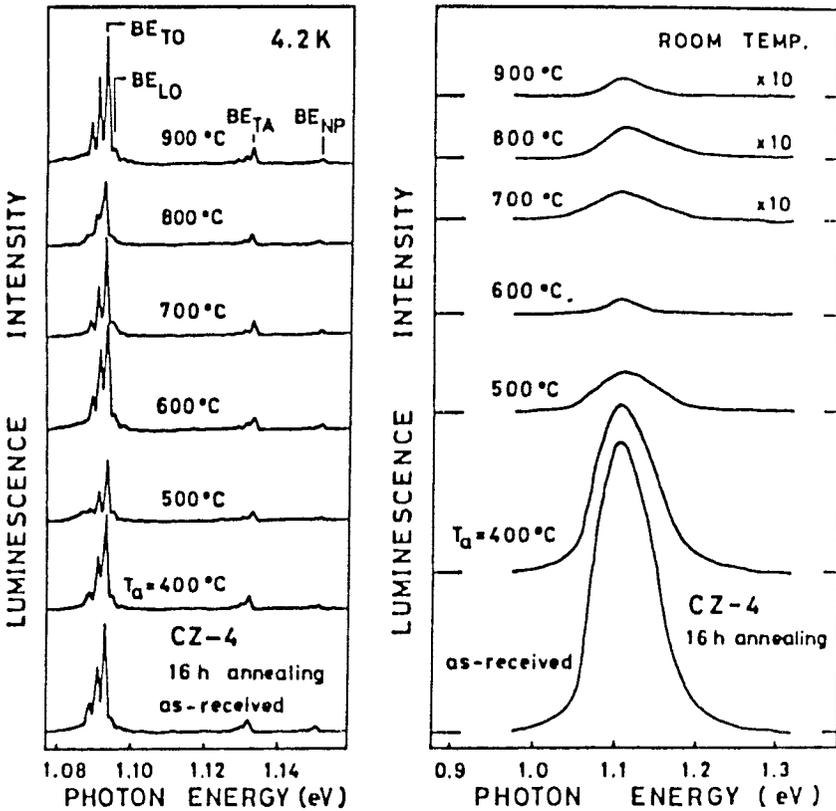


Figure 56: PL spectra measured at liquid helium temperature (a), and at room temperature (b) as a function of annealing temperature.

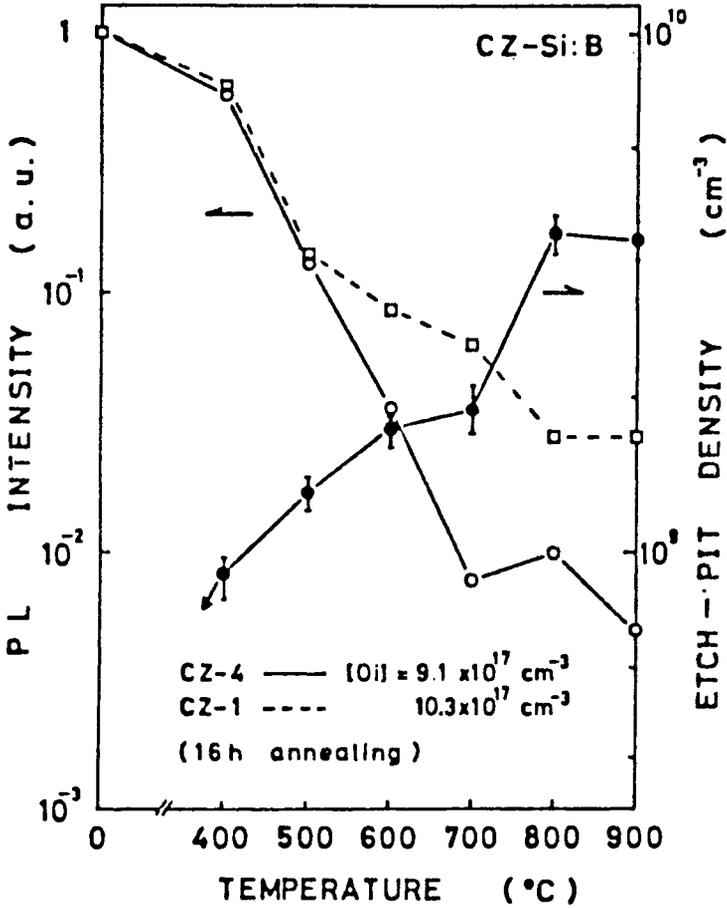


Figure 57: Relation between room temperature PL intensity and etch-pit density estimated by the Wright-etch method.

resolution of 1 micrometer can be achieved which may be used to map the PL intensity distribution of the surface.

PL is limited to the analysis of shallow acceptors in GaAs since the binding energies of donors are too small to experimentally resolve. It is particularly difficult to analyze the dopant impurity concentration in GaAs by PL because there are many competing recombination channels due to the presence of a high background of impurities and defects.

Neutron Activation Analysis

Neutron activation analysis (NAA) is accomplished by placing the sample in the thermal neutron flux available in the core of a nuclear reactor. Radioactive isotopes are produced through nuclear reactions. The most common reaction (n, γ) produces an isotope one neutron heavier than the

target isotope. The intensity of the beta or gamma radiation from the radioactive product is measured as a function of energy. Isotope identification is based on the characteristic energy of the beta or gamma radiation. The calculations upon which NAA are based yield the number of impurity atoms in the sample. A knowledge of sample volume yields concentration.

Table 7: Neutron Activation Analysis of Silicon

 HALF-LIFE OF $^{31}\text{Si} = 2.6 \text{ hr}$

<u>Element</u>	<u>Sensitivity (Atoms/cm³)</u>	<u>Factor Affecting Sensitivity</u>
B	—	Short Half Life $^{12}\text{B} = 20 \text{ ms}$
Na	2×10^{13}	
Al	—	Short Half Life $^{28}\text{Al} = 2.3 \text{ s}$
Fe	3×10^{15}	Low Isotopic Abundance
Zn	2×10^{14}	Long Half-Life $^{65}\text{Zn} = 235 \text{ days}$
As	1×10^{13}	{ Medium Half-Life Predominately Gamma-Ray Emitters Large Cross-Section
Ag	7×10^{12}	
Au	5×10^{10}	

Table 8: Neutron Activation Analysis of GaAs

 Half-Life Of $^{76}\text{As} = 24 \text{ hr}$: $^{72}\text{Ga} = 14 \text{ hrs}$

<u>Element</u>	<u>Sensitivity (Atoms/cm³)</u>	<u>Factor Affecting Sensitivity</u>
Na	—	Half-Life $^{24}\text{Na} = 15 \text{ hrs}$
Fe	5×10^{15}	Low Isotopic Abundance
Zn	5×10^{14}	
Si	—	Half-Life $^{31}\text{Si} = 2.6 \text{ hrs}$
S	—	^{35}S has No Gamma-Rays
Se	4×10^{14}	
Cr	1×10^{16}	Low Isotopic Abundance

Table 9: Carbon and Oxygen Content of Commercially Available Silicon (Using Charged Particle Accelerators)

<u>MATERIAL</u>	<u>CONCENTRATION (X10¹⁷ at/cm³)</u>	
	<u>C A R B O N</u>	<u>O X Y G E N</u>
FLOAT ZONE (prepared under Argon)	0.08 - 0.35	0.05 - 0.2
FLOAT ZONE (prepared in vacuum)	0.03 - 0.4	0.01 - 0.2
CZOCHELSKI (grown under Argon)	0.3 - 3.0	2 - 10
POLYCRYSTALLINE	0.2 - 0.8	0.8 - 3

The measure of the analyzed volume or area is often the major limitation to accuracy and precision.⁵³

The detection limit for NAA is 5×10^{10} atoms/cm³ for Au in Si. This is the ideal case where Au has a large cross-section for activation and is predominantly a gamma-ray emitter while in a host matrix with a short half-life. Table 7 lists elements that were detected by NAA in Czochralski grown Si along with the sensitivity. Factors which affect sensitivity adversely include long half-life which lowers the count rate, short half-life where the isotope decays before sufficient time is available for getting the sample into the counting chamber and low isotopic abundance. Table 8 provides a similar listing of elements that were detected in semi-insulating GaAs. The application of NAA to GaAs is less favorable than to Si, since As and Ga have significantly longer half-lives.

NAA is one of the most sensitive and accurate bulk analysis techniques. Although it is not always readily accessible, it is frequently used to calibrate other analytical techniques. One major weakness is the insensitivity of NAA for many for light elements.

Activation analysis can be extended to light elements through the use of accelerators as was described for NRA. Carbon and oxygen in Si were studied using charged particle activation analysis.⁵⁴ The results, verify that the C and O content of float zone grown Si is lower than that of Czochralski grown material. Charged particle activation analysis like NAA may not be easily accessible but it is frequently used as a technique to calibrate other analytical tools.

REFERENCES

1. G.E. McGuire and P.H. Holloway, "Applications of Auger Spectroscopy in Materials Analysis", in *Electron Spectroscopy: Theory, Techniques and Appli-*

- ations, Vol. 4, C.R. Brundle and A.D. Baker, editors, Academic Press, New York (1981).
2. L.A. Harris, *J. Appl. Phys.* 39:1419 (1968).
 3. N.E. Erickson and C.J. Powell, *Surf. Interface Anal.* 9:111 (1986).
 4. M.P. Seah and W.A. Dench, *Surf. Interface Anal.* 1:1 (1979).
 5. S. Hofmann, "Depth Profiling" in *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, D. Briggs and M.P. Seah, editors, John Wiley and Sons, New York (1983).
 6. V.J. Kapoor, R.S. Bailey and S.R. Smith, *J. Vac. Sci. Technol.* 18:305 (1981).
 7. G.P. Schwartz, G.T. Gualtieri, G.W. Kammlott and B. Schwartz, *J. Electrochem. Soc.* 126:1737-1749 (1979).
 8. M.T. Thomas, D.R. Baer, R.H. Jones and S.M. Bruemmer, *J. Vac. Sci. Technol.* 17:25-28 (1980).
 9. G.E. McGuire, W.R. Wisseman, and P.H. Holloway, *J. Vac. Sci. Technol.* 15:1701 (1978).
 10. C.C. Chang, S.P. Murarka, V. Kumar, and G. Quintana, *J. Appl. Phys.* 46:4237 (1975).
 11. M.P. Seah, *Surf. Interface Anal.* 9:85 (1986).
 12. C.S. Fadley, "Basic Concepts of X-ray Photoelectron Spectroscopy" in *Electron Spectroscopy: Theory Techniques and Applications*, Vol. 2, C.R. Bundle and A.D. Baker, editors, Academic Press, New York (1978).
 13. U. Gelius, L. Asplund, E. Basilier, S. Hedman, K. Helenelund and K. Siegbahn, *Nucl. Inst. a. Meth.* B1:85 (1984).
 14. I.W. Drummond, T.A. Cooper and F.J. Street, *Spectrochim. Acta* 40B:809 (1985).
 15. C.D. Wagner, W.M. Riggs, L.E. Davis, J.F. Moulder and G.E. Muilenberg, *Handbook of X-ray Photoelectron Spectroscopy*, Perkin-Elmer Corp., Minnesota (1978).
 16. W.E. Swartz, K.J. Wynne and D.M. Hercules, *Anal. Chem.* 43:1884 (1971).
 17. P.J. Grunthaner, F.J. Grunthaner, and A. Madhukar, *J. Vac. Sci. Technol.* 20:680 (1982).
 18. C.D. Wagner, L.E. Davis, M.V. Zeller, J.A. Taylor, R.M. Raymond and L.H. Gale, *Surf. Interface Anal.* 3:211 (1981).
 19. C.W. Magee and R.E. Honig, *Surf. Interface Anal.* 4:35 (1982).
 20. C.A. Andersen and J.R. Hinthorne, *Anal. Chem.* 45:1421-1438 (1973).
 21. C.W. Magee, *J. Electrochem. Soc.* 126:660-663 (1979).
 22. C.W. White, J. Narayan and R.T. Young, "Laser Solid Interactions and Laser Processing", *AIP Conference Proceedings* 50, D. Ferris, H.J. Leamy and J.M. Poate, editors, American Institute of Physics, New York (1975).
 23. D.J. Coleman, D.W. Shaw and R.D. Dobrott, *J. Electrochem. Soc.* 124:239-241 (1977).
 24. C.H. Becker and K.T. Gillen, *J. Vac. Sci. Technol.* 3A:1347 (1985).
 25. J.E. Parks, D.W. Beekman, H.W. Schmitt and E.H. Taylor, *Nucl. Inst. Meth. in Phys. Res.* B10/11:280-284 (1985).
 26. G.S. Hurst, M.G. Payne, S.D. Kramer and J.P. Young, *Rev. Mod. Phys.* 51:767 (1979).
 27. S. Mayo, T.B. Lucatorto and G.G. Luther, *Anal. Chem.* 54:553-556 (1982).
 28. W.K. Chu, J.W. Mayer, and M.A. Nicolet, *Backscattering Spectrometry*, Academic Press, New York (1978).
 29. L.C. Feldman, J.W. Mayer, and S.T. Picraux, *Materials Analysis By Ion Channeling*, Academic Press, New York (1982).
 30. C.D. Lien, M.A. Nicolet and S.S. Lau, *Phys. Stat. Sol.* 81:123-128 (1984).
 31. C.R. Helms, *J. Vac. Sci. Tech.* 20:950 (1982).
 32. J. Narayan and O.W. Holland, *J. Electrochem. Soc.* 131:2651-2662 (1984).
 33. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, C. Fiori and E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press, New York (1981).

34. T. Koshikawa and R. Shimizu, *J. Phys. D: Appl. Phys.* 7:1303 (1974).
35. K.F.J. Heinrich, *X-Ray Optics and Microanalysis, 4th Int. Cong. on X-ray Optics and Microanalysis*, R. Castaing, P. Deschamps, and J. Philibert, editors, Paris, p. 1509 (1966).
36. K. Kanaya and S. Okayama, *J. Phys. D: Appl. Phys.* 5:43 (1972).
37. H.P. Feuerbaum, *Scanning Electron Microscopy*, SEM Inc., AMF O'Hare, Illinois, 285-296 (1979).
38. L.E. Murr, *Electron and Ion Microscopy and Microanalysis*, Marcel Dekker, Inc., New York (1982).
39. J.C. Bravman, G.L. Patton and J.D. Plummer, *J. Appl. Phys.* 57:2279-2782 (1985).
40. R. Beyers and R. Sinclair, "The Effect of Oxygen in Cosputtered Films", in *Defects in Semiconductors II*, S. Mahajan and J.W. Corbett, editors, North-Holland, New York 423-427 (1983).
41. R.B. Marcus, "Correlation of Structure and Properties of Silicon Devices", in *Defects in Semiconductors II*, S. Mahajan and J.W. Corbett, editors, North-Holland, New York, 441-452 (1983).
42. B.K. Tanner and D.K. Bowen, *Characterization of Crystal Growth Defects by X-ray Methods*, Plenum Press, New York (1980).
43. C.S. Baber, *Thin Solid Films* 72:1332 (1980).
44. L. Jastrzebski, P. Zangucchi, D. Thebault and J. Lagowaki, *J. Electrochem. Soc.* p. 1639 (1982).
45. W.M. Theis and W.G. Spitzer, *Applied Phys. Lett.* 41:71 (1982).
46. G.L. Miller, D.V. Lang and L.C. Kimerling, *Ann. Rev. Mater. Sci.* 7:377-448 (1977).
47. C.T. Sah and J.W. Walker, *Appl. Phys. Lett.* 22:384 (1973).
48. D.V. Lang, *J. Appl. Phys.* 45:3014-3023 (1974).
49. T. Nishino, H. Nakayama and Y. Hamakawa, "Photoluminescence Characterization of Impurities and Defects in Semiconductors", *Japan Annual Reviews in Electronics, Computers and Telecommunications: Semiconductor Technologies*, Vol. 8, J. Nishizawa, North Holland, p. 295 (1985).
50. K.K. Smith, *Thin Solid Films* 84:171 (1981).
51. M. Tajima, *Appl. Phys. Lett.* 32:719 (1978).
52. J. Katsura, H. Nakayama, T. Nishino, and Y. Hamakawa, *Japan J. Appl. Phys.* 21:712 (1982).
53. J.A. Keenan and G.B. Larrabee, *Chem. Inst.* 3:125 (1971).
54. J.L. Debrun, *Prog. in Cryst. Grow. Char.* 8:115 (1984).

Acknowledgements

Figures 2 and 9 are reprinted by permission of The Physical Electronics Division of Perkin-Elmer Inc. Figure 3 is reprinted by permission of the publisher, John Wiley and Sons Ltd. Figures 5, 7, 14, 21, 23, 53 and 55 are reprinted by permission of the publisher, The American Institute of Physics. Figures 6, 13, 22 and 32 are reprinted by permission of the publisher, The Electrochemical Society. Figure 10 is reprinted by permission of Surface Science Laboratories. Figure 18 is reprinted by permission of the publisher, The American Chemical Society. Figures 24, 26 and 45 are reprinted by permission of the publisher, Elsevier Science Publishers. Figure 25 is reprinted by permission of the publisher, The American Physical Society. Figures 28 and 30 are reprinted by permission of the publisher Academic Press. Figures 38, 41, 43 and 46 are reprinted by permission of the publisher Marcel-Dekker Inc. Figure 53 is reprinted by permission of Digilab Division of Bio-Rad Laboratories. Figures 56 and 57 are reprinted by permission of the publisher, The Japanese Journal of Applied Physics.

Index

- Acceptors - 27
- Adsorption
 - adatom - 412, 421
 - binding energy - 411
- Amphoteric - 362
- Atomic mobility - 96, 102
- Auger electron spectroscopy (AES) - 6, 611-617
 - analysis - 67, 70, 178, 294
 - chemical shift - 614
 - cylindrical mirror analyzer - 613
 - detection limit - 616
 - energy level diagram - 611
 - profiling - 66, 614
- Bipolar devices - 65, 87, 576, 581, 587, 590, 603, 605
- Boltzmann constant - 464
- Capacitance, parasitic - 106, 149, 592
- Capacitance-voltage technique - 59, 71
- Carrier
 - concentration - 27
 - velocity - 27
- Channel - 1, 4
 - current - 4
- Charge - 58, 59
 - fixed oxide - 58
 - mobile ionic - 58
 - oxide trapped - 59
- Chemical vapor deposition (CVD) - 6, 80
 - boron nitride - 115, 122
 - polycrystalline silicon - 106
 - Si - 82, 94
 - Silicon carbide - 106, 115, 122
 - SiO₂ - 75, 106, 115
 - Si₃N₄ - 75, 105, 106, 115
- Chemisorption - 333, 421
- Child-Langmuir Law - 209, 397
- CMOS - 89
- Collector - 111
- Condensation - 410
- Conductivity - 308
- Crucible - 12, 19, 20
- Crystal Growth
 - Czochralski - 9
 - float-zone - 9, 20, 475
 - seed crystal - 12, 82
- Crystallographic orientation - 84, 133

- Crystal puller - 11
- Deep level transient spectroscopy (DLTS) - 7, 71, 659-660
- Defect - 6, 94
 - density - 55, 67, 86, 172
 - monovacancy - 473
 - point - 421, 455, 470, 472-476, 490, 498, 500
 - saucer pits - 33
- Denuded zone - 42, 91, 92, 173
- Deposition
 - orientation dependent - 149, 172
- Desorption
 - association reaction - 421
 - ion-induced - 367
 - mean residence time - 411
- Diamond lattice - 26
- Die - 22
- Dielectric
 - breakdown - 67
 - constant - 4, 60
 - films - 47, 48, 58
 - interlevel - 440, 592
 - isolation - 106, 109, 112, 136, 155
- Diffusion - 6, 455-539, 595-598
 - activation energy - 464, 469, 486, 488, 491
 - Arrhenius expression - 469, 616
 - barrier - 55, 439, 602-603, 606
 - bulk - 415, 616
 - coefficient - 69, 463, 495
 - continuum theory - 456-464
 - diffusivity - 469-471, 476-477, 482, 487, 495, 501, 596
 - dopant - 67, 69
 - Fick's Laws - 456, 495
 - flux - 466
 - grain boundary - 596-599, 606, 616
 - Henry's Law - 460
 - interstitialcy - 466, 476-486, 500
 - oxidation enhanced - 70, 479-485
 - oxygen - 40, 50, 67, 74
 - predeposition - 459-462
 - surface - 415
 - vacancy - 465, 491, 496, 501
- Diode - 96
- Dislocation - 6, 13, 23, 24, 32, 475
 - misfit - 94, 178
- Donor - 27
 - oxygen - 40
- Dopant - 15, 18, 19, 20, 21, 63
 - MBE films - 361-362
 - N-type - 15, 22, 61
 - P-type - 15, 22, 61
 - segregation - 61
 - substitutional - 477, 488
- Drain current - 4
- Dynamic random access memory (DRAM) - 4
- Eddy current sensor - 22
- Effusion cell - 426
- Electric field - 73
- Electromigration - 6, 604-606
- Electronic printer printhead - 112, 114
- Electron paramagnetic resonance (EPR) - 473
- Electron spin resonance - 70
- Ellipsometry - 55, 70
- Enthalpy - 468, 488
- Entropy - 468
- Epitaxial
 - homoepitaxy - 84
 - heteroepitaxy - 84
 - silicon - 22, 82, 84, 85, 86, 89, 90, 91, 178, 183
- Etching - 126
 - anisotropic - 6, 109, 112, 126, 136, 138, 145, 149, 157, 158, 162, 172
 - chemical - 23, 24, 64
 - choline - 187, 189
 - cleanup - 186, 187
 - concentration dependent - 127
 - Dash - 173
 - defect etch - 91, 92, 128, 172
 - planar - 126
 - RCA clean - 187

- reactive ion - 6
- Schimmel - 173
- Secco - 172, 173
- Sirtl - 172, 173
- vapor - 86, 89, 90, 91
- Wright-Jenkins - 173, 178
- Yang - 173
- Evaporation - 329, 336-355
 - flash evaporation - 348
 - hot-wall evaporation - 349
 - rate - 337-339
 - rate monitors - 354
 - reactive evaporation - 352-354
 - sources - 340-345
 - sublimation - 350
 - three-temperature - 419-420
 - uniformity - 345-346
 - of alloys - 347
 - of compounds - 347-348
- Fermi level - 471
- Field effect transistor (FET) - 587-588
- Flatband voltage - 60
- Gate
 - delay - 1
 - insulator - 4
 - length - 1
 - oxide - 1, 60, 592
 - voltage - 4
- Gettering - 34, 41
 - backside - 91, 94, 173
 - oxygen precipitates - 173
 - vacuum - 335
- Gibbs free energy - 468
- Grain boundary, polycrystalline
 - silicon - 73, 91
- Hall mobility - 424
- High electron mobility transistor (HEMT) - 363
- Hydrophilic - 25
- Hydrophobic - 25
- Implantation - 6
- Impurity - 6, 15, 18, 43, 90
 - gettering - 65
- N-type - 84
- P-type - 84
 - substitutional - 477
- Inelastic mean free path - 613
- Infrared - 36, 38, 654-659
 - epitaxial layer thickness measurement - 658
 - Fourier transform (FTIR) - 7
 - sensitivity - 657
- Ingot - 11, 12, 15, 18, 20, 22
- Integrated circuits - 9
- Interface
 - SiO₂/Si - 58, 59, 61, 62, 65, 66, 67, 68, 70, 71
 - traps - 59
- Interferometry - 249
- Interstitial - 20
 - oxygen - 37, 38, 41
 - self - 455, 472-490, 499
 - silicon - 69, 70
- Ion bombardment
 - cleaning - 422-425
 - influence on microstructure - 425-428
 - influence on stress - 430-438
 - ion beam mixing - 428
- Ion implantation - 512-540
 - critical energy - 520
 - electronic stopping - 519
 - implantation masking - 522-526
 - ion channeling - 527-538
 - Lindhard, Scharff & Schiott (LSS) range theory - 528
 - nuclear stopping - 519
 - projected range - 521-522
 - range theory - 515-522
- Ion plating - 424
- Junction - 82
 - breakdown voltage - 61
 - depth - 1, 67, 506
- Knock-on sputtering - 367
- Knudsen cell - 343-344
 - effusion cell - 343-344, 351
- Lifetime, minority-carrier - 34

- Lithography - 541-574
 - Abbe principle - 557
 - contact printing - 551
 - electron beam lithography - 6, 543-551, 570
 - ion beam - 6
 - numerical aperture - 555, 558
 - optical - 6, 559
 - projection printing - 555-556
 - wafer stepper - 556-557
 - x-ray - 6, 571-573
- Melting point - 82
- Metallization - 575-607
 - multilevel - 440
 - ohmic contact - 438, 598
 - polysilicon - 592
 - Rent's rule - 581
 - silicides - 439, 601-604
 - step coverage - 440
 - wiring structure - 581
- Microstructure - 6, 415
- Mobility - 27
 - atomic - 90
- Molecular beam epitaxy - 330, 355-364
 - apparatus - 356
 - application - 362-364
 - epitaxy growth - 360-362
 - three-temperature evaporation - 420-422
- MOS device - 58, 65, 73, 87, 89, 91, 109, 133
 - gate - 601, 603-605
- Multiquantum well laser (MQW) - 363
- Neutron activation analysis (NAA) - 7, 664-666
 - charged particle activation - 666
 - sensitivity - 665
- Nitridation
 - plasma - 67
 - polycrystalline silicon - 73
 - thermal - 66
- Nuclear reaction analysis (NRA) - 639-640, 666
 - hydrogen - 295, 300
- Nucleation - 41, 96, 102, 412-415
 - preferential - 416
 - sites - 94, 415, 424
- Ohmic contacts - 438
- Optical emission - 249
- Outgassing rate - 333
- Oxidation
 - ambient - 48, 65, 67, 68
 - Anodization - 46, 73, 75
 - dopant concentration dependence - 63, 478
 - mechanism - 68, 70
 - orientation dependence - 49, 62
 - plasma assisted - 72
 - reaction kinetics - 47, 50, 51, 52, 53, 54, 67
 - silicides - 73
 - surface preparation dependence - 64
 - temperature - 49, 52
 - thermal - 46, 48, 50, 51, 52, 54, 55, 59, 70, 75
- Oxide - 70
 - capacitance - 60
 - charge - 62, 65, 69, 71
 - interlevel insulation - 592
- Oxygen - 15, 19, 20, 31, 32, 35, 37, 91
 - diffusion - 39, 40
 - dissolved - 9, 38
- Packaging, thermal conduction module - 584
- Packing density - 109
- Pascal - 330
- Periodic table - 9
 - Group III - 15, 27
 - Group V - 27
 - Group VI - 15
- Phase transitions - 6
- Phosphorus - 15, 18, 27, 33
- Photoluminescence (PL) - 7, 660-664
 - bound-exciton - 661
 - detection limit - 661
 - free exciton - 661

- Photomask - 541-544, 551
 - alignment - 22
- Photoresist, multiple level resist - 568
 - negative - 542-551, 560-567
 - photoactive compound - 560
 - positive - 542-551, 560-567
 - sensitivity - 549
- Physisorption - 332
- Plasma - 191
 - anisotropy - 191, 193, 227-229, 252
 - chemical effects - 213-217
 - DC plasma - 193, 195, 196
 - feedstock composition - 210-213
 - fundamental aspects - 192-210
 - mechanism - 230-241
 - modeling - 241-242
 - non-equilibrium - 191
 - plasma body - 193
 - plasma sheath - 193, 195, 208, 209, 393, 395
 - reactor loading - 223
 - RF plasma - 193, 194, 195, 196, 394
 - surface chemistry - 217-241
- Plasma etching - 242
 - Al - 263
 - compound semiconductors - 264
 - defects - 257
 - endpoint detection - 249
 - equipment - 244-248
 - oxide - 260
 - patterning - 242-244
 - radiation damage - 257
 - selectivity - 254
 - Si - 259
 - silicides - 259
 - silicon nitride - 264
 - throughput - 256
 - uniformity - 253
- Plasma enhanced chemical vapor deposition - 265
 - compound semiconductors - 310-311
 - film properties - 284-289
 - Ge - 310
 - interface properties - 315-318
 - metals - 311-312
 - reactors - 268-280
 - silicides - 312-313
 - silicon - 305-310
 - silicon dioxide - 301-305
 - silicon nitride - 289-301
 - source gases - 280-281
 - uniformity - 281-284
- Polishing - 25
- Polycrystalline - 9, 11
- Polycrystalline silicon - 6, 24, 74, 96, 104, 109, 112, 308
 - gettering - 91
 - interconnect - 439
 - oxidation - 73
 - polysilicon charge - 11, 21, 82
- Positron annihilation - 474
- Precipitates - 6, 42
 - carbon - 34
 - oxygen - 20, 32, 35, 40, 41, 43, 91, 173
- Preferential deposition, epitaxial Si - 96
- Quartz crucible - 9, 11, 12, 48
 - reactor - 82
- Radiation damage - 443-444
- Radiation hardened - 106, 109, 129, 136
- Radio frequency (RF), induction coil - 9
- Refractive index - 285, 292, 294, 302
- Resistivity - 15, 20, 25, 31
- Resonance ionization spectroscopy (RIS) - 7, 633-635
 - detection limits - 633
 - resonance ionization schemes - 633
- Rutherford backscattering spectroscopy (RBS) - 6, 70, 294, 316, 635-640
 - energy analyzer - 636
 - scattering cross-section - 635

- Scanning electron microscopy -
 - 96, 158, 162, 640-646
 - backscattered electrons - 641
 - energy dispersive x-ray spectrometer - 642
 - secondary electrons - 641
 - wavelength dispersive analyzer - 645
- Scanning transmission electron microscopy (STEM) - 7, 475, 646-653
 - bright field image - 647
 - darkfield image - 647
 - electron diffraction - 648
 - electron energy loss spectroscopy - 649
- Schottky barrier - 362, 598, 601
- Secondary ion mass spectroscopy (SIMS) - 6, 70, 624-632
 - detection limit - 628
 - mass analyzer - 628
 - secondary ion yield - 627
- Segregation - 63
 - coefficient - 15, 18, 20, 21, 31, 34, 35, 38
 - surface - 362
- Sheet resistance - 509-512, 514
- Silicide - 6, 73
 - molybdenum - 439
 - paladium - 439
 - platinum - 439
 - tantalum - 74
 - titanium - 439
 - tungsten - 439
- Silicon carbide - 34, 111, 122
- Silicon dioxide - 4, 12, 48, 50, 54, 74, 105, 109, 122
 - monoxide - 25
 - silica - 25
- Silicon halide
 - SiBr₄ - 85
 - SiCl₄ - 82, 85, 90, 102, 122
 - SiHCl₃ - 82, 95, 90, 102
 - SiH₂Cl₂ - 85, 90, 102
 - SiH₄ - 85, 90
 - SiHBR - 85
 - SiI₄ - 85
- Silicon monoxide - 12, 19
- Silicon nitride - 55, 105, 111, 122
- Solubility, oxygen - 39
- Spreading resistance - 18
- Sputter assisted laser ionization (SALI) - 632
- Sputtering 329, 364-410
 - alloys - 376-378
 - bias - 424, 442
 - cleaning - 422
 - compounds - 376-378
 - glow discharge - 378-391
 - ion beam - 392-394
 - magnetron - 382-391
 - mechanisms - 365-369
 - planar diodes - 378-380
 - reactive - 400-409
 - RF sputtering - 394-399
 - sputtered species - 369-371
 - triode - 381-382
 - yield - 371-376
- Stacking faults - 6, 32, 33, 69, 70, 91, 104, 155, 157, 173, 178, 183, 500, 647
 - oxidation-induced - 499-500
- Stereographic projection - 133
- Sticking coefficient - 421
- Stress, compressive - 430
 - deposited films - 429
 - intrinsic - 429-438
 - oxide - 69, 303
 - SiN_x films - 295-299
 - tensile - 430
 - thermal - 429
- Structure zone model - 415-419, 426
 - Movchan and Demchishin - 416
- Sublimation - 343, 350
- Substrate - 13
 - p⁺ - 89
- Substitutional - 34, 37, 38, 41
- Susceptor - 11
- Target poisoning
 - reactive sputtering - 402-408
- Thermal decomposition - 86
- Threshold voltage - 4, 61
- Torr - 330
- Transistor - 82

- Trap, defect - 91
- Vacancy, diffusivity - 487-490
 - silicon - 70, 455, 472-485
- Very large scale integration (VLSI) -
1, 5, 6, 7, 70, 73, 91, 181
- Ultra large scale integration (ULSI) -
1, 5, 6, 7, 91, 183
- Wafer - 22, 23, 25, 34
 - flatness - 25
- Work damage - 22, 23
- Work function - 60
- X-ray diffraction - 70
- X-ray lithography - 115
- X-ray photoelectron spectroscopy
(XPS) - 6, 70, 303, 618-624
 - Auger transition - 624
 - chemical shift - 621
 - detection limit - 624
 - energy analyzers - 618
 - monochromator - 618
 - profiling - 620
- X-ray orientation - 22
- X-ray topography (XRT) - 7,
653-654

