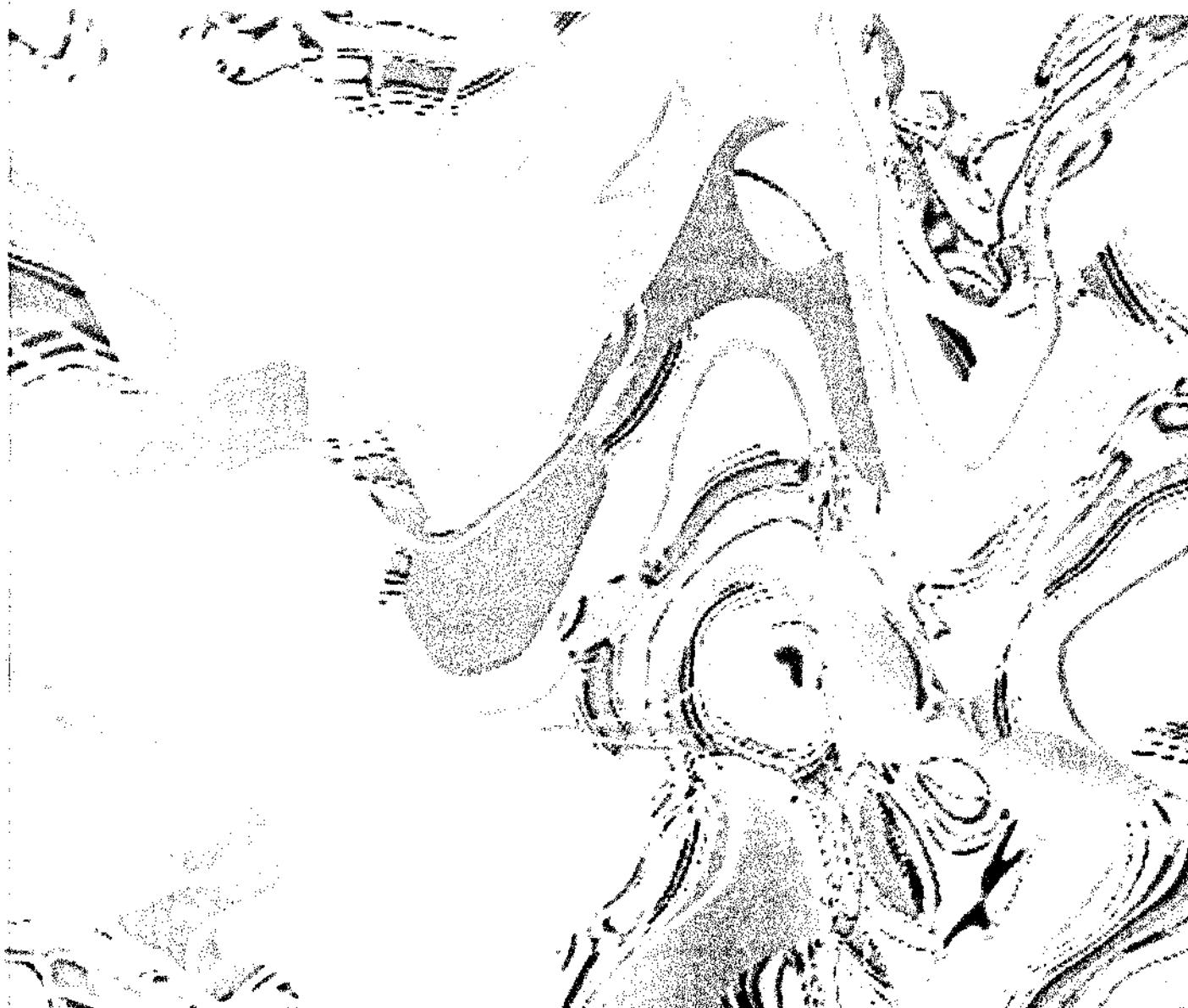


ark
on

Semiconductor Manufacturing Technology



Michael J. Quick

William Sardis

UNIVERSITY OF VIRGINIA LIBRARY



X004472523

For more information about this book
or any of Prentice Hall's other new
technology titles, visit our website at
<http://www.prenhall.com>

Semiconductor Manufacturing Technology

This book is written for students in two- and four-year technology programs at community colleges and universities. Chapters are organized around the broad technologies applicable to semiconductor manufacturing.

Chapters 1 to 8 present fundamental technical information relevant to semiconductor manufacturing. Chapter 9 presents a process model overview with a general flowchart that links the major areas in wafer fabrication. Chapters 10 to 19 cover each of the major processes. Each process chapter concludes with a summary of quality measures and troubleshooting issues to provide the student with practical challenges encountered during wafer fabrication. Chapter 20 provides an overview of the back-end process for IC assembly and packaging.

Academic and industry reviewers have applauded this book as the most comprehensive and up-to-date text currently available in the market.

ISBN 0-13-081520-9



9 780130 815200

90000>

SC
TK
78
.Q

SEMICONDUCTOR MANUFACTURING TECHNOLOGY

MICHAEL QUIRK

STAKTEK CORPORATION

JULIAN SERDA

ADVANCED MICRO DEVICES



Upper Saddle River, New Jersey
Columbus, Ohio

Library of Congress Cataloging-in-Publication Data

Quirk, Michael

Semiconductor manufacturing technology / Michael Quirk, Julian Serda.

p. cm.

ISBN 0-13-081520-9

1. Electronic industries. 2. Semiconductors—Design and construction. I. Serda, Julian.

II. Title.

TK7836.Q27 2001

00-041653

Vice President and Publisher: Dave Garza

Editor in Chief: Stephen Helba

Assistant Vice President and Publisher: Charles E. Stewart, Jr.

Production Editor: Alexandrina Benedicto Wolf

Production Coordination: Lithokraft II

Design Coordinator: Robin G. Chukes

Cover Designer: Linda Fares

Cover Image: John Foxx

Production Manager: Matthew Ottenweller

Marketing Manager: Barbara Rose

This book was set in Times Roman by Lithokraft II. It was printed and bound by R. R. Donnelley & Sons Company. The cover was printed by Phoenix Color Corp.

SCIENG Quirk, Michael, 1953-
TK Semiconductor
7836 manufacturing
.Q27 technology
2001

Copyright © 2001 by Prentice-Hall, Inc., Upper Saddle River, New Jersey 07458. All rights reserved.
Printed in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permission(s), write to: Rights and Permissions Department.



10 9 8 7 6 5 4 3 2 1
ISBN 0-13-081520-9

CONTENTS

CHAPTER 1 INTRODUCTION TO THE SEMICONDUCTOR INDUSTRY

OBJECTIVES	1
INTRODUCTION	2
DEVELOPMENT OF AN INDUSTRY	2
Industry Roots	2
The Solid State	3
CIRCUIT INTEGRATION	4
Integration Eras	5
IC FABRICATION	6
Wafer Fab	7
Stages of IC Fabrication	7
SEMICONDUCTOR TRENDS	9
Increase in Chip Performance	9
Increase in Chip Reliability	12
Reduction in Chip Price	12
THE ELECTRONIC ERA	14
The 1950s: Transistor Technology	14
The 1960s: Process Technology	14
The 1970s: Competition	14
The 1980s: Automation	15
The 1990s: Volume Production	15
CAREERS IN SEMICONDUCTOR MANUFACTURING	16
Technician	16
Job Descriptions	18
SUMMARY	19
KEY TERMS	19
REVIEW QUESTIONS	20
SELECTED INDUSTRY WEB SITES	20
REFERENCES	20

CHAPTER 2 CHARACTERISTICS OF SEMICONDUCTOR MATERIALS

OBJECTIVES	21
INTRODUCTION	21

ATOMIC STRUCTURE	22
Electrons	22
THE PERIODIC TABLE	24
Ionic Bonds	26
Covalent Bonds	28
CLASSIFYING MATERIALS	28
Conductors	29
Insulators	30
Semiconductors	32
SILICON	33
Pure Silicon	33
Why Silicon?	33
Doped Silicon	34
pn Junctions	39
ALTERNATIVE SEMICONDUCTOR MATERIALS	39
Gallium Arsenide (GaAs)	40
SUMMARY	41
KEY TERMS	41
REVIEW QUESTIONS	42
REFERENCES	42
CHAPTER 3 DEVICE TECHNOLOGIES	
OBJECTIVES	43
INTRODUCTION	43
CIRCUIT TYPES	44
Analog Circuits	44
Digital Circuits	44
PASSIVE COMPONENT STRUCTURES	45
IC Resistor Structures	45
IC Capacitor Structures	46
ACTIVE COMPONENT STRUCTURES	46
The pn Junction Diode	46
The Bipolar Junction Transistor	49
Schottky Diode	51
Bipolar IC Technology	52
CMOS IC Technology	52

IV CONTENTS

Enhancement and Depletion-Mode MOSFETs	58	Particles	87
LATCHUP IN CMOS DEVICES	59	Bulk Resistivity	87
INTEGRATED CIRCUIT PRODUCTS	59	EPIТАXIAL LAYER	87
Linear IC Product Types	60	SUMMARY	88
Digital IC Product Types	60	KEY TERMS	88
SUMMARY	62	REVIEW QUESTIONS	89
KEY TERMS	62	SELECTED INDUSTRY WEB SITES	90
REVIEW QUESTIONS	63	REFERENCES	90
IC MANUFACTURERS' WEB SITES	64		
REFERENCES	65		
CHAPTER 4 SILICON AND WAFER PREPARATION			
OBJECTIVES	67	OBJECTIVES	91
INTRODUCTION	67	INTRODUCTION	91
SEMICONDUCTOR-GRADE SILICON	68	STATES OF MATTER	91
CRYSTAL STRUCTURE	68	PROPERTIES OF MATERIALS	92
Amorphous Materials	69	Chemical Properties for Semiconductor Manufacturing	93
Unit Cells	69	PROCESS CHEMICALS	99
Polycrystal and Monocrystal Structures	70	Liquids	99
CRYSTAL ORIENTATION	71	Gases	103
MONOCRYSTAL SILICON GROWTH	72	SUMMARY	109
CZ Method	72	KEY TERMS	109
Float-Zone Method	75	REVIEW QUESTIONS	110
Reasons for Larger Ingot Diameters	76	CHEMICAL SUPPLIERS' WEB SITES	110
CRYSTAL DEFECTS IN SILICON	78	REFERENCES	111
Point Defects	78		
Dislocations	79		
Gross Defects	80		
WAFER PREPARATION	80		
Shaping Operations	81		
Wafer Slicing	82		
Wafer Lapping and Edge Contour	82		
Etching	83		
Polishing	83		
Cleaning	84		
Wafer Evaluation	84		
Packaging	84		
QUALITY MEASURES	84		
Physical Dimensions	85		
Flatness	86		
Microroughness	86		
Oxygen Content	86		
Crystal Defects	87		
		Air	120
		Humans	121
		Facility	123
		Water	126
		Process Chemicals	130

Production Equipment 131	REVIEW QUESTIONS 178
Workstation Design 131	METROLOGY EQUIPMENT SUPPLIERS' WEB SITES 179
WAFER WET CLEANING 135	REFERENCES 180
Wet-Cleaning Overview 135	CHAPTER 8 GAS CONTROL IN PROCESS CHAMBERS
Wet-Clean Equipment 138	OBJECTIVES 181
Alternatives to RCA Clean 142	INTRODUCTION 181
SUMMARY 143	VACUUM 183
KEY TERMS 143	Vacuum Ranges 183
REVIEW QUESTIONS 144	Mean Free Path 184
CHEMICAL AND EQUIPMENT SUPPLIERS' WEB SITES 145	VACUUM PUMPS 184
REFERENCES 146	Roughing Pump 185
CHAPTER 7 METROLOGY AND DEFECT INSPECTION	High Vacuum Pump 186
OBJECTIVES 149	Vacuum in Integrated Tools 188
INTRODUCTION 149	PROCESS CHAMBER GAS FLOW 189
IC METROLOGY 150	Mass Flow Controllers 189
Measurement Equipment 151	RESIDUAL GAS ANALYZER (RGA) 190
Yield 151	RGA Basics 190
Data Management 152	RGA as Real-Time Monitor 191
QUALITY MEASURES 152	PLASMA 192
Film Thickness 153	Glow Discharge 193
Film Stress 158	PROCESS CHAMBER CONTAMINATION 195
Refractive Index 158	SUMMARY 196
Dopant Concentration 159	KEY TERMS 196
Unpatterned Surface Defects 161	REVIEW QUESTIONS 196
Patterned Surface Defects 164	VACUUM EQUIPMENT SUPPLIERS' WEB SITES 197
Critical Dimension (CD) 165	REFERENCES 197
Step Coverage 167	CHAPTER 9 IC FABRICATION PROCESS OVERVIEW
Overlay Registration 167	OBJECTIVES 199
Capacitance-Voltage (C-V) Test 168	INTRODUCTION 199
Contact Angle 171	CMOS PROCESS FLOW 199
ANALYTICAL EQUIPMENT 171	Overview of Areas in a Wafer Fab 200
Secondary-Ion Mass Spectrometry (SIMS) 171	CMOS MANUFACTURING STEPS 205
Atomic Force Microscope (AFM) 173	1. Twin Well Process 205
Auger Electron Spectroscopy (AES) 174	2. Shallow Trench Isolation Process 207
X-Ray Photoelectron Spectroscopy (XPS) 174	3. Poly Gate Structural Process 210
Transmission Electron Microscope (TEM) 175	4. Lightly Doped Drain (LDD) Implant Process 211
Energy- and Wavelength-Dispersive Spectrometer (EDX and WDX) 176	5. Sidewall Spacer Formation 212
Focused Ion Beam (FIB) 176	6. Source/Drain (S/D) Implant Processes 213
SUMMARY 177	
KEY TERMS 178	

7. Contact Formation	214	Thin-Film Characteristics	260		
8. Local Interconnect (LI) Process	214	Film Growth	263		
9. Via-1 and Plug-1 Formation	216	Film Deposition Techniques	264		
10. Metal-1 Interconnect Formation	217	CHEMICAL VAPOR DEPOSITION	265		
11. Via-2 and Plug-1 Formation	218	CVD Chemical Processes	265		
12. Metal-2 Interconnect Formation	219	CVD Reaction	266		
13. Metal-3 to Pad Etch and Alloy	220	CVD DEPOSITION SYSTEMS	269		
14. Parametric Testing	221	CVD Equipment Design	270		
SUMMARY	222	APCVD (Atmospheric Pressure CVD)	271		
KEY TERMS	223	LPCVD (Low Pressure CVD)	273		
REVIEW QUESTIONS	223	Plasma-Assisted CVD	277		
REFERENCES	224	DIELECTRICS AND PERFORMANCE	282		
CHAPTER 10 OXIDATION					
OBJECTIVES	225	Dielectric Constant	282		
INTRODUCTION	225	Device Isolation	286		
OXIDE FILM	226	SPIN-ON-DIELECTRICS	287		
Nature of Oxide Film	226	Spin-On-Glass (SOG)	287		
Uses of Oxide Film	227	Spin-On-Dielectric (SOD)	288		
THERMAL OXIDATION GROWTH	231	EPIТАХY	289		
Chemical Reaction for Oxidation	231	Epitaxy Growth Methods	290		
Oxidation Growth Model	232	CVD QUALITY MEASURES	292		
FURNACE EQUIPMENT	239	CVD TROUBLESHOOTING	292		
HORIZONTAL VERSUS VERTICAL FURNACES	240	SUMMARY	294		
Vertical Furnace	241	KEY TERMS	295		
Fast Ramp Vertical Furnace	245	REVIEW QUESTIONS	295		
Rapid Thermal Processor	246	DEPOSITION EQUIPMENT SUPPLIERS' WEB SITES	296		
OXIDATION PROCESS	248	REFERENCES	296		
Pre Oxidation Cleaning	248	CHAPTER 12 METALLIZATION			
Oxidation Process Recipe	249	OBJECTIVES	299		
QUALITY MEASUREMENTS	250	INTRODUCTION	300		
OXIDATION TROUBLESHOOTING	252	TYPES OF METALS	302		
SUMMARY	252	Aluminum	302		
KEY TERMS	253	Aluminum-Copper Alloys	305		
REVIEW QUESTIONS	253	Copper	305		
FURNACE AND RTP EQUIPMENT SUPPLIERS' WEB SITES	254	Barrier Metals	307		
REFERENCES	255	Silicides	309		
CHAPTER 11 DEPOSITION					
OBJECTIVES	257	Metal Plugs	312		
INTRODUCTION	257	METAL DEPOSITION SYSTEMS	313		
Film Layering Terminology	258	Evaporation	313		
FILM DEPOSITION	260	Sputtering	314		
		Metal CVD	320		
		Copper Electroplate	323		
		METALLIZATION SCHEMES	325		
		Traditional Aluminum Structure	325		

Copper Damascene Structure	326	SUMMARY	364
METALLIZATION QUALITY MEASURES	329	KEY TERMS	364
METALLIZATION TROUBLESHOOTING	330	REVIEW QUESTIONS	364
SUMMARY	331	PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES	365
KEY TERMS	331	REFERENCES	366
REVIEW QUESTIONS	332		
METALLIZATION EQUIPMENT AND MATERIALS SUPPLIERS' WEB SITES	332		
REFERENCES	333		
CHAPTER 13 PHOTOLITHOGRAPHY: VAPOR PRIME TO SOFT BAKE			
OBJECTIVES	335	OBJECTIVES	367
INTRODUCTION	336	INTRODUCTION	367
Photolithography Concepts	336	Importance of Alignment and Exposure	368
PHOTOLITHOGRAPHY PROCESSES	339	OPTICAL LITHOGRAPHY	370
Negative Lithography	339	Light	370
Positive Lithography	340	Exposure Sources	372
EIGHT BASIC STEPS OF PHOTOLITHOGRAPHY	342	Optics	376
Step 1: Vapor Prime	343	Resolution	385
Step 2: Spin Coat	343	PHOTOLITHOGRAPHY EQUIPMENT	388
Step 3: Soft Bake	344	Contact Aligner	388
Step 4: Alignment and Exposure	344	Proximity Aligner	389
Step 5: Post-Exposure Bake (PEB)	344	Scanning Projection Aligner	390
Step 6: Develop	344	Step-and-Repeat Aligner (Stepper)	390
Step 7: Hard Bake	344	Step-and-Scan System	393
Step 8: Develop Inspect	345	Reticles	394
VAPOR PRIME	345	Optical Enhancement Techniques	398
Wafer Cleaning	345	Alignment	400
Dehydration Bake	346	Environmental Conditions	404
Wafer Priming	346	Comparison of Photo Tools	405
SPIN COAT	348	MIX AND MATCH	406
Photoresist	348	ALIGNMENT AND EXPOSURE QUALITY MEASURES	407
Photoresist Physical Properties	349	ALIGNMENT AND EXPOSURE TROUBLESHOOTING	408
Conventional I-Line Photoresists	351	SUMMARY	409
Deep UV (DUV) Photoresists	354	KEY TERMS	410
Photoresist Dispensing Methods	357	REVIEW QUESTIONS	410
SOFT BAKE	360	PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES	411
Soft Bake Equipment	360	REFERENCES	412
Process Characterization	361		
PHOTORESIST QUALITY MEASURES	362		
PHOTORESIST TROUBLESHOOTING	363		

**CHAPTER 15 PHOTOLITHOGRAPHY:
PHOTORESIST
DEVELOPMENT AND
ADVANCED
LITHOGRAPHY**

OBJECTIVES	413
INTRODUCTION	413
Advanced Lithography	414
POST-EXPOSURE BAKE	414
DUV Post-Exposure Bake (PEB)	415
Conventional I-Line PEB	416
DEVELOP	416
Negative Resist	417
Positive Resist	418
Development Methods	419
Resist Development Parameters	421
HARD BAKE	422
DEVELOP INSPECT	422
ADVANCED LITHOGRAPHY	424
Next-Generation Lithography	424
Advanced Resist Processing	428
DEVELOP QUALITY MEASURES	429
DEVELOP TROUBLESHOOTING	431
SUMMARY	432
KEY TERMS	433
REVIEW QUESTIONS	433
PHOTOLITHOGRAPHY MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES	433
REFERENCES	434

CHAPTER 16 ETCH

OBJECTIVES	435
INTRODUCTION	435
Etch Processes	436
ETCH PARAMETERS	437
Etch Rate	437
Etch Profile	438
Etch Bias	439
Selectivity	440
Uniformity	441
Residues	441
Polymer Formation	442
Plasma-Induced Damage	442
Particle Contamination	443
DRY ETCH	443
Etching Action	444

Potential Distribution	444
PLASMA ETCH REACTORS	446
Barrel Plasma Etcher	446
Parallel Plate (Planar) Reactor	447
Downstream Etch Systems	448
Triode Planar Reactor	448
Ion Beam Milling	448
Reactive Ion Etch (RIE)	450
High-Density Plasma Etchers	450
Etch System Review	453
Endpoint Detection	455
Vacuum for Etch Chambers	456
DRY ETCH APPLICATIONS	456
Dielectric Dry Etch	457
Silicon Dry Etch	459
Metal Dry Etch	462
WET ETCH	464
Types of Wet Etch	465
HISTORICAL PERSPECTIVE	466
PHOTORESIST REMOVAL	466
Plasma Ashing	466
ETCH INSPECTION	469
ETCH INSPECTION QUALITY MEASURES	469
DRY ETCH TROUBLESHOOTING	470
SUMMARY	471
KEY TERMS	472
REVIEW QUESTIONS	472
ETCH EQUIPMENT SUPPLIERS' WEB SITES	473
REFERENCES	474

CHAPTER 17 ION IMPLANT

OBJECTIVES	475
INTRODUCTION	475
Doped Regions	477
DIFFUSION	479
Diffusion Principles	479
Diffusion Process	481
ION IMPLANTATION	482
Overview	483
Ion Implant Parameters	484
ION IMPLANTERS	488
Ion Source	488
Extraction and Ion Analyzer	490
Acceleration Column	491

Scanning System 494	Li Oxide Polish 535
Process Chamber 498	Li Tungsten Polish 536
Annealing 499	ILD Oxide Polish 536
Channeling 501	Tungsten Plug Polish 536
Particles 502	Dual-Damascene Copper Polish 537
ION IMPLANT TRENDS IN PROCESS INTEGRATION 503	CMP QUALITY MEASURES 538
Deep Buried Layers 503	CMP TROUBLESHOOTING 540
Retrograde Wells 503	SUMMARY 541
Punchthrough Stoppers 504	KEY TERMS 541
Threshold Voltage Adjustment 504	REVIEW QUESTIONS 542
Lightly Doped Drain 505	CMP EQUIPMENT SUPPLIERS' WEB SITES 542
Source/Drain Implants 505	REFERENCES 543
Polysilicon Gate 506	
Trench Capacitor 506	CHAPTER 19 WAFER TEST
Ultrashallow Junctions 506	OBJECTIVES 545
Silicon-On-Insulator (SOI) 507	INTRODUCTION 545
ION IMPLANT QUALITY MEASURES 508	IC Electrical Tests 546
ION IMPLANT TROUBLESHOOTING 509	WAFER TEST 547
SUMMARY 510	In-Line Parametric Test 547
KEY TERMS 511	Wafer Sort 555
REVIEW QUESTIONS 512	Yield 560
ION IMPLANTER EQUIPMENT SUPPLIERS' WEB SITES 513	Wafer Sort Yield Models 563
REFERENCES 513	TEST QUALITY MEASURES 565
CHAPTER 18 CHEMICAL MECHANICAL PLANARIZATION	TEST TROUBLESHOOTING 567
OBJECTIVES 515	SUMMARY 567
INTRODUCTION 515	KEY TERMS 568
TRADITIONAL PLANARIZATION 518	REVIEW QUESTIONS 568
Etchback 518	TEST AND PROBER EQUIPMENT SUPPLIERS' WEB SITES 569
Glass Reflow 519	REFERENCES 569
Spin-On Films 520	
CHEMICAL MECHANICAL PLANARIZATION 520	CHAPTER 20 ASSEMBLY AND PACKAGING
CMP Planarity 521	OBJECTIVES 571
Advantages of CMP 522	INTRODUCTION 571
CMP Mechanisms 523	Packaging Levels 573
CMP Slurry and Pad 526	TRADITIONAL ASSEMBLY 574
CMP Equipment 529	Backgrind 574
CMP Clean 532	Die Separation 575
CMP Equipment Manufacturers 534	Die Attach 575
CMP APPLICATIONS 535	Wirebonding 577
STI Oxide Polish 535	TRADITIONAL PACKAGING 580
	Plastic Packaging 581
	Ceramic Packaging 584
	Final Test 586

X CONTENTS

ADVANCED ASSEMBLY AND PACKAGING 586	APPENDICES
Flip Chip 587	APPENDIX A CHEMICALS AND SAFETY 601
Ball Grid Array (BGA) 589	APPENDIX B CONTAMINATION CONTROLS IN CLEANROOMS 610
Chip on Board (COB) 590	APPENDIX C UNITS 614
Tape Automated Bonding (TAB) 591	APPENDIX D COLOR AS A FUNCTION OF OXIDE THICKNESS 617
Multichip Module (MCM) 591	APPENDIX E OVERVIEW OF PHOTORESIST CHEMISTRY 618
Chip Scale Packaging (CSP) 592	APPENDIX F ETCH CHEMISTRY 622
Wafer-Level Packaging 592	
ASSEMBLY AND PACKAGING QUALITY MEASURES 596	
IC PACKAGING TROUBLESHOOTING 597	GLOSSARY 624
SUMMARY 597	INDEX 648
KEY TERMS 598	
REVIEW QUESTIONS 598	
ASSEMBLY AND PACKAGING SUPPLIERS' WEB SITES 599	
REFERENCES 600	

PREFACE

This text started with a simple premise: as instructors, we need to teach relevant microchip technology to students and employees in semiconductor manufacturing. Unfortunately, in the semiconductor industry, changes in technology are measured in months, not years. Our challenge was to write a relevant book that would not be outdated by the time it was published. With that in mind, we researched the material and applied ourselves to writing the chapters and creating the artwork. Following the aggressive pace of Moore's law, the technical material in our book is at most only 18 to 24 months old. This permits us to keep abreast of the changing technology nodes swirling through the semiconductor industry.

This text is written for students in two-year and four-year technology programs at community colleges and universities. The text will also be a practical reference as well as a standard text in corporate and technical training classes. Students are expected to have an understanding of high school chemistry, physics, and math. Chapters are organized around the broad technologies applicable to semiconductor manufacturing.

ORGANIZATION OF THE TEXT

Our goal is to accomplish three objectives:

1. Help technology students grasp the fundamental technologies used in manufacturing semiconductor devices.
2. Present some of the many challenges in microchip fabrication.
3. Instill in the reader an appreciation of the conceptual simplicity of semiconductor manufacturing.

All fundamental technical information relevant to semiconductor manufacturing is first presented in Chapters 1 to 8. Chapter 9 presents a process model overview with a general flowchart that links the major areas in a wafer fab. Chapters 10 to 19 cover each of the major processes in the fab. Finally, Chapter 20 provides an overview of the back-end process for IC assembly and packaging. The content in the process chapters (Chapters 10 to 20) addresses critical process technology, followed by the various equipment designs needed to support this technology. Each process chapter concludes with a summary of quality measures and troubleshooting issues to familiarize the student with the practical, day-to-day challenges encountered during wafer fabrication.

The latest technologies for sub-0.25 μm processing are covered in detail. This includes chemical mechanical planarization (CMP), shallow trench isolation (STI), chemically amplified deep UV photoresists, step-and-scan systems, copper metallization with dual damascene, and the widespread move to process integration with cluster tools. Throughout the text, we explain all process and equipment technology in light of the long history of change in the industry. Early tools and processes are described to clarify the development of current technology. In some cases, the linkage between the latest equipment and earlier tools is obvious, while in other instances the change is dramatic.

Professors, students, and other readers of this book can send comments or questions about this text to the authors at the following website: <http://www.smtbook.com>. We look forward to any exchange of information that can help advance semiconductor manufacturing education.

ACKNOWLEDGMENTS

We would like to thank the following reviewers for their valuable feedback: Bruce Bothwell, Chemeketa Community College; Mike Delia, Samsung Austin Semiconductor; Chris Dennis and David Hata, Portland Community College; John Fowler, Arizona State University; Lance Kinney; Bassam Matar, Maricopa Community College; John Nistler, Advanced Micro Devices; Val Shires, Gwinnett Technical Institute; and Joseph White, Texas A&M University.

We acknowledge the support of the following colleagues, friends, and new acquaintances for their technical inputs and encouragement throughout the project.

Technical Reviewers and Editors:

- ◆ Mike D'Elia, Member of Technical Staff, Advanced Micro Devices—a tremendous inspiration to us—reviewed the process flow, oxidation and deposition chapters and was the first SMT instructor to use selected chapters.
- ◆ Mike Hillis, Module Shift Manager, Advanced Micro Devices, provided encouragement and reviewed and edited the photolithography sections.
- ◆ Steve Hymes, Ph.D., Applications Manager, Ashland Specialty Chemicals, who seemed more excited than us at times, reviewed and edited the CMP and metallization chapters.
- ◆ Rick Jarvis, Senior Member Technical Staff, Advanced Micro Devices, on assignment to International SEMATECH, with great insight to the industry, reviewed and edited the contamination, etch, and deposition sections.
- ◆ Larry Knipp, President, Nykar Inc., reviewed and edited the RF and plasma sections of the book.
- ◆ Jeremy Lansford, Module Manager, Advanced Micro Devices, very carefully reviewed and edited the CMP section, then met with us to ensure his suggestions were understood.
- ◆ Federico Miller, Chemistry Department, Austin Community College, provided advice and encouragement throughout the project and reviewed all the chemistry related information.
- ◆ Sergei Postnikov, Ph.D., Photolithography Development, Motorola, reviewed the photolithography sections and provided encouragement throughout the project.
- ◆ Alan Romriell, Senior Test Engineer, Advanced Micro Devices, reviewed the test chapter.
- ◆ Andy Shurtleff, President, Liquid Asset Services, an expert in ultra-pure water systems, reviewed and edited the section on ultra-pure water.
- ◆ Darby Webster, Senior Manager, Manufacturing Engineering, Staktek Corp., reviewed the assembly and packaging chapter and gave his support throughout the project.
- ◆ Joseph W. Wiseman, Member of Technical Staff, Advanced Micro Devices, a professional in his engineering approach and also an etch instructor, reviewed and edited the etch section.

Computer-Aided Illustrators and Writers:

- ◆ Thuat Duong, CAD specialist, Solectron, created many of the drawings used in the assembly and packaging section.
- ◆ Matt Hanson, Computer Graphics Artist, provided graphic support for the metrology section.
- ◆ Josh Hix, Technical Writer, a full-time engineering student at Georgia Tech University, produced most of the glossary.
- ◆ Trent Romriell, Computer Graphics Artist, supported us with many drawings in the beginning chapters.
- ◆ James Samara, Computer Graphics Artist, provided graphics support for the metrology chapter.
- ◆ Mark Serda, Computer Graphics Artist, provided graphics support for the gas controls chapter.

**Providers of Logistics, Technical Information,
Graphics, and Photographs:**

- ◆ Shannon Bacon, Continuous Improvement Facilitator, Advanced Micro Devices, made arrangements for us to meet with subject matter experts at AMD's facilities.
- ◆ Tony Denboer, Integrated Circuit Engineering Corporation, provided SEM micrographs of actual integrated circuit structures.
- ◆ Jim DeVries, Facilities Operations Manager, Advanced Micro Devices, directed the photo shoot of some AMD facilities.
- ◆ Charlie Earnhart, Vice-President, Staktek Corp., for his support throughout the project.
- ◆ James A. Elliott, Technical Instructor, Varian Semiconductor Equipment, assisted the writers with obtaining permission to use technical documentation and illustrations of Varian ion implanters.
- ◆ Robert Emmons, Learning and Development Manager, Advanced Micro Devices, allowed Julian to adjust his work hours so he could devote more quality time to the book.
- ◆ Kathleen Fitts, Global Quality Programs Manager, Applied Materials, supported the project with various diagrams and photographs.
- ◆ Dahlia Hernandez-Fridge, Executive Assistant to the President, CEDRA Corp., researched technical information.
- ◆ Van Pham, Process Development Engineer, Staktek Corporation, provided input regarding etch.
- ◆ Richard Rogoff, Major Accounts Manager, ASML, provided excellent diagrams and photographs of step-and-scan systems.
- ◆ Tim Roy, Director, Staktek Corp., for his support throughout the project.
- ◆ Sharon Shaw, Librarian, Advanced Micro Devices, showed us what the word "resourceful" means by always being able to locate a reference for us and for handling the logistics.
- ◆ Michael Sullivan, MEGASYSTEM Manager, Air Products, allowed us to photograph several bulk gas distribution sites.
- ◆ Chuck Tully, Austin Technical Training Center Manager, Applied Materials, who was always there when we needed him, provided technical information and graphics on electrostatic chuck and DPS etch equipment.
- ◆ Phil Ware, Director and General Manager of Marketing, Semiconductor Equipment Division, Canon USA, Inc., provided contacts, graphics, and photographs of several generations of photolithography tools.
- ◆ Scott A. Williams, Central Regional Account Manager, SpeedFam-IPEC, arranged to have photographs taken and provided diagrams of CMP equipment.

Michael Quirk
Julian Serda

This text is dedicated to our families, who endured so much during this project:

Monique, Justin, and Morgan Quirk

Dorothy, Mark, Ruben, and Elena Serda

INTRODUCTION TO THE SEMICONDUCTOR INDUSTRY

Society witnessed a technology revolution during the twentieth century with the change from products derived from mechanical technology to those centered on electronic technology. Digital compact disc players replaced record players, and automotive engines are now controlled by electronic ignition systems. Electronic computers quickly perform tasks in nearly every aspect of society, promoting efficient use of our resources. Given the breadth of changes brought by electronic technology, the revolution has just begun.

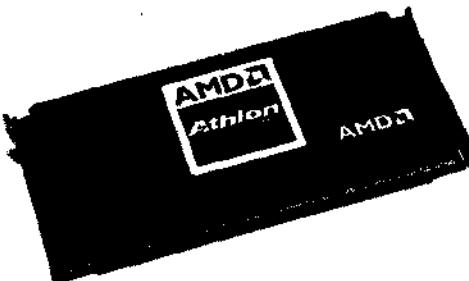
The semiconductor industry has been at the center of this technology revolution. The principal building

material, the *semiconductor*, is the main ingredient of the electronic products found throughout society. Semiconductor products are manufactured by people with diverse technical skills: designers who create new designs based on customer needs, engineers who specify improved requirements for the equipment and process, and technicians who fabricate the semiconductor products in automated factories. All the while, the growing semiconductor market continues to demand more performance at lower cost.

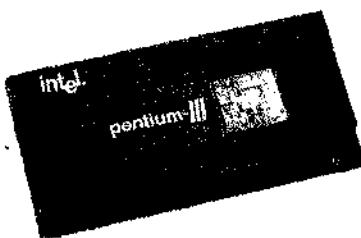
OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe the current economic state and the technical roots of the semiconductor industry.
2. Explain what an integrated circuit (IC) is and list the five circuit integration eras.
3. Describe a wafer, including how it is layered, and describe the essential aspects of the five stages of wafer fabrication.
4. State and discuss the three major trends associated with improvement in wafer fabrication.
5. Explain what a critical dimension (CD) is and how Moore's law predicts future wafer fabrication improvement.
6. Describe the different eras of electronics since the invention of the transistor up to modern wafer fabrication.
7. Discuss different career paths in the semiconductor industry.



Microprocessor Chip
(Photo courtesy of Advanced Micro Devices)



Microprocessor Chip
(Photo courtesy of Intel Corporation)

INTRODUCTION

The basic semiconductor material from which electronic devices are made comes in the form of round thin crystalline disks called *wafers*. Semiconductor products are produced from wafers in wafer fabrication (*wafer fab*) factories, and are referred to as *microchips*, or *chips*.

The technology of semiconductor fabrication is complex, requiring many specialized process steps, materials, equipment, and supplier industries. Once the microchips are fabricated, they are packaged into the various electronic and mechanical assemblies required for the numerous product applications. Examples of such applications are automotive electronics, electronic commerce, personal computers, and mobile cellular communications.

Worldwide sales of microchips are expected to exceed \$200 billion in 2001.¹ Today, semiconductors account for 30% to 40% of the cost for a personal computer, and about \$100 worth of semiconductors are found in each cellular telephone. Every automobile has approximately \$140 worth of microchips, an amount that is increasing as cars become “smarter.”

The semiconductor industry is actually a subset of a larger entity—the high-tech industry. Fabricating the microchips produces electronic hardware that is coupled with software to integrate and control the chip functions. The high-tech industry encompasses all the hardware and software technology found in all semiconductor applications (See Figure 1.1).

The high-tech industry in the United States is big. In the mid-1990s, the high-tech industry contributed 27% of the U.S. economy, compared to 14% for residential housing and 4% for the automotive industry. Why has the semiconductor industry become so large? A major factor has been the industry’s ability to consistently increase semiconductor product performance while decreasing the price. The ability to meet the market demands for high performance at low cost can be directly traced to the technology improvements regularly made in product design and manufacturing throughout the history of the industry.

DEVELOPMENT OF AN INDUSTRY

The technology used to fabricate semiconductor devices developed out of many different technological inventions from early pioneers in the electronics field.

Industry Roots

The developmental roots of the semiconductor industry grew out of technologies developed in the first half of the twentieth century.² Key technical knowledge was gained from a web of industry and academia: vacuum tube electronics, wireless communications, mechanical tabulators, and solid-state physics. These industries were the fabric of the high-tech industry at the time.

The triode vacuum tube for amplifying electrical signals was invented by Lee De Forest in 1906. It was developed out of earlier vacuum tube work by John Fleming and Thomas Edison. A triode consists of three elements: two electrodes and a grid separated in an evacuated glass enclosure. A vacuum is necessary to keep the elements from burning up and also to facilitate the transfer of electrons between the electrodes. De Forest patented and named his vacuum tube invention the audion because he thought there was some potential for amplifying and reproducing sound. He

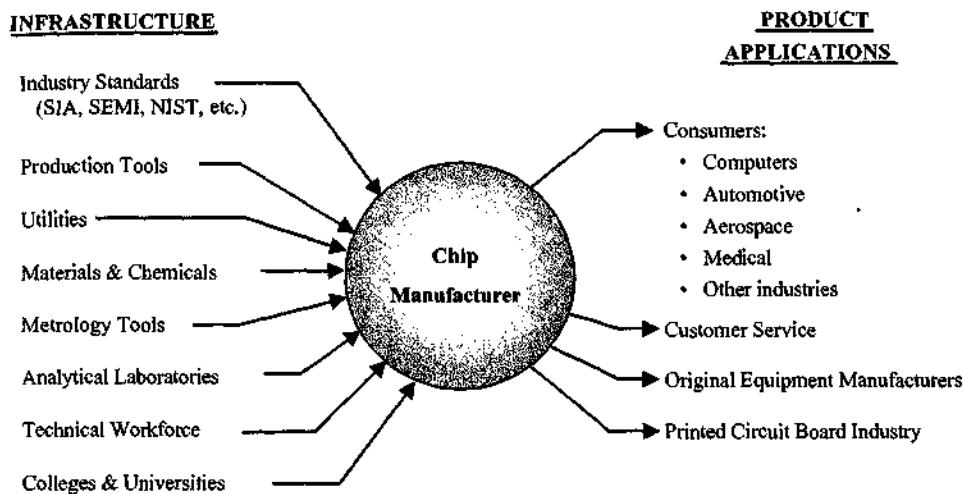


FIGURE 1.1 The Semiconductor Industry

was right, and the vacuum tube became the primary electronic device for the growth of modern radio, television, and general electronics up until the 1950s.

A relatively new material, a crystal known as silicon, was used in the early 1900s to convert wireless communications signals from alternating current to direct current. The term “semiconductor” was first adopted in Germany to cover this class of material.³ However, to truly develop the power of semiconductor technology required additional research by physical chemists and physicists around the world. This research was needed in order to understand the quantum theory of electron behavior before a full explanation of semiconductor properties was available. This substantial work extended over several decades and beyond World War II.

There was a market need to electronically calculate numbers. The first mechanical computer was the Hollerith mechanical tabulating machine, invented by Herman Hollerith, to tabulate the 1890 census figures in an amazing six weeks. This tabulator was driven by an electric motor and based on punched cards (punch cards remained the common input method for computers up through the 1970s). Vacuum tubes were used to develop the first electronic computer, the ENIAC (Electronic Numeric Integrator and Calculator) at the University of Pennsylvania during World War II. The ENIAC weighed 50 tons, required 3,000 ft² of floor space, needed 19,000 vacuum tubes, and used the equivalent electrical power of 160 lighthouses.

A major drawback to the ENIAC, besides its bulky size, was the number of problems associated with vacuum tubes. Tubes were large, unreliable and consumed large amounts of electric power. Tubes had a limited lifetime due to burnout. Vacuum tubes clearly were not the optimum technology to produce small, reliable electronic products that were needed for the rapidly developing electronic markets.

The Solid State

There was a concerted effort at Bell Telephone Laboratories after World War II to study the properties of solid silicon and germanium semiconductor crystals. The scientists conducting the research sensed the need to replace vacuum tubes and the benefits from using a solid semiconductor material in place of vacuum tubes.

The modern-day semiconductor industry was born with the invention of the *solid-state transistor* at Bell Telephone Laboratories on December 16, 1947, by William Shockley, John Bardeen, and Walter Brattain.⁴ The transistor, a name taken from the two terms *transconductance* and *varistor*,⁵ offered the same electrical functions as a vacuum tube, but with the distinct advantages of a solid state: minuscule size, no vacuum, reliable, lightweight, minimal heat production, and low power consumption. The three scientists were awarded the 1956 Nobel Prize in physics for their invention. This discovery launched the modern semiconductor industry based on solid-state materials and technology.

The semiconductor industry started growing rapidly in the 1950s to commercialize transistor technology based on silicon. Many of the early pioneers started in northern California, in the area



Vacuum Tubes

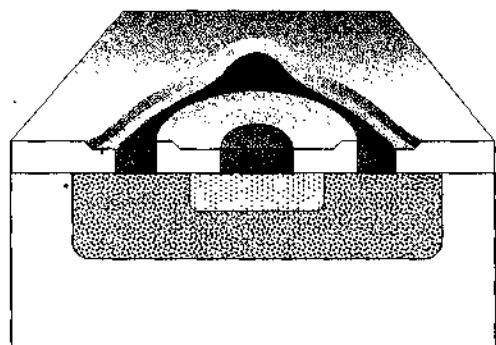


FIGURE 1.2 The First Planar Transistor

now known as Silicon Valley. The first commercial planar transistor was fabricated at Fairchild Semiconductor in Palo Alto, California, in 1957. It had a layer of aluminum interconnect material that was deposited on top of the silicon wafer to connect the different parts of the transistor (see Figure 1.2). A natural oxide layer thermally grown from the silicon was used to insulate the aluminum conductor. The use of layers was an important development in semiconductors and is the reason for the term *planar technology*.

CIRCUIT INTEGRATION

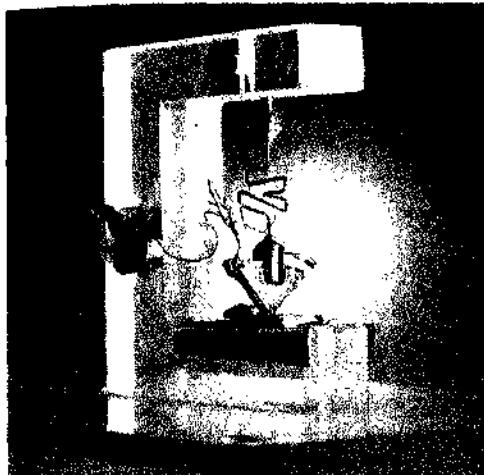
The semiconductor name derives from the material's ability to sometimes be a conductor to electricity and other times act as a nonconductor to electricity. The earliest semiconductor material was germanium, built into a single chip with one function (referred to as a discrete component). Nowadays, greater than 85% of all microchips are made from silicon semiconductor material. For this reason, we will emphasize silicon in this book.

An important step forward in the semiconductor industry was the integration of multiple electronic components on one silicon substrate. Referred to as an *integrated circuit*, or *IC*, it was co-invented independently by Robert Noyce at Fairchild Semiconductor and Jack Kilby at Texas Instruments in 1959. On the silicon surface of an IC are manufactured the many different semiconductor *devices*, such as transistors, diodes, resistors, and capacitors, which are connected into a circuit to define how the chip will function. The expression *IC* is often meant to describe the chip and all its components.

The first IC, developed in July 1958 by Jack Kilby at a Texas Instruments facility in Dallas, Texas, was made using a slice of germanium semiconductor material as the substrate.⁶ The invention combined a transistor and other components on the germanium while also using the natural resistance of the germanium as a resistor. The devices were connected with individual wires.

Robert Noyce at Fairchild Semiconductor also invented the concept of an IC by extending the idea of how to interconnect the different components on the planar silicon material. His idea was to use an aluminum metal conductor on the silicon surface to interconnect the different transistors while using a layer of oxide grown from the silicon as an insulator to separate the metal conductor from the silicon devices. This was the first practical structure of an IC as a single-structure silicon chip.⁷

Circuit integration has increased dramatically since the initial IC. Because all the components are integrated on the silicon substrate, the IC has developed into a cost-effective and reliable way to produce and interconnect many components. The ability to integrate many different components on an IC has spurred engineers to design ever more complex electronic circuits to meet new customer needs.



The First Transistor from Bell Labs
(Photo courtesy of Lucent Technologies, Bell
Labs Innovations)



Jack Kilby's First Integrated Circuit
(Photo courtesy of Texas Instruments, Inc.)

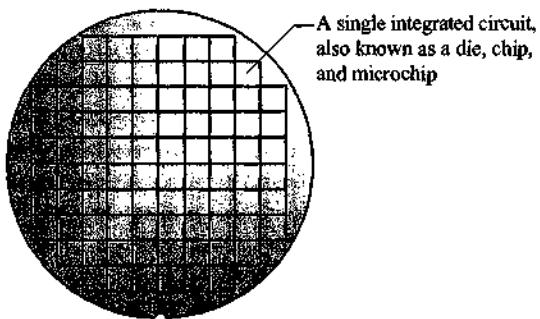


FIGURE 1.3 Top View of Wafer with Chips

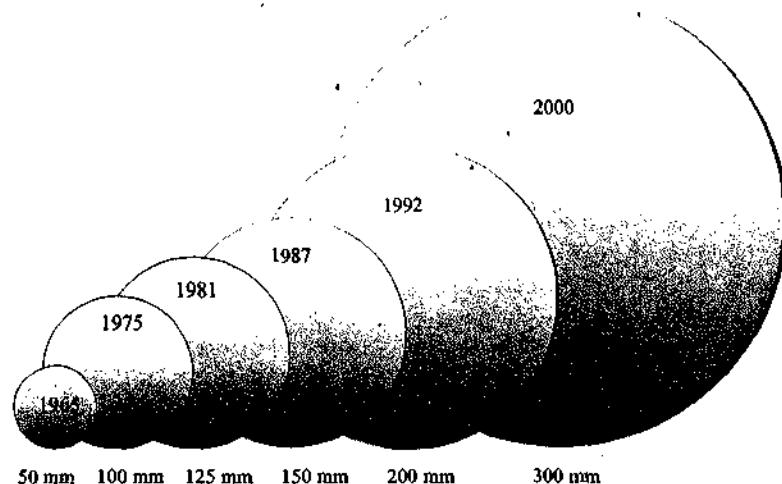
Integration Eras

We can roughly categorize the integration era by the number of components integrated on a chip. This is a useful method of organizing the development of the semiconductor industry since 1960 to the present (see Table 1.1).

TABLE 1.1 Circuit Integration of Semiconductors

Circuit Integration	Semiconductor Industry Time Period	Number of Components per Chip
No integration (discrete components)	Prior to 1960	1
Small scale integration (SSI)	Early 1960s	2 to 50
Medium scale integration (MSI)	1960s to Early 1970s	50 to 5,000
Large scale integration (LSI)	Early 1970s to Late 1970s	5,000 to 100,000
Very large scale integration (VLSI)	Late 1970s to Late 1980s	100,000 to 1,000,000
Ultra large scale integration (ULSI)	1990s to present	> 1,000,000

Circuit integration continues today as the number of devices on a chip is continually increased. A significant challenge for circuit integration is the capability of the semiconductor manufacturing process to improve the fabrication technology to produce highly integrated ULSI chips at acceptable costs. To this end, the semiconductor industry has become highly standardized, with most manufacturers using similar manufacturing process and equipment technology. The key to market success is the ability of a company to deliver the right product at the right time. Industry

**FIGURE 1.4** Evolution of Wafer Size

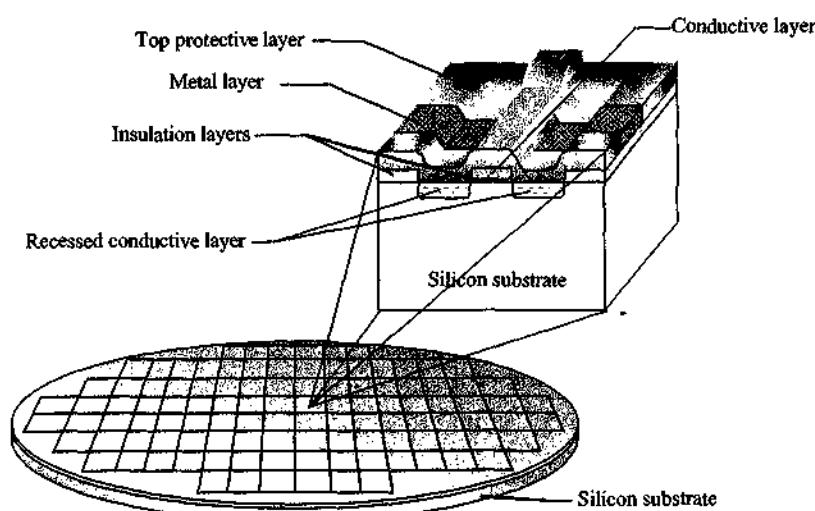
consortiums, such as International SEMATECH and the Semiconductor Industry Association (SIA), have been created to assist chip producers to be competitive in the world market of high-performance ICs.

IC FABRICATION

Tens or hundreds of identical chips are fabricated simultaneously on a silicon *wafer* (see Figure 1.3 on page 5). The number of chips on a wafer varies depending on the type of product and the size of each chip. The chip size changes depending on the level of integration on the chip.

The chip is also referred to as a *die* (singular and plural for chips or ICs), while the silicon wafer is often called the *substrate*. The wafer diameter has evolved over the years from an initial diameter of less than 1ⁱⁿ to the present common diameter of 8" (200 mm), which is undergoing a change to 12", or 300 mm (see Figure 1.4). The cost of fabricating ICs drops dramatically if there are more chips on a wafer due to benefits from the economies of scale (more chips are produced from the same effort).

Fabrication of the semiconductor devices occurs only in the first few microns of the silicon near the wafer surface. The bulk of the silicon wafer thickness is to give the wafer adequate rigidity during processing. Once the devices are fabricated in the silicon, then layers of metal circuitry are placed on the silicon to interconnect between the devices and the various electronic signals outside the chip (see Figure 1.5). The concept and materials for interconnecting a modern day IC are remarkably similar to the initial planar transistor first commercialized at Fairchild Semiconductor in 1957. The major difference is that today's chips are much more complex.

**FIGURE 1.5**
Devices and Layers
from a Silicon Chip

Wafer Fab

Early wafer fabs were simple, with operators manually processing wafers through the different operations. An essential aspect of wafer fabs is that as wafers become more densely integrated, the permissible level of contamination decreases dramatically. Contamination that can damage wafers and cause them not to function properly evolves from many sources: humans, materials, water, air, and equipment. Modern wafer fabs have become specialized facilities that provide the clean manufacturing environment and specialized equipment to produce wafers with minimal defects from contamination. This includes limited human exposure, ultrapure chemicals and utilities, and special wafer handling procedures needed to fabricate ICs in the ULSI era.

A wafer typically requires two to three months of processing to complete the 450 or more process steps in a wafer fab. At the end of the fabrication process, the individual chips are separated from the entire wafer and then prepared for packaging into the final product.

Stages of IC Fabrication

Fabrication of microchips involves five general stages of manufacturing (see Figure 1.6):

- ◆ Wafer preparation
- ◆ Wafer fabrication
- ◆ Wafer test/sort
- ◆ Assembly and packaging
- ◆ Final test

These five stages are interdependent, with a large infrastructure within the semiconductor company and a support network of industries that provide specialized chemicals and equipment. Companies that operate only in an individual stage (such as a chip company that only fabricates chips) have to meet industrywide standards to ensure that the final microchip meets performance objectives.

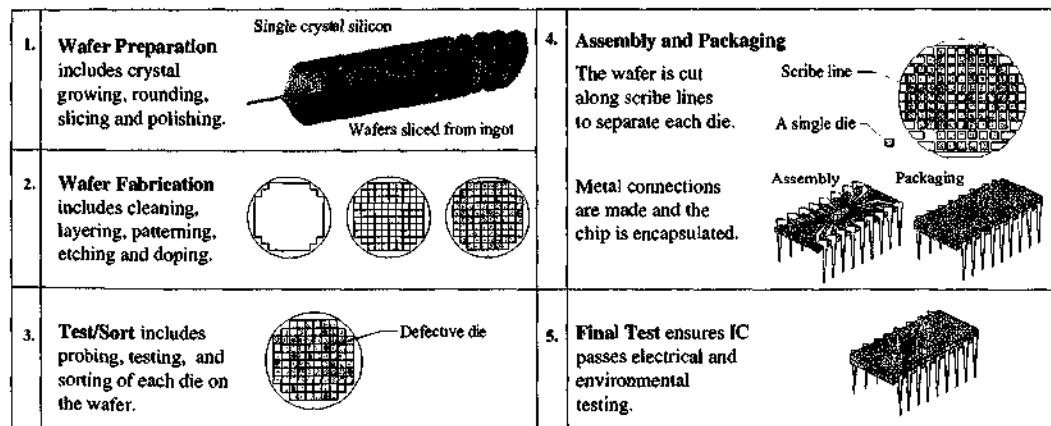


FIGURE 1.6 Stages of IC Fabrication

Wafer Preparation ■ In the first stage, *silicon* is mined from sand and purified. It undergoes specialized processing to create a silicon ingot with the appropriate diameter (see Figure 1.7 on page 8). The silicon ingot is then sliced into thin wafers that are used to fabricate the microchips. Wafers are prepared to special specifications for criteria such as flatness and contamination.

Since the 1980s, most companies that fabricate microchips purchase their wafers from suppliers who specialize in crystal growth and wafer preparation. The industry also produces wafers made from germanium or compound semiconductor materials. These are specialized applications, with the majority of all semiconductor wafers made from silicon.

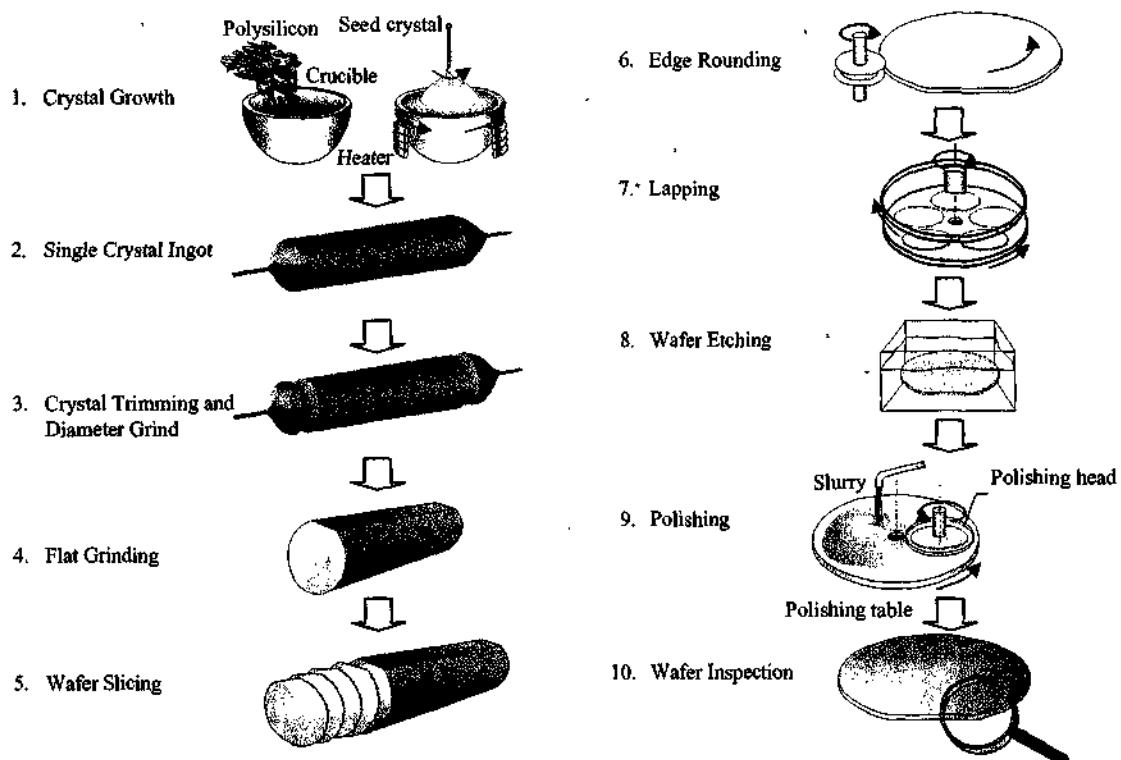
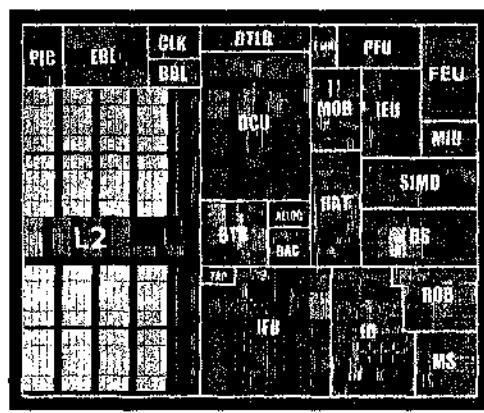


FIGURE 1.7 Preparation of Silicon Wafers
(Note: Terms in Figure 1.7 are explained in Chapter 4.)

Wafer Fabrication ■ The fabrication of microchips from wafers is the second stage, referred to as *wafer fabrication*. The bare silicon wafers arrive in the wafer fab, then they go through a variety of cleaning, layering, patterning, etching, and doping steps. The completed wafers have a full complement of integrated circuits permanently etched into each silicon wafer. Other names for wafer fabrication are *microchip fabrication*, *chip fabrication*, and *chip fab*.

Companies that fabricate chips are both merchant and captive producers. *Merchant chip suppliers* fabricate chips to sell on the open market, such as a chip manufacturer that produces memory chips for customers. *Captive chip producers* fabricate chips to be used in-house in their company's products. An example of a captive producer is a company that makes computers and also fabricates chips for their computers. Some chip makers fabricate chips for in-house use and also sell chips on the open market, while others will fabricate specialty chips and purchase other chips on the open market.

Another type of chip maker is the *fabless company*. This type of company designs chips for a particular market, such as a video microchip, while another chip maker fabricates the chips. Finally,



ULSI Chip
(Photo courtesy of Intel Corporation)

another type of semiconductor manufacturer, the *foundry*, produces chips only for other companies. Semiconductor foundries have become more common since the 1980s, with about 10% of all chips now fabricated at foundries. A major reason for the increase in fabless chip producers and foundries is the high cost of constructing and maintaining a wafer fab. Currently, a wafer fab for high-performance ICs costs around \$1.5 to \$3.0 billion, with about 75% of the total cost for equipment.

Wafer fabrication of ICs involves the interaction of many complicated process steps using automated equipment to produce the hundreds of millions of devices on one ULSI wafer. Because of the complexity associated with fabricating high-performance ICs, the semiconductor industry has always been at the leading edge of equipment design and manufacturing technology. This innovation has fostered continual improvements in wafer fabrication.

Wafer Test/Sort ■ At the completion of wafer fabrication, the wafer is then sent to the test/sort area where individual die are probed and electrically tested. Defective die are marked for sorting later into acceptable and unacceptable chips. Die that fail the wafer test will not be sent to customers, while die that pass the wafer test will continue through the process.

Assembly and Packaging ■ After wafer test/sort, the wafer enters assembly and packaging to enclose the individual chips in a protective package. The backside of the wafer undergoes a back-grind to reduce the thickness of the substrate. A thick membrane of acetate tape is attached to the backside of each wafer, then the front of the wafer is sliced along the scribe lines with a diamond-tip saw blade to separate the die from each wafer. The sticky acetate holds the silicon chips to keep them from falling off. At the assembly plant, the good die have wire bonds added or have bumps applied to form the attachment to the assembly package. Soon after, the die is hermetically sealed in a plastic or ceramic package. The actual form of the final package varies depending on the type of chip and its application (see Figure 1.8 on page 10). See Chapter 20 for more information on assembly and packaging.

Final Test ■ To ensure a chip's functionality, each packaged integrated circuit is tested to meet the manufacturer's electrical and environmental specifications. After the final test, the chips are delivered to the customer for assembly into the specific application, such as mounting memory components on a circuit board for a personal computer.

SEMICONDUCTOR TRENDS

The rapid technology changes needed to design and fabricate ICs with ULSI integration leads to the continual introduction of new equipment and processes. The semiconductor industry introduces new fabrication technology into the wafer fab every 18 to 24 months. Wafer fabrication technology changes are driven by the customers' needs. Customers require faster, more reliable, and lower cost chips. To achieve this, chip manufacturers have learned to reduce the size of components on a chip. This increases chip speed and reduces power consumption. Chips undergo extensive testing and analysis to verify long-term reliability. The number of chips on a wafer are increased to reduce cost.

There are three major trends associated with improvements in microchip technology:

- ◆ Increase in chip performance
- ◆ Increase in chip reliability
- ◆ Reduction in chip cost

As the different process areas of wafer fabrication are studied in this text, the introduction of new technologies will be discussed based on their contribution toward these trends.

Increase in Chip Performance

The performance of semiconductor microchips has dramatically improved since the SSI era of the early 1960s. A common way to judge chip performance is speed. *Chip speed* will improve if devices are made smaller and packed closer together on the chip because the electrical signal moving through the circuit has less distance to travel. Another way to increase speed is to use materials that

improve the passage of the electrical signal through the device and circuitry on the chip surface. We will study these new materials in later chapters. Microprocessor chip performance is also judged by the number of instructions the chip can perform, as measured in million instructions per second (MIPS). A faster microchip that can perform more instructions per second is beneficial to the customer.

Critical Dimension (CD) ■ The physical dimension of a feature on a chip is referred to as the *feature size*. Another term to describe feature size is *circuit geometry*. Of special note is the minimum feature size on a wafer, known as the *critical dimension*, or *CD*. We will refer throughout the text to the CD as the criteria that defines the level of fabrication complexity (i.e., if you possess the capability to fabricate the CD on a wafer, then you can fabricate all other feature sizes because the dimension is larger and therefore easier to produce). For instance, if the smallest linewidth on a chip is $0.18 \mu\text{m}$, then this dimension is the CD (see Figure 1.9). Device CDs have continually shrunk since the beginning of semiconductor fabrication, starting with a CD of about $125 \mu\text{m}$ in the early 1950s and currently at $0.18 \mu\text{m}$ and less.⁹ The semiconductor industry uses the term *technology nodes* to describe the applicable CD in use during wafer fabrication. The actual and projected industry technology nodes for CDs from $1 \mu\text{m}$ and below are shown in Table 1.2.

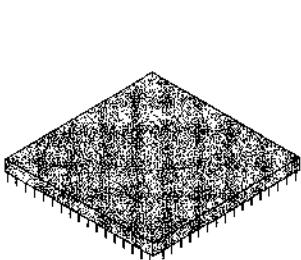


FIGURE 1.8 Sample of Microchip Packaging

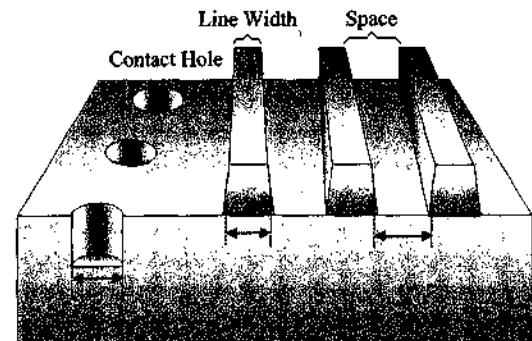


FIGURE 1.9 Critical Dimension

TABLE 1.2 Past & Future Technology Nodes for Feature Critical Dimension (CD)

	1988	1992	1995	1997	1999	2001	2002	2005
CD (μm)	1.0	0.5	0.35	0.25	0.18	0.15	0.13	0.10

Adapted from *1999 Roadmap: Solutions and Caveats, Solid State Technology*, May, 2000, p. 77.

The coordinated shrinking of device dimensions on the chip is referred to as *scaling*. It is not acceptable to only reduce one feature on a chip. This would be like designing a large pickup truck with very small tires. Normally, a large vehicle for heavy loads requires large tires. A small car with a light load is acceptable with small tires. The same goes for chip design. All dimensions must be reduced simultaneously, or scaled, in order to optimize electrical performance. For advanced semiconductor manufacturing, device scaling occurs for dimensions in both the vertical and lateral directions.

Components Per Chip ■ Reducing the feature size on a chip permits more components to be fabricated on the silicon wafer. For microprocessors, the number of transistors on the chip surface can illustrate the increased chip integration permitted by the reduction in CD. Since the number of transistors on a chip has dramatically increased over the years of wafer fabrication, there has been an improvement in chip performance (see Figure 1.10).

Moore's Law. In 1964, Gordon Moore, one of the early pioneers in the semiconductor industry and a founder of Intel, predicted that the number of transistors on a chip would double roughly every year.¹⁰ This prediction is known in the industry as *Moore's law* (it was later modified



Wafer Fab
(Photo courtesy of Advanced Micro Devices-Dresden Copyright by S. Doering)

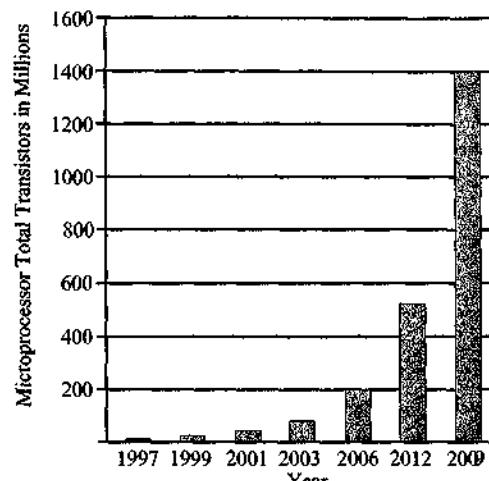


FIGURE 1.10 Increase in Total Transistors/Chip
Redrawn from Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors*, 1997

in 1975 to predict doubling every 18 months). A key contributor to Moore's law is the ability to fabricate wafers with a reduction in the device feature size to achieve a new CD and an increase in the number of transistors on a chip with the introduction of each new product generation. Moore's law for transistors on microprocessors, shown in Figure 1.11 has been surprisingly accurate.

Size reduction in semiconductor devices as predicted by Moore's law is important because it leads to smaller microchip packaging. Smaller microchips fit into small volumes, which leads to smaller commercial products. Semiconductors are the enabling technology for the development of portable electronic products such as the palmtop computer and cellular telephones (see Figure 1.12 on page 12).

Power Consumption ■ Another important aspect of chip performance is the power consumption of the device during operation. Whereas vacuum tubes were power hungry, semiconductor devices consume substantially less power. As device miniaturization proceeds, there is a corresponding reduction in power consumption.¹¹ Although the number of transistors per chip is rapidly increasing, the chip power consumption grows at a much slower rate (see Figure 1.13 on p. 12). This has become a key performance parameter given the market growth for portable electronic products.

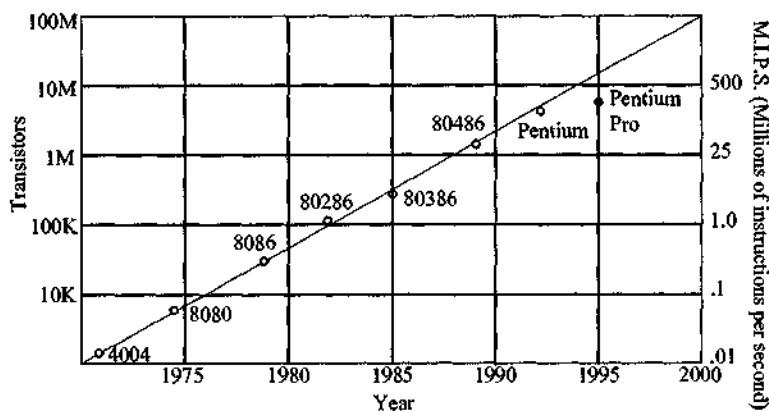


FIGURE 1.11 Moore's Law for Microprocessors
(Used with permission from *Proceedings of the IEEE*, January, 1998,
© 1998 IEEE)

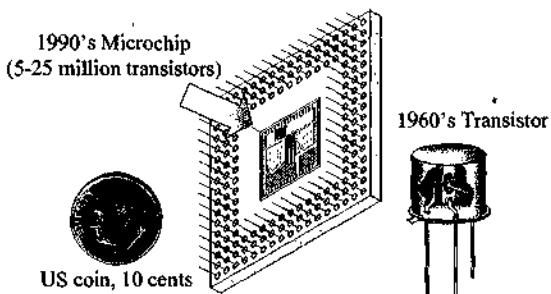


FIGURE 1.12 Size Comparison of Early and Modern Semiconductors

Increase in Chip Reliability

Chip reliability addresses the ability of the chip to function as intended over its expected life. Advances in technology have improved the product reliability of chips (see Figure 1.14). For example, contaminants are controlled through strict application of cleanroom procedures in areas such as particle-free air and control of chemical purity. Fabrication processes are constantly analyzed to improve device reliability. Wafers are monitored and microchips tested to verify acceptable performance. This translates into products that function with low failures during operation.

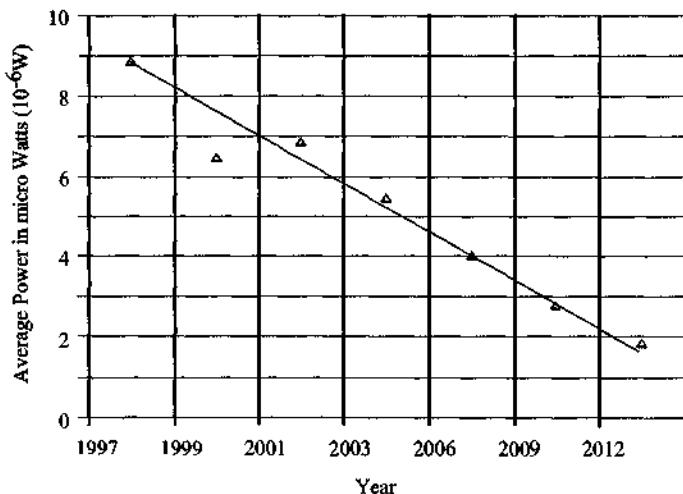


FIGURE 1.13 Reduction in Chip Power Consumption per IC
Redrawn from Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors*, 1997

Reduction in Chip Price

The price of semiconductor microchips has steadily decreased (see Figure 1.15). During the nearly 50 years prior to 1996, the price of semiconductor microchips has decreased by a phenomenal 100 million times.¹² For example, in 1958 a single silicon transistor that was of poor quality cost around \$10. Today, \$10 might buy a memory chip with over 20 million transistors, an equal number of other components, and the necessary interconnections to make a useful chip.

There are several reasons for the decreasing prices for semiconductor chips. As previously explained, factors such as reducing the feature size and increasing the wafer diameter to put more chips on a wafer are contributors to price reduction. The cost benefits from this decrease in CD are substantial. For example, if a semiconductor manufacturer decreased the CD on an 8-inch wafer from $0.35\text{ }\mu\text{m}$ to $0.25\text{ }\mu\text{m}$ in 1997, they were able to increase the number of chips per wafer from 150 to 275. In other words, they produced almost twice as many chips per wafer at nearly the same manufacturing cost. This cost benefit was achieved because the reduced CD permitted millions of circuit lines to be compressed into a smaller area.

Another reason for price reductions has been the large market growth for semiconductor products. This growth requires increased volume from chip fabrication companies, which introduces economies of scale for manufacturing. The introduction of manufacturing improvements to

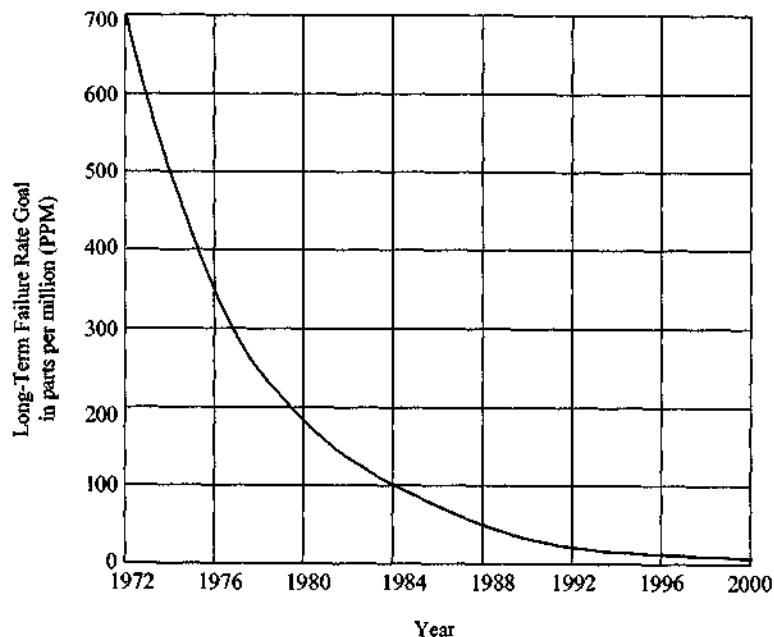


FIGURE 1.14
Reliability Improvement
of Chips

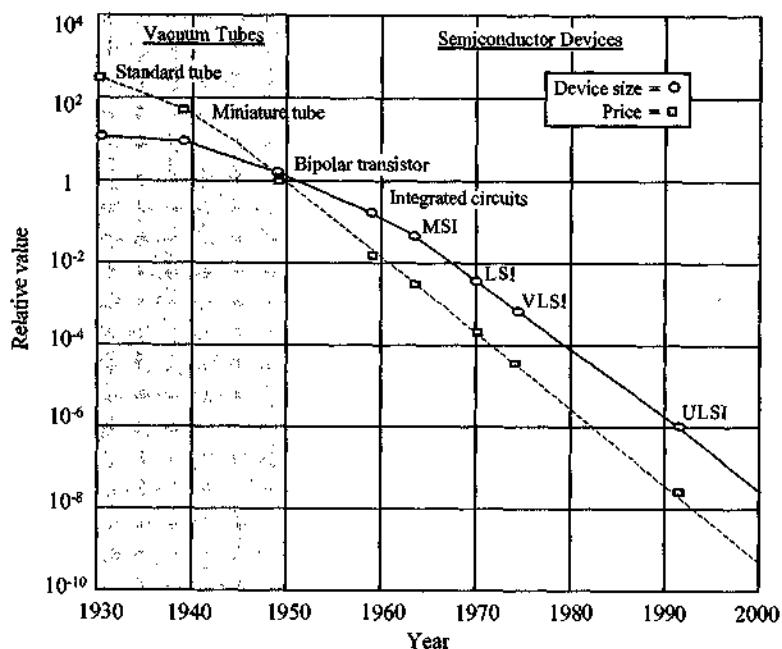


FIGURE 1.15 Price
Decrease of
Semiconductor Chips
Redrawn from C. Chang
and S. Sze, *ULSI
Technology*, (New York:
McGraw-Hill, 1996),
p. xxiii.

the equipment and process used to fabricate microchips has also reduced costs. Work teams have an important role in improving manufacturing technology and cost. They work firsthand with the chip fabrication equipment and contribute to improving manufacturing productivity. The efforts of manufacturing teams make a direct difference in lowering the cost to manufacture chips.

THE ELECTRONIC ERA

The semiconductor industry has undergone intensive growth since the invention of the first transistor at Bell Labs in 1947. It is beneficial to review the major industry changes since the beginning of solid state electronics to gain insight from the foundations of modern wafer fabrication.

The 1950s: Transistor Technology

The 1950s were a rich era in the development of different types of semiconductor transistor technology. Many different types of transistors were developed at Bell Labs and the early semiconductor efforts in northern California and Texas. The first transistors were developed from germanium semiconductor material. In 1954 the first manufacturable silicon transistor was made by Gordon Teal at Texas Instruments.¹³

With the invention of the IC in 1959, the industry was poised to solidify transistor technology as the replacement for vacuum tubes and to develop new market applications. The basic elements needed to commercialize the transistor were in place. The semiconductor industry was growing in high technology centers such as Silicon Valley in northern California; Route 128 near Boston, Massachusetts; and Texas Instruments in Dallas, Texas. This concentration of ideas and technology was important for the development of semiconductor technology and the new manufacturing techniques necessary to move wafer fabrication out of the laboratory processes of the 1950s into production processes of the 1960s. The growing dominance of semiconductors and silicon in all aspects of society gave credence to the expression, "the silicon age."

The 1960s: Process Technology

In the 1960s, the semiconductor industry entered an era oriented toward solving the basic problems of producing semiconductor ICs. This was the beginning of integration and the SSI era. Semiconductor manufacturers were proliferating. The market for semiconductor ICs was growing rapidly, with sales exceeding \$1 billion in 1961. The subsequent growth of the many chip manufacturers served to increase IC performance and reduce cost. Because of process commonality across semiconductor companies, special supplier industries developed to provide the chemicals and equipment needed for wafer fabrication.

There was a proliferation of engineers wanting to exploit the new semiconductor technology, which led to the formation of many new high-tech companies in the 1960s. Signetics was founded in 1961 by a group of Fairchild engineers to commercialize ICs. Robert Noyce, Gordon Moore, and Andrew Grove left Fairchild to found Intel in 1968. Jerry Sanders and other scientists, also from Fairchild, founded Advanced Micro Devices in 1969. There were also some large captive supplier companies in the IC business, such as IBM and Digital Equipment Corporation, and midsize companies such as Hewlett-Packard. The exploding demand for semiconductor products helped the many small start-up companies overcome the hurdles of constant price reductions, rapid company growth, and financial challenges of starting new companies.

The 1970s: Competition

The early 1970s was the period of medium-scale integration (MSI) for chip design. The manufacturing processes were largely manual operations based on batch processing. A typical wafer fab was fortunate to start a new product with a yield of 5 or 10% (yield is the percent of acceptable die produced at the completion of the process) and after an intense effort possibly raise it to 30%.¹⁴

Microprocessors were invented separately in the early 1970s at both Texas Instruments and Intel, and with its wide acceptance in market applications came the need for more chip integration. LSI-level integration remained only a few years and then was quickly replaced by VLSI, the integration standard by the end of the 1970s.

The rapidly changing semiconductor industry became chaotic. Much of the equipment and many of the processes used in the wafer fabs were developed by the same companies that manufactured the semiconductor devices, with few industrywide standards. This created inefficiencies for manufacturers and suppliers. Asia emerged as a formidable competitor, with Japan as a powerhouse of semiconductor innovation and manufacturing. Asian electronic giants challenged

American dominance in semiconductors. By 1979, Japan had captured more than 40% of the world's demand for memory microchips.¹⁵

As the demand for more complex chips grew, equipment technology changed from the manual tools of the 1960s to semiautomatic operation with buttons for operator control and onboard solid-state controllers. Equipment suppliers also encountered stiff competition from Japan and Asia for key semiconductor tools.

The cost of constructing a wafer fab became extremely expensive. Control of wafer contamination levels became critical for shrinking device feature sizes, requiring special wafer fab cleanliness standards that exceeded nearly every other industry. Special purity requirements were needed for water, air, and the many chemicals and gases used during fabrication. By the end of the 1970s, the cost of a wafer fab was around \$30 million and rising. Market demand and the need to continually replace equipment to maintain the advancing technology encouraged the construction of new fabs.

The industry attempted to form collaborative ventures to address the disorder. In 1970, SEMI (Semiconductor Equipment and Materials Institute) was founded to standardize and promote equipment, materials, and services in the industry. The Semiconductor Industry Association (SIA) was started in 1977, with leadership by Robert Noyce. The goal was to have more industry coordination about mutual problems created by rapid growth.

The 1980s: Automation

By the 1980s, the widespread acceptance of solid-state electronics in society was established, driving the growth of chip fabrication in the semiconductor industry. This was the VLSI era, with a high degree of component integration on a chip. The growth of the personal computer industry fueled hardware and software demand. At the same time, the industry faced competitive pressures by Japanese IC manufacturers that expanded their manufacturing capacity while at the same time driving yield and quality to unprecedented levels. U.S. semiconductor companies were clearly shaken by these developments and reached near panic because of their inability to compete. By the mid-1980s, Japan almost totally dominated the fast growing and technologically demanding DRAM (dynamic random access memory) market segment.¹⁶

The semiconductor industry formed SEMATECH in 1987, under the guidance of the Department of Defense, with Robert Noyce as its first Chief Executive Officer. The goal was to develop industrywide policies with respect to the specification and evaluation of manufacturing equipment. These organizations were formed, in part, because Japanese competition threatened the very existence of U.S. semiconductor manufacturers.¹⁷ Almost simultaneously SEMI/SEMATECH was formed of U.S. equipment and materials suppliers in support of SEMATECH's mission. The names of the consortiums would be changed in 1999 to International SEMATECH and the Semiconductor Industry Suppliers Association, respectively.

Improvements in semiconductor equipment, manufacturing efficiency, and product quality were emphasized in U.S. companies. Fabrication tools were automated to include all significant wafer processing steps. The intent was to largely remove the operator from the process, especially since the human was a major source of contamination in the cleanroom.

The industry was able to maintain the predicted growth of Moore's law by means of ongoing chip design changes and reduction in chip feature sizes. These design changes presented manufacturing challenges that led to the development of sophisticated processes. The equipment and cleanroom controls necessary to fabricate the new ICs escalated the cost of building wafer fabs, which rose to nearly \$1 billion by the late 1980s (see Figure 1.16 on page 16). To overcome this high cost of entry into new technology with new fabs, many U.S. semiconductor companies entered into joint venture agreements with Japanese and European chip manufacturers to share the cost of constructing wafer fabs.

The 1990s: Volume Production

Feature size dimensions shrunk below 1 μm for production chips during the late 1980s and early 1990s. By the end of the 1990s, the minimum feature size was at 0.18 μm . Microchip design experts predict that at some point there will be a physical limitation to further reductions in feature size. At this time, there is no barrier anticipated until the CD goes well below 0.1 μm . CDs

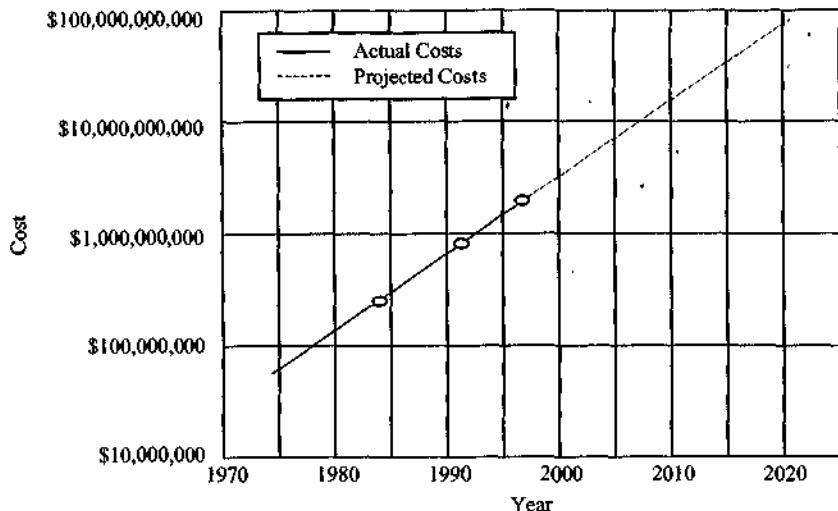


FIGURE 1.16
Start-Up Cost of
Wafer Fabs
(Used with per-
mission from
*Proceedings of
IEEE*, January,
1998, © 1998
IEEE)

continue to shrink because of the cost pressures for manufacturing efficiency resulting from putting more chips on the same wafer.

Submicron geometry processing ushered in the ULSI era of integration, with high-performance ICs containing 10 million transistors or more. The two types of chips with the most advanced technology are microprocessors and memory chips. Highly integrated chips require multiple layers of circuitry to interconnect the devices (up to eight layers and increasing) and up to 450 or more process steps to fabricate the chips.

The semiconductor industry has become increasingly competitive in the 1990s. To survive in the worldwide chip market, it is critical for manufacturers to produce complex, high quality chips in record time. Delivery of the right microchip to the market has a small window of opportunity based on customers and their need for advanced technology. Semiconductor manufacturers try to be the first (or nearly the first) with the new technology. If a company misses this product window, then it risks spending substantial money to develop the chip technology without realizing any sales potential.

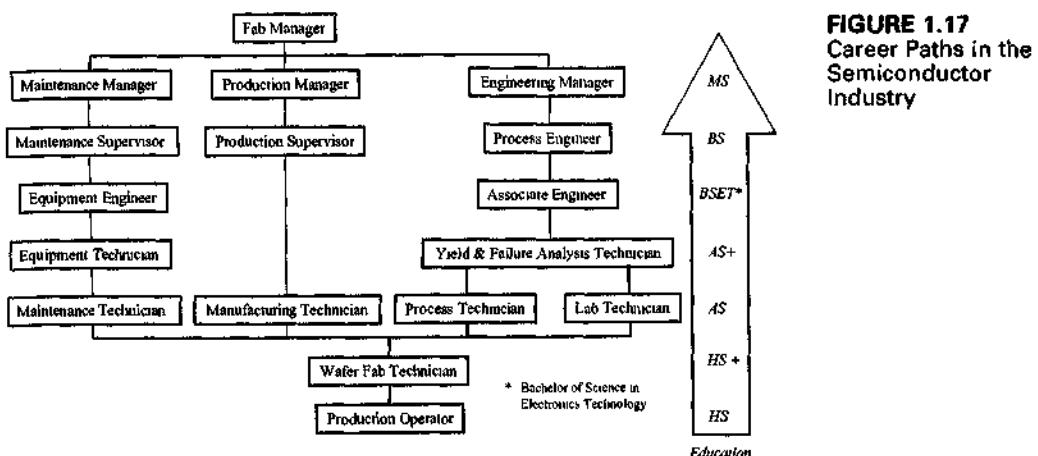
Semiconductor equipment is highly automated. Advanced material handling systems move wafers between workstations with no required human intervention. Expert software systems control nearly all equipment functions, including troubleshooting diagnostics. Technicians and engineers intervene to download product recipes into the equipment software database and interpret software diagnostic commands to take corrective action for equipment repair.

CAREERS IN SEMICONDUCTOR MANUFACTURING

The career paths in semiconductor manufacturing are grouped into three main areas: technician, engineering, and management. The choice of a particular career path generally depends on an individual's technical knowledge, education, and personal goals. Examples of different career paths are shown in Figure 1.17.

Technician

The technician has the technical skills to operate, troubleshoot and maintain the advanced equipment used to fabricate microchips in a production environment. Equipment operation requires knowledge to set up and operate the automated equipment so that it functions properly in the manufacturing process. Most equipment operation is done through software interfaces to the computer-controlled equipment. Troubleshooting equipment problems requires expertise in equipment systems (e.g., electronics, mechanical systems, motors, etc.) with the ability to interpret software diagnostic routines. Equipment maintenance involves extensive training and knowledge of electrical, mechanical, chemical, and computer equipment systems. An equipment technician often performs the troubleshooting and maintenance of wafer fab equipment.



Technicians must be capable of performing their technical role as a manufacturing team member, while contributing to the productivity objectives of the wafer fab (see Figure 1.18). Technicians are knowledgeable about safety procedures for chemicals and equipment. As equipment automation continues toward more integrated equipment across different operations, the skills of the technician in the wafer fabs must also advance.

Technicians in a wafer fab are referred to by different names at the various chip manufacturers, such as a wafer fab technician (WFT), self sustaining technician (SST), manufacturing association (MA), or manufacturing technician (MT). Technicians are also used in the semiconductor industry, such as for field service representatives, to install and service sophisticated wafer fab equipment. The skills required of technicians include computer usage, knowledge of equipment procedures and support, analytical skills to solve problems for process improvement, and critical thinking skills for decision-making. Due to the advanced sophistication of wafer fab equipment, it is clear that people who merely operate equipment are in less demand in the wafer fab. Furthermore, humans create too many problems with respect to wafer contamination, manufacturing repeatability, and inefficiency.

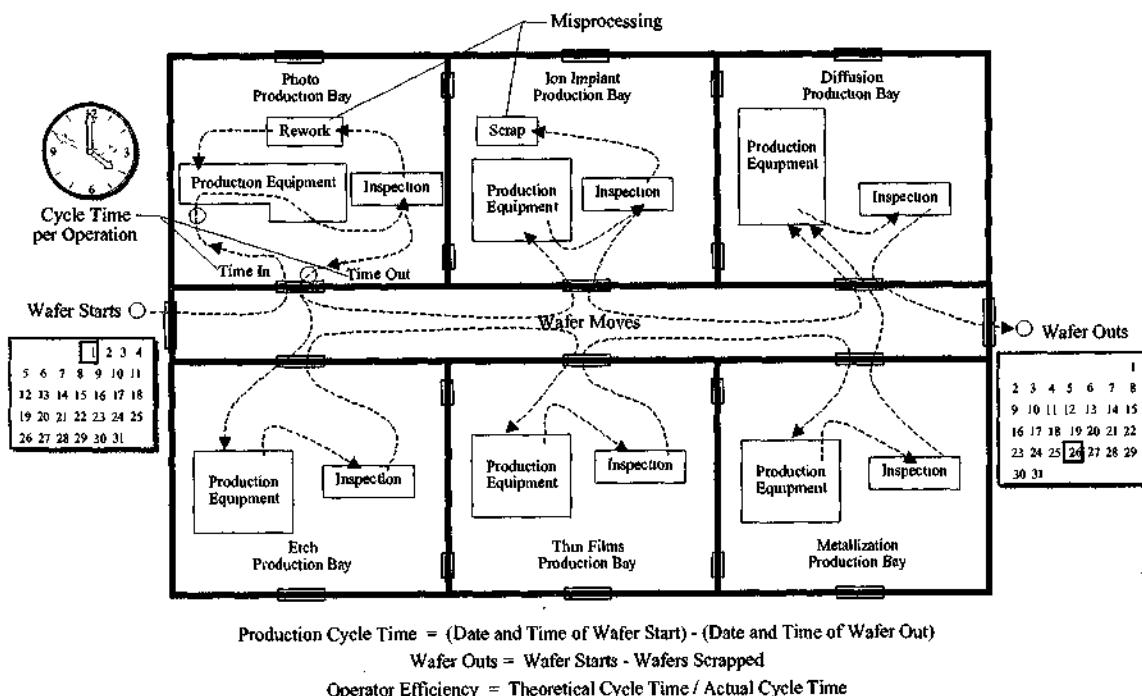
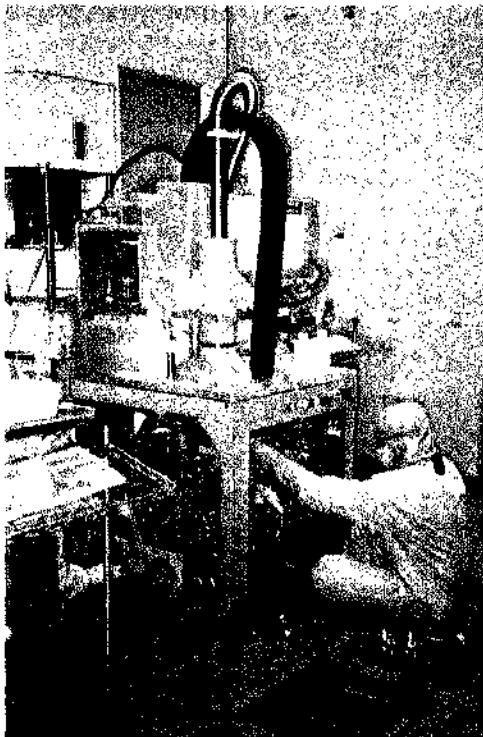
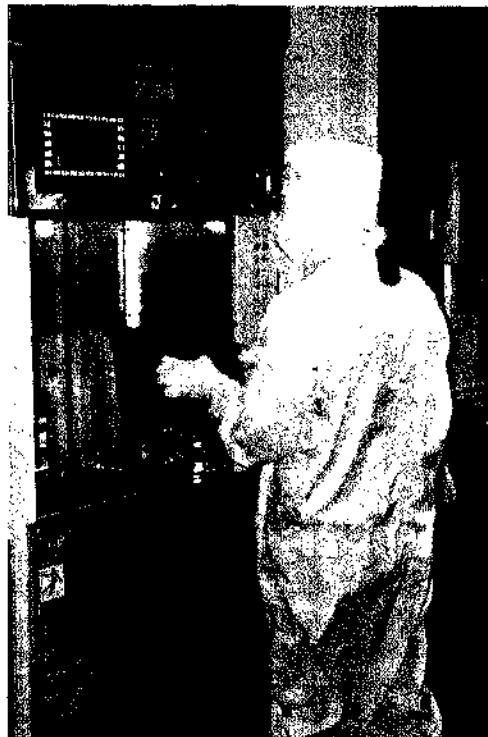


FIGURE 1.18 Productivity Measurements in a Wafer Fab



A. Equipment Technician in a Wafer Fab

B. Technician in a Wafer Fab
(Photos courtesy of Advanced Micro Devices)

Job Descriptions

Within each of the three major career paths, there are various options available for different jobs. A description of some different technician, engineer and management positions in the wafer fab is given.

- ◆ **Wafer Fab Technician:** The wafer fab technician is responsible for operating the wafer fab equipment. The wafer fab technician performs some equipment maintenance and basic troubleshooting of process and equipment. Semiconductor companies often prefer high school or some college (e.g., a certificate). In some cases, a two-year associate of applied science degree may be desirable.
- ◆ **Equipment Technician:** The equipment technician troubleshoots and maintains advanced equipment systems so that the equipment performs correctly during wafer manufacturing. This position typically requires at least a two-year associate of applied science degree and extensive hands-on training.
- ◆ **Equipment Engineer:** The equipment engineer specifies the equipment design parameters and optimizes equipment performance for wafer production. This career typically requires at least a four-year engineering degree.
- ◆ **Process Technician:** The process technician supports the production equipment and process engineering teams by troubleshooting process-related problems. A two-year associate of applied science degree or four-year engineering technology degree is helpful; however, good writing, math, science, and analytical thinking skills are necessary.
- ◆ **Process Engineer:** The process engineer analyzes the performance of the fabrication process and equipment to determine optimum parameter settings. This position typically requires a four-year engineering or science degree.
- ◆ **Field Service Representative:** A field service representative in a wafer fab installs the fabrication equipment. The field service representative also performs equipment maintenance, diagnostics, and repairs to ensure the equipment is production-worthy. This position requires at least a two-year associate degree or four-year engineering technology degree.
- ◆ **Lab Technician:** Lab technicians can work in development laboratories setting up and running experiments. This job skill usually requires some experience and people who are self-motivated.

- ◆ **Yield/Failure Analysis Technician:** These technicians perform the tasks associated with defect analysis, such as preparing the materials to be analyzed and operating analytical equipment to determine the source of problems that arise during wafer fabrication.
- ◆ **Yield Improvement Engineer:** The yield improvement engineer collects and analyzes yield and test data to improve wafer fabrication performance. This position often requires a four-year science or engineering degree.
- ◆ **Facilities Technician:** Facilities technicians support the facility equipment and utilities required in a wafer fab, including chemical management and cleanroom utilities.
- ◆ **Facilities Engineer:** This engineer provides engineering design support for the wafer fab utility infrastructure for chemicals, clean air and equipment that are used in the facility. The position typically requires a four-year engineering degree.
- ◆ **Supervisor/Manager:** Supervisory and management positions in a wafer fab combine technical skills with people management skills to achieve the organizational goals of the company. These positions often will require either a four-year science, engineering, or management degree or an additional two years of college after the associate degree.

SUMMARY

The semiconductor industry is a major part of the high tech industry in the U.S. The invention of the solid-state transistor launched the semiconductor industry in 1947, followed by the introduction of silicon and the development of the integrated circuit (IC). The IC combines multiple components on one chip to increase chip performance and lower cost. There are IC integration eras, with the most recent era being ULSI (ultra large scale integration). IC fabrication occurs on a wafer, with the devices on the top of the silicon and layers of circuitry deposited on the substrate. There are five stages of microchip fabrication: wafer preparation, wafer fabrication, wafer test/sort, assembly and packaging and final

test. The three major trends for improvement in microchip technology are: increase in performance, increase in reliability, and reduction in cost. Chip speed is important for performance and is improved by reducing the critical dimension (CD), or minimum feature size, on a chip. Moore's law predicts that the number of components on a chip will double every 18-24 months, which has remained relatively true since the 1960s. The reliability of chips has improved while the price of chips has substantially decreased since 1947. The semiconductor industry has become increasingly competitive and the wafer fab technician plays an important role in a manufacturer's ability to fabricate microchips.

KEY TERMS

semiconductor
wafers
wafer fab
microchips (chips)
solid-state transistor
planar technology
integrated circuit (IC)
devices
wafer
die
substrate
wafer fab
fab
silicon
wafer fabrication

microchip fabrication
chip fabrication
chip fab
merchant chip suppliers
captive chip producers
fabless company
wafer sort
packaging
chip speed
feature size
circuit geometry
critical dimension (CD)
technology node
scaling
Moore's law
wafer fab technician (WFT)

REVIEW QUESTIONS

1. How big is the high-tech industry relative to the U.S. economy? What has been a major factor in helping the semiconductor industry become so large?
2. List four different industries from the first half of the twentieth century that contributed to the development of the semiconductor industry.
3. When, where, and by whom was the solid-state transistor invented?
4. What is an integrated circuit (IC)? When was it invented, and by whom?
5. List the five integration eras, provide the time period for each era, and give the number of components on a chip in each era.
6. What is a wafer? What is a substrate? What is a die?
7. Give a short description of a wafer fab.
8. List the five major stages of IC fabrication, and give a short description of each stage.
9. Describe merchant chip supplier, captive chip producer, fabless company, and foundry.
10. List the three major trends associated with improvement in microchip fabrication technology, and give a short description of each trend.
11. What is the chip critical dimension (CD)? Why is this dimension important?
12. Describe scaling and its importance in chip design.
13. What is Moore's law and what does it predict? Has this law been basically true?
14. By what factor have chip prices reduced since 1947? Give two reasons for this change.
15. Provide a short description of how the electronic era developed from 1950 through 2000.
16. Describe the responsibilities of a wafer fab technician and an equipment technician.
17. List and give a short description of eight different job descriptions in the semiconductor industry.

SELECTED INDUSTRY WEB SITES

Advanced Micro Devices
 AT&T Tech. History
 Semiconductor International Magazine
 Fairchild Semiconductor
 IBM Microelectronics
 Intel
 International SEMATECH
 MATGEC, Maricopa Advanced Technology Education Center
 Motorola Semiconductor
 Mitsubishi
 National Semiconductor
 NIST, National Institute of Standards and Technology
 Selete, Semiconductor Leading Edge Technologies, Inc.
 SEMI, Semiconductor Equipment and Materials International
 Semiconductor Search Engine
 SIA, Semiconductor Industry Association
 SISA, Semiconductor Industry Suppliers Association
 Solid State Technology Magazine
 Texas Instruments

<http://www.amd.com>
<http://akpublic.research.att.com/>
<http://www.semiconductor.net/>
<http://www.fairchildsemi.com/>
<http://www.chips.ibm.com/>
<http://www.intel.com/>
<http://www.sematech.org/public/index.htm>
<http://matec.org/>
<http://mot-sps.com/>
<http://www.mmc-sil.com/>
<http://www.national.com/>
<http://www.nist.gov/>
<http://www.selete.co.jp>
<http://www.semi.org/>
<http://www.semiseek.com/>
<http://www.semichips.org/>
<http://www.sisa.org/>
<http://sst.pennwellnet.com/home/home.cfm>
<http://www.ti.com/>

REFERENCES

1. *Chip Scale Review*, Volume 3, Number 4 (July-August 1999); p. 25.
2. Historical information generally taken from F. Seitz and N. Einspruch, *Electronic Genie: The Tangled History of Silicon* (Urbana: University of Illinois Press, 1998), p. 52.
3. F. Seitz and N. Einspruch, *Electronic Genie*, p. 52.
4. P. Bondyopadhyay, "In the Beginning," *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 63.
5. J. Pierce, "The Naming of the Transistor," *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 37.
6. Texas Instruments, *The Chip That Jack Built Changed the World*, Texas Instruments, (September 1997).
7. G. Moore, "The Role of Fairchild in Silicon Technology in the Early Days of 'Silicon Valley,'" *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 59.
8. Ibid., p. 54.
9. Ibid., p. 56.
10. G. Moore, "Cramming More Components onto Integrated Circuits," *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 84. Reprinted from *Electronics*, April 19, 1965.
11. C. Chang and S. Sze, *ULSI Technology* (New York: McGraw-Hill, 1996), p. xxi.
12. Ibid., p. xix.
13. I. Ross, "The Invention of the Transistor," *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 17.
14. F. Seitz and N. Einspruch, *Electronic Genie*, p. 94.
15. C. Melliar-Smith, M. Borrus, D. Haggan, T. Lowrey, A. Vincentelli, W. Troutman, "The Transistor: An Invention Becomes a Big Business," *Proceedings of the IEEE* 86, no. 1 (January 1998); p. 93.
16. Ibid., p. 94.
17. F. Seitz and N. Einspruch, *Electronic Genie*, p. 238.

CHAPTER 2

CHARACTERISTICS OF SEMICONDUCTOR MATERIALS

Understanding the structure and bonding of atoms provides the key to knowing how silicon performs its vital role as a semiconductor. Once unlocked, this knowledge

provides the foundation for understanding how the simplicity of a semiconductor device can be a part of the complex world of the microchip.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe the atom, including the valence shell, band theory, and ions.
2. Interpret the periodic table of the elements with regards to main-group elements, and explain how ionic and covalent bonds are formed.
3. State the three classes of materials, and describe each one with regards to current flow.
4. Explain resistivity, resistance, and capacitance, and discuss their importance to wafer fabrication.
5. Describe pure silicon and give four reasons why it is the most common semiconductor material.
6. Explain doping and discuss how the trivalent and pentavalent dopant elements make silicon a useful semiconductor material.
7. Discuss the difference between p-type (acceptor) silicon and n-type (donor) silicon, describe how silicon resistivity changes with the addition of a dopant, and explain the pn junction.
8. Discuss alternative semiconductor materials, with emphasis on gallium arsenide.

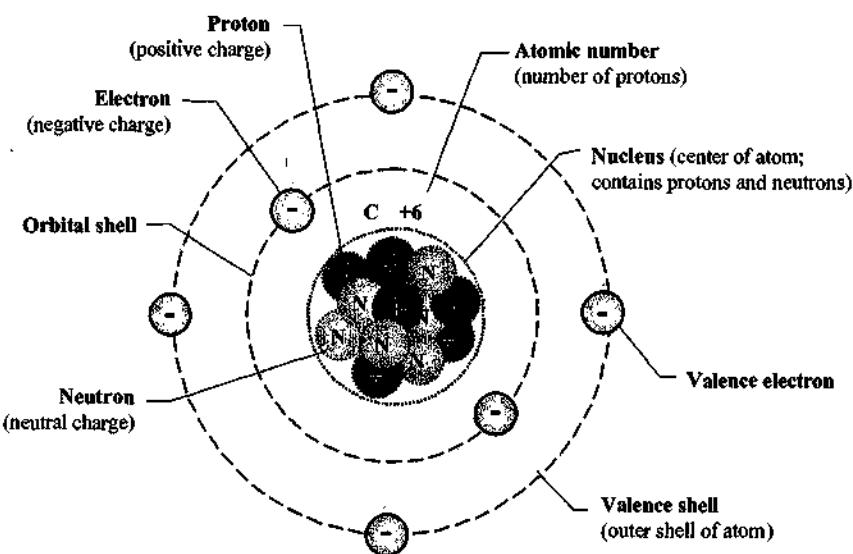
INTRODUCTION

The most important semiconductor material in use today is silicon. Devices are fabricated in silicon and function as a semiconductor because of the unique properties of this material. In this chapter, we will first review some physical and chemical concepts associated with the individual atom. We then will use this information to study properties of silicon material to learn its unique physical and chemical structure and how it permits a semiconductor to function. This chapter will establish basic material and semiconductor concepts that will be used throughout the different process steps in wafer fabrication.

ATOMIC STRUCTURE

Matter is everything that has shape, size, and occupies space in the universe. If we take a material, such as a grain of sand, and continually divide it into smaller and smaller particles, we eventually arrive at the smallest particle, an atom. In the atomic model, atoms consist of three types of particles: the neutral neutrons, the positively charged protons that together make up the nucleus of the atom, and the negatively charged electrons that orbit the nucleus (see Figure 2.1). The number of protons in the atom equals the number of electrons, which makes the atom electrically neutral.

An *element* is a substance formed from one type of atom and is the simplest substance with unique physical and chemical properties. A *molecule* consists of two or more atoms in a structure that are chemically bound together, behaving as an independent unit. A *compound* is a substance composed of two or more atoms (similar or dissimilar) that are chemically bound together and have formed a new substance which has different properties from its individual component atoms.



Carbon atom: The nucleus contains an equal number of protons (+) and neutrons (6 each). Six electrons (-) orbit around the nucleus.

FIGURE 2.1 Elementary Model of the Carbon Atom

Electrons

The basic law of attraction and repulsion that occurs at the microscopic level states that like charges repel and unlike charges attract. Orbiting negative electrons are distributed in a space cloud around the nucleus and held in place by their attraction toward the positive nucleus. Individual electrons occupy orbitals within seven distinct shells. In a hydrogen atom, for example, each shell corresponds to a specific energy level and is identified using the letters K through Q (see Figure 2.2). For all other atoms, referred to as many-electron atoms, the orbital energies within each shell are not the same.

Electron Energy ■ The unit of energy at the atomic level is the electron volt (eV). It represents the kinetic energy (energy associated with motion) gained by an electron in passing from a point of low potential to a point 1 volt higher in potential. One electron volt is equal to 1.6×10^{-19} joules of energy. The electron volt is used to describe electron energy in different processes of semiconductor manufacturing (e.g., ion implant energy levels, as described in Chapter 17).

Valence Shells ■ The outermost electron shell for a given atom is the *valence shell*. Valence electrons are found in the valence shell and have considerable influence on the chemical and physical properties of the atom and exist at the highest energy level for a given atom. The number of

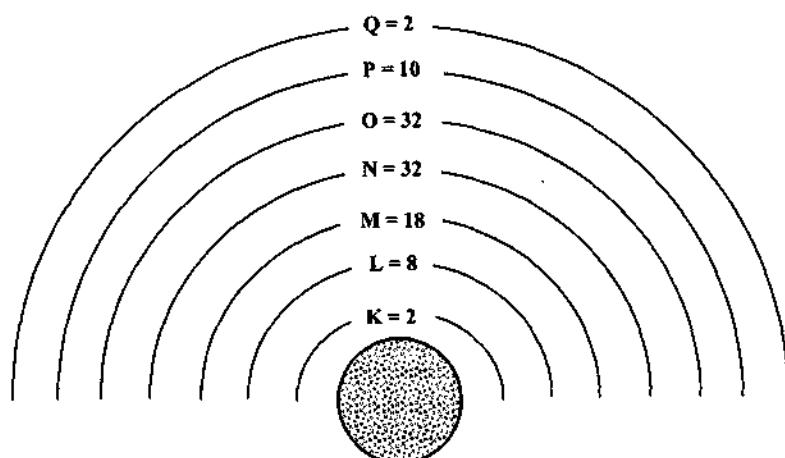


FIGURE 2.2 Electron Shells in Hydrogen Atom

electrons in a valence shell varies depending on the shell level, K to Q. Atoms with one electron in a valence shell that permits eight electrons easily give up this electron. On the other hand, atoms with seven electrons in a valence shell have an affinity for electrons and readily accept one electron to fill the valence shell. The electrons and valence shells for sodium and chlorine are shown in Figure 2.3.

Energy-Band Theory of Solids ■ Energy-band theory explains how electrons change orbital levels in a solid material. The valence electrons are in the valence band. There is a band gap between the valence band and the conduction band (see Figure 2.4 on page 24). The band gap has very high energy levels in some materials that creates a forbidden gap (usually > 2 eV). These materials are referred to as *insulators* because it is hard to move an electron from the valence band to the conduction band. Other materials, called *conductors*, have a valence band that overlaps the conduction band, requiring very little energy for electrons to move into the conduction band. A third class of materials has a band-gap energy level found somewhere between insulators and conductors. These materials are referred to as *semiconductors*. The forbidden gap of semiconductors is moderate. The band-gap energy of silicon is 1.11 eV.

Ions ■ Ions are formed when an atom gains or loses one or more electrons. An atom becomes positively charged if it loses an electron and negatively charged if it gains an electron. Ions with opposite charges are mutually attracted and can form a chemical bond as an ionic compound.

An example of an ionic compound is common table salt, sodium chloride (see Figure 2.5 on page 24). A sodium atom, which is a metal, is initially neutral because it has the same number of

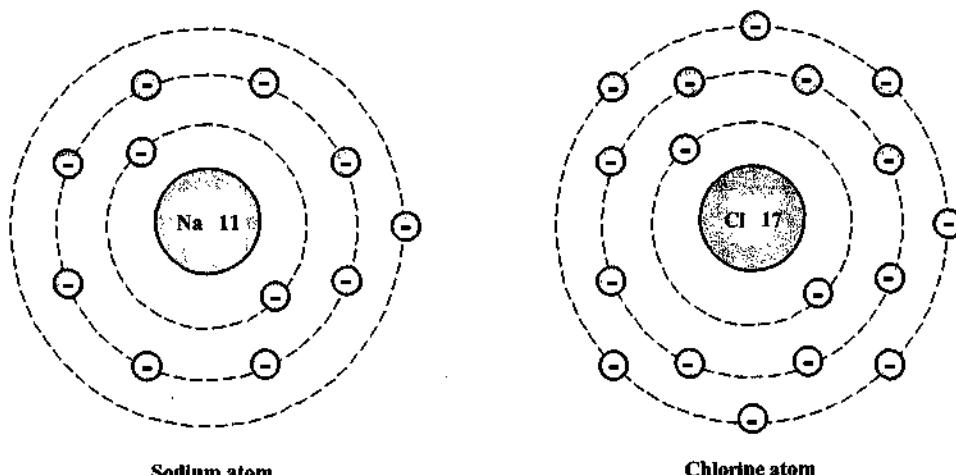
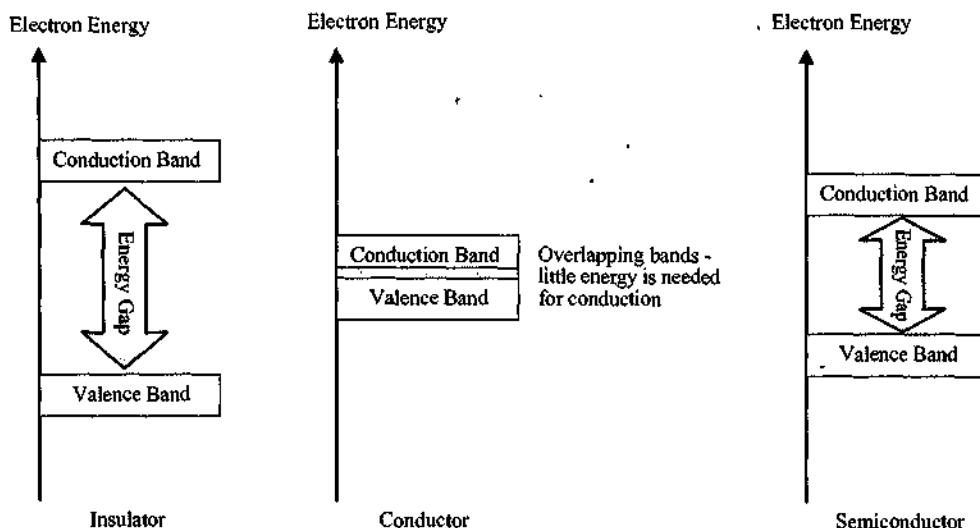


FIGURE 2.3 Electron Shells for Sodium and Chlorine Atoms

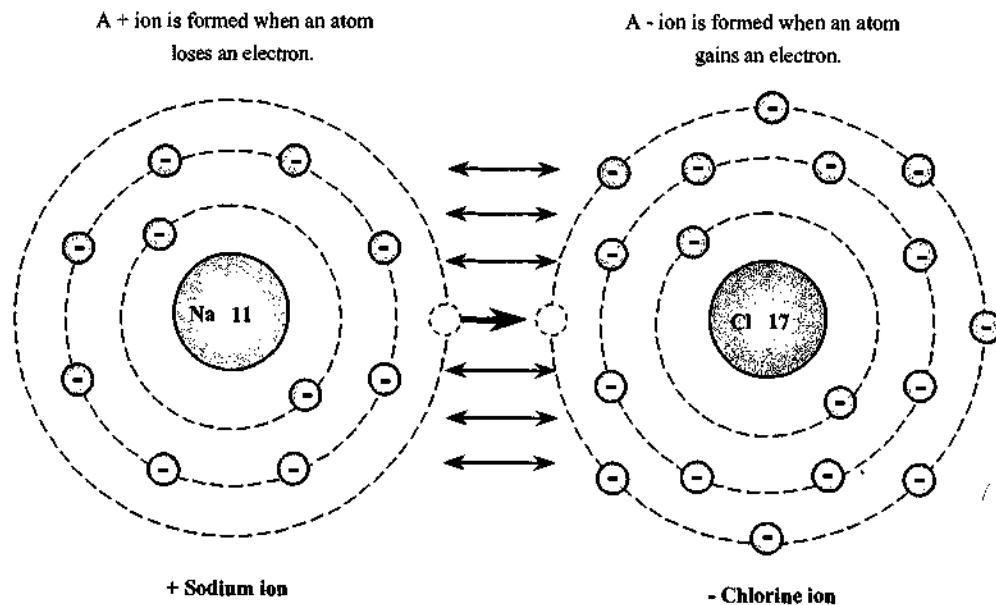
**FIGURE 2.4** Energy Band Gaps

protons as electrons. If it loses one electron, it becomes positively charged as Na^+ and is referred to as a *cation*.¹ A chlorine atom, which is a nonmetal, gains an electron and becomes negatively charged as Cl^- , referred to as the *anion*. The Na^+ and Cl^- are attracted to one another due to opposite charges and form an ionic compound.

Many gases exist as molecules. *Ionization* is the removal of electrons from atoms, creating positively charged atoms or molecules. Ionization is used in many process areas in semiconductor fabrication. Once gas particles are charged by ionization, the gas flow and motion of atoms can be controlled by the use of electrostatic and magnetic fields.

THE PERIODIC TABLE

The periodic table lists all elements known at this time (see Figure 2.6). It is organized in such a manner because of certain periodic patterns of behavior among the elements.

**FIGURE 2.5** Sodium Chloride

IA								VIIA		VIII A			
1 1.008 H								2 4.0028 He					
		IIA											
3 6.939	4 9.012	Li	Be	Lithium	Beryllium								
		Transition Metals											
11 22.989	12 24.312	Na	Mg	Sodium	Magnesium	Al	Si	P	S	Cl	Ar	VIIA	VIII A
19 39.102	20 40.08	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
Potassium	Calcium	Scandium	Titanium	Vanadium	Chromium	Manganese	Iron	Dobalt	Nickel	Copper	Zinc	Gallium	Boron
37 85.47	38 87.62	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
Rubidium	Strontium	Yttrium	Zirconium	Nobium	Molybde-	Techneium	Ruthenium	Rhodium	Palladium	Silver	Cadmium	In	Sn
55 132.90	56 1137.34	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg
Ce	Barium	Lanthanum	Hafnium	Tantalum	Tungsten	Rhenium	Osmium	Irindium	Platinum	Gold	Mercury	Thallium	Lead
87 223	88 226	89 227	104	105	106	107	108	109	110	110	110	110	110
Ft	Ra	Ac	Rf	Ha	Sg	Uhs	Uno	Une	Unu				
Francium	Radium	Actinium											
Metalloids (semimetals)													
58 140.12	59 140.91	60 144.24	61 147	62 150.35	63 151.96	64 157.25	65 158.92	66 162.50	67 164.93	68 167.26	69 168.93	70 173.04	71 174.97
Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Ce	Praseodym-	Neodim-	Prometh-	Samarium	Europium	Gadolin-	Terbium	Dyspro-	Holmium	Erbium	Thulium	Ytterbium	Lutetium
90 232.0491	231 922.238.03	93 237.94	242	95 243	96 247	97 247	98 249	99 254	100 253	101 256	102 253	103 257	
Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
Thorium	Protactinium	Uranium	Neprium	Plutonium	Amerium	Curium	Berkelium	Californium	Einsteinium	Fermium	Mendelevium	Nobelium	Lanthanum

FIGURE 2.6 The Periodic Table of the Elements

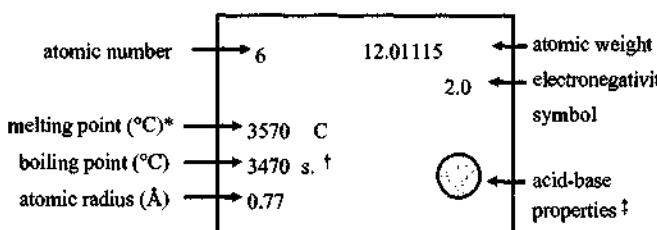
The periodic table has a box for each element (see Figure 2.7). Each element box on the periodic table provides information about the element, including the following:

Atomic symbol: The symbol for each element (e.g., Si is silicon).

Atomic number: Equals the number of protons in the nucleus. All atoms of a particular element have the same atomic number (and therefore the same number of electrons).

Atomic mass number: The sum of the protons and neutrons in an atom. Isotopes of the same element have the same number of protons but different mass numbers (therefore, each one has a different number of neutrons).

Atomic weight: Also referred to as atomic mass. This figure is the average of the masses of the naturally occurring isotopes of an element and is weighted according to the abundance of these isotopes. One *atomic mass unit (amu)* is a relative measure of mass and is exactly equal to $\frac{1}{12}$ the mass of a carbon-12 atom.



* Based on carbon-12. () indicates the most stable or best-known isotope.

† s. indicates sublimation

‡ For representative oxides of group. Oxide is acidic if color is red, basic if color is blue, and amphoteric if both colors are shown. Intensity of color indicates relative strength.

FIGURE 2.7 Element Box of the Periodic Table

We will be concerned primarily with the main-group elements of the periodic table found in the columns with group numbers from IA through VIIA. In the periodic table, the group number is given for each column and represents the number of valence-shell electrons. For instance, Boron (B) is in Group IIIA (found in column IIIA) and therefore has three valence electrons. The group number characteristics of elements commonly used in the semiconductor industry are outlined in Table 2.1.

Ionic Bonds

An ionic bond is formed when valence-shell electrons are transferred from the atoms of one element to another. Unstable atoms easily form ionic bonds. An example is sodium chloride. Sodium (Na) is located in Group IA, indicating there is one electron in its valence shell. Na is an unstable element that is highly corrosive. It is carefully controlled in wafer fabrication to avoid device contamination (see Chapter 6). Chlorine (Cl) is found in Group VIIA and has seven valence-shell electrons. This atom lacks one electron for a fully occupied valence shell; therefore, it also is unstable. Because of their instability, these two elements (Na and Cl) have an affinity for each other. Na readily gives up its valence-shell electron to Cl, forming an ionic bond (see Figure 2.8).

In the formation of ionic bonds, the transfer of electrons is such a common process that we define two terms that apply specifically to this type of bond:

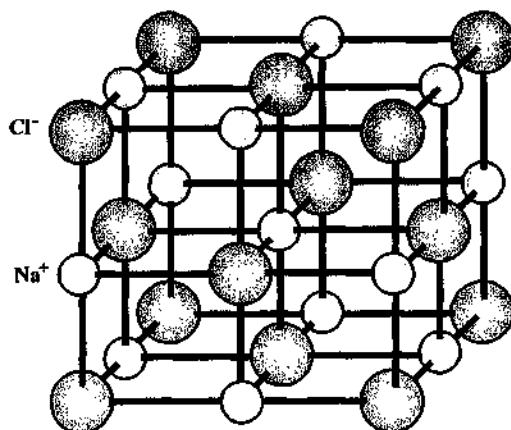
Oxidation: loss of electrons

Reduction: gain of electrons

Thus, in the formation of NaCl, the neutral Na atom is oxidized into a positive Na^+ ion when it loses an electron. A neutral Cl atom is reduced into a negative Cl^- ion by gaining an electron.

TABLE 2.1 Group Number Characteristics of Commonly Used Chemical Elements in Wafer Fabrication

Group Number	Characteristics
I A	<ul style="list-style-type: none"> • 1 valence electron that is easily given up; low electronegativity • Highly unstable • Very reactive; explosive • Form ionic bonds
II A	<ul style="list-style-type: none"> • Prefer not to use the metals in this group due to contamination issues • 2 valence electrons • Somewhat unstable • Quite reactive • Prefer not to use metals in this group
III A	<ul style="list-style-type: none"> • 3 valence electrons • Dopant elements (primarily B) added to semiconductor material • Common interconnect conductor material (Al)
IV A	<ul style="list-style-type: none"> • 4 valence electrons • Semiconductor materials • Form covalent bonds
V A	<ul style="list-style-type: none"> • 5 valence electrons • Dopant elements (primarily P and As) added to semiconductor material
VIA	<ul style="list-style-type: none"> • 6 valence electrons
VIIA	<ul style="list-style-type: none"> • 7 valence electrons; readily accept electrons; high electronegativity • Corrosive • Very reactive • Form ionic bonds • Useful in some semiconductor applications; used as etching and cleaning compounds
VIIIA	<ul style="list-style-type: none"> • 8 valence electrons • Stable; nonreactive • Inert gas • Safe to use in some aspects of semiconductor manufacturing
IB	<ul style="list-style-type: none"> • Best metal conductors • Cu is replacing Al as primary interconnect conductor material
IVB – VIB	<ul style="list-style-type: none"> • Refractory (high melting temperature) metals commonly used in semiconductor manufacturing to improve metallization (especially Ti, W, Mo, Ta, and Cr) • React well with silicon to form stable compound with good electrical characteristics

**FIGURE 2.8** Ionic Bond for NaCl

Covalent Bonds

Another form of chemical bond is a covalent bond, where atoms of different elements share valence-shell electrons. The atoms share these valence electrons in order to attain a full valence shell and become stable. An example of a covalent bond is hydrogen chloride (HCl), where molecules consist of one hydrogen (H) atom forming a covalent bond with one chlorine (Cl) atom to form HCl (see Figure 2.9).

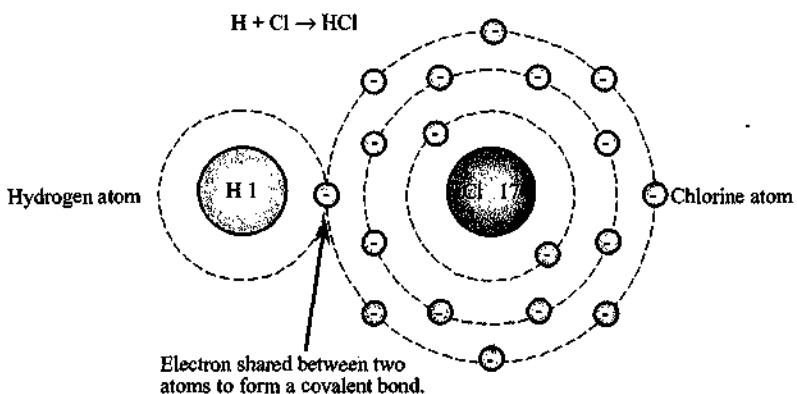


FIGURE 2.9 Covalent Bond for HCl

CLASSIFYING MATERIALS

Electrical current is the movement of charge carriers (electrons in metals, electrons or holes in semiconductors) from one point to another in a material under the influence of applied electric fields.² The unit for current is amps (see Figure 2.10). A major way of classifying electronic materials is based on how current flows through the material.

There are three different classes of materials based on the flow of electric current through these materials:

- ◆ Conductors
- ◆ Insulators
- ◆ Semiconductors

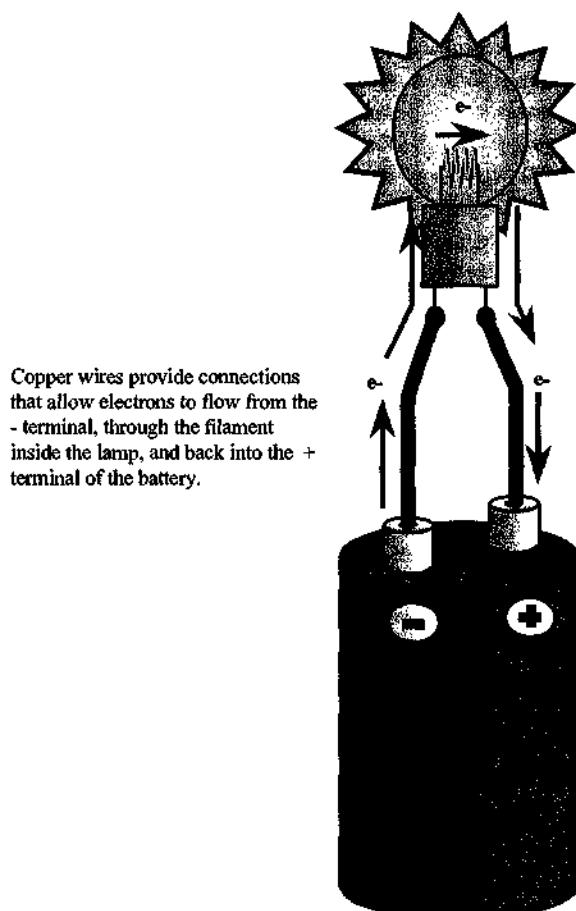


FIGURE 2.10 Electrical Current Flow
(Used with permission from Advanced Micro Devices)

Conductors

A material is a *conductor* if electrons readily flow through the material as electrical current. A good conductor has high current-carrying capability, referred to as current density. A conductor generally has few valence electrons in its outermost shell that are loosely bound and easily given up by the atom. Metals typically have this valence-shell configuration. Common conductor metals used in semiconductor manufacturing are aluminum, used for interconnect wiring between devices, or tungsten, used as the material for forming electrical interconnections between metal layers.

An example of a good metal conductor is copper (Cu), which recently has been introduced as a replacement for aluminum metallization for interconnecting the different devices on a microchip during wafer fabrication. Cu has twenty-nine electrons, with one electron in the valence shell that is relatively far from the nucleus. This electron is easily freed from the atom and is available to conduct electricity (see Figure 2.11).

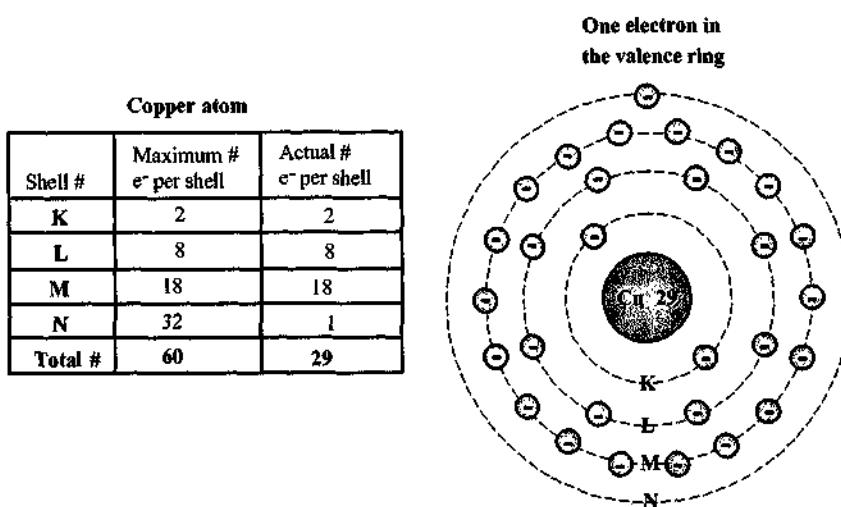


FIGURE 2.11 Flow of Free Electrons in Copper

Conductors require little energy to move a valence electron from its valence band into the conduction band. This electron movement leaves a gap in the covalent bond of the atom, which is referred to as a hole. For every conduction-band electron there must exist a valence-band hole. Valence electrons can jump into the position of the hole, which promotes electron movement and conduction. This condition is referred to as the *electron-hole pair*.

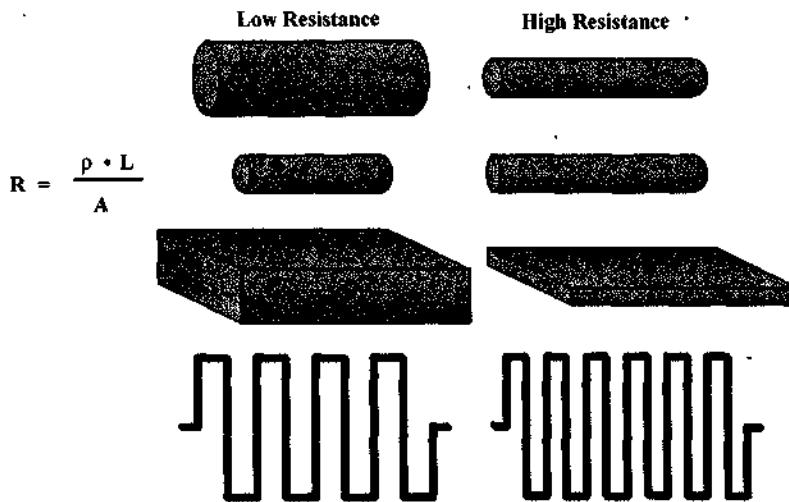
Very quickly after becoming a free electron in the conduction band, an electron gives up its energy and falls into a hole in the covalent bond of the valence band. This process is known as *recombination*. The time period from when an electron moves from the conduction band until recombination is known as the *lifetime* of the electron-hole pair. Thermal energy leads to the continual creation of electron-hole pairs followed by their recombination.

Conductivity and Resistivity ■ The property of a material to conduct electricity is its *conductivity*. Materials are usually characterized by their *resistivity* (ρ), or resistance to current flow. A material with lower resistivity has better conductivity. These two properties of resistivity and conductivity depend only on the material, not its geometry, and are related by the following formula:

$$\rho = \frac{1}{C}$$

where, ρ = resistivity in ohms-centimeter ($\Omega\text{-cm}$)
 C = conductivity

Resistivity of materials is important in semiconductor products because of its effects on the electrical operation of an integrated circuit. Resistivity can be deliberately designed into an IC in the form of an electronic component called the resistor. In this case the resistor serves as a desired control for current in a particular part of an electronic circuit. In some cases, resistivity is an

**FIGURE 2.12** How Wiring Sizes Affect Resistance

undesirable characteristic of materials that introduce too much resistance to the flow of current, resulting in power losses and overheating within the IC.

Resistance. *Resistance* is the opposition to current flow and is accompanied by the dissipation of heat. A material with high resistance has high opposition to current flow. Resistance depends on both resistivity and the physical geometry of the material and is calculated by:

$$R = \frac{\rho l}{\text{area}}$$

where, R = resistance of the conductor material, in ohms

ρ = resistivity of the conductor material, in ohms-cm

l = length of the conductor, in cm

area = cross-sectional area of conductor = width \times thickness

Reduced feature sizes in wafer fabrication makes resistance an important parameter. Smaller geometries cause increased resistance in interconnect lines, which is undesirable due to increased heat losses. Lower resistance is why copper is replacing aluminum as the primary interconnect wiring material. Another example of the application of resistance in semiconductor manufacturing is the sheet resistance measurement to control the thickness of conductive films (see Chapter 7). Since resistance is dependent on geometry, the sheet resistance of a thin film changes depending on the film's thickness.

Insulators

An insulator is a material that has a high resistance to current flow. Another term for an insulator is a *dielectric*. Insulators have no loosely bound electrons in the valence shell available for electrical conduction, with a high energy gap to keep valence band electrons out of the conduction band. Examples of insulators in everyday life are rubber, plastics, glass, and ceramics. Some insulator materials in semiconductor manufacturing are silicon dioxide (SiO_2), silicon nitride (Si_3N_4), and polyimide (a plastic material).

Purified and deionized (DI) water is a good example of an insulator, with a resistivity of about 18×10^6 ohms-cm, or 18 megohms-cm. There are insufficient free electrons in DI water to be able to sustain current flow from a small battery. However, the conductivity of an insulator can be improved by adding an impurity to it (see Figure 2.13). In the case of water, we can add common table salt. The salt dissociates in water into its basic ionic elements of sodium (Na^+) and chlorine (Cl^-), forming an electrolyte (a conducting solution). The net effect of these charged atoms is the same as having free electrons in a piece of copper wire—if there are sufficient charged atoms, current flow can be sustained. As we shall learn later, adding an impurity to alter an insulator material's conductivity is an important aspect of semiconductor technology.

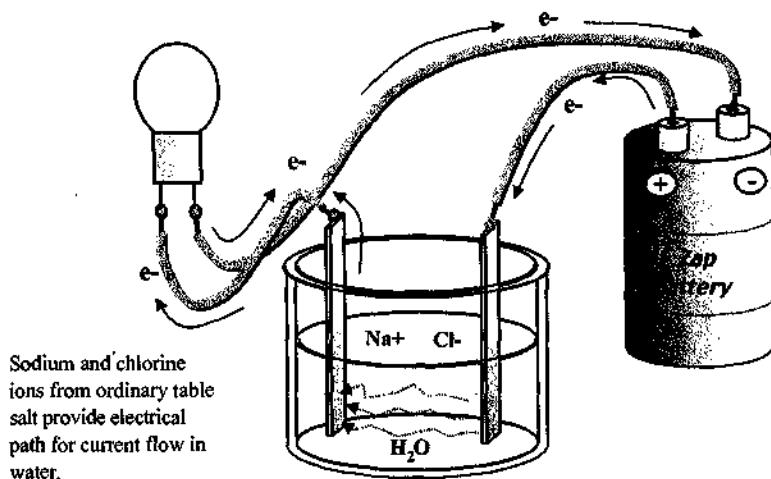


FIGURE 2.13 Adding an Impurity to Water to Improve its Conductivity

Capacitance ■ *Capacitance* is the storage of electrical charge on two conductive plates separated by a dielectric material (see Figure 2.14). The unit of measure for capacitance is farads and is usually expressed as picofarads (10^{-12} farads) for capacitance in integrated circuit structures. The amount of charge that can be stored by a capacitor varies depending on certain physical characteristics. These characteristics include the area of the plates, the spacing between the plates, and the quality of the insulation material between the plates, better known as the *dielectric constant*, k (farads/cm). The k value of air is 1, while that of glass is between 4 and 7.

We can use a simple circuit to explain how a capacitor holds an electrical charge (see Figure 2.15 on page 32). In this circuit, a 1.5 V battery is connected to a capacitor. When the switch is closed, electrons from the left metal plate are attracted to the positive (+) side of the battery. At the same time, electrons from the negative side (-) of the battery flow to the right plate to balance the deficiency of electrons from the left side. The initial current stops due to the dielectric material between the plates. The result, however, is a difference in potential between the two plates. An electrostatic field exists between the charged plates.

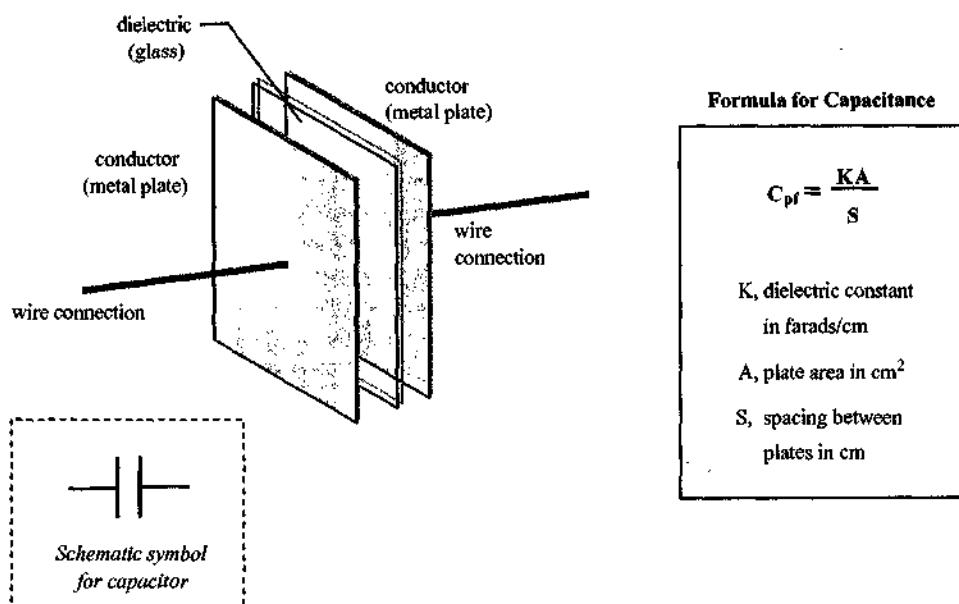


FIGURE 2.14 Basic Capacitor Structure
(Used with permission from Advanced Micro Devices)

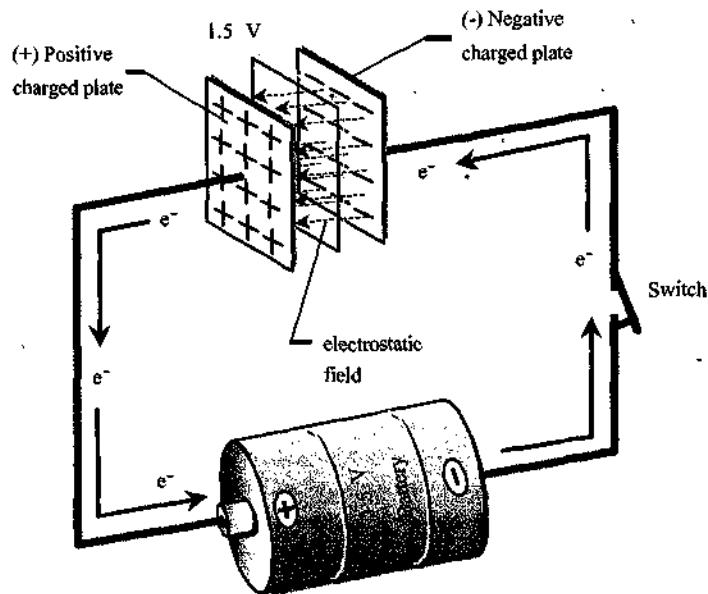


FIGURE 2.15 Battery Charges a Capacitor
(Used with permission from Advanced Micro Devices)

If the battery and circuit are removed from the capacitor, the left plate is positively charged and the right plate is negatively charged (see Figure 2.16). The overall difference in potential is equal to the original battery value and remains as long as there is no other way for electrons to flow to the opposite side of the capacitor (in reality, some leakage does occur through the dielectric). Understanding these concepts regarding how a capacitor works is crucial for understanding how a field-effect transistor (FET) works (see Chapter 3).

Dielectric Constant. The dielectric material is a key component of capacitors. This material can change the capacitance of a capacitor by more effectively concentrating the electric field occurring between the two conductors. This dielectric constant, k , has become an important parameter in semiconductor performance. As current flows through adjacent metal conductors that interconnect the devices on the chip (referred to as wiring), it is desirable to have a low- k dielectric material to minimize capacitance losses (see Figure 2.17).

Semiconductors

The third type of material is *semiconductors*. These materials are special because they can function as either conductors or insulators. A semiconductor material has a small energy gap (i.e., 1.11 eV for silicon) that is a value between an insulator (>2 eV) and a conductor. This energy gap permits the electrons to jump from the valence band into the conduction band when energy is supplied. This

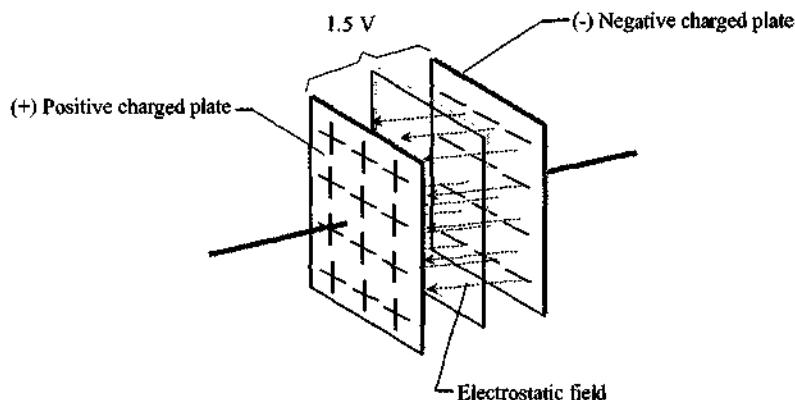
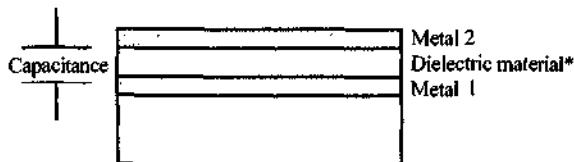


FIGURE 2.16 Capacitor Holds a Charge



* Low k dielectric reduces capacitance between metal layers.

FIGURE 2.17 Low- k Dielectric Material

action occurs when a semiconductor is heated, thus improving conductivity with an increase in temperature (which is the opposite condition for a conductor).

The most important semiconductor material in wafer fabrication is silicon, used as the semiconductor wafer substrate in over 85% of all chips.³ This text will only address silicon because of its predominant use in the semiconductor industry. A short review of some of the other semiconductor materials and their applications is given at the end of this chapter.

SILICON

Silicon is an elemental semiconductor material because it has four valence-shell electrons, along with the other elements located in Group IVA of the periodic table (see Figure 2.18 on page 34). The number of valence-shell electrons in silicon places it midway between the valence-shell condition of a good conductor (one valence electron) and an insulator (eight valence electrons).

Silicon is not found pure in nature. It must be refined and purified to become pure for semiconductor manufacturing (see Chapter 4). It usually occurs as silica (silicon dioxide, or SiO_2) and other silicates. Silica is sand and is a primary ingredient of glass. Other forms of SiO_2 are rock crystal, quartz, agate, and opal.

The melting point of silicon is 1412°C. Silicon is a hard and brittle material that fractures easily if deformed, much like a piece of glass. It can be polished to a mirrorlike finish. Silicon exhibits many of the same characteristics as metals and at the same time exhibits characteristics of nonmetals. This is why silicon is classified as a semiconductor, midway between the conductors (metals) and insulators (nonmetals) on the periodic table.

Pure Silicon

Pure silicon is referred to as *intrinsic silicon*, with no contaminants or impurities from other substances. The silicon atoms in pure silicon bond together through covalent bonds to share electrons and complete their valence shells (see Figure 2.19 on page 34).

Many of silicon's properties result from this strong covalent bonding. The covalent bonds in pure silicon hold the atoms together to form a solid, electrically stable material that is an insulator. Pure silicon is a poor conductor because all valence shells are fully occupied in covalent bonds. In this pure form, silicon is not useful as a semiconductor.

The solid material formed when two or more silicon atoms bond together in this set, repeatable pattern, as the one shown in Figure 2.19 on page 34, is referred to as a *crystal*. A crystal is a smooth, glassy solid that forms a lattice structure with a three-dimensional shape. An example of a crystal material is window glass. We will study more about crystals in Chapter 4 when we discuss silicon wafer preparation.

Why Silicon?

Germanium was the first material used for semiconductors in the 1940s and early 1950s, but as we learned in Chapter 1, it was soon replaced by silicon. Why was silicon chosen as the predominant semiconductor material? There are four main reasons for the selection of silicon as the primary semiconductor material:

- ◆ Abundance of silicon
- ◆ Higher melting temperature for wider processing range

- ◆ Wider temperature range of operation
- ◆ Natural growth of silicon dioxide

Semiconductors	
Group IVA	
C, Carbon	6
Si, Silicon	14
Ge, Germanium	32
Sn, Tin	50
Pb, Lead	82

FIGURE 2.18 Group IVA Elemental Semiconductors

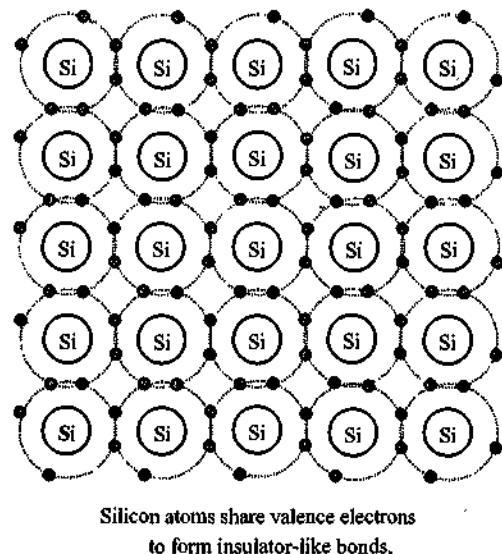


FIGURE 2.19 Covalent Bonding of Pure Silicon

Silicon is the second most abundant element on earth and makes up about 25% of the earth's crust. If processed properly, silicon can be refined into ample quantities of the very pure form necessary for semiconductor fabrication which leads to lower costs. Silicon's melting temperature of 1412°C is much higher than the melting temperature for germanium which is 937°C. This higher melting temperature permits silicon to withstand high-temperature processing. Another advantage to using silicon is that a semiconductor device made from silicon can function over a wider temperature range than germanium, which increases the semiconductor's applications and reliability.

Finally, an important reason for using silicon as a semiconductor material is the ability to naturally grow *silicon dioxide* (SiO_2) on the silicon surface (see Figure 2.20). SiO_2 is a high-quality, stable electrical insulator material that also serves as a good chemical barrier to protect silicon from external contaminants.⁴ Electrical stability is important to avoid leakage between adjacent conductors in an IC. The ability to grow stable, thin SiO_2 material is fundamental to the fabrication of high-performance metal-oxide semiconductor (MOS) devices. SiO_2 has mechanical properties similar to silicon, which allows high temperature processing without excessive wafer warpage.

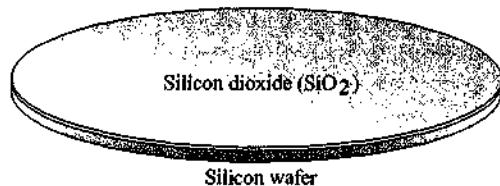


FIGURE 2.20 SiO_2 on Silicon Wafer

Doped Silicon

Silicon in its pure state has little use in semiconductor technology. However, the structure of silicon can be altered to greatly enhance its conductivity by adding small amounts of other elements to the material through a process known as doping. *Doping* is the process of adding certain elements to pure silicon to improve the conductivity of the semiconductor (see Figure 2.21). For example, the electrical resistivity (ρ) of pure silicon is approximately 2.5×10^5 ohms-cm. If only one in

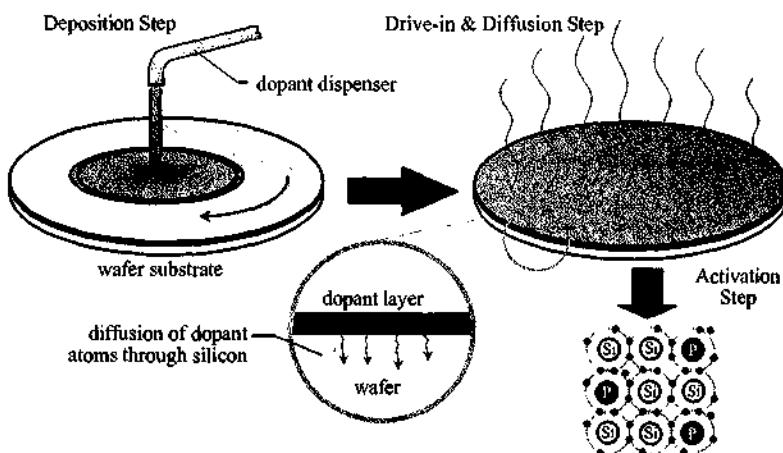


FIGURE 2.21 Doping of Silicon
(Used with permission from Advanced Micro Devices)

every one million silicon atoms is replaced by one atom of arsenic, resistivity will drop to 0.2 ohms-cm.⁵ This is an improvement in conductivity of 1,250,000 times.

The elements added during doping are referred to as *dopants* or *impurities* because the silicon is no longer pure. In other words, we dope the silicon with an impurity so that it will conduct electricity. The more impurity added then the higher the conductivity (or the lower the resistivity). Note that we use the term *impurity* to indicate that we have added another element to the silicon. We intentionally add the impurity to increase the conductivity of silicon. Doped silicon is also known as *extrinsic silicon*.

The concept of doping the silicon with an impurity to improve electrical conductivity is a critical aspect of semiconductor fabrication. If we can introduce impurities that alter the silicon's resistivity and then control when the silicon performs as an insulator or as a conductor, then we have the essence of solid-state technology.

Dopant Materials ■ Silicon is located in Group IVA of the periodic table and has four valence electrons. Elements from the two adjacent groups are commonly used for doping: Group IIIA and Group VA (see Figure 2.22). Group IIIA elements are referred to as *trivalent* because of three valence electrons, whereas elements from group VA are *pentavalent* because of five valence electrons. The trivalent dopant increases the number of holes (positive doping or p-type), whereas the pentavalent will increase the number of free electrons (negative doping or n-type).

Acceptor Impurities		Semiconductor		Donor Impurities	
Group III (p-type)	Group IV			Group V (n-type)	
<u>Boron</u> 5	Carbon 6			Nitrogen 7	
Aluminum 13	<u>Silicon</u> 14			<u>Phosphorus</u> 15	
Gallium 31	Germanium 32			<u>Arsenic</u> 33	
Indium 49	Tin 50			<u>Antimony</u> 51	

* Elements underlined are the most commonly used in silicon-based IC manufacturing.

FIGURE 2.22 Silicon Dopants

When trivalent dopant atoms are added to silicon, the resulting material is called a *p-type silicon*. The trivalent dopants are referred to as *acceptors* (they accept an extra mobile electron), with the most common acceptor element being boron. When a pentavalent element is added to pure silicon, the resulting material is referred to as an *n-type silicon*. Pentavalent dopants are referred to as

donors (they donate an extra mobile electron), and are typically either phosphorous, arsenic, or antimony. The different doping elements are listed in Table 2.2.

TABLE 2.2 Dopant Elements Commonly Used in Semiconductor Manufacturing

Trivalent Dopants (p-type, positive doping, acceptor)	Pentavalent Dopants (n-type, negative doping, donor)
Boron (B)	Phosphorus (P) Arsenic (As) Antimony (Sb)

N-Type Silicon. For an n-type silicon, there are more electrons than valence-band holes. This is shown in Figure 2.23 for silicon with a pentavalent phosphorous dopant atom.

The silicon atoms will form covalent bonds with the donor phosphorous atom, each sharing one phosphorous electron. However, the fifth phosphorous electron is not bound to any of the surrounding silicon atoms. Because of this, the fifth phosphorous electron needs little energy to break away and enter the conduction band. For n-type silicon, conduction-band free electrons are the majority carriers, of which the material has an abundance. There are also many fewer minority carrier valence-band holes.

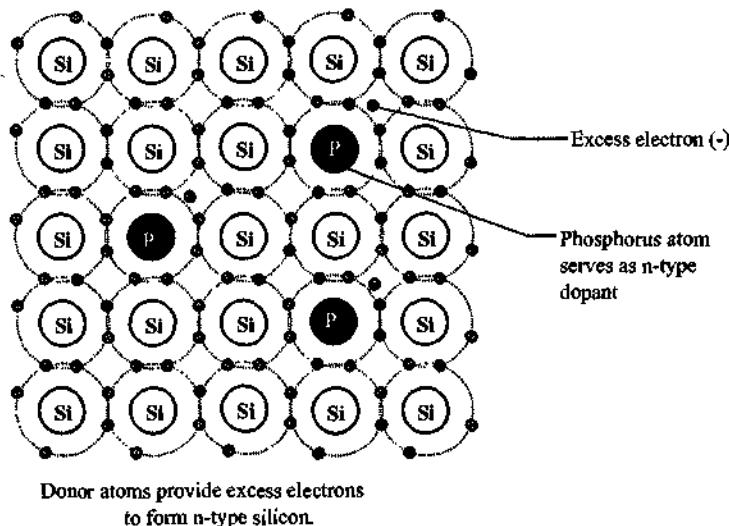


FIGURE 2.23 Electrons in n-Type Silicon with Phosphorous Dopant

One electron in the conduction band is not significant for conduction. However, when we dope silicon with a dopant, we add literally millions of dopant atoms, producing many electrons that are not part of the covalent bonds. There is a high amount of movement between electrons and holes. Negative electrons are attracted to the positive holes. Electrons enter the conduction band relatively easily. If a voltage is applied to this material, the electrons can be made to flow through the material as electric current (see Figure 2.24).

Note that the doped silicon is still electrically neutral (this is true for n-type or p-type silicon). In the case of n-type silicon, this is because each phosphorous atom still has the same number of protons as electrons, as do the silicon atoms. Thus the overall number of protons and electrons in the semiconductor is still equal and the result is a net charge of zero. What is not equal is that there are many more conduction-band electrons (majority carriers) than there are valence-band holes (minority carriers).

P-Type Silicon. In the p-type silicon shown in Figure 2.25, the boron atom is a p-type acceptor dopant that forms a covalent bond with four adjacent silicon atoms. The boron acceptor atoms provide a deficiency of electrons caused by the lack of a fourth electron in the boron

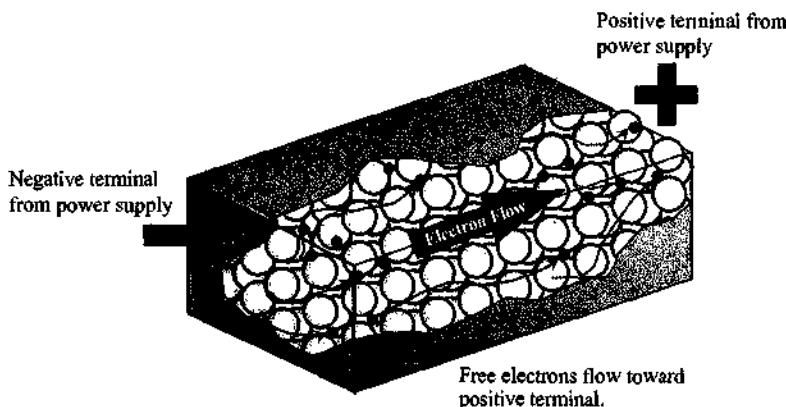


FIGURE 2.24 Flow of Free Electrons in n-Type Silicon

atom, thus creating the p-type silicon. There is an excess of holes (absence of electrons) in the valence shell.

Since there are many more valence-band holes than conduction-band electrons, the holes are the major current carriers in p-type material. The holes are referred to as *majority carriers* and electrons are the *minority carriers*. If a dc voltage is applied across a p-type semiconductor, then the large number of holes in the material will attract electrons from the negative terminal of the voltage source into the p-type semiconductor. This is current flow in a p-type material (see Figure 2.26 on page 38). The holes appear to move because each time an electron moves into a hole it serves to create a hole in its previous position. The holes seemingly move in the opposite direction as the electrons.

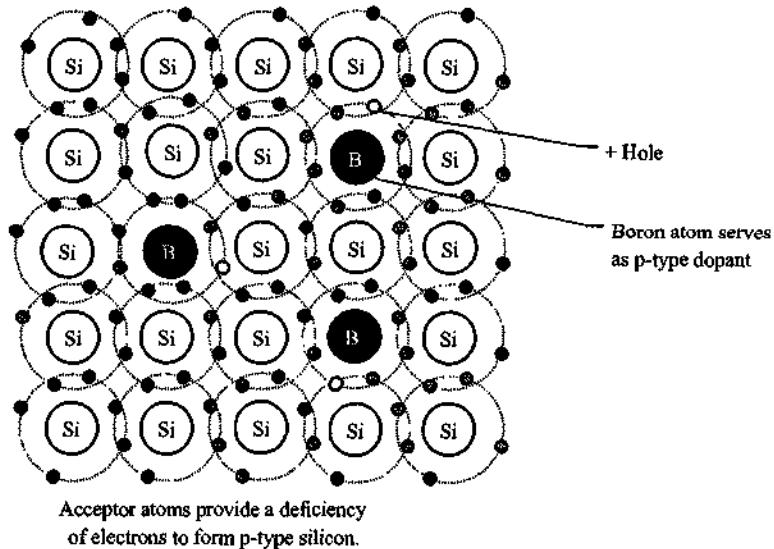
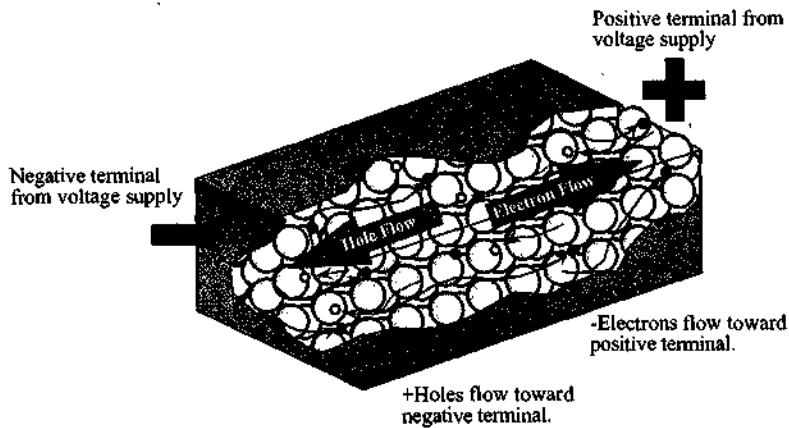


FIGURE 2.25 Holes in p-Type Silicon with Boron Dopant

Resistivity of Doped Silicon. We obtain precise control of silicon's resistivity by introducing dopants into the silicon crystal structure. The concentration of the dopant atom placed in the silicon crystal defines how well the material will conduct electrical current. Pure silicon has a resistivity of about $250,000 \Omega\text{-cm}$ and is an insulator. Compare this to copper, a good conductor, with a resistivity of $1.7 \mu\Omega\text{-cm}$ ($0.0000017 \Omega\text{-cm}$). By adding the proper type and concentration of dopants to pure silicon, the doped silicon's resistivity is decreased and conductivity is improved (see Figure 2.27 on page 38). For a given resistivity there is less concentration of n-type dopants than p-type dopants. This is because it takes less energy to move an electron than to move a hole.

**FIGURE 2.26** Flow of +Holes in p-Type Silicon

It only takes a small amount of dopant (from as little as 0.000001% to 0.1%) to make silicon a useful conductor. This is important for the fabrication of semiconductor devices on wafers. The amount of dopant in the silicon, or concentration, is carefully controlled during semiconductor fabrication to attain a precise resistivity. The ion implantation process for adding dopants to silicon will be covered in Chapter 17.

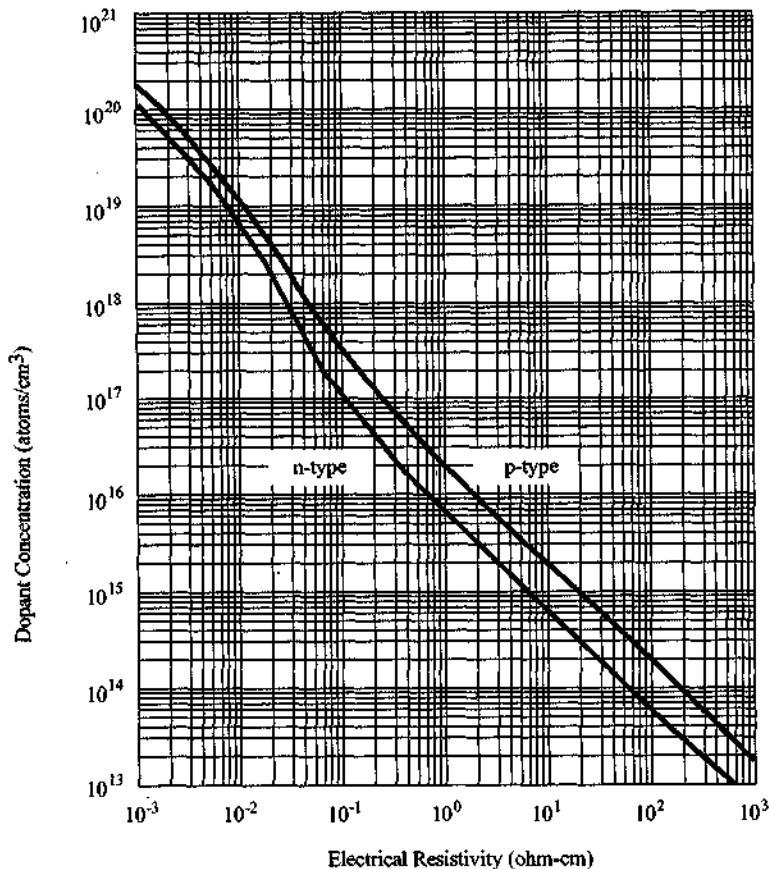


FIGURE 2.27 Silicon Resistivity Versus Dopant Concentration
Redrawn from S. Ghandi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, John Wiley & Sons, 1994.

pn Junctions

We dope pure silicon with pentavalent or trivalent elements to obtain either an n-type or p-type semiconductor. The type and concentration of the dopant element determines whether conduction occurs by hole or by electron and determines the final value of the silicon's resistivity. The number of carriers, and therefore the resistivity, is determined by the net donor and acceptor atom concentration in the silicon crystal. Because of this, an n-dopant can be implanted or diffused into a p-region to convert that region into an n-region (or vice versa). To achieve this, the concentration of the n-type donor dopant must exceed that of the p-type acceptor dopants in the region.

The ability to incorporate both n-type and p-type regions in the silicon crystal is beneficial because semiconductor devices require both of these regions in order to function as a semiconductor. It is the junction between the n-type and p-type regions that is important and creates the useful characteristics of silicon as a semiconductor. This junction is referred to as a *pn junction* (see Figure 2.28).

The pn junction is the essence of solid-state electronics and is the basis for how semiconductor wafers can achieve their unique ability to act as an insulator and a conductor, depending on the bias voltage applied to the junction. The specifics of how a pn junction functions as a semiconductor are discussed in Chapter 3. The creation of a pn junction is almost universally done in wafer fabrication with ion implantation (see Chapter 17).

Remember that a pn junction is formed between two pieces of essentially the same material. The p-type and the n-type materials scarcely differ except for a minute amount of a dopant. The n-type material has extra mobile electrons from the donor impurity, whereas the p-type material has extra mobile holes. It is deceptive to speak of one material meeting the other.⁶ The junction is so intimate that the n-type and p-type materials are formed from one continuous solid. The silicon crystal with a pn junction continues to look like and in most ways behave like the solid crystal material that it is.

The depth and definition of this junction is critical in semiconductor manufacturing. As device critical dimensions are reduced, the ability to precisely control both the pn junction in silicon (e.g., junction depth) and the concentration of dopants is becoming a major challenge for semiconductor chip fabrication.

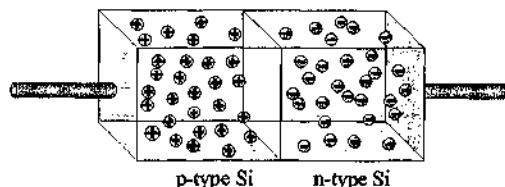


FIGURE 2.28 Cross Section of Planar pn Junction

ALTERNATIVE SEMICONDUCTOR MATERIALS

Germanium and silicon are the two elemental semiconductor materials from Group IVA with four valence electrons. We have learned that germanium was the first semiconductor material used for transistor fabrication and was replaced by silicon in the 1950s for process and performance reasons.

There are alternative semiconductor materials used for specific market applications. These are primarily the *compound semiconductors*. A major class of compound semiconductors are formed from Group IIIA and Group VA of the periodic table (often referred to as III-V compounds). An example is gallium arsenide (GaAs). In addition, other semiconductor compounds come from elements found in Groups IIA and VIA, and are referred to as II-VI compounds.

The need for alternative semiconductor materials is driven by the growth of markets that require IC performance beyond that capable by silicon semiconductors. The key IC performance factor is speed. Wireless and high-speed digital communications, space applications, and consumer markets such as the automotive industry, are developing special niche semiconductor markets that

are centered on high speed that can support higher signal frequencies. In addition, III-V compound semiconductors are used for light-emitting diodes (LEDs).

GaAs is the most common compound III-V semiconductor and will be discussed in the next section. There are variations of this compound, such as gallium nitride (GaN) for use in the fabrication of blue semiconductor lasers and LEDs. Another compound semiconductor is silicon germanium (SiGe), which has been researched for more than two decades and may compete against GaAs for market applications.

There is also some growth in II-VI semiconductors, with their market revenue expected to be \$179 million in 2001.⁷ The two principal II-VI materials are cadmium telluride (CdTe) and zinc selenide (ZnSe). CdTe semiconductors are used primarily in infrared (IR) detection systems. ZnSe is another II-VI compound material used to fabricate blue LEDs.

Gallium Arsenide (GaAs)

Gallium arsenide (GaAs) is the most common III-V compound semiconductor material and is formed by combining gallium (Ga) from Group IIIA with the element arsenic (As) from Group VA. GaAs has greater electron mobility than silicon, so the majority carriers move faster than in silicon. There are also some attributes of GaAs semiconductor material that reduce parasitic capacitance (see Chapter 3) and signal losses. These result in ICs that are faster than those made with silicon.⁸ The improved signal speed of GaAs devices permits them to react to high-frequency microwave signals and accurately convert them into electrical signals for communication systems. Semiconductors based on silicon are too slow to react to the microwave frequency. For these reasons, products for wireless and high-speed digital communications and high-speed optoelectronics devices are made from GaAs and other compound semiconductors.

An advantage of GaAs is the increased resistivity of the material, with values up to $10^8 \Omega\text{-cm}$. This makes it easier to isolate semiconductor devices fabricated in the GaAs substrate without loss in electrical performance. GaAs devices also exhibit a higher radiation hardness than silicon, which makes this material attractive for military and space applications.

A summary of some different semiconductor material properties is shown in Table 2.3. The properties of SiO_2 have been included for comparison.

TABLE 2.3 Comparison of Some Physical Properties for Semiconductor Materials

Property	Si	Ge	GaAs	SiO_2
Melting point (°C)	1412	937	1238	1700 (approx.)
Atomic Weight	28.09	72.60	144.63	60.08
Atomic Density (atoms/cm ³)	4.99×10^{22}	4.42×10^{22}	2.21×10^{22}	2.3×10^{22}
Energy Band Gap (eV)	1.11	0.67	1.40	8 (approx.)

A. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, 1967.

The primary disadvantage of GaAs semiconductor material is the lack of a natural oxide. This feature hinders the development of standard MOS devices that require the ability to grow a surface dielectric. Another problem with GaAs is the fragility of the material, which makes wafer handling a major issue in wafer fabrication. The cost of GaAs is as much as ten times higher than silicon because of the relative scarcity of gallium and the energy used in its purification process. A final point to note is that the extreme toxicity of arsenic requires special controls for the equipment, process, and waste disposal facilities. These precautions have a significant effect on the manufacturing cost of GaAs semiconductors.

SUMMARY

All matter is made up of atoms. The number of valence-shell electrons in an atom is one of several factors that influences its chemical and physical properties and its ability to bond with other atoms. The energy-band theory of solids explains how electrons change orbital levels between the valence band and conduction band to be an insulator, conductor or semiconductor. Chemical elements are described in the periodic table, which is used to explain how ionic and covalent bonds form and how certain elements are used in semiconductor manufacturing. All material is classified as either a conductor, which conducts current, an insulator, which does not pass current, or semiconductor, which conducts current under certain conditions. Resistivity is a material property, while resistance depends on both resistivity and the geometry of the material. A capacitor uses a dielectric to store an electric charge.

KEY TERMS

element
molecule
compound
valence shell
insulators
conductors
semiconductors
cation
anion
ionization
atomic mass unit (amu)
electron-hole pair
recombination
lifetime
conductivity
resistivity
resistance
capacitance

Silicon is the most common semiconductor material because of its abundance, high melting temperature, wide temperature range, and ability to oxidize. Oxidation permits masking of the silicon surface and the ability to dope silicon with impurities to make it conductive. Silicon is doped with trivalent (p-type acceptor dopant) and pentavalent (n-type donor dopant) elements. Introducing a specific concentration of dopants into the silicon crystal permits the precise control of silicon's resistivity. When p-type and n-type silicon form a junction in the silicon crystal, then silicon can act as either an insulator or a semiconductor. There are alternative semiconductor materials, primarily the compound semiconductors, of which gallium arsenide is the most common.

silicon
intrinsic silicon
crystal
silicon dioxide
doping
dopants (or impurities)
extrinsic silicon
trivalent
pentavalent
p-type silicon
acceptors
n-type silicon
donors
majority carriers
minority carriers
pn junction
compound semiconductors
gallium arsenide

REVIEW QUESTIONS

1. What is matter?
2. Describe the model for atomic structure, including the proton, neutron, and electron, and state the charge of each.
3. Describe an element, a molecule and a compound.
4. What is a valence electron and why are valence electrons important for forming compounds?
5. How does energy-band theory explain the difference among insulators, conductors, and semiconductors? How does the energy gap change for each material?
6. How is an ion formed? What is a cation and an anion?
7. What does the group number on the periodic table represent in terms of valence electrons? What is the difference between the atomic number and the atomic mass number?
8. Define the atomic mass unit (amu).
9. How is an ionic bond formed? How is a covalent bond formed? What is oxidation and reduction and how are these processes related to the formation of a bond?
10. State the three classes of materials and explain how electric current flows in each one.
11. What is resistivity? What is conductivity? How is resistivity related to conductivity?
12. What is resistance? State the resistance formula.
13. What is another term for an insulator material?
14. What is capacitance? What is the dielectric constant and why is this concept important to semiconductor manufacturing?
15. Name the most common semiconductor material and give four reasons why it is so commonly used.
16. What is the difference between intrinsic and extrinsic silicon?
17. Describe silicon dioxide. Discuss why the oxidation of silicon is important.
18. What is doping? Why is doping important for semiconductor silicon?
19. Give one example of a trivalent doping element and three examples of a pentavalent dopant element. From what group number of the periodic table is each of these elements?
20. Describe the n-type doping of silicon with phosphorus.
21. n-type silicon has what type of majority and minority carriers? Give the majority and minority carriers for p-type silicon.
22. What characteristics determine how well silicon will conduct electricity?
23. Describe a pn junction and explain why it is important for a semiconductor.
24. Compound semiconductors come from what group numbers of the periodic table?
25. What is the advantage of gallium arsenide over silicon?
26. What is the primary disadvantage of gallium arsenide over silicon?

REFERENCES

1. Basic chemistry concepts referenced from M. Silberberg, *Chemistry, The Molecular Nature of Matter and Change*, (St. Louis: Mosby, 1996).
2. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era, Vol. 1—Process Technology* (Sunset Beach: Lattice Press, 1986), p. 118.
3. *Solid State Technology* (January 1998); p. 18.
4. F. Seitz and N. Einspruch, *Electronic Genie: The Tangled History of Silicon*, Urbana, (University of Illinois Press, 1998), p. 175.
5. G. Aner, *Planar Processing Primer*, (New York: Van Nostrand Reinhold, 1990), p. 57.
6. A. Holden, *Conductors and Semiconductors*, (Bell Telephone Laboratories, 1964), p. 109.
7. "Growth Seen for II-VI Materials," *Semiconductor International* (March 1997); p. 57.
8. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: Wiley, 1994), p. 2.

CHAPTER 3

DEVICE TECHNOLOGIES

There are many different types of semiconductor devices, each meeting a functional need for users. This chapter introduces various electronic components and

explains how they are constructed in silicon for use in the evolution of specific IC technologies.

OBJECTIVES

After studying the material in this chapter, you will be able to:

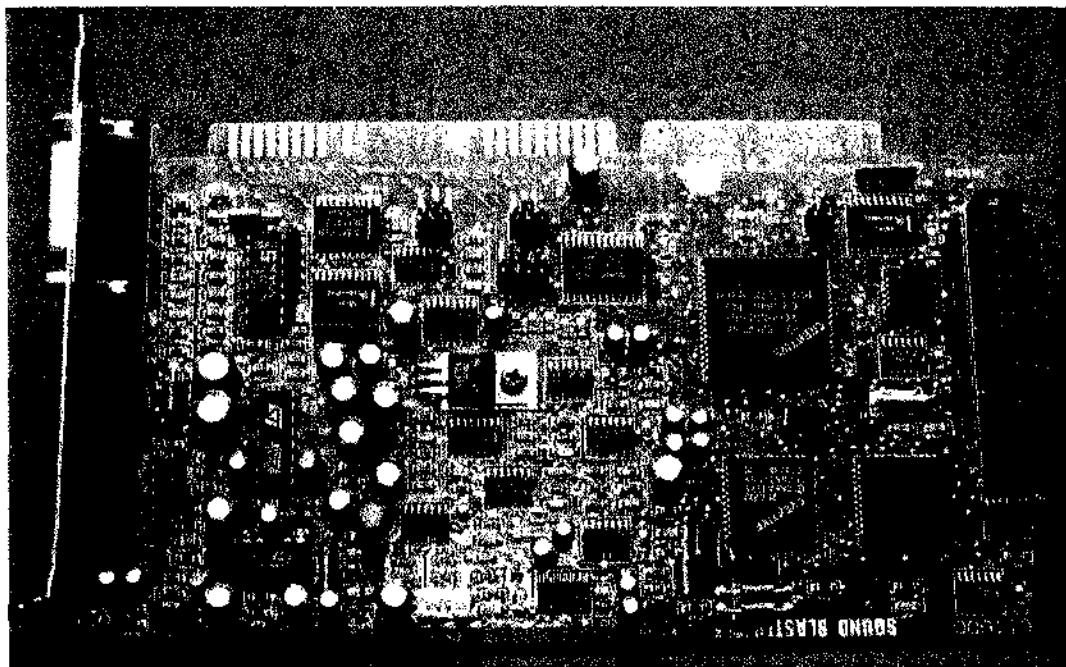
1. Identify differences between analog and digital devices and passive and active components. Explain the effects of parasitic structures in passive components.
2. Describe the pn junction, discuss why it is important, and explain reverse and forward biasing.
3. State the characteristics of bipolar technology and the bipolar junction transistor in terms of function, biasing, structure, and applications.

4. Explain the basic characteristics of CMOS technology, including the field-effect transistor, biasing, and the CMOS inverter.
5. Explain the difference between enhancement mode and depletion mode MOSFETs.
6. Explain the effects of parasitic transistors and the implications for CMOS latchup.
7. Give examples of IC products and state some applications of each.

INTRODUCTION

Electronic devices used in microchips are constructed in the silicon substrate. Common microchip devices include resistors, capacitors, fuses, diodes, and transistors. Their integration on a silicon substrate is the basis for IC wafer fabrication technology.

How a particular electronic device is formed in silicon is referred to as a *structure*. There are literally thousands of different semiconductor device structures and we will evaluate only a few of them. In this chapter, we discuss the actual formation of the devices in order to understand how they can perform in their many applications. Also covered in this chapter is a review of the various classifications of IC products.



Components on PC Motherboard

CIRCUIT TYPES

Circuits constructed from electronic components can be classified into two basic types: analog and digital circuits. Each circuit type has applications where it is most beneficial, as illustrated on page 30.

Analog Circuits

In electronics technology an *analog circuit* is one in which the electrical data varies continuously over a range of voltage, current, or power values. Analog circuits can be designed to operate with direct current (DC), alternating current (AC), or a combination of the two, pulsating DC. Some examples of electronics products that operate primarily as analog (also known as *linear*) circuits are radio transmitters and receivers, audio recording and playback systems, X-ray machines, and automotive ignition systems. The magnitude of the input and output signals, however, may not always be set at any predetermined level. For example, when scanning radio stations on an AM/FM radio receiver, not all radio signals have the same signal strength. As a result, the volume control may have to be adjusted according to the strength of the incoming signal. A volume control on a radio, a manual thermostat on a wall heater, the light control for a household light fixture—these are all examples of common analog devices.

Digital Circuits

Digital circuits have operating signals that vary about two distinct voltage levels—a high and a low. A logic high represents a binary bit = 1 and a logic low represents a binary bit = 0. Digital circuits are associated with *logic devices* such as computers and calculators. Other applications of digital logic devices include clocks, handheld computer games, and barcode readers. Digital devices can be used to measure and/or control events that require either an on/off type of command or that can be controlled in discrete incremental changes that approximate the actions of a linear circuit. That is why today it is sometimes difficult to tell the difference between an analog device and a digital system. Exactly what voltage levels a high and a low are depends on a specific *device technology*. Here are two examples of these logic voltage levels:

Logic Family	High State = 1	Low State = 0
TTL	5 VDC	0.0 VDC
CMOS	3.5 VDC	0.0 VDC

Note: VDC is volts DC.

PASSIVE COMPONENT STRUCTURES

You learned in Chapter 2 about the differences in materials categorized as insulators, conductors, or semiconductors. Now you will begin to see how these basic materials are used in the construction of some very basic electronic components—for example, a resistor and a capacitor. These components are referred to as passive components for various reasons. *Passive components* can conduct electrical current regardless of how the component is connected across a voltage supply. A resistor, for example, will conduct current equally as well regardless of which end of the resistor is connected to the positive or negative terminal of a battery.

IC Resistor Structures

An IC resistor can be created out of metal films or from doped polysilicon or by diffusion of dopant impurities into specific areas of the substrate (see Figure 3.1).¹ The resistors are microstructures and, therefore, occupy very small areas of the substrate. The connection of the resistors to the chip circuitry is formed from conductor metals such as aluminum or tungsten in the form of contacts (see Chapter 12).

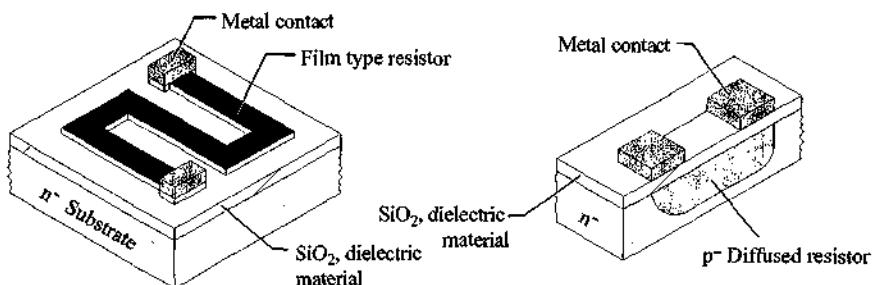


FIGURE 3.1 Examples of Resistor Structures in ICs

Parasitic Resistance Structures ■ *Parasitic resistance* is an unwanted resistance found in the design of components of an IC. It may exist in structures simply for reasons such as the size, shape, material type, dopant species, and quantity of dopant in the material. Parasitic resistance is an undesirable condition because it reduces the operational performance of IC devices. Figure 3.2 illustrates the locations of parasitic resistances in a transistor.

Parasitic resistances are cumulative, which means that the overall effect of these resistors in series is greater than a single resistance alone. The effect of parasitic resistances in IC devices becomes a detriment to the ability to reduce device feature sizes on chips. With higher circuit density comes higher resistances and overall degradation of electrical performance. Designers are aware of this problem and automatically calculate resistance losses into their design equations as well as select low-resistance metals for contacts and design process specifications to reduce the bulk resistance within active devices.

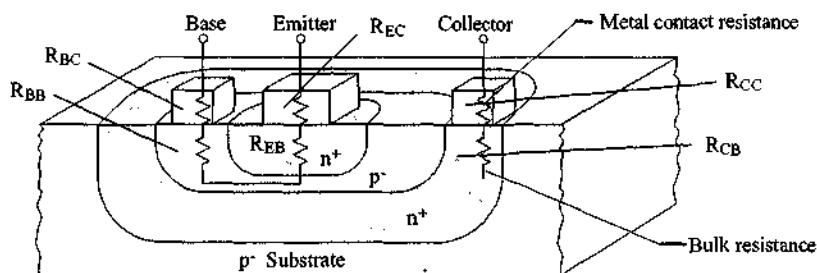


FIGURE 3.2 Cross Section of Parasitic Resistances in a Transistor

IC Capacitor Structures

You may recall in Chapter 2 that a simple capacitor is formed when two conductors are separated by a dielectric (insulator) material. In microchip manufacturing the dielectric material is usually silicon dioxide (glass, SiO_2), more commonly referred to as simply *oxide*. *Planar capacitors*, which are built laterally on the substrate, may be formed from metallic films, doped polysilicon, or diffused areas of the silicon substrate. In general, capacitors are formed on substrates using four basic techniques (see Figure 3.3).

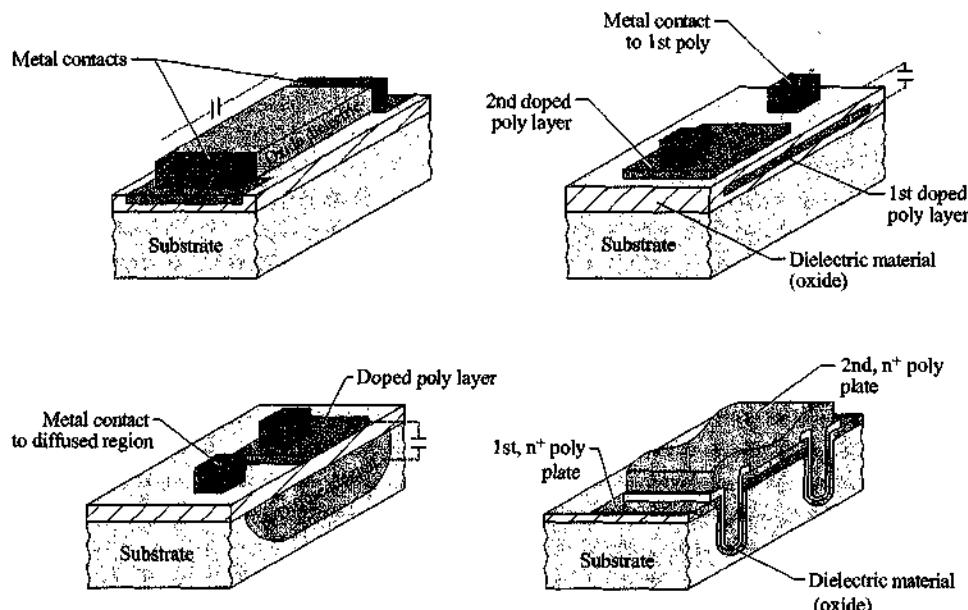


FIGURE 3.3 Examples of Capacitor Structures in ICs

Parasitic Capacitance Structures ■ Quite frequently capacitance occurs unintentionally as a result of the way materials are constructed on the substrate. This condition is referred to as *parasitic capacitance*.² An example of this would be two adjacent metal conductors with a dielectric material between them. This type of capacitance is undesirable for IC performance. In fact, parasitic capacitance interferes with the ability of electronic circuits to operate at faster switching speeds. In some cases parasitic capacitance may create instability in circuits, cause parasitic oscillations, and even create short-circuit paths for AC signals where they are not needed. Figure 3.4 illustrates the location of parasitic capacitance in between the electrodes of bipolar junction transistors and field-effect transistors.

ACTIVE COMPONENT STRUCTURES

Active components, such as diodes and transistors, offer some very different electronic control attributes that are not characteristic of passive components. They can be used to control the direction of current flow. Furthermore, active components can amplify small signals and can be used to create more complicated circuits, such as current and voltage regulators, oscillators, and logic gates. Active components require definite polarity (+ or -) when connecting these devices to power supplies. Active components utilize the flow of both electrons and holes in their operation.

The pn Junction Diode

A *pn junction diode* is formed any time there is a region of n-type semiconductor adjacent to a region of p-type semiconductor. A pn junction may be deliberately designed as a functional part of a larger integrated circuit or it may exist as a nonfunctioning diode in another part of the integrated

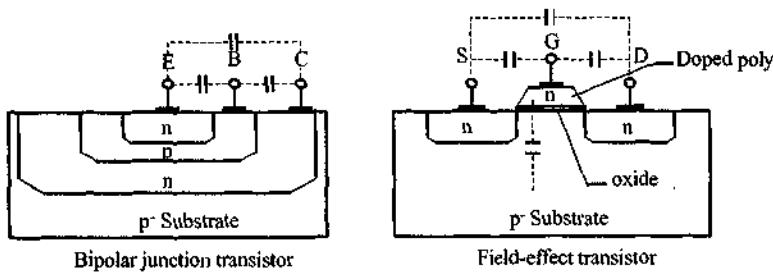


FIGURE 3.4 Examples of Parasitic Capacitance in Transistors

circuit. pn junctions are important for ICs because they are fundamental to the operation of both bipolar junction transistors and field-effect transistors.³

The pn junction diode is made in a single crystal of semiconductor material, e.g. silicon. As shown in Figure 3.5, one region of the substrate is doped heavily with donor dopant, such as arsenic, phosphorus, or antimony, to form an n-type silicon region. On the other side, acceptor impurity (e.g., boron) is used to create an p-type silicon region. The actual manufacturing procedures will be described later in Chapter 9. Typical metal contact materials used to connect the diode to the chip circuitry are aluminum, tungsten, titanium, or copper.

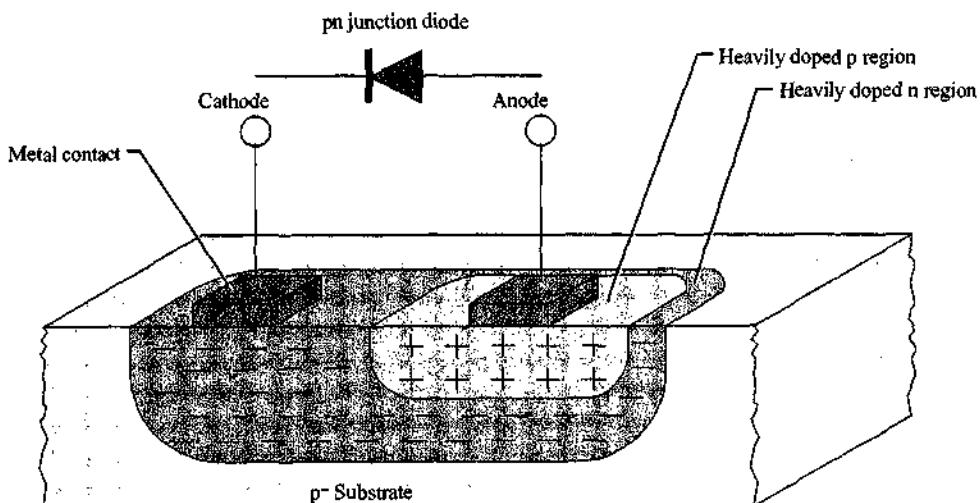


FIGURE 3.5 Basic Symbol and Structure of the pn Junction Diode

The open-circuit condition of the diode is shown in Figure 3.6 on page 48. Holes (+) make up the majority carriers in the p-type silicon while electrons (-) are the majority carriers in the n-type silicon. Initially, both sides of the diode are electrically neutral; however, some holes from the p region diffuse over to the n region, while electrons from the n region diffuse into the p region. These holes and electrons recombine at the junction of the diode. This recombination process has the net result of depleting each region of its majority current carriers. The p region has fewer holes than previously and the n region has been depleted of electrons. Thus, the term *carrier-depletion region* is used to describe the recombination area. The depletion of holes from the p region creates a negative charge in this area. Similarly, the loss of electrons from the n region results in a positive charge. The resulting charges from each region produce corresponding electric fields that oppose any further diffusion of carriers from the opposite side of the silicon crystal.

The net effect of the difference in charges across the depletion region is to create a potential difference across the junction. This results in a *barrier voltage* that has to be overcome before the diode can be operated.

Reverse Biasing the pn Junction Diode ■ Assume a voltage or bias supply (see the battery symbol in Figure 3.7) is connected across the pn junction as shown in Figure 3.7 on page 48. This bias

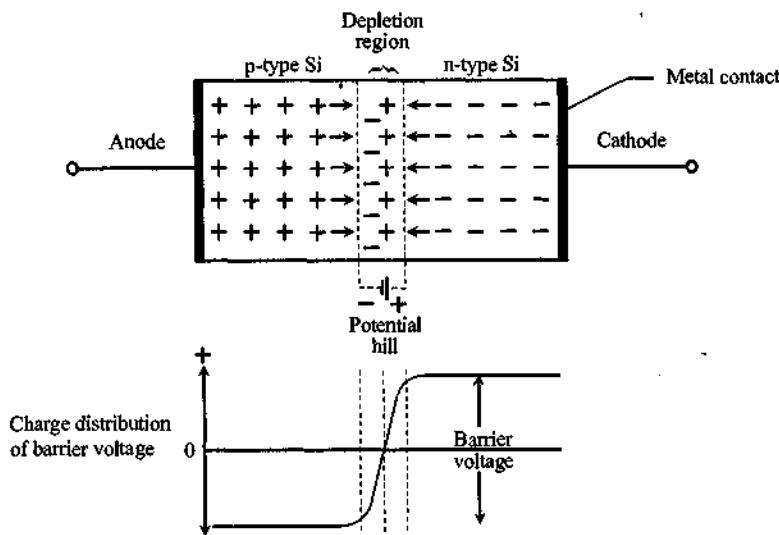


FIGURE 3.6 Open-Circuit Condition of a pn Junction Diode

configuration, called *reverse bias*, results in little or no conduction through the diode. The lamp is placed in the circuit as an indicator of current flow. Any appreciable current through the device can be attributed to minority current carriers in the diode.

Forward Biasing the pn Junction Diode ■ Figure 3.8 illustrates the effects of applying a *forward bias* to a pn junction diode. In this circuit electrons in the n region are repelled away from the negative terminal of the bias supply. Additional electrons are injected from the negative terminal to fill holes left vacant by the electrons in the n region. Similarly, holes in the p region are repelled away from the positive terminal of the bias supply. Holes are supplied from the positive side of the bias supply to balance the electrons supplied from the opposite side of the bias supply. Holes recombine with electrons at the junction and overcome the barrier voltage, greatly reducing its action as a barrier to current flow. Current flow will continue through the circuit as long as the bias supply can maintain a constant injection of holes and electrons through the diode.

Holes and electrons are drawn toward the junction to overcome the potential hill, thus, allowing current to flow through the circuit. Note that in Figure 3.8 solid arrows illustrate the direction of hole flow and broken arrows depict the direction of electron flow. Solid arrows also indicate conventional current flow.

The electrical characteristics of the diode are represented by the current-versus-voltage curve (see Figure 3.9). The graph in Figure 3.9 represents the forward and reverse bias characteristics of a typical silicon diode. Note the breakdown voltage at the knee of the forward bias curve. This voltage is characteristic of silicon diodes and is generally in the range of 0.6 to 0.8 volts. The reverse bias curve shows a point when the diode will conduct in the opposite direction when the junction voltage is exceeded.

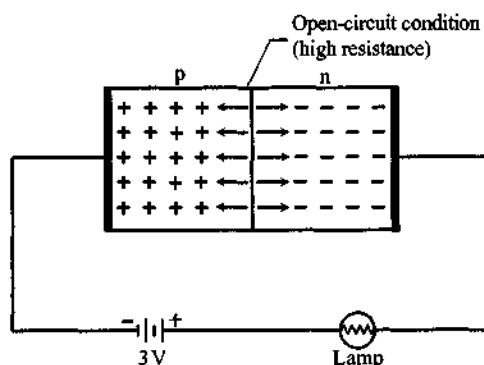


FIGURE 3.7 Reverse-Biased pn Junction Diode. Holes and electrons are drawn away from the pn junction, essentially creating an open-circuit condition.

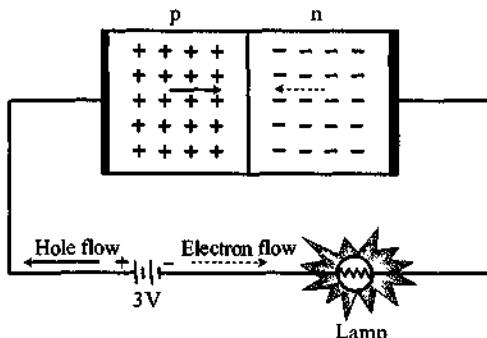


FIGURE 3.8 Forward-Biased pn Junction Diode

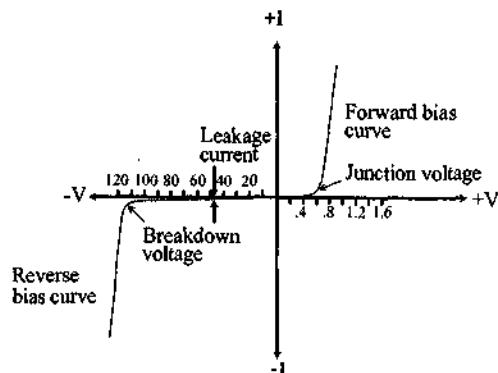


FIGURE 3.9 Forward and Reverse Electrical Characteristics of a Silicon Diode

The Bipolar Junction Transistor

The *bipolar junction transistor* (BJT) has three electrodes and two pn junctions. The entire transistor is constructed from a single semiconductor substrate. There are two variations of the bipolar junction transistor—*npn* and *pnp*. The simplified structures and schematic symbols of these transistors are shown in Figure 3.10. The electrodes are labeled *emitter*, *base*, and *collector*. The emitter arrows indicate the direction of hole flow through the respective transistor.

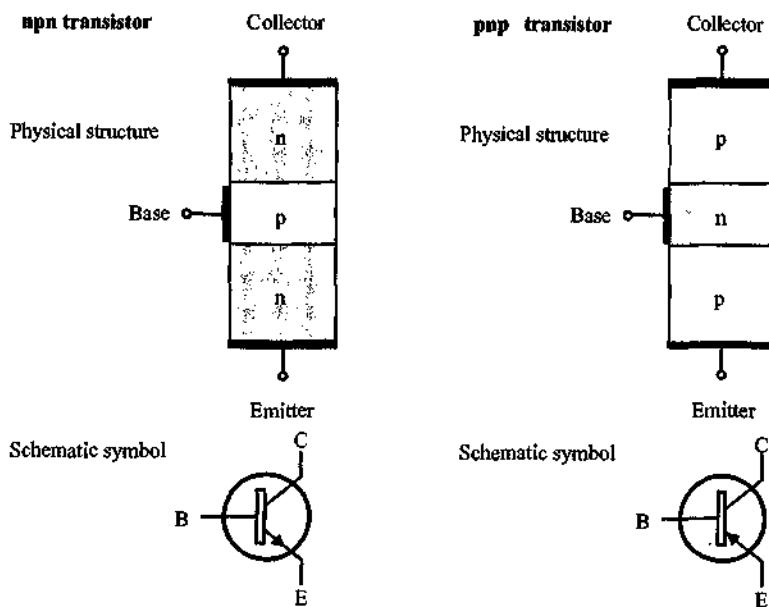


FIGURE 3.10 Two Types of Bipolar Transistors

Biassing the npn Transistor for Conduction Mode ■ The biassing scheme for proper operation of the npn transistor is shown in Figure 3.11 on page 50. Note the polarity of the two bias supplies. The emitter-base junction is forward-biased by a smaller bias supply, i.e., a simple 1.5-volt cell. The collector bias supply, a 3-volt battery, is connected in reverse polarity relative to the n-type collector. The emitter is shown as the reference point (ground) for all voltages. In this circuit configuration, the base serves as the input to the transistor and the collector is the output. The lamp serves the purpose as an output load and as an indicator of current flow through the collector.

There can be no conduction through the collector of the transistor without a completed circuit path through the emitter-base (E-B) junction. This nonconducting mode is shown in Figure 3.11 on page 50. As long as switch S_1 remains open, the pn junction of the emitter and base remains in a nonconducting mode (open circuit). This condition will not allow for current flow from emitter to

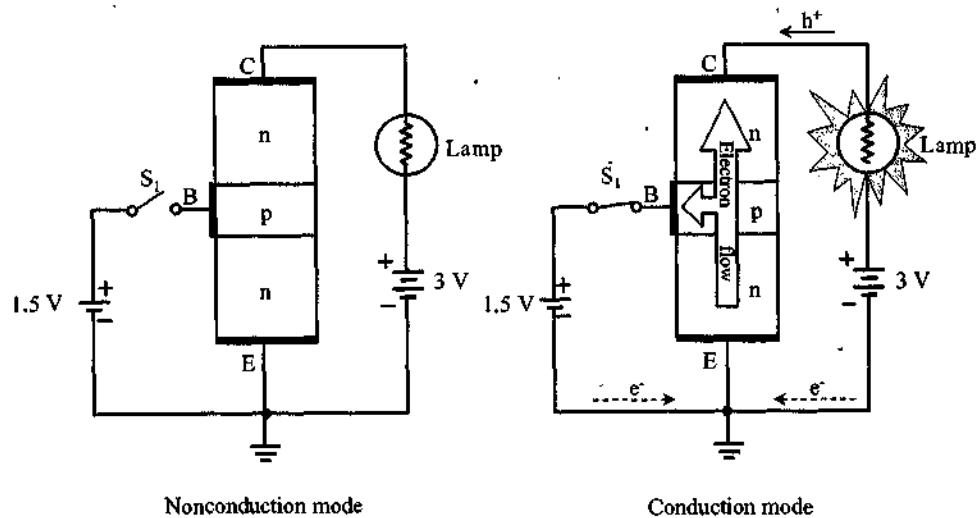


FIGURE 3.11 npn Transistor Biasing Circuit

collector. Closing switch S_1 forward-biases the E-B junction (the same as a diode). Conduction through the E-B junction immediately reduces the barrier and allows electrons to be injected from the negative side of the 3-volt battery through the E-B region and into the collector. At the same time holes are injected from the positive side of the collector battery through the lamp and then into the top of the collector to recombine with electrons. The lamp now indicates current flow through the output section of the transistor. This condition will continue until a change in S_1 or one of the circuit elements is disconnected.

Note that in Figure 3.11 solid arrows illustrate the direction of hole flow and broken arrows depict the direction of electron flow. Solid arrows also indicate conventional current flow.

Biasing the pnp Transistor for Conduction Mode ■ The operation of the pnp transistor is similar to the npn except that the bias supply connections are reversed. Conduction of a pnp transistor requires the same conditions as with the npn: first, the E-B junction must be forward biased (See Figure 3.12), and second, the collector supply voltage, V_{CC} , must be reversed relative to the collector. Upon initial conduction through the E-B junction, holes are injected from the positive side of the 1.5-volt supply through the E-B junction and into the collector. Electrons from the 3-volt battery leave the negative side of the battery, flow through the lamp, and enter the top of the collector to recombine with holes.

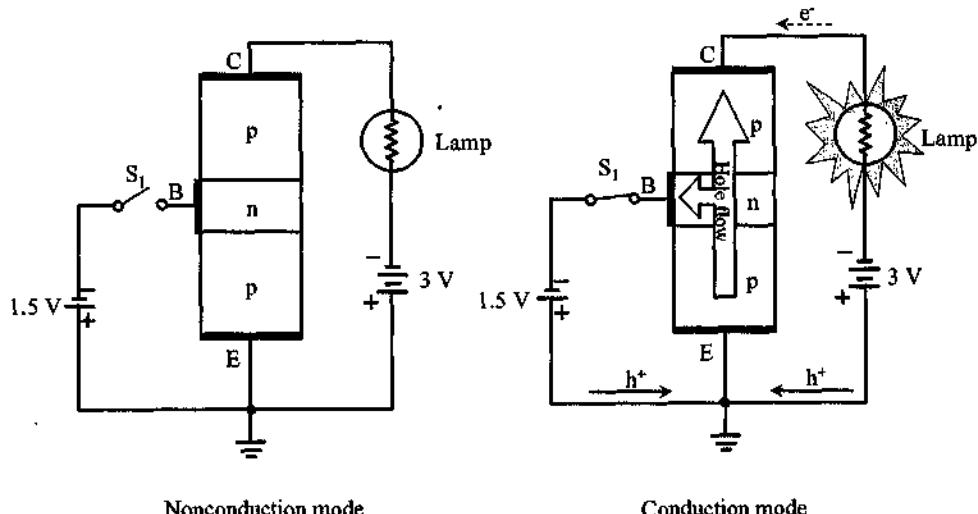


FIGURE 3.12 pnp Transistor Biasing Circuit

Structure of a Bipolar Junction Transistor ■ Figure 3.13 illustrates the structural features of the cross section of an npn bipolar junction transistor. Note the differences in the amount of dopant concentration in each of the electrodes. The emitter (E) and collector (C) are both heavily doped with n-type dopant, such as arsenic or phosphorus. The base (B) is lightly doped with boron, a p-type dopant. With fewer current carriers in the base, the current drawn through the base is significantly smaller than the current drawn through the collector. This difference accounts for the *gain* (amplification) in current from the input to the output of the transistor. This feature is the major difference between the transistor and the diode. The transistor is able to amplify small input signals literally hundreds of times to drive output devices such as speakers, motors, lamps, relays, and other electromechanical devices. The BJT is a *current-driven current amplifier* device. Thus, bipolar transistors are commonly used in radios, tape recorders, automotive electronics, aircraft control systems, biomedical instrumentation, robotics, and wherever there is a need for high power control.

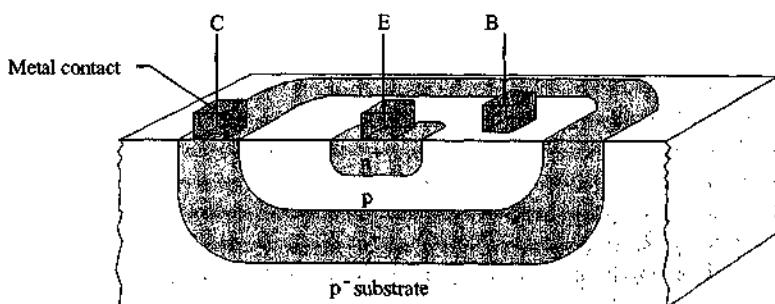


FIGURE 3.13 Cross Section of an npn BJT

Schottky Diode

The *Schottky diode* is formed when metal is brought in contact with lightly doped n-type semiconductor material. The resulting device operates just like an ordinary pn junction diode—low resistance when forward-biased and high resistance when reverse-biased. The forward junction voltage drop of the silicon Schottky diode (0.3 to 0.5 volts) is nearly half that of the silicon pn junction diode (0.6 to 0.8 volts). The Schottky diode's biggest advantage is that its conduction is due entirely to electrons, which results in faster switching time from on to off.⁴ Figure 3.14 shows the schematic symbol and structured cross section of the Schottky diode.

The discovery of the Schottky diode has helped extend the usefulness of bipolar IC technologies into the twenty-first century. The Schottky diode concept has been utilized in the development of faster and more power efficient bipolar integrated circuits.

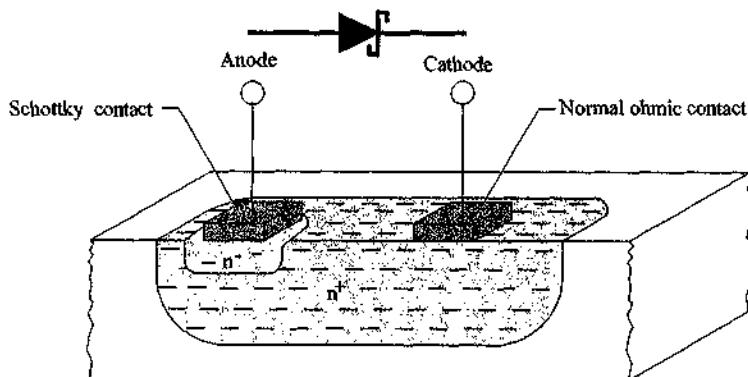


FIGURE 3.14 Schematic Symbol and Structural Cross Section of the Schottky Diode

Bipolar IC Technology

Diodes and bipolar transistors—along with supporting components of resistors, capacitors, insulators, and conductors—are used in the development of a breed of integrated circuits called *bipolar technology*. Bipolar technology was the first of many IC technologies that was used in the production of analog and digital integrated circuits. For many years bipolar devices have been noted for fast speeds, durability, and power-controlling abilities. The biggest drawback, however, has been their high power consumption. This, of course, meant higher costs for operating these devices in terms of electrical utilities and batteries for portable equipment. Some families of the bipolar technology era have since become obsolete. Other bipolar manufacturers have managed to maintain their business based on a continued need for bipolar devices for use in high-power applications. A list of the bipolar logic families is shown in Table 3.1. Further information about bipolar technology can be obtained by consulting the reference list at the end of this chapter.

TABLE 3.1 Bipolar Logic Families

Bipolar Logic Family	Abbreviation
Direct-Coupled Transistor Logic	DCTL**
Resistor-Transistor Logic	RTL**
Resistor-Capacitor-Transistor Logic	RCTL**
Diode-Transistor Logic	DTL**
Transistor-Transistor Logic*	TTL**
Schottky TTL Logic*	STTL†
Emitter-Coupled Logic*	ECL†

*Some forms of TTL, STTL, and ECL still in use through were 2000. **From G. Deboo and C. Burrous, *Integrated Circuits and Semiconductor Devices: Theory and Application*, 2nd ed. (New York: McGraw-Hill, 1977), p. 192. †From A. Sedra and K. Smith, *Microelectric Circuits* (Oxford: Oxford University Press, 1998), p. 1187, 1196.

CMOS IC Technology

When energy conservation became an international concern in the late 1970s, the semiconductor industry responded with the *field-effect transistor (FET)*, which benefits from more compact and power-efficient electronic devices. Although early experiments with FETs date back to the 1930s, the first mass-produced FETs became available in the 1960s. Several versions of the FET have been used since then. Today the most popular IC technology, *CMOS (complementary metal-oxide semiconductor)*, revolves around the improvements that have been made in FET design and manufacturing. The remainder of this text concentrates on the development and manufacturing processes related to CMOS devices.

The Field-Effect Transistor ■ The development of the field-effect transistor essentially created a new era in the history of the semiconductor industry. As opposed to the current-amplifying bipolar junction transistor, the FET is a *voltage-amplifying device*. The only similarities between the BJT and the FET are in the types of materials used to build the transistors and the number of electrodes. Both have three electrodes and both types are constructed of a single crystalline substrate. The greatest advantage of the FET is its low voltage and low power requirements. Whereas the BJT requires input current at the base to turn on the transistor, the FET turns on as a result of an electric field created when an input voltage is applied to the *gate*—thus the name field-effect transistor.

The FET has seen applications as an amplifier in linear/analog circuits and as a switching component in digital electronics. Its high input resistance and moderate gain characteristics make it an excellent device for use in instrumentation and communications. Its low power consumption and compactness make it extremely suitable for the ever-shrinking dimensions of VLSI and ULSI technology.

There are two basic types of FETs: the junction (*JFET*) and the metal-oxide (*MOSFET*) semiconductor. The major difference between these two types is that the gate on the MOSFET, which is the input to the FET, is insulated by a thin dielectric (silicon dioxide, referred to as *gate oxide*) from the other two electrodes of the transistor. The gate of the JFET actually forms a physical pn junction with the other electrodes of the transistor. JFETs are used extensively in GaAs integrated circuits. When metal gates are used in GaAs JFETs, the term *MESFET* is used. Because of the

popularity of MOSFETs in silicon VLSI applications, the remainder of our discussion of FETs will focus on the MOSFET. Consult the reference list at the end of this chapter for further information on the JFET.

MOSFETs. MOSFETs gained wide acceptance in logic applications in the 1970s and have been the mainstay transistor in IC products ever since. There are two categories of MOSFETs: *nMOS* (*n*-channel) and *pMOS* (*p*-channel). Each type is distinguishable by the majority current carriers in each device. Figure 3.15 shows some common schematic symbols and physical cross sections of the two types of MOSFETs.

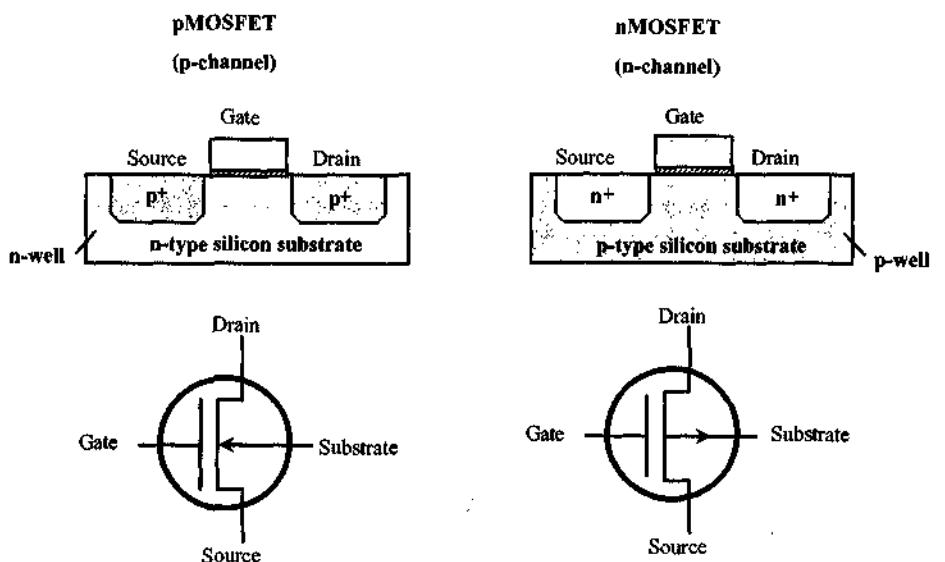


FIGURE 3.15 Two Types of MOSFETs

Each MOSFET has an input electrode called the gate. The term *metal oxide* refers to the material the gate is made of. However, the “metal” descriptor in MOS derives from the early days of MOS technology and is no longer true. The most popular material used in the formation of gates for MOSFETs is *polysilicon*, a polycrystalline silicon material that is deposited on the substrate during IC fabrication (see Chapter 11). The polysilicon, however, must be doped with one of the common p- or n-type dopants to give the material its conductive characteristics.

The *source* and the *drain* electrodes are heavily doped, respectively, with n-type or p-type dopant depending on the category of transistor being manufactured. The supply of majority current carriers comes from these two electrodes. The *nMOSFET* uses electrons as the majority carriers; therefore, the channel is n-type. The *pMOSFET*, then, has a p-channel formed by holes from the source and drain. When in the nonconduction mode, the channel is an open circuit consisting of an opposing doped region called the *well*. The *n-channel MOSFET* is built inside a *p-well*, while the *p-channel MOSFET* is built inside an *n-well*. When in the conduction mode, opposing carriers in the upper part of the well move away from the gate oxide interface and a channel of majority current carriers forms across from the source to the drain to complete the circuit (very similar to closing a switch).

Biasing the nMOSFET for Conduction Mode. Figure 3.16 on page 54 illustrates the biasing scheme for operating an n-channel MOSFET. The lamp in the circuit is an indicator of output current and normal operation of the transistor. At the moment there is no conduction in the n-channel MOSFET in Figure 3.16 because there is no input voltage applied to the gate. The condition of the channel is that of an open circuit. The majority of carriers in the area immediately beneath the gate oxide and between the source and drain electrodes are holes. At the moment the source and the p-well are at the same potential. As a pn junction there is no forward bias applied, so the pn junction remains in a nonconducting mode.

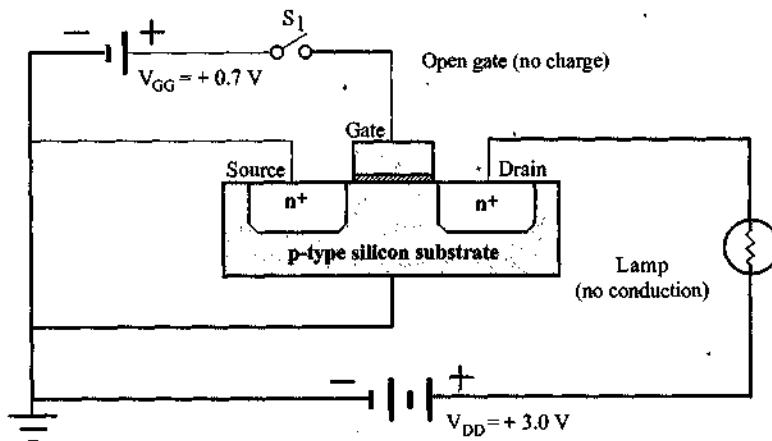
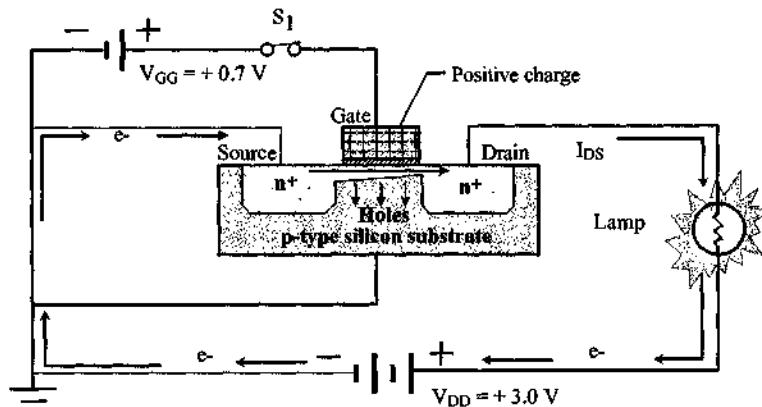


FIGURE 3.16 Biasing Circuit for an nMOS Transistor

When the switch, S_1 , is closed as in Figure 3.17, certain conditions begin to happen. The small bias voltage of 0.7 V places a charge equal to the bias supply directly across the gate and source electrodes. An electrostatic charge is created with positive holes on the gate and electrons flowing up from the source to the gate oxide interface. At the same time an electric field is created by the positive charges at the gate. This field opposes the electric field of the holes in the upper part of the p-well. These opposing fields force the holes in the upper part of the p-well to move away from the gate oxide interface. As a result, electrons move from the source and drain to fill the gaps left vacant by the holes. The electric field from the positive charge on the gate continues to attract and hold the electrons in the channel area, thus closing the gap between the source and the drain with only electrons. Now, the electrons form a continuous crystalline structure of n-type silicon. Electrons from the 3-volt battery flow freely from the negative terminal through the source, to the n-channel, into the drain, and then through the lamp and back to the positive side of the battery. This condition will remain the same until any part of the input or output circuit is changed.

FIGURE 3.17 nMOS Transistor in Conduction Mode. Closing S_1 allows V_{GG} to place a positive charge on the gate of the transistor that causes an n-channel to form; therefore, allowing electrons to flow from the source to the drain.

Increasing the voltage at the gate increases all the electrical activities just mentioned. The electrostatic charge on the gate increases, which increases the electric field strength and forces the holes in the p-well to move further away from the gate-oxide interface. This results in an increase in the size of the n-channel and the number of electrons flowing through the channel. The net effect is for more drain current to flow, thus increasing the power delivered to the load (increased brilliance of the indicator lamp).

Figure 3.18 provides an example of a graphical representation of the electrical characteristics of an n-channel MOSFET. The graph represents drain current, I_{DS} , as a function of V_{DS} and V_{GS} . Each curve labeled V_{GS} represents a specific setting of the gate-source voltage. Using the curves it is possible to determine the value of I_{DS} for a given value of V_{GS} and V_{DS} . For example, assume $V_{GS} = 4\text{ V}$ and $V_{DS} = 2\text{ V}$. The resulting drain current, $I_{DS} = 0.28\text{ mA}$. The *threshold voltage* is the lowest attainable V_{GS} value that will turn on the FET. This value is usually less than 1.0 V and is dependent on several variables, including structural sizes and doping characteristics of the electrodes and p-well.

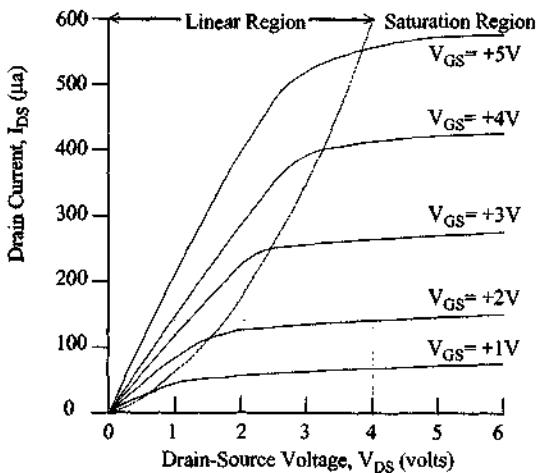


FIGURE 3.18 Example of Characteristics Curves of an n-channel MOSFET

Biasing the pMOSFET for Conduction Mode. Figure 3.19 illustrates the biasing circuit for a p-channel MOSFET. The operation of the p-channel transistor is similar to the operation of the n-channel type. The only difference is that the majority current carriers in the p-channel MOSFET are holes and the bias supplies are reversed. The other major difference between these two transistors is in terms of performance. The speed of the p-channel transistor is slower than that of its counterpart, primarily due to the holes, which move slower than electrons.

Closing S_1 , allows V_{GG} to place a negative charge on the gate of the transistor that causes a p-channel to form, therefore allowing holes to flow from the source to the drain (see Figure 3.20 on page 56).

CMOS Technology ■ Manufacturers of MOSFET-based integrated circuits for many years concentrated on the development and manufacturing of products based solely on n-channel MOSFET technology. Whereas some semiconductor manufacturers produced both discrete nMOS and pMOS transistors, few, if any, manufactured integrated circuits made primarily with pMOS transistors.

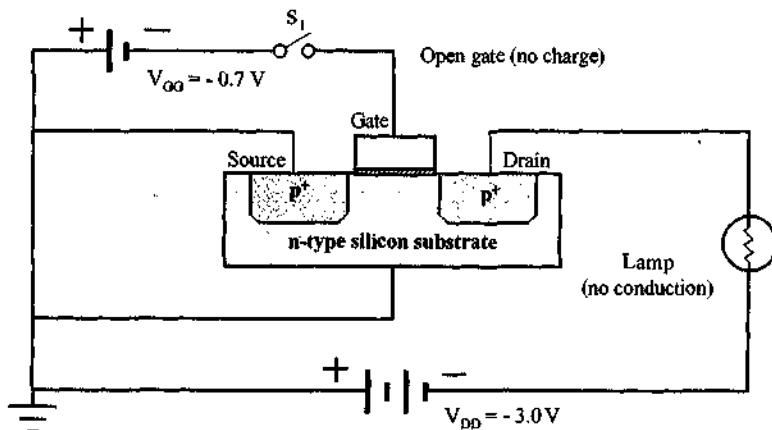


FIGURE 3.19 Biasing Circuit for a p-Channel MOSFET

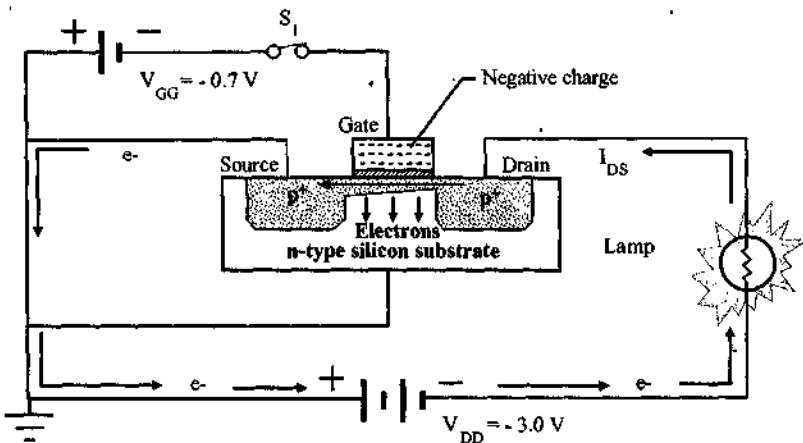


FIGURE 3.20 pMOS Transistor in Conduction Mode

Discrete pMOS transistors have served many usable functions in specific electronic applications, but in general, nMOS IC devices outperformed the pMOS technology. Thus, nMOS became the choice of most IC manufacturers.

CMOS incorporates both nMOS and pMOS transistors in the same integrated circuit. The combination of power efficiency, design scaling techniques, and improved manufacturing processes have made CMOS one of the most popular device technologies since the early 1980s. The term *scaling* has been commonly used to describe the process of *shrinking* the overall dimensions and operating voltages of existing ICs to attain improved operating performance as well as compactness. All dimensions and voltages must shrink in unison (thus the source of the word scaling) through the use of a design model employed by the IC designer during the circuit design and layout stage.

A schematic diagram of a simple CMOS inverter circuit is shown in Figure 3.21. The schematic shows the gates of both transistors connected together. These gates serve as a single input to the inverter. The output of the inverter is taken out of the two drains, which are tied together. The source of the n-channel transistor is grounded, while the source of the p-channel transistor is connected to the bias supply, V_{DD} . A signal applied to the input of the CMOS is inverted at the output, as shown in Figure 3.21. Normal operation has the n-channel operating during the positive transition of the input signal and the p-channel operating during the negative transition. The efficiency of the CMOS inverter circuit occurs during the transition period when the input signal is at zero. There is no power consumed by the transistors when the input signal is at zero. nMOS, TTL, and ECL circuits differ from CMOS in that those logic families all dissipate power even with no signal applied. This fact is one major reason why CMOS is currently the preferred IC technology for use in the manufacture of portable electronics products such as calculators, clocks, cellular telephones, and notebook computers.

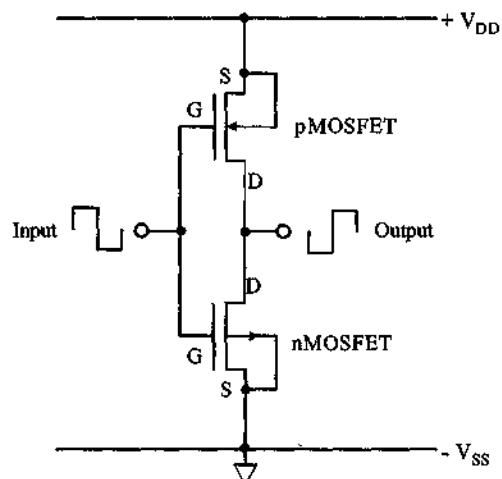


FIGURE 3.21 Schematic of a CMOS Inverter

The physical structure of a simple CMOS inverter is shown in the top view in Figure 3.22 and cross-sectional view in Figure 3.23. The n-substrate serves as the n-well for the p-channel transistor. A separate p-well must be created for the n-channel transistor. This well is created by a selective doping process called *ion implantation*. Isolation regions separate transistors from each other and from other transistors (not shown). The isolation regions are referred to as *field oxide*, which is manufactured from silicon dioxide (also called *glass*). The field oxide insulates transistors from each other in order to prevent the flow of undesirable leakage current between transistors. The actual electrical connections between the two transistors are made possible with metal deposited over the insulated transistors. Later, a patterning process is used to define the exact locations where the metal connections are to be made, followed by an etching process to remove the excess metal. Further details of the CMOS manufacturing process are beyond the scope of this chapter. Chapter 9 focuses on the specific physical and chemical requirements and process flow in CMOS IC manufacturing.

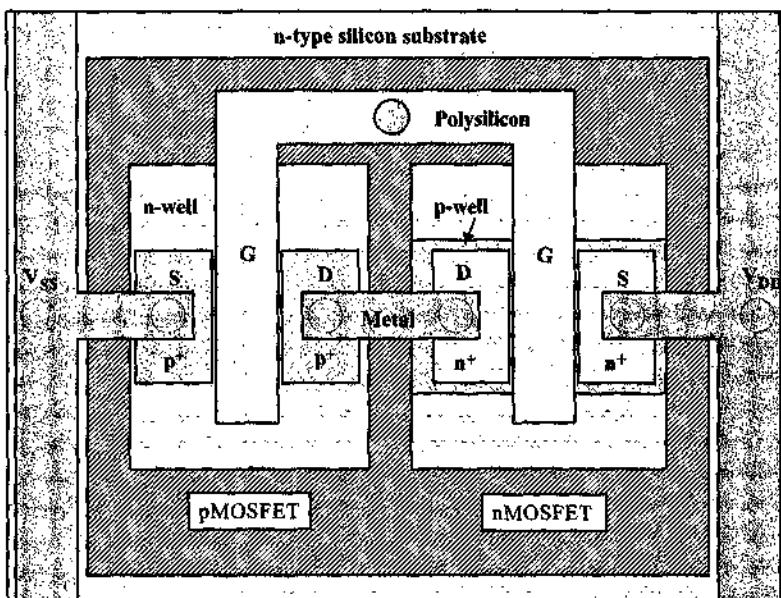


FIGURE 3.22 Top View of CMOS Inverter

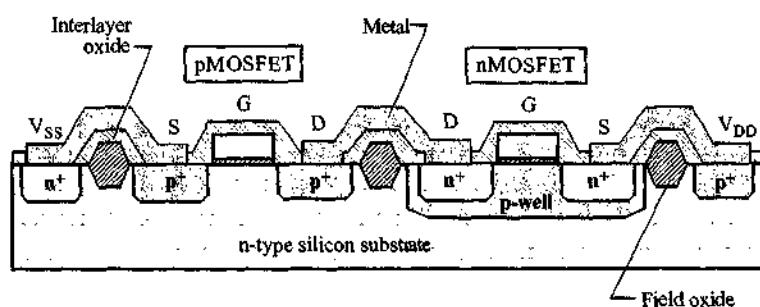


FIGURE 3.23 Cross-section of CMOS Inverter

BiCMOS Technology ■ *BiCMOS technology* makes use of the best features of both CMOS and bipolar technology in the same IC device. BiCMOS incorporates the low-power, high-density CMOS structures with the high-current drive capabilities of TTL or ECL device structures. Applications of BiCMOS products can be found wherever the need for complex digital control of high-power loads is desired. In which case, *digital/analog* (D/A) converter chips may be used to provide the analog drive signals that are used to control electromechanical equipment. On the instrumentation side, *analog/digital* (A/D) chips may be used to measure the outcome of analog drive

signal. Figure 3.24 shows a basic example of BiCMOS chips used in an instrumentation and control application. Other applications for BiCMOS include automotive electronics, aerospace, robotics, and industrial equipment.

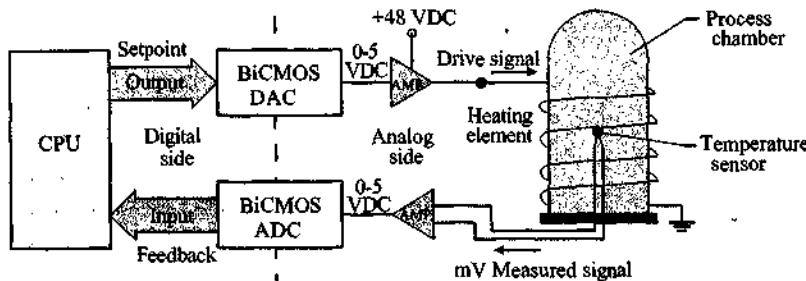


FIGURE 3.24 BiCMOS Chips used in the Control of a Simple Heating System

Figure 3.25 shows additional bipolar transistors added to the CMOS inverter mentioned earlier in Figure 3.21 on page 56. The bipolar transistors (Q3, Q4) provide greater current drive than the CMOS inverter (Q1, Q2) is capable of.

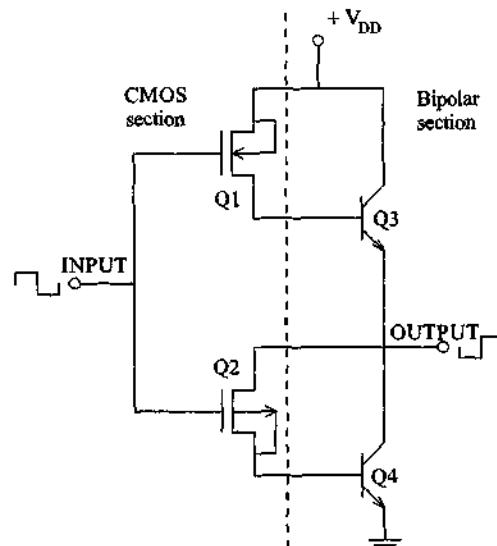


FIGURE 3.25 Simple BiCMOS Inverter
Redrawn from H. Lin, J. Ho, R. Iyer, and K. Kwong, "Complementary MOS-Bipolar Transistor Structure," *IEEE Transactions Electron Devices*, ED-16, 11 Nov. 1969, p. 945-951.

Enhancement and Depletion-Mode MOSFETs

Another way to categorize FETs is in terms of *enhancement mode* and *depletion mode*. Up to now all n-channel and p-channel MOSFETs discussed in this chapter have been enhancement-mode transistors. Enhancement-mode MOSFETs are the most commonly used transistors in the industry. The difference between the enhancement-mode and depletion-mode transistors is shown in Figure 3.26.

In general, the major difference between enhancement- and depletion-mode transistors is in the way the channels are doped. While the enhancement-mode channels are doped opposite in polarity to the source and drain regions, the depletion-mode channels are doped the same as their respective source and drain regions. This setup creates open and closed circuit conditions, respectively, for the enhancement- and depletion-mode transistors. Consequently, the enhancement-mode transistor is regarded as *normally-off* and the depletion mode as *normally-on*.

The enhancement-mode transistor works very well in digital logic applications and requires only a single polarity input signal (V_{GS}) to operate the FET. On the other hand, the depletion-mode transistor is partly turned on by the already existing closed channel. The input voltage can swing in one direction to increase the current through the channel or reverse in the opposite direction to decrease the current through the channel. The depletion-mode transistor will turn off completely if the input voltage to the gate is further increased in the opposite direction. Both types of transistors can be designed into practical applications in both analog and digital circuit applications in either discrete component applications or in integrated circuit applications.

MOSFET Type	Mode	Standby Condition	V_{GS} Switching Requirements	Physical Structure
nMOS	Enhancement	Off	+	
nMOS	Depletion	On	-	
pMOS	Enhancement	Off	-	
pMOS	Depletion	On	+	

FIGURE 3.26 Comparison of Enhancement- and Depletion-Mode MOSFETS

LATCHUP IN CMOS DEVICES

Just as undesirable parasitic resistances and capacitances exist in semiconductor devices, sometimes the pn junctions in CMOS devices can produce *parasitic transistors* that can create a *latchup* condition in CMOS ICs that causes transistors to unintentionally turn on. Figure 3.27 illustrates the parasitic transistors in a CMOS inverter structure. Complementary junction transistors are formed as a result of normal manufacturing of MOSFETs in a CMOS structure. Given certain operating conditions it is possible for the parasitic junction transistors to turn on and create a low resistance path for current to flow across the CMOS structure.⁵ The transistors become latched, thus preventing any further control of the MOSFETs in the CMOS device.

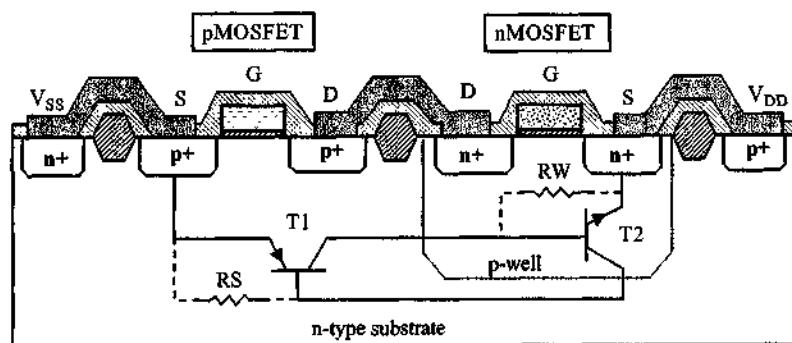


FIGURE 3.27 Parasitic Junction Transistors within a CMOS Structure

Indeed, the latchup phenomenon is a very complex concept. Any further discussion is beyond the scope of this chapter. Suffice it to say that knowing about its existence will help you understand the design and manufacturing steps that are taken to prevent this condition. Later chapters describe several manufacturing techniques used in the prevention of latchup, namely creating isolation barriers between transistors, placing an epitaxial layer explained in Chapter 11 between the substrate and the CMOS structures, and using ion implantation to create retrograde wells explained in Chapter 17.

INTEGRATED CIRCUIT PRODUCTS

Semiconductor products can be found in most electrical and electronic apparatuses throughout the world. The most familiar applications of ICs are found in the personal computer (PC). Such a computer would employ a microprocessor unit (MPU), read-only memory (ROM), random-access memory (RAM), communications interface chips, video graphics controller chips, disk drive controller chips, and more. For the most part, these chips are built using CMOS technology, followed by BiCMOS and bipolar technologies.

The types of ICs used vary depending on customer needs, environment, cost, power requirements, and other specifications. Most ICs, as was mentioned earlier, fall into two basic categories—analog or digital. The following sections list the most common types of ICs found in most consumer and industrial applications.

Linear IC Product Types

The *linear IC* family consists of devices that are designed to operate primarily in analog circuit applications. Linear ICs are typically used in audio systems, radio communications, industrial controls and instrumentation, aerospace, and in automotive electronics. The following are a few examples of linear ICs.

Operational Amplifier ■ The *op-amp* is a high-gain and high-input impedance amplifier that can be adapted to a variety of electronic control applications. It can be used as an amplifier, oscillator, voltage regulator, voltage limiter, sample-and-hold amp, voltage rectifier, integrator, and differentiator.

Voltage Regulator ■ The *voltage regulator* is a single-chip device containing diodes, transistors, resistors (VR) and capacitors for the sole purpose of regulating the voltage delivered to a load. VRs can be used in any electronic subassembly or system where a constant voltage source is needed to maintain a continuous voltage over a specific range of load resistances. Some applications include computers, peripherals, and various types of instrumentation.

Stepper Motor Driver ■ The *stepper motor driver (SMD)* is a bipolar IC used for controlling stepper motors. An SMD can be used any place where a stepper motor must be accurately controlled, such as in laser printers, photocopiers, scanners, and robotics, as well as aerospace, automotive, and industrial applications.

Digital IC Product Types

The digital IC family includes devices that operate with binary (1s and 0s) bits of data signals, such as those used in calculators, computers, digital pagers, cellular phones, and many others similar applications.

Volatile Memory ■ *Volatile memory* is a semiconductor device that allows data to be stored and changed as desired. Data in the volatile memory is lost when the power is turned off. Volatile type memory devices are used in computers, calculators, and appliances as well as automotive, aerospace, medical, military, and industrial equipment—in any application where logic instructions must be stored and changed as needed by the user. All memory devices contain millions of individual memory cells. The probability of defects increases with device density; therefore, redundant

memory cells are built into memory chips. Semiconductor fuses are built into the memory cells that are deliberately blown during the initial product test to disable defective cells.

RAM. A *random-access memory* device is one that can be accessed to read its stored data or it can be erased and new data rewritten into it. RAM chips do not need to be removed from the printed circuit board to change the data held in storage within its memory cells. Its contents can be easily changed in the normal operations of the logic system.

DRAM. The *dynamic RAM* is the most common and least expensive of the RAM family. The term *dynamic* refers to the refresh voltage that must be applied regularly to the storage capacitors to retain the data. DRAMs require more power to operate the capacitors.

SRAM. The *static RAM* uses flip-flops as data storage registers. SRAMs do not require refreshing; thus, they use less power than the DRAMs. The data is also lost when the power is removed.

MPU or CPU. *Microprocessor units* (also known as *central processing units*), are complex logic ICs that are capable of executing instructions programmed into a separate or internal ROM. The MPU is capable of decision making and of performing math functions. Microprocessors are used in computers, calculators, and appliances as well as automotive, aerospace, medical, military, and industrial equipment—in any application where a controlling function is needed or where computer functions are a requirement.

Nonvolatile Memory ■ The *nonvolatile memory* is a semiconductor device designed to store digital data in the form of an electrical charge. The charge remains in storage even after the power is turned off. Nonvolatile type memory devices are used in computers, calculators, and appliances as well as automotive, aerospace, medical, military, and industrial equipment—in any application where logic instructions must be stored for later recall.

ROM. A *read-only memory* IC is nonvolatile and is programmed directly in the IC manufacturing process. It is also referred to as a *mask-programmable ROM*. The mask sets used during manufacturing contain code patterns for a specific ROM. ROMs are very expensive to produce. ROMs were the earliest version of nonvolatile memory devices. The consumer can read electronically what has been coded into the ROM but cannot change its contents.

PROM. A *programmable read-only memory* is an IC that can be reprogrammed in the field and is less expensive than the mask-programmable ROM. This tool applies large voltage pulses that allow memory cells to be changed as needed. The data programmed into this memory cannot be changed during normal operation of a computer. There are two types of *field-programmable* proms—the EPROM and the EEPROM.⁶

EPROM. An *erasable PROM* can be erased and reprogrammed in the field. The chip must be removed from the circuit board and erased by exposure to UV light. The EPROM can then be reprogrammed. The EPROM is an improvement over its two predecessors; however, any change in the stored code requires complete erasure of the PROM followed by reprogramming.

EEPROM. The *electrically erasable PROM* IC can be electrically erased and reprogrammed in the field without having to remove it from the circuit board. The EEPROM is the most convenient form of ROM. The device operates more like a random-access memory in that it can be erased and rewritten without the use of any special equipment. EPROMs have many useful applications in the computer and peripheral industry.

Flash RAM. Flash memory is a type of nonvolatile memory that can be erased and reprogrammed—similar to electrically-erasable programmable read-only memory (EEPROM), but faster to update. Flash memory is often used to store the operating code such as the Basic Input/Output System (BIOS) in a personal computer. Flash memory is used in computers, digital cellular phones, digital cameras, embedded controllers, and others products.

ASIC. *Application-specific ICs* are fully custom-designed and manufactured to meet an individual customer's needs. The ASIC device may include existing logic circuitry with new design features as requested by the customer. ASIC chips are also nonvolatile. All logic functions of the IC

are determined by specific masks that have been designed to achieve unique structural features given in the specific ASIC design specifications. While ASIC devices provide the exact product to meet the user's needs, the cost of manufacturing them is high.

PLD. *Programmable logic devices* are ICs that utilize a wide variety of logic components. The actual logic function implemented is determined by the user, using some form of design software to specify the state of the internal programming points.⁷ During the programming process, higher than normal voltages are applied to specific semiconductor fuses. The fuses heat up and vaporize leaving behind a desirable logic circuit configuration as programmed by the user.

PAL. A *programmable array logic* IC contains a network of programmable logic gates used to create a custom logic circuit. The PAL has an array of input AND gates that drive an array of hard-wired output OR gates. The AND gates can be field-programmed as required, but the output OR gates are fixed.⁸

PLA. A *programmable logic array* differs from the PAL in that both the input AND gates and the output OR gates are programmable. PLAs are available in both mask-programmed and field-programmed versions.

MPGA. A *mask-programmable gate array* chip can be customized to meet the functional needs of individual customers. Initially during the manufacturing process all gate array products passing through a specific manufacturing facility may contain the same basic number and configuration of transistor circuitry. It is during the interconnect layers of the manufacturing process that specific masking steps are used to define the function of thousands of transistors within individual chips. The mask sets are designed in cooperation with customers to define the proper interconnects for a specific gate-array application.⁹

FPGA. A *field-programmable gate array* is an IC device in which the final logic product can be defined in the field without the use of an IC manufacturing facility.¹⁰ Programmable logic switches within the FPGA can be activated to create connections between desired logic circuits. The convenience of customizing in the field and its low manufacturing cost make the FPGA a viable alternative to the MPGA. However, the MPGA device operates faster than the FPGA since all interconnects are hard-wired internally during the IC manufacturing process.

SUMMARY

Electronic components are assembled into analog or digital circuits. Passive components, such as resistors and capacitors, conduct current without specific polarity requirements and are constructed on silicon substrates. These components have parasitic losses that are detrimental to IC performance. Active components, such as transistors and diodes, control the direction of current flow. The pn junction diode can form an open circuit (reversed biased) or closed circuit (forward biased). The bipolar junction transistor (BJT) can be biased to conduct current and also amplify input signals. The structure of the BJT is based on n-type or p-type dopant regions that form two pn junctions per transistor as either an npn or pnp. The Schottky diode has faster switching times, which has helped bipolar IC technology extend its applications in to the twenty-first century.

CMOS IC technology is more power efficient than bipolar and is one of the most popular device technologies with the most common transistor being the MOSFET as either an nMOS or pMOS. The MOSFET has three electrodes, the gate, source and drain that are attached to a biasing circuit for conduction mode. The minimum voltage to turn on a MOSFET is the threshold voltage. An example of CMOS technology is the inverter circuit. BiCMOS technology combines both CMOS and bipolar technology. MOSFETs can be in enhancement (normally-off) or depletion (normally-on) mode for both analog and digital applications. Parasitic transistor effects can occur in pn junctions that create a latchup condition that causes the transistor to turn on with no further control. There is a wide range of IC products, including analog and digital.

KEY TERMS

- structure
- analog circuit
- linear circuit
- digital circuit
- logic devices
- device technology
- passive components
- parasitic resistance
- planar capacitors
- parasitic capacitance
- active components
- pn junction diode
- carrier-depletion region
- barrier voltage
- reverse bias
- forward bias
- bipolar junction transistor (BJT)
- npn
- pnp
- emitter
- base
- collector
- current-driven current amplifier
- well
- n-channel MOSFET
- p-well
- p-channel MOSFET
- n-well
- scaling
- shrinking
- field oxide
- glass
- BiCMOS technology
- digital/analog (D/A)
- analog/digital (A/D)
- enhancement mode
- depletion mode
- normally-off
- normally-on
- parasitic transistors
- latchup
- op amp
- voltage regulator (VR)
- stepper motor driver (SMD)
- volatile memory
- random-access memory (RAM)
- gain
- Schottky diode
- bipolar technology
- field-effect transistor (FET)
- CMOS (complementary metal-oxide semiconductor)
- voltage-amplifying device
- gate
- JFET
- MOSFET gate oxide
- MESFET
- nMOS (n-channel)
- pMOS (p-channel)
- polysilicon
- source electrodes
- drain electrodes
- nMOSFET
- pMOSFET
- dynamic RAM (DRAM)
- static RAM (SRAM)
- microprocessor units (MPU)
- central processing units (CPU)
- nonvolatile memory
- read-only memory (ROM)
- programmable read-only memory (PROM)
- field-programmable PROM
- erasable PROM
- electrically erasable PROM
- flash RAM
- application specific IC
- programmable logic devices (PLD)
- programmable array logic (PAL)
- programmable logic array (PLA)
- mask-programmable gate array (MPGA)
- field-programmable gate array (FPGA)

REVIEW QUESTIONS

1. What is an analog circuit? Provide two examples of applications.
2. What is a digital circuit? Provide two examples of applications.
3. What is a passive component? Give two examples of this type of component.
4. List three ways to make a resistor on a single crystalline substrate.
5. What is parasitic resistance? Why is it undesirable in ICs?
6. What is a planar capacitor? Describe four techniques for building this component on a silicon substrate.
7. What is parasitic capacitance? What problems can this condition create in ICs?
8. What is an active component? Give two examples of this type of component.
9. When is a pn junction diode formed? Why is this junction important for an IC?
10. What is the barrier voltage of a pn junction, and how is it formed?
11. Explain what happens when a pn junction is reversed biased.
12. Explain what happens when a pn junction is forward biased.
13. A bipolar junction transistor (BJT) has how many electrodes, junctions, and variations? What are the names of the electrodes? What are the names of the variations?
14. Refer to Figure 3.11. What happens when switch S1 is open? What happens to the E-B junction when switch S1 is closed? When does conduction occur and why?
15. Describe the dopant concentrations in the emitter, collector, and base in an npn bipolar junction transistor.
16. What type of amplifier device is the BJT? How does this affect the applications in terms of power requirements?
17. How is the Schottky diode formed? What is its biggest advantage?
18. What are some notable characteristics of bipolar technology? What is biggest drawback to bipolar technology?
19. What are the benefits of the field-effect transistor (FET)?
20. What is the most popular IC technology today?
21. Is the FET a voltage-amplifying or current-amplifying device?
22. What is the greatest advantage of the FET?
23. How does the FET turn on? What voltage level must be applied to turn on the FET and where should the voltage be applied in the FET?
24. Why is the FET suitable for VLSI and ULSI technology?
25. What are the two basic types of FETs? What is the major difference between them?
26. What are the two categories of MOSFETs? How are they distinguishable from one another?
27. What is the most popular conductor material used in the formation of gates for MOSFETs? How is this material made a conductor?
28. What purpose does the source and drain serve in a MOSFET?
29. What is the majority carrier for a nMOSFET, and what type of channel does it have?
30. What is the majority carrier for a pMOSFET, and what type of channel does it have?
31. What is the well in a MOSFET? Explain what happens in the well region in the conduction mode.
32. If an nMOSFET is in an open-circuit mode, explain what is happening in the three electrode regions (i.e., the source, the drain and the channel below the gate).
33. When switch S1 is closed on the nMOSFET transistor as shown in Figure 3.17, explain what happens with respect to holes and electrons in the electrode regions.
34. What is the threshold voltage in a FET?
35. What is the major difference in performance between a pMOSFET and nMOSFET?
36. What does scaling mean with regards to IC design?
37. What is the efficiency of a CMOS inverter circuit with respect to power consumption?
38. If a p-substrate is used for a CMOS inverter, how is the n-well created?
39. What is the purpose of the field oxide? What is another name for silicon dioxide?
40. What two IC technologies are used in BiCMOS?
41. What could a digital/analog (D/A) converter chip be used for? What could an analog/digital (A/D) chip be used for?
42. Explain the difference between an enhancement-mode transistor and depletion-mode transistor with regards to their standby condition.
43. Which type of MOSFET, enhancement-mode or depletion-mode, is the most commonly used in the semiconductor industry?
44. What is latchup in CMOS devices? What undesirable condition can cause this?
45. Describe three manufacturing techniques used to prevent latchup. What is a linear IC and when would it be used?
46. What is an op amp?
47. What is a voltage regulator?
48. What is a stepper motor driver?
49. What is volatile memory?
50. Describe RAM, DRAM, and SRAM memory.
51. What is MPU?
52. What is nonvolatile memory?
53. Describe the differences among ROM, PROM, EPROM, and EEPROM memory. What is flash RAM, and what are some applications?
54. What is ASIC?
55. What is PLD?
56. What are PAL and PLA?
57. What is an MPGA and a FPGA?

IC MANUFACTURERS' WEB SITES

Actel	http://www.actel.com/
Advanced Micro Devices	http://www.amd.com/
Altera	http://www.altera.com/
Analog Devices	http://www.analog.com/
AT&T Tech. History	http://www.akpublic.research.att.com/
Burr Brown	http://www.bbrown.com/
Cirrus Logic	http://www.cirrus.com/
Cypress Semiconductor	http://www.cypress.com/
Dallas Semiconductor	http://www.dalsemi.com/
Fairchild Semiconductor	http://www.fairchildsemi.com/
Fujita Laboratory	http://www.fujita3.iis.u-tokyo.ac.jp/
Fujitsu	http://www.fujitsu.com/
General Semiconductor	http://www.gensemi.com/
Hitachi Semiconductor	http://semiconductor.hitachi.com/
IBM Microelectronics	http://www.chips.ibm.com/
Intel	http://www.intel.com/
International SEMATECH	http://www.sematech.org/public/index.htm
Intersil Corporation	http://www.intersil.com/
Lattice Semiconductor	http://www.latticesemi.com/
LSI Logic	http://www.lsilogic.com/
Micron Semiconductor	http://www.micron.com/mti/
Mitsubishi Silicon Amer.	http://www.mmc-sil.com/
Motorola Semiconductor	http://mot-sps.com/
National Semiconductor	http://www.national.com/
NEC Semiconductor	http://www.nec.com/semiconductors/
Philips Semiconductor	http://www-us2.semiconductors.philips.com/
QuickLogic	http://www.quicklogic.com/
Rockwell International	http://www.rockwell.com/
Samsung Semiconductor	http://www.usa.samsungsemi.com/
SEMATECH	http://www.sematech.org/public/index.htm
ST Microelectronics	http://www.st.com/stonline/index.shtml
Texas Instruments	http://www.ti.com/
Vishay Siliconix	http://www.vishay.com/brands/siliconix/
Xicor	http://www.xicor.com/
Xilinx	http://www.xilinx.com/
Zilog	http://zilog.com/

Note: Web site addresses are subject to change. The Web site addresses listed here were current at the time of publication.

REFERENCES

1. S. Wolf, *Silicon Processing for the VLSI Era*, vol. 2 of *Process Integration, IC Resistor Fabrication*, (Sunset Beach, CA: Lattice Press, 1990), Appendix A.
2. R. Jacob Baker, H. Li, and D. Boyce, chap. 7 in *CMOS Circuit Design, Layout and Simulation*, (IEEE Press, 1998).
3. A. Sedra and K. Smith, *Microelectronic Circuits*, (New York: Oxford University Press, 1998), p. 138.
4. Ibid., p. 197.
5. R. Jacob Baker, H. Li, and D. Boyce, *Circuit Design*, p. 212–215.
6. S. Brown et al., *Field-Programmable Gate Arrays*, (Norwell, MA: Kluwer Academic Publishers, 1992), p. 2.

66 CHAPTER 3

7. R. Seals and G. Whapshott, *Programmable Logic PLDs and FPGAs*, (New York: MacMillan Press Ltd., 1997), p. 1.
8. J. Carter, *Digital Designing With Programmable Logic Devices*, (Upper Saddle River, NJ: Prentice-Hall, 1997), p. 17.
9. M. Smith, *Application-Specific Integrated Circuits*, (Reading, MA: Addison-Wesley Longman, 1997), p. 4-15.
10. S. Brown et al., *Field Programmable Gate Arrays*, (Norwell, MA: Kluwer Academic Publishers, 1992), p. 1.

CHAPTER 4

SILICON AND WAFER PREPARATION

The primary semiconductor material used to fabricate microchips is silicon, the most important material in the semiconductor industry. In order for silicon to be acceptable for the fabrication of semiconductor devices, a special purity grade is used that meets stringent material and physical conditions.

The typical U.S. semiconductor manufacturer does not produce silicon wafers. The production of the silicon material and its preparation into wafers occurs in a highly specialized factory dedicated to the production

of silicon wafers. These wafers are then supplied to wafer fabrication facilities for processing into various types of microchips by semiconductor manufacturers.

The quality of the final microchip produced in a wafer fab depends directly on the quality of the starting silicon wafer. If defects exist in the incoming wafer, then there will surely be defects in the microchip. An understanding of the silicon wafer and its preparation process will assist us in comprehending the silicon wafer's importance to the complete microchip process.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe how raw silicon is refined into semiconductor-grade silicon.
2. Explain the crystal structure and growth method for producing monocrystalline silicon.
3. Discuss the major defects in silicon crystal.

4. Outline and describe the basic process steps for wafer preparation, starting from a silicon ingot and finishing with a wafer.
5. State and discuss seven quality measures for wafer suppliers.
6. Explain what epitaxy is and why it is important for wafers.

INTRODUCTION

To exploit silicon's desirable properties as a semiconductor, the natural silica must be purified into an extremely clean silicon material. Pure silicon is required to minimize microdefects at the atomic level of the silicon that are detrimental to semiconductor performance. Once pure silicon is obtained, it is then transformed into the physical wafer with the desired crystal orientation, proper amount of dopants, and physical dimensions necessary for semiconductor wafer fabrication.

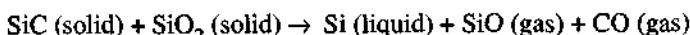
SEMICONDUCTOR-GRADE SILICON

The highly refined silicon used for wafer fabrication is termed *semiconductor-grade silicon* (SGS), which is also sometimes referred to as electronic-grade silicon. Obtaining SGS from natural silicon with the acceptable purity level needed for semiconductor device performance is a multistep process (see Table 4.1).¹ There are alternative methods for obtaining SGS, but the following description is the predominant production process.

TABLE 4.1 Steps to Obtaining Semiconductor-Grade Silicon (SGS)

Step	Description of Process	Reaction
1	Produce metallurgical-grade silicon (MGS) by heating silica with carbon.	$\text{SiC}(\text{s}) + \text{SiO}_2(\text{s}) \rightarrow \text{Si}(\text{l}) + \text{SiO}(\text{g}) + \text{CO}(\text{g})$
2	Purify MG silicon through a chemical reaction to produce a silicon-bearing gas of trichlorosilane (SiHCl_3).	$\text{Si}(\text{s}) + 3\text{HCl}(\text{g}) \rightarrow \text{SiHCl}_3(\text{g}) + \text{H}_2(\text{g}) + \text{heat}$
3	Using the Siemens process, SiHCl_3 and H_2 react to produce pure semiconductor-grade silicon (SGS).	$2\text{SiHCl}_3(\text{g}) + 2\text{H}_2(\text{g}) \rightarrow 2\text{Si}(\text{s}) + 6\text{HCl}(\text{g})$

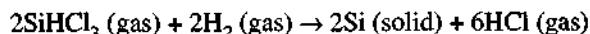
The first step to achieving SG silicon is to produce a metallurgical-grade silicon (MGS) by heating silica (SiO_2), a pure sand, with carbon in a reducing-gas atmosphere.



The resulting metallurgical-grade silicon on the right side of the reaction is 98% pure. Because MG silicon has unacceptable levels of contaminants, it is of no use for semiconductor fabrication. This metallurgical-grade silicon is then crushed and purified through a chemical reaction that produces a silicon-bearing gas of trichlorosilane (SiHCl_3).



The silicon-bearing gas of SiHCl_3 undergoes additional chemical processing and reduction in hydrogen to produce 99.999999% (nine nines) pure, semiconductor-grade silicon (SGS).² This rate equation is,



The pure SGS is produced in a process referred to as the Siemens process.³ The SiHCl_3 gas and hydrogen are injected into the Siemens reactor (See Figure 4.1) and chemically react on heated rods of ultrapure silicon (the rod temperature is about 1100°C). After several days, the process is complete and the rods of deposited SGS are cut up into smaller pieces for growth into the silicon crystal.

Semiconductor-grade silicon has the ultrahigh purity required for semiconductor manufacturing, with less than two parts per million (ppm) of carbon and less than one part per billion (ppb) for elements from Groups III and V (the critical doping elements) of the periodic table.⁴ However, the silicon produced by the Siemens process does not have atoms arranged in a desirable crystal order. We will now analyze crystal structure to understand the correct atomic order for semiconductor-grade silicon.

CRYSTAL STRUCTURE

It is not only critical that the semiconductor-grade silicon used to fabricate semiconductor devices have ultrahigh purity, but it must also have a near-perfect crystal structure. This condition is necessary to avoid electrical and mechanical defects that are detrimental to device performance.

A *crystal* is a solid material that has an ordered, repeatable three-dimensional pattern over a long range of many atoms. Figure 4.2 shows how atoms are related in a crystal structure, referred to as the *crystal lattice*. The crystal lattice represents the repeatable order within a crystal at the atomic level of the internal structure (atomic order), even though the surfaces of the crystal may be abraded or rough. An example is the uneven surface of common beach sand, which has an internal crystalline structure with atomic order. Our goal is to obtain a specific crystal lattice necessary for the specific requirements for SG silicon used to fabricate wafers.

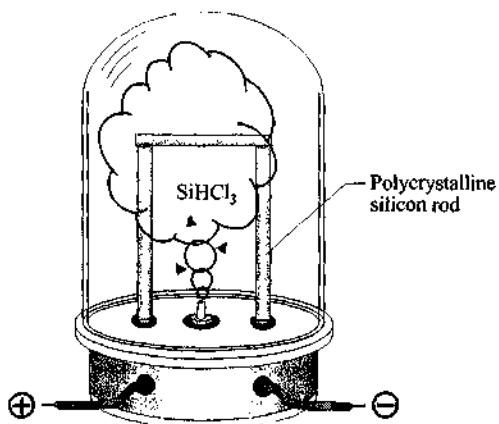


FIGURE 4.1 Siemens Reactor for SG Silicon

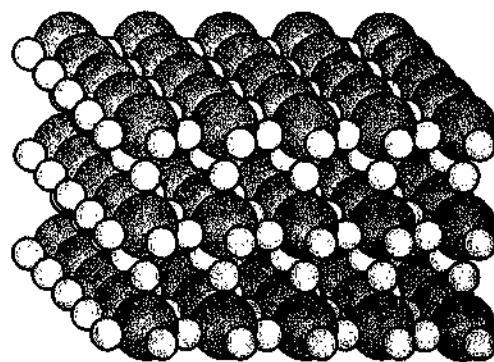


FIGURE 4.2 Atomic Order of a Crystal Structure

Amorphous Materials

Amorphous materials are noncrystalline solids that lack a repetitive structure and demonstrate structural disorder at the atomic level (see Figure 4.3). Amorphous materials are the opposite of crystalline materials. Plastic is an example of an amorphous material. Amorphous silicon would be an unacceptable structure for the silicon used in wafers to fabricate semiconductor devices. This is because many of the electrical and mechanical aspects of a device occur at the atomic level in the silicon, which requires order and predictability to yield repeatable results from chip to chip.

Unit Cells

The most fundamental entity for the long-range order found in the atomic pattern of a crystal material is the *unit cell*. A unit cell is the simplest arrangement of atoms that, when repeated in a three-dimensional framework, gives the crystal structure. We need to study unit cells in order to understand how silicon replicates itself in its crystal structure. A two-dimensional analogy to a unit cell can be seen in a checkerboard or a section of a tiled floor. A three-dimensional analogy would be a neat arrangement of a child's building blocks. An illustration of a unit cell in a three-dimensional structure is shown in Figure 4.4.

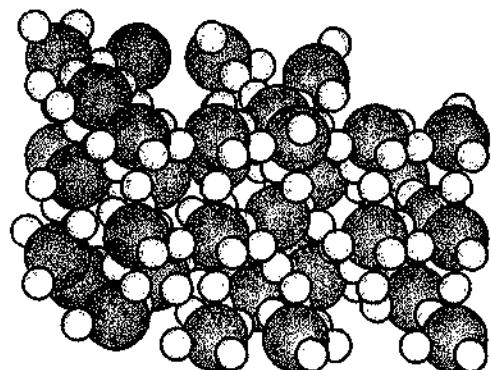


FIGURE 4.3 Amorphous Atomic Structure

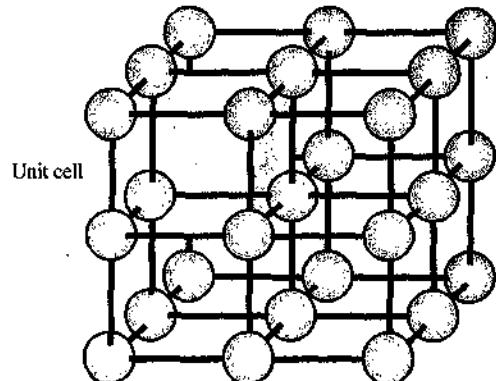


FIGURE 4.4 Unit Cell in 3-D Structure

Since crystal structures are identical in three dimensions, unit cells have a framework structure, such as a cube. There are seven possible crystal systems in nature.⁵ The crystal system relevant to silicon technology is the cubic system. We will only consider one of the basic crystal structures because it is applicable to silicon: *face-centered cubic*, or FCC. The FCC unit cell is shown in Figure 4.5.

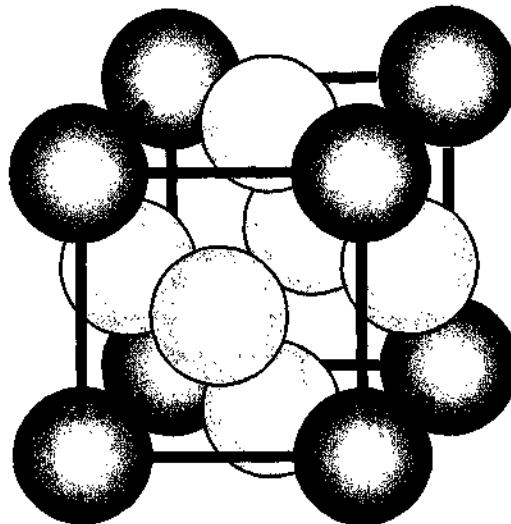


FIGURE 4.5 Face-Centered Cubic (FCC) Unit Cell

Within a crystal structure, unit cells are tightly packed and, therefore, share atoms. Sharing atoms is important because this is how unit cells build up into a cohesive crystal lattice structure. In an FCC unit cell, each corner shares an atom with eight other unit cells, while each face shares with one other unit cell. For an FCC unit cell, there is the equivalent of four shared atoms.

For a silicon crystal, the unit cell is a variation of the FCC known as the *FCC diamond structure* (see Figure 4.6). Atoms are shared as with the FCC unit cell, plus there are four complete atoms situated inside the cubic structure. For a silicon unit cell, there are a total of eight complete atoms, with four shared and four unshared.

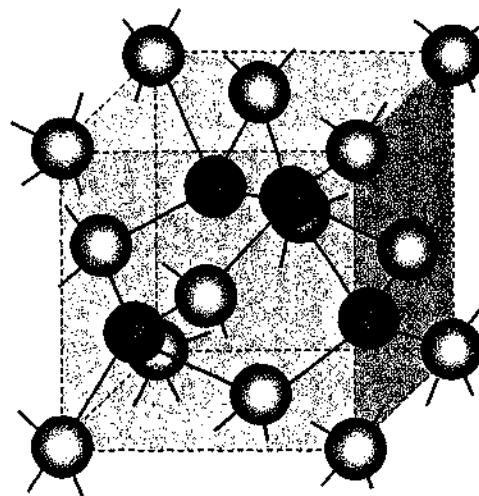


FIGURE 4.6 Silicon Unit Cell: FCC Diamond Structure

Polycrystal and Monocrystal Structures

Now that the unit cell is defined for a crystal structure, we can define how the unit cells are organized. If the unit cells are not in a regular arrangement, then the material is a *polycrystal*. The

silicon produced from the semiconductor-grade purification process is polycrystalline silicon. Another name for polycrystalline silicon material is *polysilicon*. An analogy for a polycrystal structure is a pile of bricks. Each individual brick represents the unit cell and the pile is polycrystalline because there is no repetitive arrangement (see Figure 4.7).

If the unit cells are neatly arranged in a three-dimensional, repeatable manner, then the crystal structure is *monocrystal*. Another term often used for monocrystal is *single crystal*. For our brick analogy, if the bricks are now stacked neatly in repeating rows next to one another, then the bricks represent a monocrystal. A monocrystal structure is also shown in Figure 4.7.

Semiconductor wafer processing requires a pure, monocrystalline silicon structure (single crystal). This is because the repeatable unit cells of a monocrystal structure provide the desirable electrical and mechanical properties necessary for silicon wafer processing and performance. Unacceptable crystal structure and defects influence the formation of microdefects that affect wafer processing (see the following section on crystal defects in silicon).

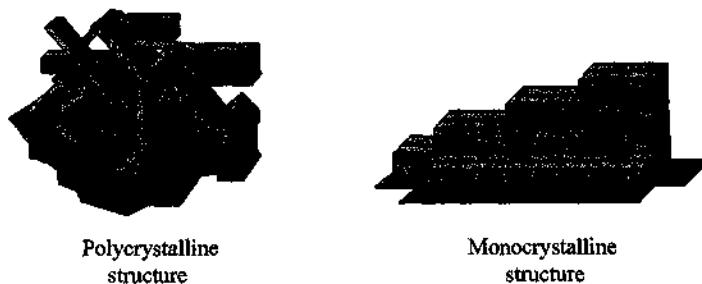


FIGURE 4.7 Polycrystalline and Monocrystalline Structures

CRYSTAL ORIENTATION

Before studying the production process to obtain a monocrystal structure for silicon, let's first examine how unit cells are oriented in silicon. Unit cell orientation is important because it defines how the silicon crystal structure is physically aligned with the silicon wafer. Different crystal orientations change the chemical, electrical, and mechanical properties of the silicon, which affects the manufacturing process conditions and final device performance.

To define the silicon unit cell orientation, we need a coordinate system. In crystals, the coordinate system has three axes, *x*, *y*, and *z*, shown in Figure 4.8. We will arbitrarily assign the number 1 at an equal distance along each axis, while the center point is given the value of 0. These are referred to as unit values. If the crystal is a monocrystal, then all unit cells are aligned in a repetitive manner along these three coordinate axes.

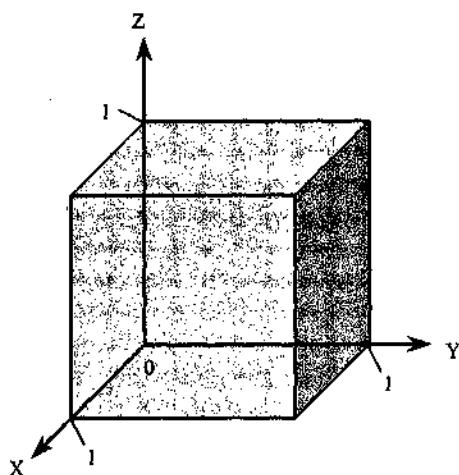


FIGURE 4.8 Axes of Orientation for Unit Cells

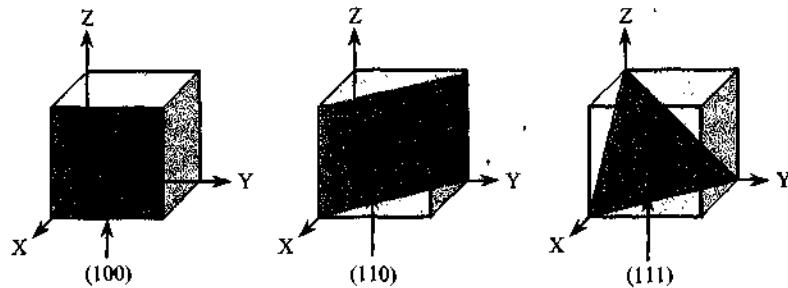


FIGURE 4.9 Miller Indices of Crystal Planes

The orientation of planes in the silicon crystal is described by a set of numbers known as *Miller indices*.⁶ It is beyond the scope of this book to completely describe the Miller notation system for crystal planes and their directions. In the Miller system of notation, parentheses () are used to denote a specific plane, whereas brackets < > denote groups of equivalent directions.⁷

From our standpoint, what is important is to recognize that each set of Miller indices enclosed in a parenthesis specifies a unique plane in the crystal structure. For semiconductor fabrication, the Miller notations applicable to the most common crystal planes on a wafer are (100), (110), and (111). These three crystal planes are shown in Figure 4.9 on page 72. They are obtained in the silicon crystal by maintaining precise control of orientation during the crystal growth process (described in the next section). You interpret each Miller notation based on where the plane intersects the axes.

The (100) crystal plane is parallel to the y-z plane and intersects the x-axis at the unit value of one. The (110) intersects the x- and y-axes only, whereas the (111) intersects the x-, y-, and z-axes. Wafers with a (100) crystal plane orientation are most common for fabricating MOS devices. The reason for this is that the surface state condition for (100) silicon is more conducive toward controlling the threshold voltage required to turn MOS devices on and off.⁸ The (111) crystal plane orientation has a tighter packing density at the atomic level, making it easier to grow. These are the least expensive crystals to grow and are often used for bipolar devices.⁹ Gallium arsenide (GaAs) technology also uses wafers with (100)-oriented planes.

MONOCRYSTAL SILICON GROWTH

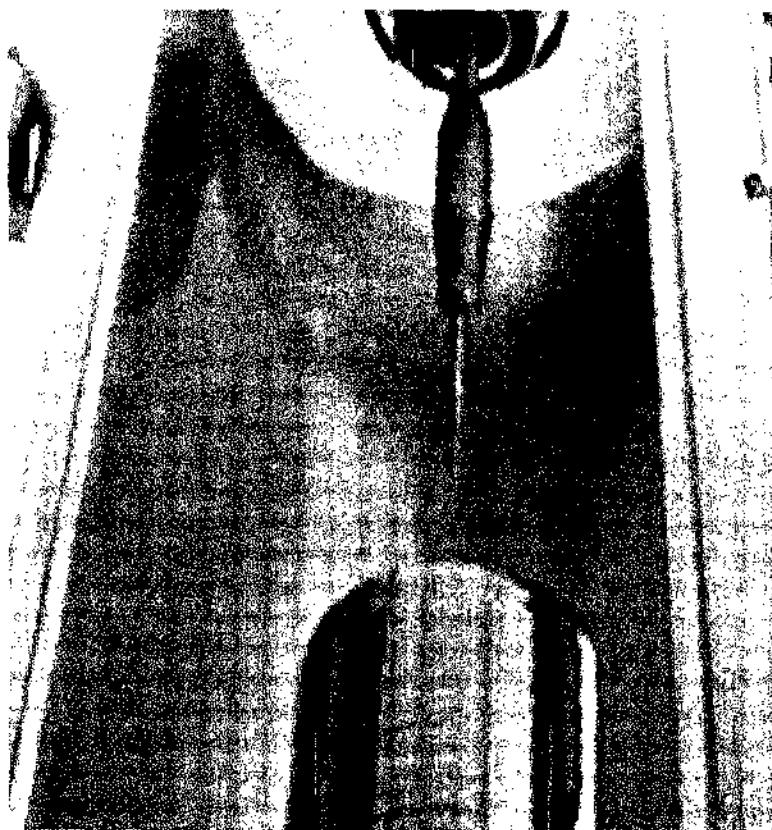
Crystal growth is the process of converting the polysilicon chunks of semiconductor-grade silicon into a large monocrystal of silicon. The grown silicon monocrystal is referred to as an *ingot*. The most common technique today for growing a silicon monocrystal ingot to be used in wafer fabrication is the *Czochralski (CZ) method*, which is named for the inventor who developed the process in the early 1900s.

CZ Method

The Czochralski (CZ) growth of a silicon monocrystal involves the transformation of molten SG silicon liquid into a solid silicon ingot that has the correct crystal orientation and is doped either as n-type or p-type. Over 85% of all silicon crystals are grown according to the CZ method.¹⁰

A piece of monocrystalline silicon having the desired crystal orientation is used as a starter seed to grow the silicon ingot that is a replica of the original seed crystal. To achieve crystal growth with the CZ method, precise conditions are controlled at the contact interface between molten silicon and a single monocrystal seed of silicon. These conditions permit thin films of silicon to accurately replicate the seed crystal structure and grow into a large ingot. This is done with equipment known as a *CZ crystal puller*.

CZ Crystal Puller ■ To grow a silicon ingot, chunks of SG silicon are placed in a fused silica (amorphous quartz) crucible holder, along with small amounts of dopant to create either n-type or p-type silicon. The crucible is large, with crucible diameters for a 300-mm wafer at 32 inches or larger. These crucibles must hold from 150 to 300 kg of silicon.¹¹ An alternative approach for larger diameter wafers is to use granular polysilicon in the crucible, which permits the gradual



Silicon Ingot Grown by CZ Method
(Photo courtesy of Kayex Corporation)

introduction of the silicon during the meltdown to minimize stresses on the large crucible during the melting process.¹² The crucible is positioned in the crystal puller, which is where the silicon crystal ingot is grown (see Figure 4.10).

The crucible of silicon is heated in the furnace of the puller using either resistance heaters or RF (radio frequency) heating coils. Resistance heaters are used for larger diameter ingots. As the silicon is heated, it turns to liquid and is referred to as the *melt*. A seed of a perfect silicon crystal is attached to a pull mechanism to start the structure of the new crystal. The seed is placed at the surface of the melt and is slowly drawn away while rotating in a direction that is opposite to the rotation of the crucible.

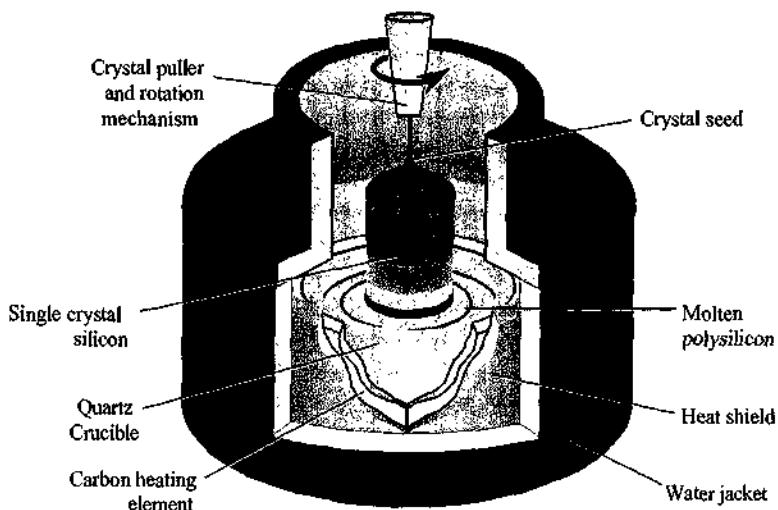
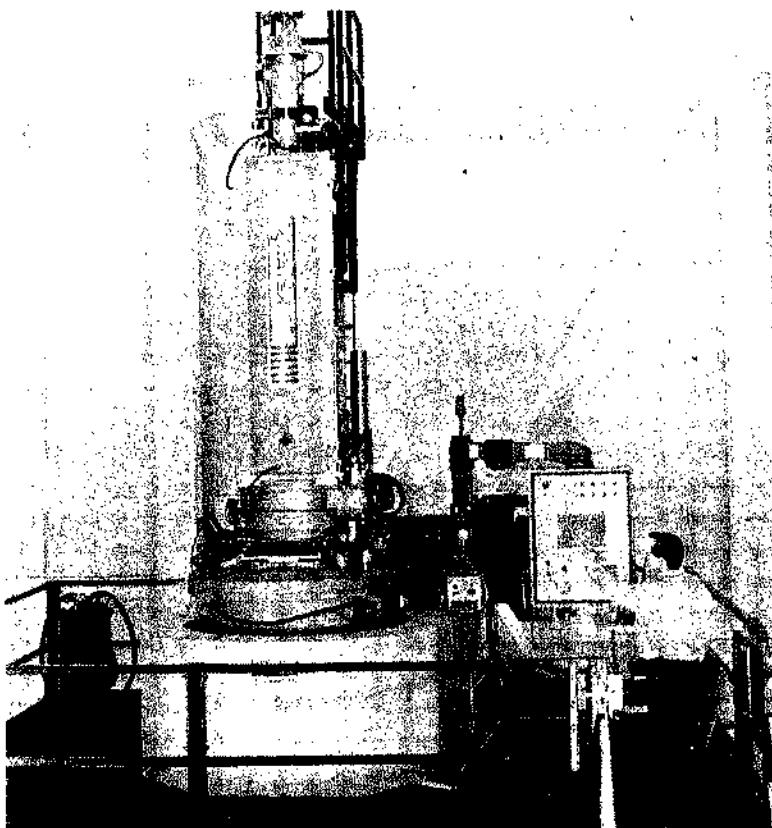


FIGURE 4.10 CZ Crystal Puller



CZ Crystal Puller
(Photo courtesy of Kayex Corporation)

As the single seed is drawn away from the melt during the pull process, liquid from the melt is raised with it by surface tension. The interface at the single seed dissipates heat and solidifies downward toward the melt. As the seed is steadily pulled out of the melt and rotated, a single crystal grows with the same crystal orientation as the seed. While the seed is rotated, the crucible is also rotated. Different ingot growth patterns result depending on the speed of the seed and crucible rotation and their respective directions.

The objective of the pull process is to precisely replicate the seed structure while obtaining dopant uniformity, achieving the correct ingot diameter, and limiting introduction of impurities into the silicon. The two main parameters that affect the pull process are the pull rate and crystal rotation.¹³ The pull starts at a relatively high speed and then slows down considerably. This action forms a neck in the ingot, since the diameter of the growing crystal is directly related to the rate of pull.

A goal of the crystal growth process is to have uniform (or homogeneous), large-diameter crystals. This result gives more predictable crystal properties with respect to parameters such as resistivity. A technique developed to achieve this with the CZ process is to use a magnetic field around the silicon melt to stabilize the crystal during its growth.¹⁴ This condition is referred to as *magnetic CZ (MCZ)*.

Doping ■ Dopant material is added to the melt in the crystal puller in order to obtain the desired electrical resistivity in the finished crystal. The resistivity of pure silicon is approximately $2.5 \times 10^5 \Omega\text{-cm}$. The most common dopants used in crystal growth are trivalent boron to create p-type silicon or pentavalent phosphorus to create n-type silicon. Dopant concentration ranges in the silicon are indicated through letter designations and superscripts, as shown in Table 4.2. Usually raw dopants are not added directly to the melt because the dopant amounts are extremely small. It is typical to add the dopant in the form of a highly doped powder of crushed silicon.¹⁵

TABLE 4.2 Dopant Concentration Nomenclature in Silicon

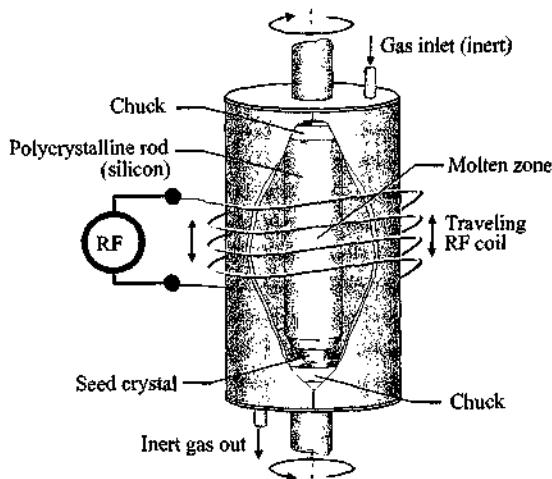
Dopant	Material Type	Concentration (Atoms/cm ³)			
		< 10 ¹⁴ (Very Lightly Doped)	10 ¹⁴ to 10 ¹⁶ (Lightly Doped)	10 ¹⁶ to 10 ¹⁹ (Doped)	> 10 ¹⁹ (Heavily Doped)
Pentavalent	n	n ⁻	n ⁻	n	n ⁺
Trivalent	p	p ⁻	p ⁻	p	p ⁺

Impurity Control ■ Impurity control is important during crystal growth since unacceptable impurities affect the device performance. One impurity that must be controlled but that can also be beneficial is oxygen. The primary source of oxygen in the CZ method results from the dissolution of the crucible during crystal growth.¹⁶

Small amounts of oxygen in the ingot are desirable because oxygen can serve as an internal getter to tie up metallic contaminants that are introduced into the wafer during the fabrication process. Note that the term *getter* is used to describe any process that immobilizes or ties up impurities. Most of the oxygen in the silicon crystal near the wafer surface is removed during the many heating processes that take place during wafer fabrication, leaving higher oxygen concentrations deeper in the wafer. This oxygen then serves as the getter that attracts the contaminants away from the devices located on the surface. The amount of oxygen required to serve as a getter is continually being reduced as the device fabrication process is improved to reduce the sources of contaminants.¹⁷

Float-Zone Method

An alternative crystal growth method is the *float-zone* method, which produces a silicon monocrystal ingot with significantly lower oxygen content.¹⁸ The float-zone method was developed in the 1950s and produces the purest bulk silicon single crystals known to date. A schematic of the float-zone process is shown in Figure 4.11.

**FIGURE 4.11** Float Zone Crystal Growth

The float-zone method of growing a monocrystal ingot of silicon starts with a bar of doped polysilicon that has been cast in a mold. A monocrystal seed is attached to one end of the bar and then placed in the grower. An RF (radio frequency) coil applies heat at the contact area of the bar and seed. Heating of the polysilicon bar is the most important aspect of the float-zone process, as each section of the bar is molten only for about 30 minutes before it solidifies again at the monocrystal interface.¹⁹ This monocrystal growth process progresses along the bar as the heater is moved along its axis.

The float-zone process typically makes smaller-diameter wafers than the CZ process and was producing predominantly 125-mm wafers at the turn of the century. By not using a crucible, the float-zone process results in high-purity silicon with lower oxygen content.

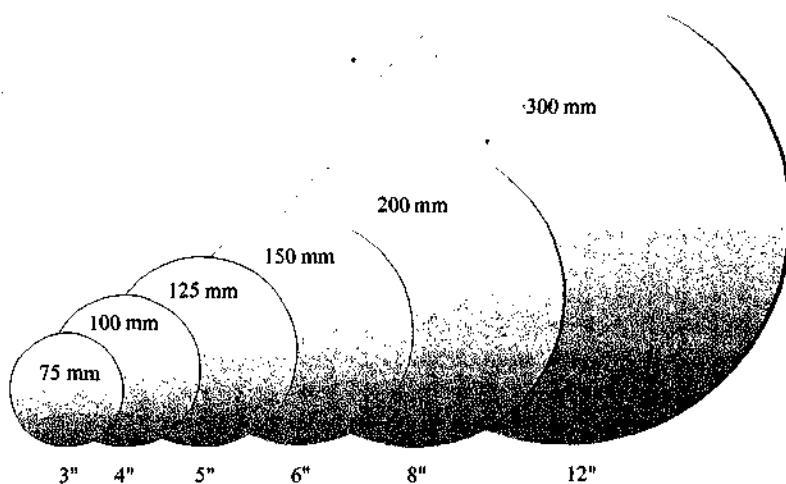


FIGURE 4.12 Wafer Diameter Trends

Reasons For Larger Ingot Diameters

Silicon ingot diameters have increased from the early diameters of less than 25 mm in the 1950s to the most recent diameter of 300 mm. The historical trend for larger semiconductor wafer diameters is shown in Figure 4.12.

There are wafer fabrication facilities still in use that produce wafers with diameters of 75 mm, 100 mm, 125 mm, and 150 mm. The most common practice is to build a new fab to introduce a new wafer diameter, since upgrading a facility to a larger wafer diameter can cost hundreds of millions of dollars. The semiconductor industry began to transition to 300-mm wafers around the year 2000. There are also early evaluations underway to understand the possibility of increasing wafer diameters to 400 mm. Table 4.3 highlights different attributes of various wafer sizes.

TABLE 4.3 Wafer Dimensions & Attributes

Diameter (mm)	Thickness (μm)	Area (cm ²)	Weight (grams/lbs)	Weight/25 Wafers (lbs)
150	675 ± 20	176.71	28 / 0.06	1.5
200	725 ± 20	314.16	53.08 / 0.12	3
300	775 ± 20	706.86	127.64 / 0.28	7
400	825 ± 20	1256.64	241.56 / 0.53	13

From H. Huff, R. Goodall, R. Nilson, and S. Griffiths, "Thermal Processing Issues for 300 mm Silicon Wafers: Challenges and Opportunities," *ULSI Science and Technology* (New Jersey: The Electrochemical Society, 1997), p. 139.

Larger ingot diameters bring new challenges to achieving correct crystal growth and maintaining correct process control during ingot growth. Ingots with a 300-mm diameter are about 1 meter long and require between 150 to 300 kg of SG silicon in the crucible for the melt. With the increased complexity needed to manufacture the silicon ingot, what is the main reason to keep increasing the diameter? The answer lies in the cost benefits to the wafer fabrication process due to increasing the wafer diameter.

Larger-diameter wafers have a greater surface area for chips. For 300-mm wafers, there is 2.25 times the surface area than that available on a 200-mm wafer (see Figure 4.13). This increase translates into more chips that can be produced from a single wafer.

An analogy would be driving a car 200 miles from city A to city B. If the driver travels alone, then this establishes a certain cost for making the trip (e.g., fuel, wear on the vehicle, and so on). However, if a passenger goes with the driver, then the driver and the passenger make the trip for less cost. If three passengers make the trip with the driver, then there are even more savings per passenger. This approach to efficiency is called *economies of scale*.

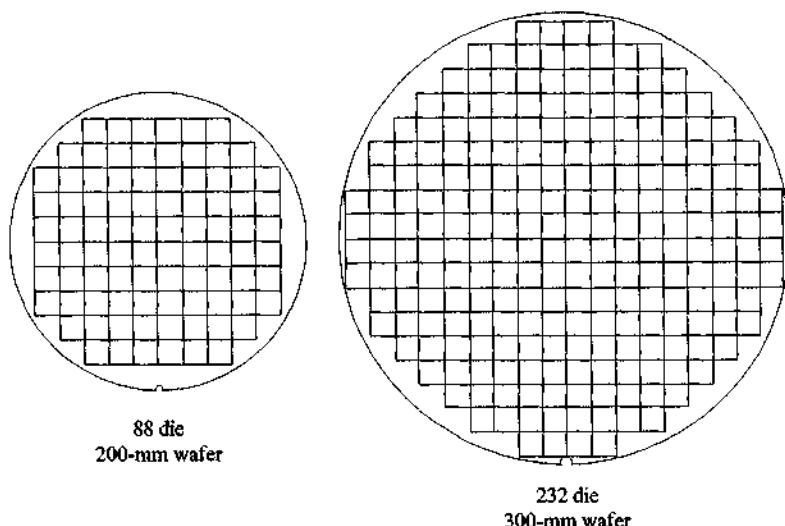


FIGURE 4.13 Increase in Number of Chips on Larger Wafer Diameter

Larger wafers means more chips per wafer, leading to improved equipment productivity due to economies of scale for less handling per chip and less processing time per chip. It is estimated that conversion to a 300-mm wafer diameter will reduce the fabrication cost per chip by approximately 30% through equipment utilization improvements.²⁰ Another benefit from larger-diameter wafers is that fewer chips are near the edge of the wafer, which translates into higher production yield. There is also a benefit from tool repeatability from chip to chip since more chips are exposed to the same process conditions.

The total industry cost to convert the standard wafer diameter from 200 mm to 300 mm has been estimated at \$13 billion to \$15 billion. No one firm or country has the resources to make this change alone. Global 300-mm standards have been developed and 23 semiconductor companies worldwide have joined two separate consortia to coordinate the conversion to 300-mm wafers. The two consortia are the International 300 mm Initiative (I300I), which includes the U.S., Europe, Korea and Taiwan, and the Semiconductor Leading Edge Technology (Selete), which represents ten IC companies from Japan.²¹ Ultimately, cost savings will be the major factor that drives the industry toward a larger wafer diameter. A summary of dimensional specifications for 300-mm wafers is provided in Table 4.4.

TABLE 4.4 Developmental Specifications for 300-mm Wafer Dimensions and Orientation Requirements

Parameter	Units	Nominal	Some Typical Tolerances
Diameter	mm	300.00	± 0.20
Thickness (center point)	μm	775	± 25
Warp (max)	μm	100	
Nine-Point Thickness Variation (max)	μm	10	
Notch Depth	mm	1.00	+ 0.25, -0.00
Notch Angle	Degree	90	+5, -1
Back Surface Finish		Bright Etched/Polished	
Edge Profile Surface Finish		Polished	
FQA (Fixed Quality Area—radius permitted on the wafer surface)	mm	147	

From H. Huff, R. Goodall, R. Nilson, and S. Griffiths, "Thermal Processing Issues for 300 mm Silicon Wafers: Challenges and Opportunities," *ULSI Science and Technology* (New Jersey: The Electrochemical Society, 1997), p. 139.

CRYSTAL DEFECTS IN SILICON

Semiconductors require a nearly perfect silicon crystal structure in order to function optimally for advanced ICs. A *crystal defect* is any interruption in the repetitive nature of the unit cell crystal structure. Another term used for a crystal defect is *microdefect*. It is essentially impossible to grow or process silicon without making a defect. However, progress has been made on growing nearly defect-free silicon with a very low defect density.

Defect density is a commonly used term in crystal growth and wafer fabrication. It is defined as the number of defects per cm^2 of wafer surface that may occur during processing due to all kinds of causes. Reducing defect density is a critical aspect for increasing wafer yield (see Figure 4.14). Yield is the percentage of good chips produced out of the total chips on a wafer.

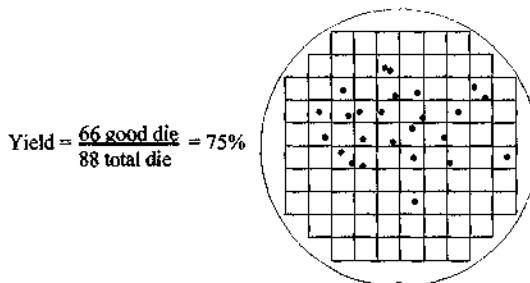


FIGURE 4.14 Yield of a Wafer

It is important to study silicon crystal defects because they have a damaging effect on the electrical properties of the semiconductor. These problems includes SiO_2 dielectric breakdown and leakage current failures. As device geometries shrink and active gate areas increase due to more transistors on a chip, the probability of a defect occurring in a sensitive area of the chip increases. Such a defect can negatively affect the device yield for advanced ICs.

Crystal defects can occur during crystal growth and the subsequent processing of the silicon ingot and wafers. Some crystal defects occur due to surface damage, including cracks and surface defects from mechanical operations on the wafer. There are three general forms of crystal defects in silicon:

1. Point defects: Localized crystal defect at the atomic level
2. Dislocations: Displaced unit cells
3. Gross defects: Defects in crystal structure

Point Defects

A *point defect* occurs at a particular location in the crystal lattice. Three types of point defects are shown in Figure 4.15. The most basic type of point defect is a *vacancy* (also referred to as a *void*).

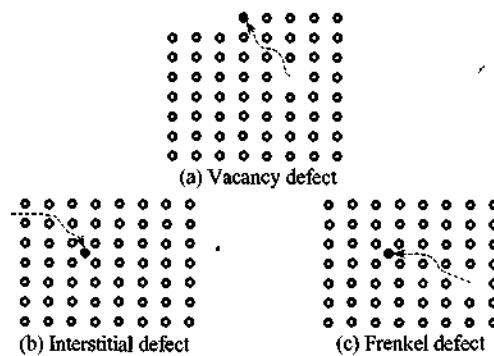


FIGURE 4.15 Point Defects

Redrawn from S. Ghandi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed. (New York: Wiley, 1994), p. 23.

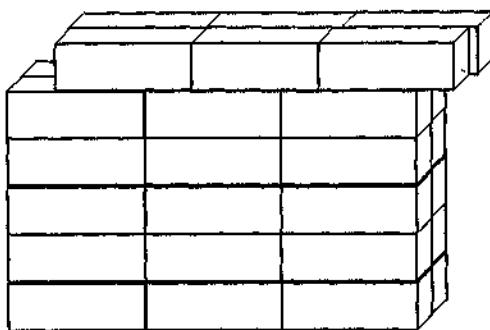


FIGURE 4.16 Dislocations

This defect occurs when an atom is removed from its lattice site and moves to the surface of the crystal. Another type of point defect is an *interstitial*, which occurs when an atom becomes located in one of the many voids within the crystal structure. When an atom leaves its lattice site and takes up position in a void, it creates a vacancy-interstitial pair, or *Frenkel defect*.

Point defects in semiconductor silicon have become increasingly important as device technology becomes more complex. Crystal growth conditions that affect the creation of point defects during wafer growth are the growth rate (how fast the crystal is pulled) and the thermal gradient (the temperature difference between the melt and the solid crystal) at the crystal-melt interface.²² The silicon crystal cooling rate is controlled, which effectively reduces the defect generation. High-temperature processing during semiconductor manufacturing is also a source of point defects.

Another type of point defect is due to chemical impurities that are introduced into the lattice. Impurities are introduced intentionally or unintentionally during fabrication. They can take up locations at sites ordinarily occupied by atoms, known as *substitutional impurities*, or locate themselves in voids as *interstitial impurities*.

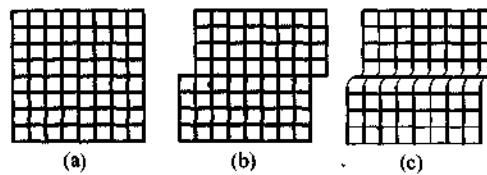
The most significant impurities not introduced intentionally into the silicon are oxygen and carbon. It is important in VLSI and ULSI device fabrication to control these impurities. Oxygen enters the silicon melt via dissolution of the crucible during the CZ process, as discussed earlier in this chapter. Most of this oxygen is removed, but less than 5% remains in the growing crystal. Oxygen present in silicon can create crystal defects such as dislocations (see the next section). The presence of donor electrons from oxygen impurities in the silicon lattice can also affect the electrical parameters of a pn junction. Carbon can act as a nucleation site for crystalline defect formation in silicon.

Dislocations

Unit cells form a repetitive structure in a monocrystal. If unit cells are displaced, this condition is known as a *dislocation* (see Figure 4.16). An analogy would be a neat stack of bricks that has a group of bricks displaced along a row. A form of dislocation is *stacking faults* which is due to layer stacking errors. Dislocations can be introduced at any stage of the crystal growth or wafer fabrication process. However, dislocations occurring after crystal growth are often associated with mechanical stress to the wafer, such as that caused by uneven heating and cooling or excessive force applied to the wafer.

In some instances, dislocation faults are induced following the thermal oxidation of the silicon wafer surface during device fabrication (see Chapter 10).²³ The result is referred to as *oxygen-induced stacking faults (OISF)*. OISF defects appear as shallow, saucer-shaped depressions in the lattice. These defects can be detected through X-ray analysis or surface etching. Specialized thermal treatments (referred to as annealing) or gettering can be used to minimize stacking faults and dislocations in silicon crystal structures.

Crystal growth dislocations have been reduced for large-diameter wafers through use of a neck-down procedure during ingot growth.²⁴ This procedure consists of necking down the cross section of the single seed at the beginning of the pull so that the dislocations grow out to the surface, and then growing several centimeters at sufficiently high speed so that high-vacancy densities remove the edge dislocations. However, as described above, dislocations can still occur from other

**FIGURE 4.17** Crystal Slip

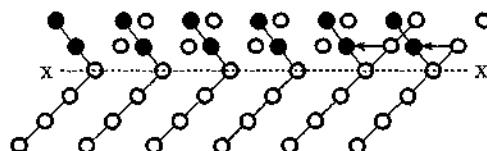
Redrawn from S. Ghandi. *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed. (New York: Wiley, 1994), p. 49.

sources during the wafer fabrication process, such as from wafer edge chipping or high temperature processing.

Gross Defects

Gross defects are related to the structure of the crystal and typically occur during crystal growth. A *slip* is a gross defect in a crystal that occurs when there is slippage of the crystal along one or more crystal planes (see Figure 4.17).

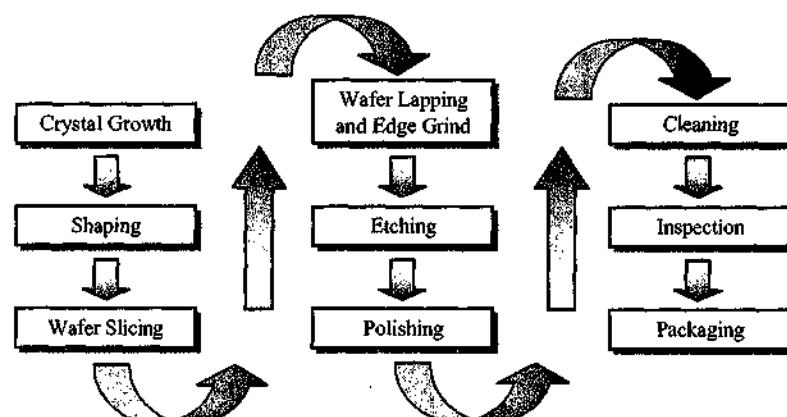
Another gross defect is *twin planes* (or *twinning*), where the crystal grows in two different directions from the same plane (see Figure 4.18). The source of twin planes could be thermal or mechanical shock during the growth process. The crystal on either side of the defect may be perfect. Crystals with either slip or twin planes are unacceptable for semiconductor manufacturing.

**FIGURE 4.18** Crystal Twin Planes

Redrawn from S. Ghandi. *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed. (New York: Wiley, 1994), p. 55.

WAFER PREPARATION

Silicon is a hard, brittle material and of little use for semiconductor manufacturing in its form as an ingot after crystal growth. The cylindrical, single-crystal ingot (also called a *boule*) must undergo a series of process steps to transform it into wafers that meet stringent specifications for semiconductor manufacturing. These wafer preparation steps include machining operations, chemical operations, surface polishing, and quality measures. The basic process flow for wafer preparation is shown in Figure 4.19.

**FIGURE 4.19** Basic Process Steps for Wafer Preparation

Because of the continued advances in chip design and fabrication requirements, the wafer preparation process must be capable of delivering wafers that meet tighter specification requirements. Wafer preparation requirements include the wafer geometry (diameter, flatness, and warpage), surface perfection (roughness and light scattering), and cleanliness (sources of particles). These wafer specifications address issues such as dimensional control for robotic material handling or surface conditions required for critical process steps during IC fabrication.

Shaping Operations

After growth of the ingot in the crystal puller, the shaping operations are the first process steps. Shaping operations include all preliminary steps to prepare the monocrystal silicon ingot prior to slicing into wafers.

End Removal ■ The first operation is to remove the two ends of the ingot. The ends are referred to as the *seed end* (where the seed was located) and the *tang end* (opposite to the seed end). Once the ends are removed (cropped off), a four-point probe resistivity check is done to confirm the proper dopant uniformity throughout the ingot (see Chapter 7 for the resistivity measurement procedure).

Diameter Grinding ■ The next operation is *diameter grinding* to create the precise diameter of the material. Ingots are grown slightly oversize to permit this grinding step, since diameter control and roundness cannot be adequately controlled during crystal growth. Precise diameter control is critical for semiconductor manufacturing given the automated wafer-handling steps in wafer fabs. Figure 4.20 illustrates the diameter-grinding process.

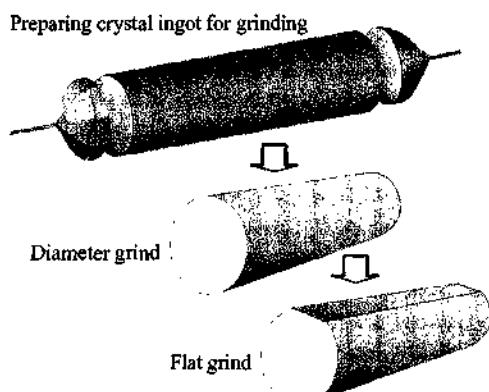


FIGURE 4.20 Ingot Diameter Grind

Wafer Flat or Notch ■ The semiconductor industry has traditionally placed *flats* on the ingot to identify the crystal structure and to orient the wafer. The primary flat orients the wafer to the crystal structure, as shown in Figure 4.21 on page 82. There is a secondary flat which identifies the orientation and conductivity type of the wafer.

Wafer flats have been replaced in the U.S. with a *notch* for wafers 200 mm and larger. Notched wafers have information about the wafer laser-scribed in a small region on the wafer. The location and depth of the laser-scribing initially created some fear of contamination collecting in the small laser marks. For 300-mm wafers, a standard has been approved for laser-scribing all wafers on the back surface in an exclusion zone near the wafer's edge.²⁵ The notch and laser scribe are illustrated in Figure 4.22 on page 82.

For 300-mm wafers, the exclusion zone is outside an area known as the fixed-quality area (FQA), which is where chips are permissible on the wafer surface. The exclusion zone is currently 3 mm, but may be reduced to 2 mm at some time in the future.

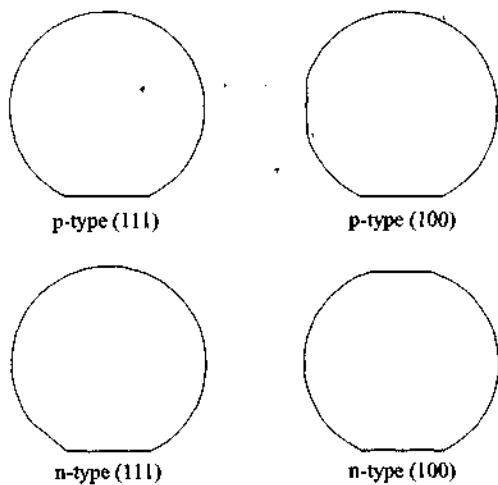


FIGURE 4.21 Wafer Identifying Flats

Wafer Slicing

Once the shaping operations are complete, the ingot is ready to undergo *wafer slicing*. This is the first major process step after ingot growth. For wafers up to 200 mm in diameter, slicing is done with an internal diameter saw that has a diamond cutting edge, as seen in Figure 4.23. The internal diameter saw is used because the cutting edge is more stable to yield a flat cutting surface.

For 300-mm wafers, internal diameter saws are not desirable for slicing wafers due to the larger diameter. Wafer slicing for 300-mm ingots is currently being done with wire saws.²⁶ Wire saws yield more wafer slices per inch of crystal than conventional ID saws due to thinner kerf (saw blade thickness) losses associated with using a slurry-coated wire instead of a diamond-coated blade.²⁷ Wafer sawing reduces mechanical damage to the wafer surface during the slicing process, but there are still concerns about inadequate wafer-flatness control with sawing.²⁸ Equipment development is ongoing for wafer slicing of 300-mm and larger wafers to obtain precise wafer dimensions.

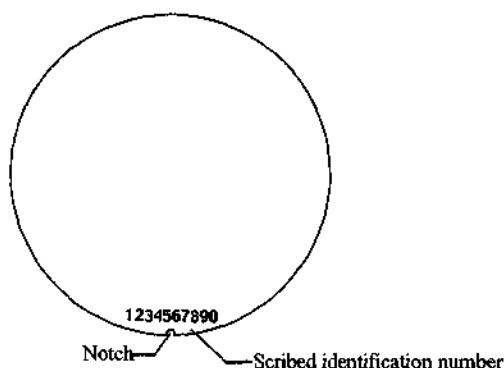


FIGURE 4.22 Wafer Notch and Laser Scribe

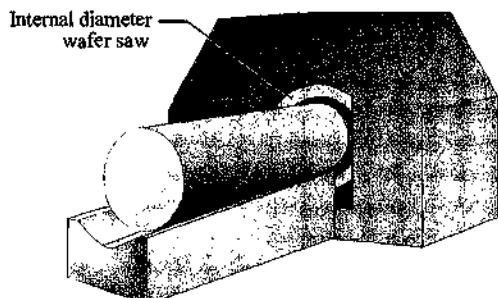


FIGURE 4.23 Internal Diameter Saw

The thickness of the wafer is carefully controlled during the slicing process. The 300-mm wafer is currently specified at a thickness of 775 ± 25 microns. Thicker wafers are better able to withstand thermal and mechanical shocks introduced during high temperature processing in the semiconductor fabrication process.

Wafer Lapping and Edge Contour

Following wafer slice, there is traditionally a mechanical, two-sided *lapping* operation to remove damage left by slicing and to achieve a high degree of parallelism and flatness of the wafer. Lapping is performed under rotational pressure with pads and an abrasive slurry mixture that typically consists

of alumina or silicon carbide and glycerine. Flatness is a critical wafer parameter for many process steps in the wafer fabrication process.

A *polished wafer edge* finish (also called *edge grind*) is applied to the wafer to contour a smooth radius on the edge of the wafer (see Figure 4.24). This step may take place either before or after lapping. Cracks and small crevices at the edge of the wafer create mechanical stress in the wafer that activates crystal dislocations, especially during thermal process steps that occur in wafer fabrication. Crevices can also be the source of unwanted contamination buildup as well as particulate flaking during fabrication. A smooth edge radius is important to minimize these concerns.²⁹ Furthermore, chipped edges are a source of edge dislocation growth during thermal cycles in the wafer fabrication process.

Etching

Shaping the wafer leaves the surface and edges damaged and contaminated. The depth of the wafer damage depends on the particular process at a manufacturer, but it is usually on the order of several microns deep. To remove the damaged surface from the wafer, wafer suppliers use a technique known as wafer etch or chemical etching. *Wafer etch* is a process involving the chemical removal of selective surfaces of a material (see Figure 4.25). Wafers undergo a wet chemical etch process to remove the damage and contamination on the wafer. Typically about 20 microns of silicon wafer surface is removed during the etch process to ensure that all damage is removed.³⁰ The etching is done with either acid or alkaline chemicals, depending on the process in place at the wafer manufacturer. The subject of etching will be covered in detail in Chapter 16.

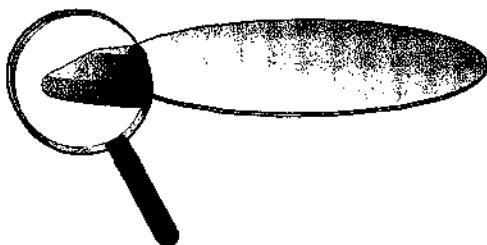


FIGURE 4.24 Polished Wafer Edge

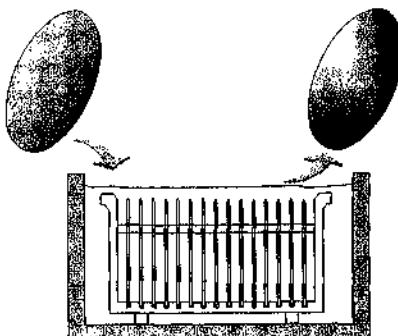


FIGURE 4.25 Chemical Etch of Wafer Surface to Remove Damage
Redrawn from M. S. Bawa, E. F. Petro, H. m. Grimes, "Fracture Strength of Large Diameter Silicon Wafers," *Semiconductor International* (November 1995), p. 115.

Polishing

The final step in the wafer preparation is a *chemical mechanical planarization (CMP)* to achieve a smooth wafer surface with a high degree of flatness. CMP is also referred to as polish (see Chapter 18 for a discussion of CMP). This CMP step has traditionally been done only on the topside of the wafer for 200-mm and earlier wafers, leaving the chemically-etched surface on the backside. This etched backside left a relatively rough surface, being approximately three times rougher than a surface processed with CMP.³¹ Its purpose was to give a rougher surface for handling devices. However, there has been concern about the ability of an etched surface to meet the flatness requirements for deep submicron photolithography, plus the possibility of introducing particulate contamination into the wafer fabrication process.

For 300-mm wafers, double-sided polishing (DSP) with CMP is the final major manufacturing step. The planetary motion of the wafer between the polishing plates produces flat and parallel surfaces while improving surface roughness (see Figure 4.26 on page 84). Since this is the final wafer preparation step, flatness across the large wafer diameter can be readily maintained. The polished back surface also permits the wafer manufacturer to characterize wafer cleanliness prior to delivery to the wafer fab. The final wafer surface has a mirrorlike finish on both sides of the wafer.

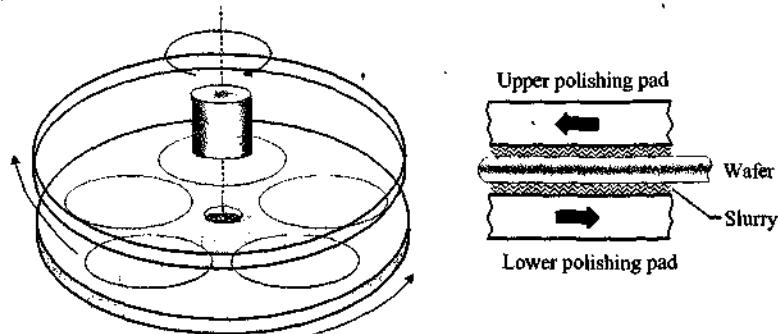


FIGURE 4.26 Double-Sided Wafer Polish

Cleaning

Semiconductor wafers must be cleaned to achieve an ultraclean state of cleanliness prior to shipment to wafer fabrication facilities. Cleaning specifications have undergone considerable development in the last few years, to the extent that wafers should be free of particles and contamination. The different types of wafer contamination, the appropriate cleaning steps, and the process procedures for cleaning are discussed in Chapter 6.

Wafer Evaluation

Prior to packaging the wafer, it is inspected for quality criteria that are specified by the customer. Standard quality measures are discussed later in this chapter. The most critical criteria relate to surface defects such as particulate contamination and stains.

Packaging

Wafer suppliers must carefully package wafers for shipment to the wafer fabrication facilities. Considerable effort can be lost if the wafers are damaged from transit or from the packaging material. The wafers are stacked in a plastic cassette or "boat" that has grooved slots to hold the wafers. A fluorocarbon resin material such as Teflon is often used as the cassette material to minimize particle generation. In addition, the Teflon is made conductive so that it does not generate an electrostatic discharge. All equipment and operators are grounded to drain off any charge buildup that could attract particles.

Once filled with wafers, the cassette is then placed in a nitrogen-filled, sealed canister that prevents oxidation and other contamination during transit. When the wafers are received at the wafer fab, they are transferred to other standardized cassettes used for transport and handling during wafer processing inside the fabrication facility. The shipping container is designed to minimize the need for wafer handling. One shipping concept under development is a 25-wafer container referred to as a front-opening shipping box (FOSB) which would interface with automated material handling systems in the wafer fab.³²

QUALITY MEASURES

Silicon wafer suppliers produce wafers to requirements that are becoming more stringent, which represents improved control over wafer quality. A sampling of some wafer specifications and their rate of improvement relative to the critical dimension on a wafer is shown in Table 4.5.

For wafer quality measures, uniformity throughout the wafer is critical. Important silicon wafer quality requirements are:

- ◆ Physical dimensions
- ◆ Flatness
- ◆ Microroughness
- ◆ Oxygen content
- ◆ Crystal defects
- ◆ Particles
- ◆ Bulk resistivity

The suppliers of silicon wafers must control their wafer quality by performing ingot and wafer quality inspections to demonstrate that quality specifications are met. Defective wafers shipped unknowingly to a semiconductor manufacturing wafer fab facility would be catastrophic.

TABLE 4.5 Improving Silicon Wafer Requirements

	Year (Critical Dimension)			
	1995 (0.35 μm)	1998 (0.25 μm)	2000 (0.18 μm)	2004 (0.13 μm)
Wafer diameter (mm)	200	200	300	300
Site flatness ^A (μm)	0.23	0.17	0.12	0.08
Site size (mm \times mm)	(22 \times 22)	(26 \times 32)	26 \times 32	26 \times 36
Microroughness ^B of front surface (RMS) ^C (nm)	0.2	0.15	0.1	0.1
Oxygen content (ppm) ^D	$\leq 24 \pm 2$	$\leq 23 \pm 2$	$\leq 23 \pm 1.5$	$\leq 22 \pm 1.5$
Bulk microdefects ^E (defects/ cm^2)	≤ 5000	≤ 1000	≤ 500	≤ 100
Particles per unit area (#/ cm^2)	0.17	0.13	0.075	0.055
Epilayer ^F thickness (\pm % uniformity) (μm)	3.0 ($\pm 5\%$)	2.0 ($\pm 3\%$)	1.4 ($\pm 2\%$)	1.0 ($\pm 2\%$)

Adapted from K.-M. Kim, "Bigger and Better CZ Silicon Crystals," *Solid State Technology* (November 1996), p. 71.

Notes:

A: Flatness is the linear thickness variation across the wafer or a site on a wafer (see the following).

B: See the following for a description of microroughness.

C: RMS is a method for determining the best estimate of group of measurements—in this case, the surface finish measurements (see the following). It is calculated by taking the root-mean-square (square root of the average of all measurements squared). Surface finish measurements are obtained by measuring the highest point relative to the lowest point on a surface.

D: ppm is part per million.

E: Bulk microdefects represent all defects within a square centimeter.

F: See the following section for a description of epilayer.

Physical Dimensions

Wafers must be physically dimensioned in order to meet the requirements for device fabrication and to accommodate the automated wafer handling equipment used in the wafer fab. The wafer physical dimensions controlled and inspected during wafer preparation include measurements such as diameter, thickness, location and size of orientation, flat (or notch), and wafer deformation. Figure 4.27 shows an example of wafer deformation. The most likely source of wafer deformation is the slicing process.

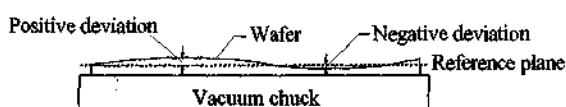


FIGURE 4.27 Wafer Deformation

Flatness

Flatness is one of the most critical wafer parameters specified, primarily because the process of photolithography (Chapters 13 to 15) is highly sensitive to local-site flatness. *Wafer flatness* is the linear thickness variation across the wafer. It is measured by determining the distance that the front surface of the wafer deviates from a specified reference plane. For a wafer, the reference plane is theoretically defined by the back wafer surface if it were held perfectly flat, such as when pulled down by a vacuum onto a clean, flat surface (see Figure 4.28). Note that the front surface of the wafer is the wafer side where chips will be placed.

Flatness can be specified as site flatness for a specific site on the wafer or specified as global flatness, which is the total wafer flatness across the fixed-quality area (FQA) of the wafer surface. The FQA does not include the exclusion zone on the periphery of the wafer surface. A larger area for measuring flatness is more difficult to control than a smaller area. Typical flatness specifications for site flatness are shown in Table 4.5 on page 85.

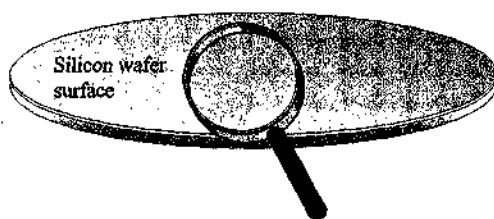


FIGURE 4.28 Flatness of Wafer Front Surface

Microroughness

Microroughness is the small-scale deviation of the actual surface from a nominal plane surface, with many small, closely spaced peaks and valleys.³³ It is an indication of the surface texture on the wafer. Surface microroughness measures the maximum and minimum height deviation on the wafer surface and is specified in units of nanometers (10^{-9} meters). The roughness criteria is usually specified as a root-mean-square (RMS), which is the square root of the average of the square of all deviation measurements from a defined plane. This is a general statistical technique for determining the most probable measurement. Measurement of wafer surface microroughness is done using one of several types of optical surface profile instruments.

Surface microroughness is important to control for wafer fabrication because it is known to have a negative effect leading to the breakdown of very thin dielectric films that are placed on the wafer during device fabrication.³⁴ Wafers undergo a lapping step followed by an etching step to remove surface microroughness. Typical specification values for microroughness are shown in Table 4.5 on page 85.

Oxygen Content

Control of the oxygen level and uniformity in the silicon ingot is important and becoming increasingly difficult with larger crystal diameters. Small amounts of oxygen can have beneficial gettering effects, which can tie up contaminants in the wafer, as noted previously in this chapter. However, excessive oxygen in the silicon ingot can affect the mechanical and electrical properties of the silicon. For instance, oxygen can lead to an increase in the leakage current at the pn junction and also to increased MOS leakage.³⁵

Oxygen content in silicon is verified by using cross-sectioning, which permits composition analysis of the silicon crystal structure. A representative piece of the silicon is potted in epoxy and then ground and polished flat to expose the grain structure of the solid silicon. A chemical etchant is used to darken and highlight certain elements for identification. Once the sample is prepared, special microscopes such as transmission electron microscopy (TEM) are used to characterize the crystal structure. The use of TEM is discussed in Chapter 7. The oxygen content is currently controlled in silicon wafers within a range of 24 to 33 ppm.³⁶ Typical specifications are shown in Table 4.5 on page 85.

Crystal Defects

It is necessary to control the silicon in order to minimize the various crystal defects previously discussed in this chapter. As shown in Table 4.5 on page 85, the current requirement is for the number of crystal defects per square centimeter to be less than 1000. Cross-sectioning is the technique used to control bulk microdefects in the crystal.

Particles

The number of surface particles are controlled on a wafer to minimize yield loss during wafer fabrication. The primary means of reducing particles is to minimize the generation of particles during wafer processing, followed by effective cleaning steps to remove particles. A typical wafer-cleanliness specification is to have less than 0.13 particles per square centimeter (1300 particles per square meter) of wafer surface area on a 200-mm wafer. This particle size measured is greater than or equal to 0.08 micron.

Bulk Resistivity

The resistivity of the bulk silicon ingot depends on the density of the dopant material that was added to the silicon melt prior to crystal growth. Recall that the most common dopant materials are either boron to create p-type silicon or phosphorus to make n-type silicon. The result of adding these Group IIIA or VA dopants to silicon is to decrease its resistivity due to increased carrier mobility.

It is important to achieve uniform resistivity throughout the bulk silicon. During the actual crystal growth process, there is a radial temperature gradient that reaches a maximum at the center of the ingot and decreases toward the outer edge. This radial temperature gradient produces variations in the radial doping concentration of the ingot.³⁷

The bulk silicon ingot is checked for correct resistivity and uniformity after the ends of the ingot have been removed. The four-point probe measurement tool is used to measure wafer resistivity. This measurement tool is described in Chapter 7.

EPITAXIAL LAYER

In some cases, it is desirable to have wafers with a very pure silicon surface of the same crystal structure as the substrate wafer (monocrystalline), yet retain specific control over the doping type or concentration. This condition is achieved by depositing an *epitaxial layer* on the surface of the silicon (referred to as an *epilayer*). Epitaxial is a combination of two Greek words, *epi*, meaning "upon", and *taxis*, meaning "ordered."

In epitaxial silicon, the base wafer is used as a seed crystal to grow a thin layer of silicon on the wafer. The crystal structure of the new epitaxial layer will duplicate that of the wafer. Since the substrate wafer is monocrystal, the epitaxial layer is monocrystal. Furthermore, the dopant of the epitaxial layer can be n-type or p-type and is independent of the initial wafer's dopant type. For example, it is possible to grow a p-type epilayer on a p-type wafer with a lower concentration of electrically active dopant in the epilayer than what is found on the wafer (see Figure 4.29).

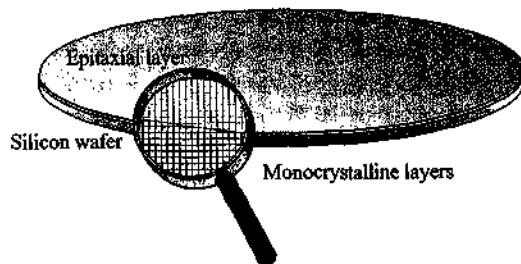


FIGURE 4.29 Formation of Epitaxial Silicon Layers

The original reason for the development of silicon epitaxy was to attain improved performance in bipolar transistors and integrated circuits. Epitaxy permitted the growth of a lightly doped epilayer over a heavily doped substrate, which optimized for a high breakdown voltage in the pn junction while permitting low collector resistance, yielding high device operating speeds at moderate currents. Epitaxy has become important in CMOS ICs because it minimizes latchup problems as device geometries continue to shrink. An epitaxial layer is generally free of contaminants (such as no oxygen particles, which is not true for CZ-grown silicon).

The thickness of the epitaxial layer can vary, with a typical thickness of 0.5 to 5 μm for high speed digital circuit applications, and an epilayer as thick as 50 to 100 μm for silicon power devices.³⁸ The methods for depositing an epitaxial layer on silicon will be discussed under the topic of deposition in Chapter 11.

SUMMARY

Naturally occurring silicon is used to produce ultra-pure, semiconductor-grade polycrystalline silicon. Silicon is a crystal with a repetitive FCC diamond unit cell structure at the atomic level. The crystal orientation is defined by Miller indices, with (100) most common for MOS devices. Polycrystalline silicon is converted into the monocrystal (single-crystal) ingot necessary for wafer fabrication through crystal growth using the CZ method and equipment referred to as a crystal puller. Dopant material is added to the liquid silicon during the CZ process to achieve proper doping levels. Undesirable impurities are tightly controlled in order to grow pure silicon. An alternative process, the float-zone method, produces silicon with very low oxygen content.

Ingot diameters have been increasing over the years to place more chips on a wafer and lower cost through economies of scale. The major crystal defects controlled during growth are point defects, dislocations, and gross defects. Ingots undergo a multistep process to make acceptable wafers, including diameter grinding, wafer notch, wafer slicing, wafer lapping, edge contour, etching, polishing, cleaning, inspection, and packaging.

Wafer suppliers have specific quality measures to control the wafer quality, for characteristics such as physical dimension, flatness, microroughness, oxygen content, crystal defects, particles, and bulk resistivity. An epitaxial layer is sometimes grown on the silicon surface to achieve a very pure silicon layer with the same crystal structure yet have precise control of resistivity by the controlled addition of dopants.

KEY TERMS

- | | |
|-----------------------------|--------------------------------|
| semiconductor-grade silicon | float-zone method |
| crystal | economies of scale |
| crystal lattice | crystal defect |
| amorphous materials | microdefect |
| unit cell | yield defect density |
| face-centered cubic (FCC) | point defect |
| FCC diamond structure | vacancy (or void) |
| polycrystal | interstitial |
| polysilicon | Frenkel defect |
| monocrystal | substitutional impurities |
| single crystal | dislocation |
| Miller indices | stacking faults |
| crystal growth | oxygen-induced stacking faults |
| ingot | gross defect |
| Czochralski (CZ) method | slip |
| CZ crystal puller | twin planes (twinning) |
| melt | boule |
| magnetic CZ (MCZ) | seed end |
| getter | tang end |

diameter grinding

flats

notch

fixed-quality area (FQA)

wafer slicing

lapping

polished wafer edge (edge grind)

wafer etch

chemical mechanical planarization (CMP)

double-sided polishing (DSP)

wafer flatness

microroughness

epitaxial layer (epilayer)

REVIEW QUESTIONS

1. List the three steps to obtaining semiconductor-grade silicon. How pure is semiconductor-grade silicon?
2. Name the process for producing pure semiconductor-grade silicon?
3. What is a crystal? What is a crystal lattice?
4. Describe an amorphous material. Why is this unacceptable for the silicon used in wafers?
5. Define a unit cell. What type of a unit cell does silicon form?
6. How many complete atoms are in the silicon unit cell? How many atoms are shared and how many are unshared in the silicon unit cell?
7. Describe a polycrystal. What is another name for polycrystalline silicon?
8. Describe a monocrystal. What is another name for a monocrystal?
9. Why is it necessary to have monocrystal silicon for wafer fabrication?
10. What are Miller indices, and what do they indicate?
11. Draw a picture of the following three crystal planes: (100), (110) and (111).
12. Which crystal plane orientation is most common for MOS? Which is most common for bipolar?
13. Define crystal growth.
14. What is the CZ method for crystal growth?
15. Describe the silicon seed for the CZ method and how it is used.
16. What is the name of the main equipment used in the CZ method?
17. Describe a crystal puller. Specifically, describe the crucible holder, seed, melt, pull process and ingot growth.
18. What is the objective of the pull process and what are the two main parameters that affect the pull process in the CZ method?
19. What is the purpose of magnetic CZ?
20. Why is dopant material added to the melt during the CZ method?
21. How can small amounts of oxygen in the ingot be beneficial?
22. Why is the float-zone method desirable to grow silicon crystal?
23. Describe the float-zone process.
24. What is the main reason for the continued increase in wafer diameter?
25. Give three benefits to using larger-diameter wafers.
26. What is a crystal defect?
27. Define defect density. How is defect density related to wafer yield?
28. What are the three general forms of crystal defects in silicon?
29. List and describe three types of point defects.
30. Which crystal growth parameters can affect the occurrence of point defects?
31. What is a dislocation? Explain a stacking fault.
32. What is an oxidation-induced stacking fault?
33. Explain how crystal growth dislocations have been reduced for large-diameter wafers.
34. What is a gross defect? Explain the difference between crystal slip and crystal twin planes.
35. What is the purpose of diameter grinding of the ingot?
36. Describe or draw a picture of the four types of wafer flats. What have the wafer flats been replaced with on 200-mm diameter and larger wafers?
37. How is wafer slicing done on wafers up to 200 mm in diameter?
38. What is the purpose of wafer lapping?
39. Why is a polished wafer edge (edge grind) done to the wafer? Give three reasons why this grinding benefits the wafer quality.
40. Why is chemical etching done on the wafer?
41. Why is chemical mechanical planarization (CMP) done on the wafer surface?
42. How are wafers packaged for shipment to the wafer fab?
43. List seven wafer quality requirements for a silicon wafer.
44. What is wafer flatness, and how is it measured?
45. What is surface microroughness, and why is it important to control?
46. What is an epitaxial layer, and why is it used on wafers?

SELECTED INDUSTRY WEB SITES

GT Equipment Technologies, Inc.
 International SEMATECH
 Kayex CZ Crystal Growers
 Mitsubishi
 NIST, National Institute of Standards and Technology
 Selete, Semiconductor Leading Edge Technologies, Inc.
 SEMI, Semiconductor Equipment and Materials International
 Semiconductor International Magazine
 Semiconductor Search Engine
 SIA, Semiconductor Industry Association
 SISA, Semiconductor Industry Suppliers Association
 Solid State Technology Magazine
 Wafer World, Inc.

<http://www.gtequipment.com/>
<http://www.sematech.org/public/index.htm>
<http://www.kayex.com/>
<http://www.munc-sil.com/>
<http://www.nist.gov/>
<http://www.selecte.co.jp>
<http://www.semi.org/>
<http://semiconductor.net/>
<http://www.semiseek.com/>
<http://www.semichips.org/>
<http://www.sisa.org/>
<http://sst.pennwellnet.com/home/home.cfm>
<http://www.waferworld.com/>

REFERENCES

1. C. Pearce, *Crystal Growth and Wafer Preparation*, VLSI Technology, 2nd ed., ed. S. Sze (Boston: McGraw-Hill, 1988), p. 11.
2. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: Wiley, 1994), p. 102.
3. K. Bachmann, *The Materials Science of Microelectronics* (New York: VCH Publishers, 1995), p. 102.
4. Society of Chemical Engineers of Japan, eds., *Introduction to VLSI Process Engineering* (New York: Chapman and Hall, 1993), p. 148.
5. L. Van Vlack, *Elements of Materials Science and Engineering*, 4th ed., (Reading: Addison-Wesley, 1980), p. 73.
6. S. Ghandhi, *VLSI Fabrication Principles*, p. 16.
7. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1 of *Process Technology* (Sunset Beach: Lattice Press, 1986), p. 2.
8. S. Ghandhi, *VLSI Fabrication Principles*, p. 21.
9. Ibid.
10. S. Takasu and W. Zulehner, "Silicon Crystal Growth," *Semiconductor Silicon Proceedings 94-10* (Pennington: The Electrochemical Society, 1994): p. 55.
11. K.-M. Kim, "Bigger and Better CZ Silicon Crystals," *Solid State Technology* (November 1996), p. 73.
12. K. Bachmann, *Materials Science of Microelectronics*, p. 248.
13. S. Sze, *VLSI Technology*, 2nd ed. (New York: McGraw-Hill, 1988), p. 26.
14. K. Bachmann, *Materials Science of Microelectronics*, p. 270.
15. S. Ghandhi, *VLSI Fabrication Principles*, p. 120.
16. K. Bachmann, *Materials Science of Microelectronics*, p. 248.
17. K.-M. Kim, "Bigger and Better CZ Silicon Crystals," p. 74.
18. Ibid., p. 74.
19. H. Wolfgang, "High Purity Silicon IV," *International Symposium*, (Pennington, NJ: The Electrochemical Society, 1996): p. 5.
20. H. Huff et al., "Thermal Processing Issues for 300 mm Silicon Wafers: Challenges and Opportunities," *ULSI Science and Technology: Proceedings of the Electrochemical Society*, 97-3: p. 136.
21. Ibid., p. 136.
22. K.-M. Kim, "Bigger and Better CZ Silicon Crystals," p. 74.
23. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, (Boston: Kluwer Academic Publishers, 1995), p. 68.
24. A. Takao, "Innovative Silicon Crystal Growth and Wafering Technologies," *ULSI Science and Technology: Proceedings of the Electrochemical Society*, 97-3: p. 123.
25. S. Brunkhorst and D. Sloat, "The Impact of the 300-mm Transition on Silicon Wafer Suppliers," *Solid State Technology* (January 1998): p. 90.
26. Ibid.
27. G. Fisher, "Challenge for 300mm Polished Wafer Manufacturers," *Semiconductor International* (September 1998): p. 98.
28. A. Takao, *Innovative Silicon Crystal Growth*, p. 124.
29. H. Huff, et al., "Thermal Processing Issues," p. 140.
30. S. Sze, *VLSI Technology*, p. 40.
31. S. Brunkhorst and D. Sloat, "Impact of the 300-mm Transition," p. 88.
32. G. Fisher, "Challenges," p. 102.
33. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., (New York: McGraw-Hill, 1997), p. 214.
34. W. Bullis, "Microroughness of Silicon Wafers," *Proceedings of Semiconductor Silicon 94-10* (Pennington, NJ: The Electrochemical Society, 1994): p. 1156.
35. J. Vanhellemont et al., "On the Electrical Activity of Oxygen-Related Extended Defects in Silicon," *Semiconductor Silicon Proceedings 94-10* (Pennington, NJ: The Electrochemical Society, 1994): p. 670.
36. K.-M. Kim, "Bigger and Better CZ Silicon Crystals," p. 73.
37. S. Ghandhi, *VLSI Fabrication*, p. 258.
38. S. Ghandhi, *VLSI Fabrication*, p. 258.

CHEMICALS IN SEMICONDUCTOR FABRICATION

The fabrication of semiconductors is largely a chemical process. Chemical manufacturers obtain raw chemicals from nature, refine the chemical makeup of these materials to achieve ultrahigh purity, and transport them to the fabrication area, where these chemicals are introduced to the surface of the wafer. Chemical reactions alter the silicon to create semiconductors and then apply many layers of interconnections. These chemical reactions form the basis for semiconductor device fabrication and microchip performance.

The semiconductor industry is undergoing rapid changes to introduce fundamentally new materials and processes into wafer fabrication. The goal is to achieve improved microchip performance and productivity. Examples of new technologies are insulators with a low dielectric constant (see Chapter 11) and copper metallurgy (see Chapter 12). Chip design and process improvements create the ongoing need for new and improved chemicals.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Identify and discuss the four states of matter.
2. Describe the important chemical properties relevant to semiconductor manufacturing.
3. State how the different process chemicals are categorized and used in a wafer fab.

4. Explain how an acid, a base, and a solvent are used in chip manufacturing.
5. State whether a gas is a bulk or specialty gas and explain how each type of gas is delivered and used in wafer fabrication.

INTRODUCTION

Semiconductor manufacturing makes use of large quantities of chemicals to fabricate wafers. There is also a high usage of chemicals for cleaning wafers and the processing tools used in the manufacturing processes. Chemicals used in wafer fabrication are referred to as *process chemicals*. They exist in various chemical states and are controlled to strict purity requirements. To understand process chemicals, we will first review some concepts in basic chemistry.

STATES OF MATTER

All matter in the universe exists in three basic states: solid, liquid, or gas. There is an additional fourth state that is not widely understood, known as plasma. We can describe each state by the way the substance fills a container (see Figure 5.1 on page 92).

A *solid* has its own fixed shape and will not conform to the shape of the container. A *liquid* conforms to the container shape. It will fill the container to the extent of the liquid's volume, and has a surface. A *gas* also conforms to the container shape, but it fills the entire container volume.

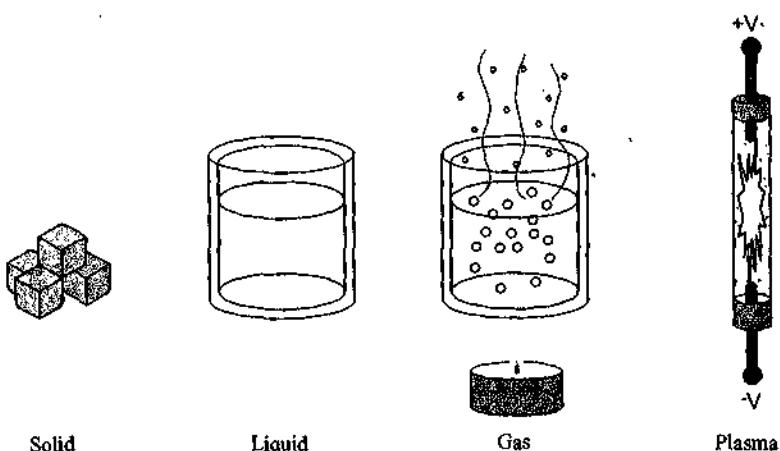


FIGURE 5.1 Physical States of Matter

without forming a surface. Gas particles are small and freely moving. Some gases, like hydrogen and oxygen, are reactive and easily form stable chemical compounds with other gases or elements. Other gases, like helium and argon, are inert. An *inert gas* (also called a *noble gas*) does not readily form compounds. Inert gases are widely used in semiconductor manufacturing because they do not react with other chemicals.

A fourth state of matter is *plasma*. A plasma state exists when there is a high-energy collection of ionized molecules or atoms. This state is considerably different than the three basic states of matter. Examples of a plasma state are a star, a fluorescent light, and neon signs. Plasma states can be induced in certain gases through exposure to high-energy electric fields. As we shall see in later chapters, plasma states are widely used in semiconductor manufacturing.

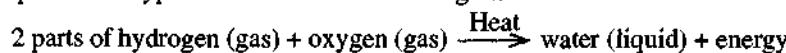
Many substances can exist in any one of the three basic physical states. For example, water undergoes changes of state depending on the temperature and pressure of its surroundings. As temperature increases, solid water (ice) melts to a liquid water, and then boils into a gas state (water vapor). Many substances behave in this same manner.

PROPERTIES OF MATERIALS

We can understand how to use materials to our benefit in semiconductor manufacturing by investigating their properties. *Properties* are the characteristics of a material that describe its unique identity. For instance, a solid material can have different properties, such as a rigid solid (iron), a soft solid (wax) at room temperature (25°C, 77°F) and a flexible solid (lead). Identifying properties for materials used in semiconductor manufacturing is important for understanding how to properly fabricate the silicon wafer to build chips.

There are two types of material properties: physical and chemical. *Physical properties* of a material are those that reflect a material by itself without its interacting with another substance. Some of these physical properties are the melting point, boiling point, resistivity, and density. *Chemical properties* of a material result from an interaction or transformation with another substance. Examples of chemical properties are flammability, reactivity, and corrosiveness. A *chemical reaction* occurs when a material is converted into a different material that has a different composition and properties. The starting material is called a *reactant*, and the resultant material is called a *product*.

An example of a chemical reaction is the combustion of hydrogen in the presence of oxygen. Both chemicals exist as gases in their natural state. Hydrogen will undergo a chemical reaction with oxygen when the temperature of hydrogen exceeds 600°C. The result is an explosive heat reaction with water vapor as the byproduct. The chemical change is:



Chemical Properties for Semiconductor Manufacturing

There are many different types of chemicals and materials used to fabricate silicon microchips. Manufacturers of advanced ICs use new materials to improve chip performance and reduce device feature sizes. Some chemical properties are important to understanding how existing and new semiconductor process materials perform. These properties are:

- ◆ Temperature
- ◆ Pressure and vacuum
- ◆ Condensation
- ◆ Vapor pressure
- ◆ Sublimation and deposition
- ◆ Density
- ◆ Surface tension
- ◆ Thermal expansion
- ◆ Stress

Temperature ■ *Temperature* is a measure of how hot or cold a substance is relative to another substance. Thus it is a measure of the average kinetic or thermal energy (molecular or atomic motion) of a substance. The transfer of energy between objects at different temperatures is *heat*. Wafer fabrication requires extensive processing at high temperatures for reasons such as using heat to affect chemical reactions (e.g., changing the rate of a reaction) or annealing the silicon crystal structure to rearrange atoms.

Three temperature scales exist: Fahrenheit in °F, Celsius in °C (commonly called centigrade), and Kelvin in K. Figure 5.2 illustrates how these three scales are related.

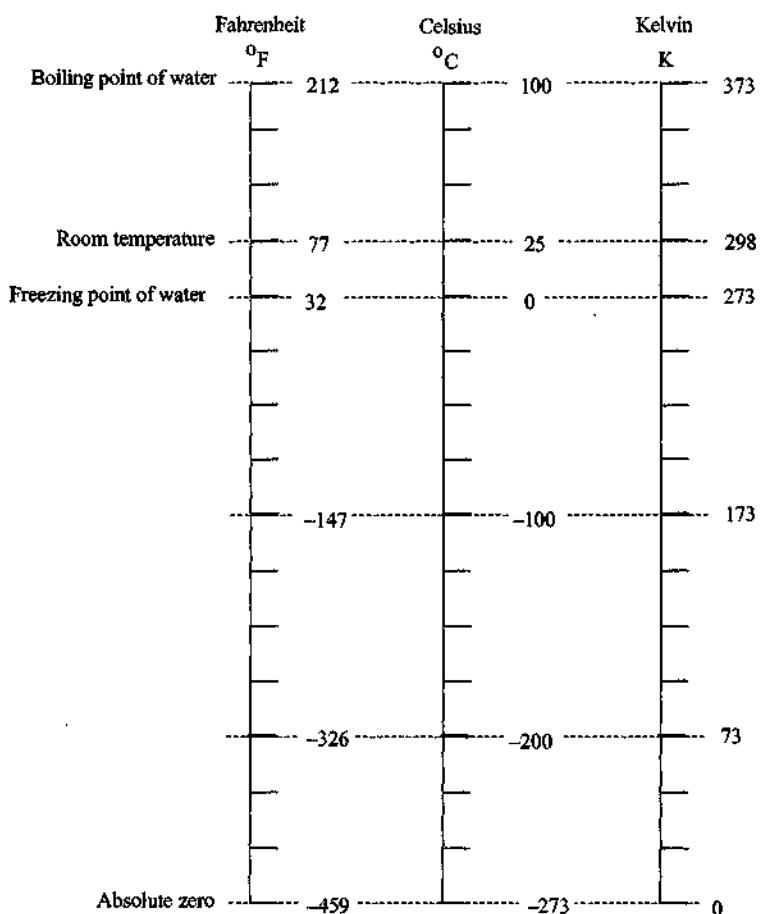


FIGURE 5.2 Temperature Scales

The most common temperature scale in scientific work is the *Celsius* scale. This scale is based on changes in the physical state of water: 0°C is set at the water's freezing point and 100°C at its boiling point when measured at standard atmospheric pressure at sea level. To convert between Fahrenheit and Celsius,

$$\begin{aligned}^{\circ}\text{F} &= \frac{9}{5}^{\circ}\text{C} + 32 \\^{\circ}\text{C} &= \frac{5}{9}[^{\circ}\text{F} - 32]\end{aligned}$$

The *Kelvin* scale, or absolute scale, is the base unit of temperature for the metric unit system. The temperature 0 K (absolute zero) is -273°C (more precisely, -273.15°C), which is the lowest temperature attainable and the theoretical temperature at which all atomic motion would cease. To convert between Kelvin and Celsius,

$$\begin{aligned}\text{K} &= ^{\circ}\text{C} + 273 \\^{\circ}\text{C} &= \text{K} - 273\end{aligned}$$

Pressure and Vacuum of a Gas ■ Gas fills the entire volume of a container and exerts uniform pressure on its walls. *Pressure* (P) is defined as the force exerted per unit area against a surface:

$$\text{Pressure} = \frac{\text{Force}}{\text{Area}} \text{ (pounds per square inch, or psi)}$$

The pressure of any gas is dependent on the number of gas molecules present, the temperature, and the volume of the chamber. If there are more gas molecules striking the sides of the container per unit time, then pressure increases (see Figure 5.3). This is also evident when blowing up a balloon.

The English unit of pressure common in the United States is pounds per square inch (psi), specified as either psia (absolute pressure, which includes atmospheric pressure of 14.7 psi) or psig (gauge pressure). These two gauges are illustrated in Figure 5.4. Different units of pressure at standard atmosphere and temperature (sea level and 23°C) are shown in Table 5.1.

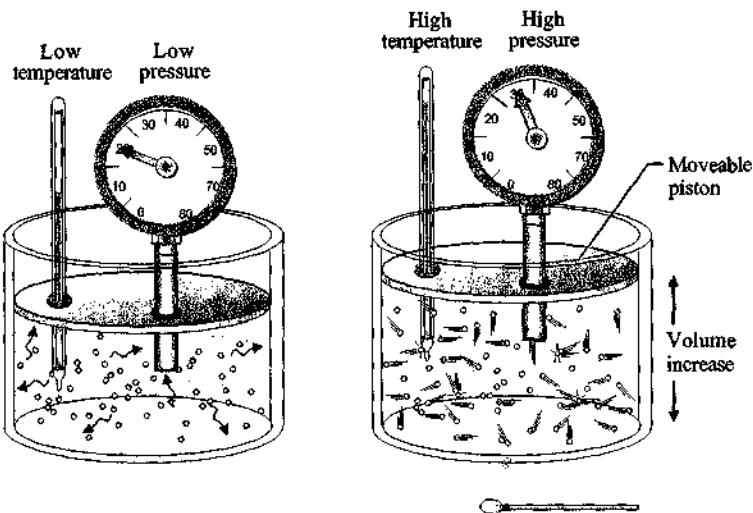


FIGURE 5.3 Pressure Against a Container Wall

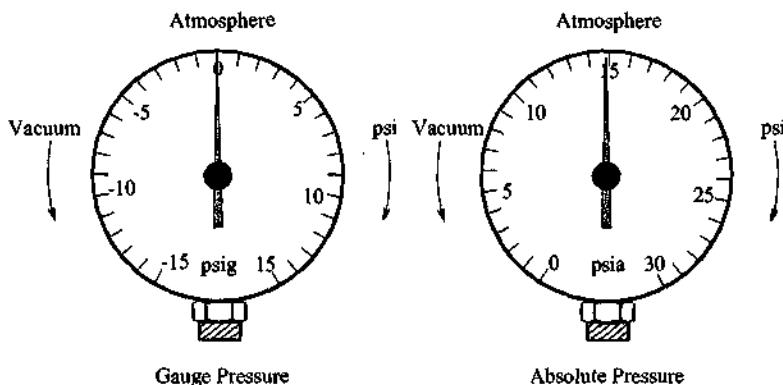


FIGURE 5.4 Gauge Pressure (PSIG) Versus Absolute Pressure (PSIA)

TABLE 5.1 Units of Pressure at Sea Level and 23°C

Unit of Pressure	Sea Level and 23°C
Psig (gauge)	0 psi
Psia (absolute)	14.7 psi
Atmosphere	14.7 psi
Inches of mercury	29.92 inches
Millimeters of mercury	760 mm
Torr	760 torr
mtorr	760,000 mtorr
Bar	1.013 bar
millibar	1013 mbar
Pascal	101,325 pascal

Pressure is a widely used property in semiconductor manufacturing. Chemicals and gases flow from regions of higher pressure to lower pressure. Some fabrication processes occur at atmospheric pressure, while some require higher pressures. Other processes occur in chambers at less than atmospheric pressure (vacuum).

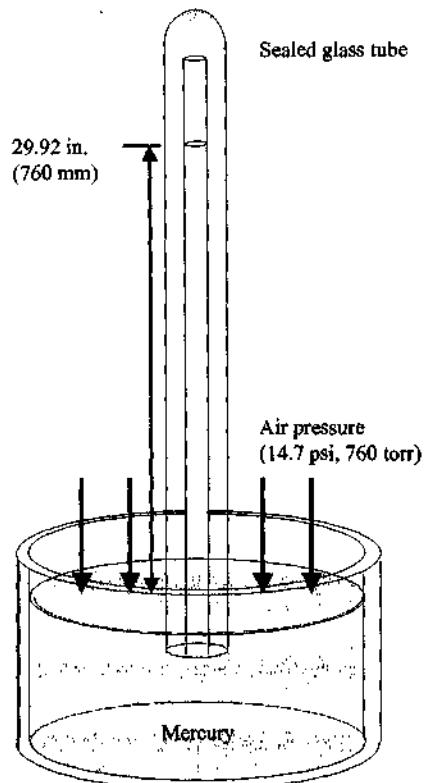
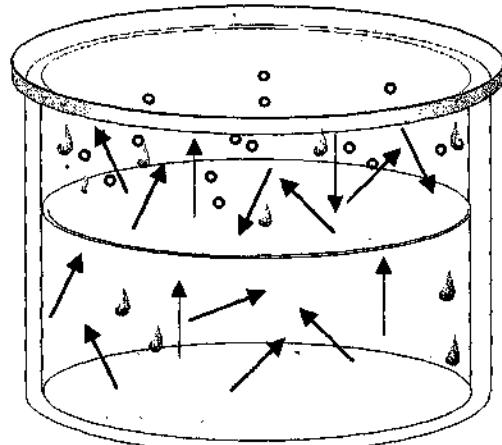
Vacuum. If the gas pressure in a container is less than 14.7 psi, then a vacuum exists. Vacuum is the removal of gas molecules (e.g., air, moisture, and gas residues) in a closed container to achieve a pressure less than atmosphere. Many semiconductor operations occur in vacuum conditions. Vacuum is commonly measured in units of torr. A torr is defined as the equivalent of one mm of mercury in a pressure-measuring device known as a barometer (see Figure 5.5 on page 96).

At one atmosphere, the weight of the air will push down on the mercury in the bowl and cause it to rise 760 mm (760 torr or 29.92 inches) in the vacuum of the column. The amount of mercury rise is proportional to the pressure bearing down on the mercury in the dish. Less pressure (or vacuum) makes the mercury rise less. For the sensitivity required to measure and control vacuum conditions in many semiconductor processes, the units of millitorr are required.

Condensation and Vaporization ■ The process of a gas changing into a liquid is called *condensation*. When water vapor cools, a mist appears as the particles form tiny droplets of liquid that then collect into a bulk sample with a single surface. The opposite process, changing from a liquid into a gas, is called *vaporization*.

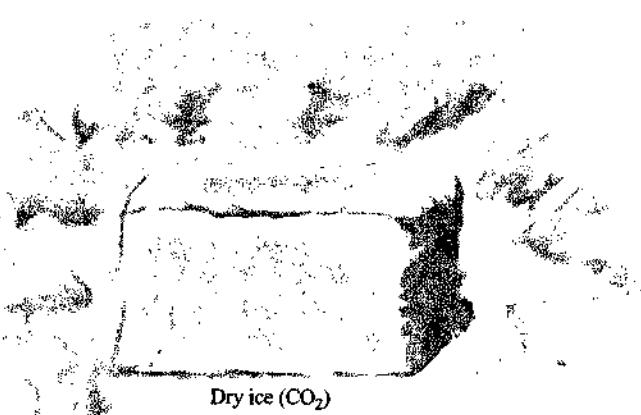
Liquids and gases interact with materials in different ways. *Absorption* is the taking up of a liquid or a gas into the bulk of another material, as in the dissolving of a gas in a liquid. *Adsorption* is the condensing of a gas or a liquid on the surface of a solid. The adsorbed molecules actually stick to the surface through a chemical bond or through a weaker bond of physical attraction.

Vapor Pressure ■ *Vapor pressure* is the pressure exerted by vapor in a closed container when there is an equilibrium condition between the rate of vaporization and rate of condensation. Vapor pressure is shown in Figure 5.6 on page 96.

**FIGURE 5.5** Barometer at Atmospheric Pressure**FIGURE 5.6** Vapor Pressure

High vapor pressure materials are *volatile* (tend to become a gas). A material with a high vapor pressure will release vapors more easily when exposed to vacuum conditions. Examples are solvents, perfumes, and lotions. These materials typically exhibit odors that are readily detectable through human nasal passages.

Sublimation and Deposition ■ A solid can change directly to a gas through a process termed *sublimation*. Have you ever put an ice tray in the freezer and left it there for six months? The ice cubes are smaller because the ice sublimes. Other examples of sublimation are dry ice (shown in Figure 5.7) and moth balls.

**FIGURE 5.7** Sublimation

The opposite process, changing from a gas into a solid, is called *deposition* (see Figure 5.8). This process is how ice forms on a cold window from the deposition of water vapor. As we shall see in a later chapter, deposition is an important process in semiconductor manufacturing.

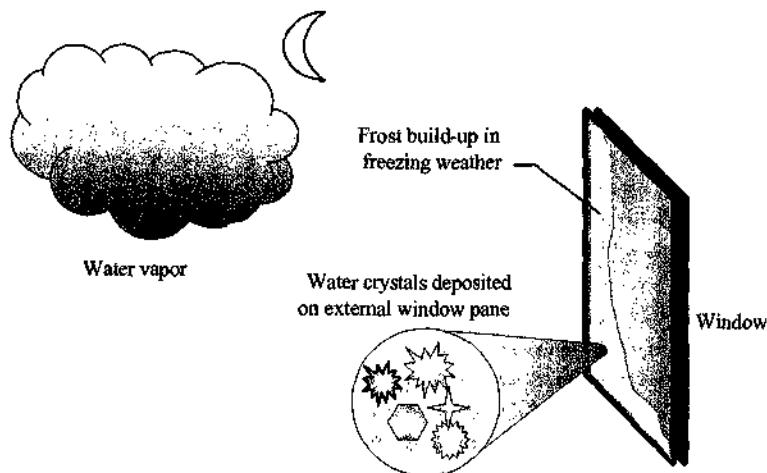


FIGURE 5.8 Deposition

Density ■ The *density* of a substance is defined as its mass (or weight) divided by its volume:

$$\text{Density} = \frac{\text{Mass}}{\text{Volume}} \text{ (grams/cm}^3\text{)}$$

A dense material is thought of as being heavy. If two objects have the same volume and one object is heavier than the other is, then it is denser (see Figure 5.9). A sponge has a lower density than an equal volume of steel. We know this because the sponge weighs less than the steel.

Table 5.2 on page 98 gives the density for different substances at standard temperature and pressure.¹ Water is the standard, with a density of 1 gram/cm³. The densities of other substances are often expressed as a ratio of their density to that of a comparable volume of water. For instance, silicon has a density of 2.3. Interpret this number to mean that a piece of silicon occupying a volume of one cm³ will weigh 2.3 grams.

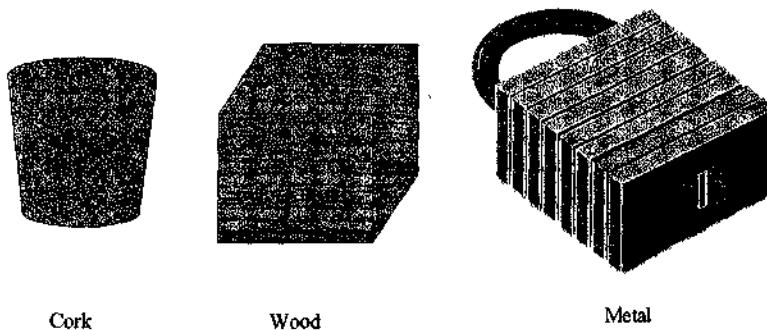


FIGURE 5.9 Density of Objects

TABLE 5.2 Densities of Some Common Substances at Standard Temperature and Pressure

Substance	Physical State	Density (g/cm ³)
Hydrogen	Gas	0.000089
Oxygen	Gas	0.0014
Water	Liquid	1.0
Table salt	Solid	2.16
Silicon	Solid	2.33
Aluminum	Solid	2.70
Gold	Solid	19.3

A term related to density is specific gravity. *Specific gravity (SG)* refers to the density of liquids and gases at 4°C and is the ratio of the density of a substance to the density of water.

Surface Tension ■ When a liquid is on a flat surface, it has a surface area of contact (see Figure 5.10). The *surface tension* of the liquid is the energy required to increase the surface area of contact. For the surface area to increase, liquid molecules must break intermolecular attraction forces and move away from the interior of the liquid and toward its surface. This movement requires energy. The concept of surface tension is used in semiconductor manufacturing to measure the ability of liquid coatings to uniformly adhere to the wafer surface.



FIGURE 5.10 Surface Tension of a Liquid on a Wafer

Thermal Expansion ■ When an object is heated, it expands due to the increased vibrations of the atoms. Due to this thermal expansion, the dimensions of a heated object will actually increase (see Figure 5.11). When the same object is cooled, its dimensions will decrease.

Some materials will expand thermally more than others. The amount a material expands due to heating is known as its *coefficient of thermal expansion*, or *CTE*. Amorphous materials that have a cubic crystal orientation expand to thermal exposure in all directions. For all other crystals, such as monocrystal silicon, thermal expansion varies with the type of crystal orientation.

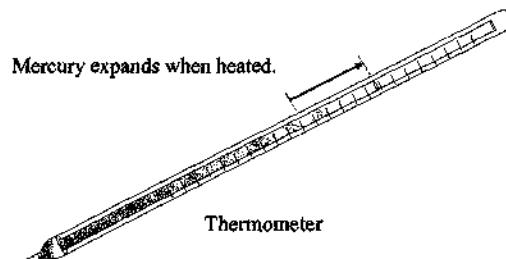


FIGURE 5.11 Thermal Expansion of a Heated Object

Stress ■ Stress occurs when an object is exposed to a force. The magnitude of the stress depends on two factors: the amount of force, and the area over which the force is applied. The units of stress are pounds per square inch (psi) or pascal (Pa) in the metric system. The formula for stress is,

$$\text{Stress} = \frac{\text{Force}}{\text{Area}} \text{ (psi)}$$

The force that creates stress in a wafer can come from various sources. Stress can come from work damage on the surface of the wafer; from internal forces due to dislocations, excess vacancies and impurities; and from growth around included foreign material.² If two dissimilar materials with significantly different coefficients of thermal expansion (CTE) are bonded together and heated, then

a force occurs because the two materials expand at different rates and pull against each other to cause a stress. This stress can create a warped wafer due to the CTE mismatch. Such stress is a concern in semiconductor manufacturing because a microchip is a planar structure with layers of different materials with different CTEs. Deposited films usually have some form of stress that can be compressive or tensile, and the nature of the stress depends on process conditions (see Figure 5.12). The reliability of a chip is improved by ensuring materials have minimal stress.

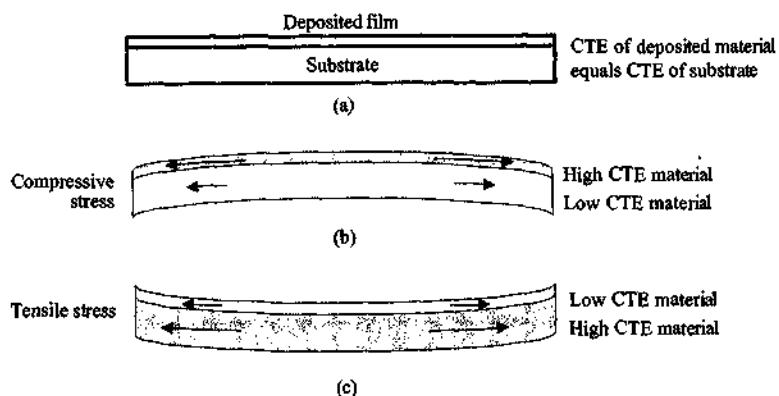


FIGURE 5.12 CTE Mismatch of Two Materials

PROCESS CHEMICALS

Semiconductor manufacturing is a chemically intensive process involving many different types of process chemicals of ultrahigh purity. The process chemicals are used in all three states: liquids, solids, and gases. Chemicals are used in the process of manufacturing semiconductors to:

- ◆ Clean or prepare the wafer surface with wet chemical solutions and ultrapure water rinse.
- ◆ Dope the wafer with energetic atoms to create p-type and n-type silicon.
- ◆ Deposit the different metal conductor layers and the necessary dielectric layers between the conductors.
- ◆ Grow the thin silicon dioxide film to be used as the critical MOS gate dielectric.
- ◆ Etch thin films with plasma or wet chemicals to selectively remove material and form the required pattern in the film.

There is a vast array of chemicals used in the fab, yet some predominant chemicals are repeatedly used in the different process areas of the fab. Cleaning is one of the most chemically intense steps of semiconductor manufacturing. It is estimated that 30% of all processing steps in the wafer fabrication process are for wafer cleaning or wafer surface preparation.³

Liquids

Liquids can be either a pure substance, such as pure water, or a chemical mixture. Automotive gasoline is an example of a chemical mixture of hydrocarbons and additives blended together to produce efficient combustion.

If a chemical mixture is so well mixed that the molecules or ions of the components are the same throughout, then we call this a *chemical solution*. Gasoline is a chemical solution. Another example of a solution is the household first aid antiseptic hydrogen peroxide, which consists mostly of water, with only 5% hydrogen peroxide. The component of the solution present in the larger amount (such as water in the hydrogen peroxide solution) is called the *solvent*. The dissolved substances in the solution are the *solutes*. Solutions in water are called *aqueous solutions*, meaning water is the solvent.

There are many types of liquids used in the wet processes of semiconductor manufacturing. All liquids used in wafer fabs must have extremely high purity, with no contamination from particulates, metallic ions, or unwanted chemicals. Contamination of chemicals is a relative

concept. A term that is frequently used to describe small concentrations of an impurity is *parts per million (ppm)*, which is measured by volume or by mass (weight). An example of ppm is if you consider the amount of an impurity in air. To calculate the ppm of an impurity by mass, you determine the weight of the impurity in a fixed volume of air, divide this number by the weight of the air, and multiply by one million. Chemical purity specifications have become more stringent so that ultra-pure chemicals used in the fab often have less than a *parts per billion (ppb)* or *parts per trillion (ppt)* of an impurity.⁴ The amount of impurity measured in a wafer fab process chemical is often limited by the measurement capability of the instrument.

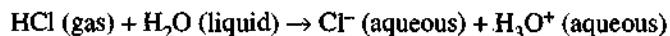
There has been a long-term effort to reduce the quantity of liquid chemicals in the wafer fab. Many liquid chemicals used in semiconductor manufacturing are hazardous and require special handling and disposal procedures. Furthermore, chemical residues can contaminate wafers, as well as produce vapors that diffuse through the air and deposit on wafer surfaces. Although it has not been possible to eliminate wet chemicals from the wafer fab, there has been a significant reduction in their use, such as in reducing the number of wet cleaning steps.⁵

Liquid process chemicals in a wafer fab are grouped in the following manner:

- ◆ Acid
- ◆ Base
- ◆ Solvent

Acid ■ In the classical definition, *acids* are solutions that contain hydrogen and dissociate in water (meaning bonds break down) to yield hydronium ions, H_3O^+ . Since an acid contains hydrogen, it contains H in its formula, such as phosphoric acid (H_3PO_4) or hydrochloric acid (HCl).

As an example of how acid dissociates in water, consider the following reaction for hydrochloric acid:



We can see that when HCl is dissolved in water, it reacts with the water to produce the hydronium ion, H_3O^+ . This defines HCl as an acid.

There are many different acids used in semiconductor manufacturing. Table 5.3 lists some common acids and their typical use in a wafer fab.

TABLE 5.3 Common Acids Used in Semiconductor Manufacturing

Acid	Symbol	Example of Use*
Hydrofluoric acid	HF	Etching silicon dioxide (SiO_2) and cleaning quartzware
Hydrochloric acid	HCl	Wet cleaning chemical that is part of the standard clean 2 (SC-2) solution to remove heavy metals from wafer
Sulfuric acid	H_2SO_4	Solution known as "Piranha" [7 parts H_2SO_4 to 3 parts of 30% hydrogen peroxide (H_2O_2)] used to clean wafers
Buffered oxide etch (BOE): Solution of hydrofluoric acid and ammonium fluoride	HF and NH_4F	Etching of silicon dioxide (SiO_2) film
Phosphoric acid	H_3PO_4	Etching of silicon nitride (Si_3N_4)
Nitric acid	HNO_3	Used in mixture of HF and HNO_3 to etch phosphosilicate glass (PSG)

*Explanation of wafer cleaning is in Chapter 6 and etching in Chapter 16.

Acids are further divided into two categories: organic and inorganic. Organic acids such as carboxylic acids contain hydrocarbons, while inorganic acids such as hydrofluoric acid (HF) do not.

Base ■ A *base* is a substance that contains the OH chemical group (e.g., NaOH, sodium hydroxide and KOH, potassium hydroxide) and dissociates in water to yield the hydroxide ion, OH^- . Another name for a base substance is *alkali* or an *alkaline substance*. A base will increase the

hydroxide ion in an aqueous solution. For example, sodium hydroxide is an ionic compound containing metal ions and hydroxide ions. It is a base because when it dissolves in water it dissociates to yield Na^+ and OH^- ions:



Common alkaline substances used in semiconductor manufacturing are shown in Table 5.4.

TABLE 5.4 Common Bases Used in Semiconductor Manufacturing

Base	Symbol	Example of Use
Sodium hydroxide	NaOH	Wet etchant
Ammonium hydroxide	NH_4OH	Cleaning solution
Potassium hydroxide	KOH	Positive photoresist developer
Tetramethyl ammonium hydroxide	TMAH	Positive photoresist developer

pH ■ An acid or a base varies in strength and is classified as strong or weak. The *pH scale* is used to assess how strong or weak a solution is as an acid or base. This scale ranges from 0 to 14, with 7 being the neutral point. Acids have pH values below 7 and bases have pH values above 7. Pure water is the reference substance for the pH scale and is neutral with a pH of 7. Strong acids, such as sulfuric acid (H_2SO_4) have low pH values between 0 and 3. Strong bases, such as sodium hydroxide (NaOH) have pH values greater than 7 and approaching 14. Figure 5.13 shows some common chemicals and their location on the pH scale.

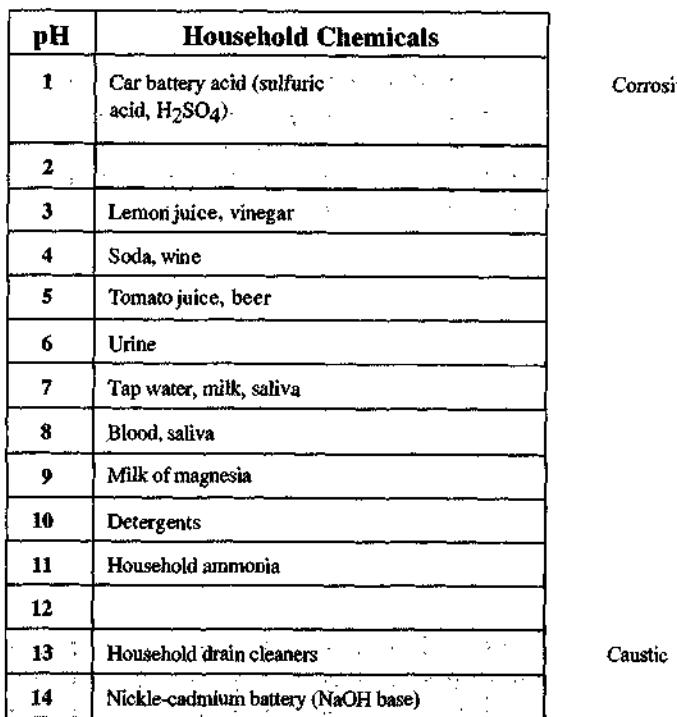


FIGURE 5.13 The pH Scale for Different Chemicals

Solvents ■ A *solvent* is a substance capable of dissolving another substance to form a solution. A good solvent will dissolve (solubilize) a broad range of substances. Most solvents, such as alcohol and acetone, are volatile and flammable. Common solvents used in a wafer fab are listed in Table 5.5 on page 102.

TABLE 5.5 Common Solvents Used in Semiconductor Manufacturing

Solvent	Common Name	Example of Use
Deionized water	DI Water	Widely used to rinse wafers and to dilute cleaning solutions.
Isopropyl alcohol	IPA	General-purpose cleaning solvent.
Trichloroethylene	TCE	Solvent used for wafer and general cleaning.
Acetone	Acetone	General-purpose cleaning solvent (stronger than IPA).
Xylene	Xylene	Strong cleaning solvent; may also be used for photoresist edge bead removal.

Deionized water (DI water) is a widely used solvent in semiconductor manufacturing that has all conductive ions removed. With a pH of 7, which is neutral, DI water is neither an acid nor a base. It has the ability to dissolve other substances, including many ionic and covalent compounds. When H₂O molecules dissolve an ionic compound, they separate, surround, and disperse the ions into the liquid. Water does this by overcoming the electrostatic force of attraction between the ions.

Chemical Distribution ■ There is a wide range of chemicals used in the semiconductor industry, with many being toxic and hazardous. A review of key chemical safety issues is provided in Appendix 1. Safe, high-purity, uninterrupted delivery of chemicals from storage vessel to process tools is critical. For liquid chemicals, this delivery is often accomplished through bulk chemical distribution (BCD).

A *bulk chemical distribution* system consists of a chemical source, such as a storage vessel, a chemical delivery module, and a piping system (see Figure 5.14).⁶ The BCD storage vessel is often housed below the main production floor in the groundlevel sub-fab. The chemical delivery module will filter, blend, and transport the chemicals through the piping system. A piping system delivers the chemical to the individual process stations. Modern BCDs are integrated into a fully computerized and networked system for real-time chemical monitoring and control.

The decision regarding how to store and deliver process chemicals depends on factors such as chemical compatibility, reduction of chemical contamination, and safety. Purity requirements of

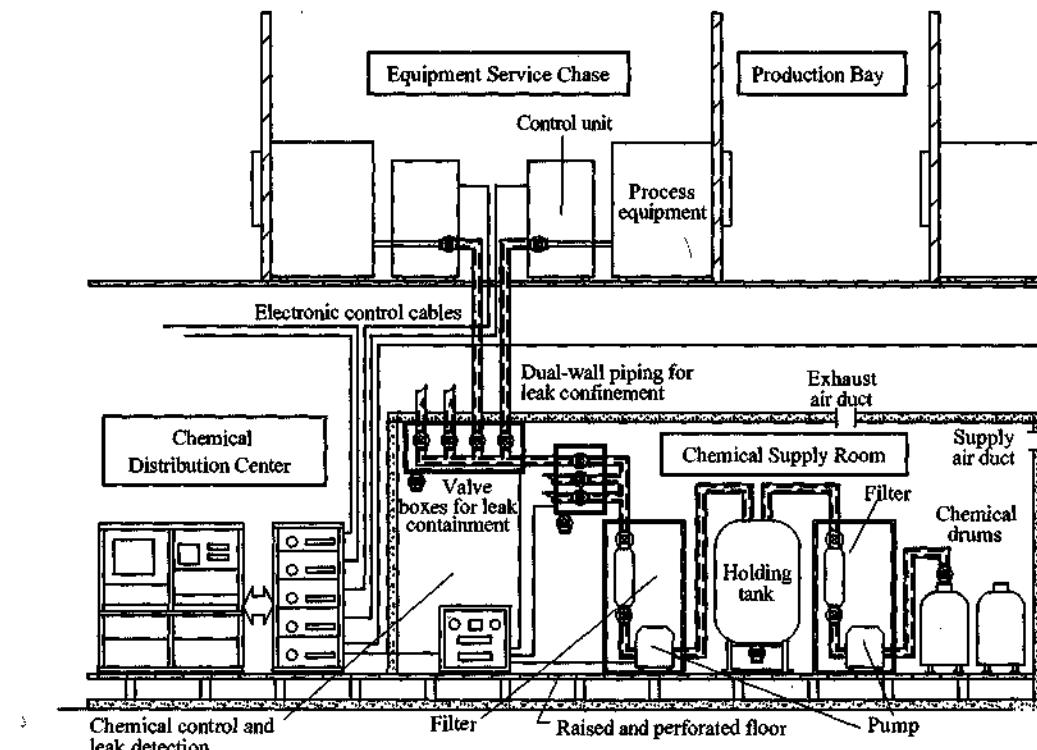


FIGURE 5.14 Bulk Chemical Distribution



Bulk Chemical Distribution
(Photo courtesy of Advanced Micro Devices)

chemicals used in semiconductor manufacturing are described as *ultrahigh purity (UHP)*, with contaminants being controlled to the parts per billion (ppb) to parts per trillion (ppt) range.

Some chemicals are not suitable for BCD. They may be used in small quantities or have a limited shelf life (how long they can be stored before use). In this case, these chemicals will have special packaging systems for *point-of-use (POU)* delivery, which means they are stored and used at the process station. An example of this type of chemical is the photoresist chemical used in photolithography (discussed in Chapter 13).

Gases

There are about 50 different types of gases used throughout the 450 or so process steps in semiconductor manufacturing.⁷ The types and quantities of gases are changing because there are new materials being introduced into the fab process, including copper metallurgy and new diffusion barriers (see Chapter 12). Gases are usually categorized as one of two types: bulk gases or specialty gases. *Bulk gases* are traditionally defined as oxygen (O₂), nitrogen (N₂), hydrogen (H₂), helium (He), and argon (Ar). The *specialty gases* are also referred to as process gases, and are the other important gases used to manufacture semiconductor microchips.

Extremely high purity is required of all gases, with bulk gases controlled to seven nines purity (99.99999%) and specialty gases controlled at least to four nines (99.99%). Particulate contamination in gases is controlled to the 0.1-μm range. Other controlled contaminants are oxygen, moisture content, and trace impurities such as metals.⁸ Many process gases are toxic, corrosive, reactive, and pyrophoric (they combust or burn when exposed to air). For these reasons, gases are contained in a gas delivery system that delivers the gases in a safe, clean, and accurate manner to the various process stations in the wafer fab.

Bulk Gases ■ Bulk gases are relatively simple gases for the gas suppliers to manufacture and are stored outside of the wafer fab manufacturing area in large storage tanks or 1000-lb bulk tube trailers. These gases are distributed to the workstation through the bulk gas distribution (BGD) system. Benefits from the centralized gas control of BGD are a reliable gas supply, less sources of particulate contamination, and less human involvement in the daily delivery of the gas. Bulk gas is often the lowest cost method for gas delivery with higher gas purity. Furthermore, on-site production of the bulk gases is done at large wafer fabs in order to reduce costs. Bulk gases are divided into inert, reducing, and oxidizing gases (see Table 5.6 on page 104).

TABLE 5.6 Bulk Gases

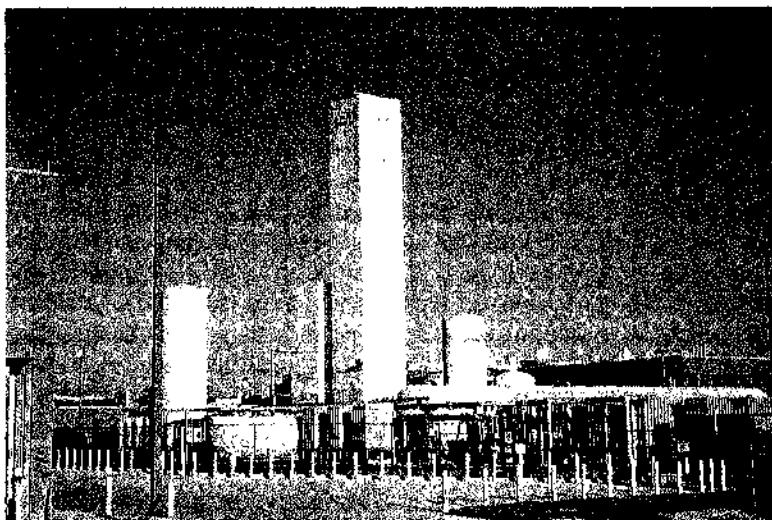
Type of Gas	Gas	Symbol	Example of Use
Inert	Nitrogen	N ₂	Purge gas lines and process chambers of moisture and residual gas. N ₂ is sometimes used as a process gas during some deposition processes.
	Argon	Ar	Used in process chambers during wafer processing.
	Helium	He	Process-chamber gas also used to leak check vacuum chambers.
Reducing	Hydrogen	H ₂	Carrier gas for epitaxial layer process. Also used in combination with O ₂ to form steam during some furnace oxidation processes. Used in many wafer fabrication processes.
Oxidizing	Oxygen	O ₂	Process-chamber gas.

In recent years, safety concerns have driven the development of on-site generation for other gases, in particular arsine and phosphine.⁹ This arrangement permits these highly toxic gases to be produced close to the process tool.

Specialty Gases ■ Specialty gases are those gases supplied in relatively low volumes. These gases are frequently more hazardous than the bulk gases. The specialty gases are the source of many of the materials needed to manufacture a microchip. The greatest challenge for dealing with specialty gases is that many of them are a chemical hazard. They are corrosive (e.g., HCl and Cl₂), pyrophoric (e.g., silane), toxic (e.g., arsine and phosphine), and extremely reactive (e.g., WF₆). The specialty gases are typically used in process chambers at the workstation tools.

Specialty gases traditionally have been transported to the wafer fab in 100-lb metal containers called cylinders for distribution of the gas to the workstations. The cylinders are placed in a gas cylinder cabinet containing a control panel with regulators to control pressure, a flow control, shut-off valves, and a purge panel to control purge sequences needed between cylinder changes. The cabinet also may contain filters, purifiers, or equipment to monitor gas purity, as well as safety equipment such as fire sensors and leak-detection sensors.

A local gas distribution system in the process area is used to deliver the specialty gas from the cylinder to the work chamber of the process tool (see Figure 5.15). This piping can cover hundreds of feet with many permanent welds, bends and fittings and is connected to the process tool at the tool "drop."



Bulk Gas Distribution System
(Photo courtesy of Air Products and Chemicals, Inc.)

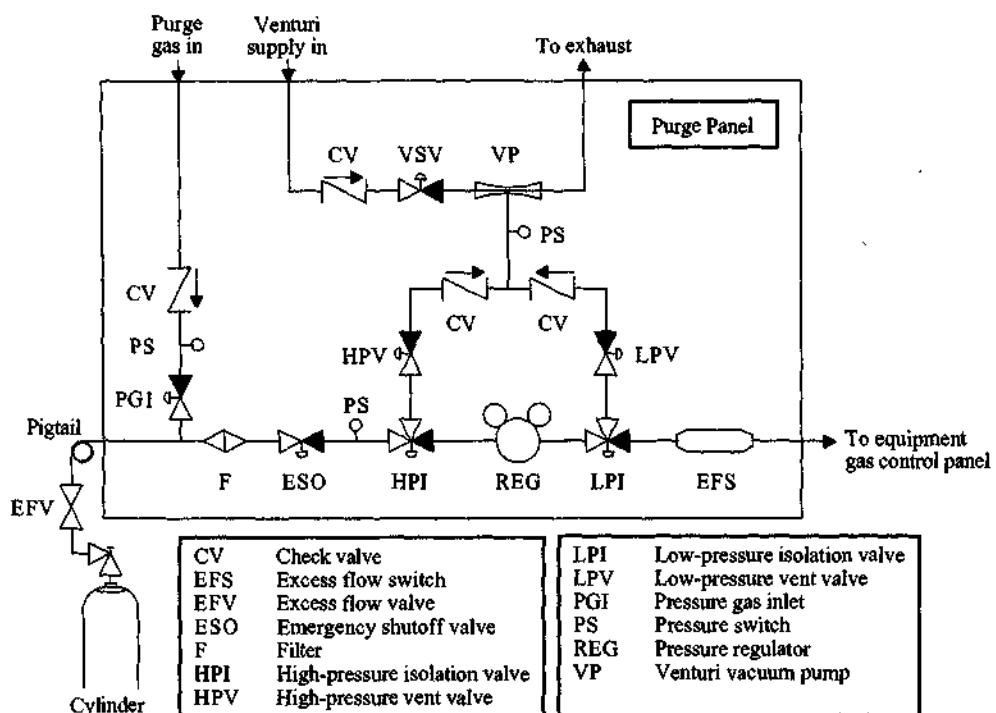


FIGURE 5.15 Typical Specialty Gas System Design
(Schematic used with permission from International SEMATECH)

Gas Purge. A *gas purge* is a method of flushing undesirable residual gases, atmospheric gases, or water vapor from a process chamber and the gas delivery system. Purging can eliminate stagnation points and contamination in gas lines. It is done with an inert gas such as nitrogen and involves replacing the undesirable gas with the purge gas, either through displacement or by pulling the gas out of the system by vacuum flow. Purging of the system on automated equipment is done automatically through software control of gas line valves before and after situations such as a gas cylinder change or opening a process chamber.

Gas Piping. Gas delivery piping is constructed with electropolished 316L stainless steel tubes to transport the gas (316L is a specific type of stainless steel). There are no plastic parts in a gas piping system, except for some membrane gas filters. Double-walled tubing is often used for hazardous gases (see Figure 5.16). The inner walls of the double-walled tubing are electropolished to minimize contamination. *Electropolishing* is a chemical process done to remove about 30 microns of the pipe's inner surface, creating a clean, smooth surface that minimizes the possibility of chemical reactions that produce contaminants. An electropolished surface finish brings a thin layer of chromium to the steel's surface, which emits very few particles.¹⁰

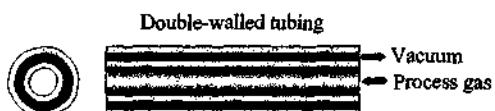


FIGURE 5.16 Double-Walled Tubing
(Used with permission from International SEMATECH.)

Gas Line Connections. For safety reasons, gas lines will have a 360-degree bend, referred to as a pigtail, to make the line more flexible. The gas line terminates in a CGA (Compressed Gas Association) connector for connection to the gas cylinder valve (see Figure 5.17 on page 106). A recently developed DISS (Diameter Index Safety System) cylinder valve has been introduced with specific connector diameters for different gases, which creates a better seal with less abrasion to reduce the potential for contamination. DISS valves are generally used with high-purity products, toxics, and corrosives. Gas lines need a continuous flow through the piping system with no dead spots (known as a deadleg) to trap gas and cause undesirable chemical degradation due to stagnation.

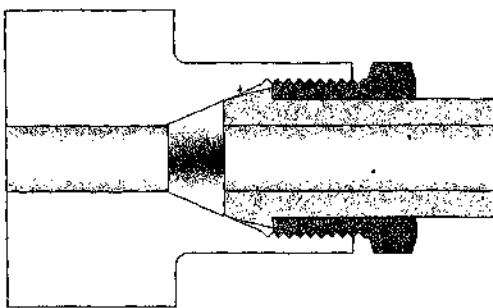


FIGURE 5.17 CGA Gas Line Connector
(Used with permission from International SEMATECH)

Gas Stick. The gas line from the local gas distribution system connects into another gas panel at the process tool. The tool panel is made up of a series of gas “sticks,” each controlling one type of incoming gas. Each stick will typically have an on/off valve, flow controller, pressure regulator, and filters (see Figure 5.18). The number of incoming gas lines depends on the process, with a typical number being from six to 30 or more lines for multi-chamber process tools.

Cylinder Change-Out. Cylinders for specialty gases require technicians to change them when they are empty. This is referred to as a *cylinder change-out* and is sometimes done as often as several times per week. Because specialty gases are often hazardous, gas cylinder change-outs require care to avoid safety and product problems.¹¹ One safety concern during a cylinder change is incorrectly purging gas lines, which causes residual gas to escape, leading to a flame or vapor cloud. Another safety concern is an improperly supported cylinder that may fall over, causing a leak. With regard to the product performance, an incorrect cylinder change can lead to contamination in the gas line and affect product yield. Furthermore, a cylinder change may require the process chamber to be requalified to ensure the gas is acceptable, which interrupts manufacturing production.

Due to safety hazards, contamination concerns, and high maintenance costs, there has been an effort to convert some specialty gases to bulk distribution. For example, a tube trailer of silane is estimated to eliminate the change-out of 300 to 400 cylinders per year, while the tube trailer is changed only once per year.¹² This is a substantial reduction in risk for a problem during change-outs. Eventually more bulk specialty gas distribution systems similar to those used for wet chemicals and bulk gases may replace the cylinders commonly used to deliver specialty gases.¹³ This

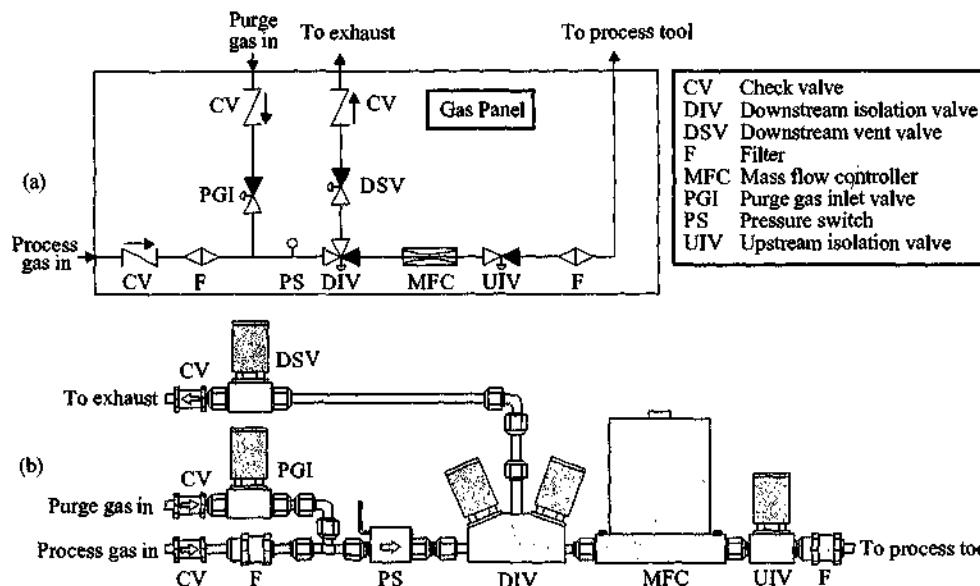
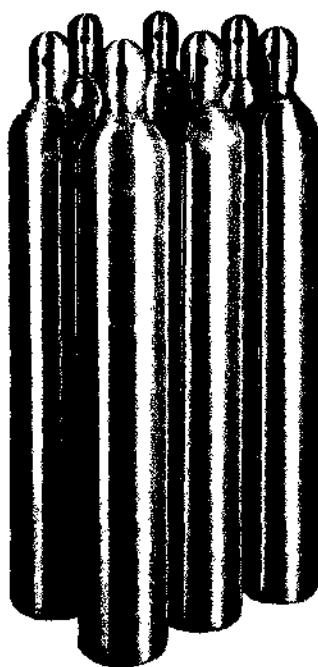


FIGURE 5.18 Gas Stick at Process Tool.
(Schematic (a) used with permission from International SEMATECH)
Component diagram (b) redrawn from Swagelok components, Swagelok Co. Catalog provided by Arthur Valve & Fitting Co., Austin, TX.

trend initially started with silane because it is pyrophoric (it burns upon exposure to air) and now encompasses other specialty gases used in high volume, such as nitric oxide (NO). The goal is to reduce the human involvement in the gas delivery system and the potential for error.



Specialty Gas Cylinder
(Photo courtesy of Praxair)

Classifying Specialty Gases ■ Specialty gases may be classified as hydrides (or similar gases), fluorinated compounds, or acid gases. Hydride gases contain hydrogen, while fluorinated compounds contain fluorine. Some common specialty gases are listed in Table 5.7 on page 108, with an example of their use.

TABLE 5.7 Some Common Specialty Gases for Semiconductor Manufacturing

Class of Gas	Gas	Symbol	Example of Use
Hydrides	Silane	SiH_4	Source of silicon in deposition processes (Chapter 11)
	Arsine	AsH_3	Source of arsenic for n-type doping of silicon wafers (Chapter 17)
	Phosphine	PH_3	Source of phosphorous for n-type doping of silicon wafers (Chapter 17)
	Diborane	B_2H_6	Source of boron for p-type doping of silicon wafers (Chapter 17)
	Tetraethyl orthosilicate (TEOS)	$\text{Si}(\text{OC}_2\text{H}_5)_4$	Source of silicon dioxide used in deposition processes (Chapter 11)
	Silicon tetrachloride (also tetrachlorosilane)	SiCl_4	Source of silicon used in deposition processes (Chapter 11)
Fluorinated Compounds	Dichlorosilane (DCS)	SiH_2Cl_2	Source of silicon used in deposition processes (Chapter 11)
	Nitrogen trifluoride	NF_3	Source of fluoride ions used in plasma etch processes (Chapter 16)
	Tungsten hexafluoride	WF_6	Source of tungsten used in metal deposition processes (Chapter 12)
	Tetrafluoromethane	CF_4	Source of fluoride ions used in plasma etch processes (Chapter 16)
	Carbon tetrafluoride	CF_4	Source of fluoride ions used in plasma etch processes (Chapter 16)
	Silicon tetrafluoride	SiF_4	Source of silicon and fluorine ions for use in deposition, implant, and etch processes (Chapters 11, 16, 17)
Acid Gases	Chlorine trifluoride	ClF_3	Process chamber cleaning gas
	Boron trifluoride	BF_3	Source of boron for p-type doping of silicon wafers (Chapter 17)
	Chlorine	Cl_2	Source of chlorine used in metal etching (Chapter 16)
	Boron trichloride	BCl_3	Source of boron for p-type doping of silicon and a source of chlorine for metal etching (Chapters 16, 17)
Other Gases	Hydrogen chloride	HCl	Process chamber cleaning gas and contamination getterer
	Ammonia	NH_3	Process gas used with SiH_2Cl_2 for deposition of SiN_3 (Chapter 11)
	Nitrous oxide	N_2O	Source of oxygen for reaction with silicon to form silicon dioxide (Chapter 11)
	Carbon monoxide	CO	Used in etch processes (Chapter 16)

SUMMARY

There are four states of matter: solid, liquid, gas, and plasma. Important chemical properties for wafer fabrication are: temperature, pressure and vacuum, condensation, vapor pressure, sublimation and deposition, density, surface tension, thermal expansion, and stress. Process chemicals are widely used to manufacture microchips. Liquids are grouped as acid, base, or solvent. An acid or a base varies in strength and is measured on the pH scale. Chemicals are delivered to the workstation through bulk chemical distribution or point-of-use

delivery. Gases used in the wafer fab are classified as either bulk or specialty. Bulk gases are the relatively simple gases used to manufacture chips, such as nitrogen and oxygen. Specialty gases are often a chemical hazard and are transported and stored in the fab in metal cylinders. There are special procedures and delivery systems to ensure that the specialty gases are cleanly and safely used at the workstations. Specialty gases are generally classified as either hydrides, fluorinated compounds, or acid gases.

KEY TERMS

process chemicals
solid
liquid
gas
inert gas
plasma
properties
physical properties
chemical properties
chemical reaction
temperature
heat (thermal energy)
Celsius
Kelvin
pressure
vacuum
condensation
vaporization
absorption
adsorption
vapor pressure
volatile
sublimation

deposition
density
specific gravity
surface tension
coefficient of thermal expansion (CTE)
stress
chemical solution
solute
aqueous solutions
parts per million (ppm)
parts per billion (ppb)
parts per trillion (ppt)
acids
base
pH scale
solvent
bulk chemical distribution (BCD)
ultrahigh purity (UHP)
bulk gases
specialty gases
gas purge
electropolishing
cylinder change-out

REVIEW QUESTIONS

1. What are the four states of matter? Describe each state.
2. What is a material property?
3. Describe the two types of properties of materials.
4. What happens in a chemical reaction? Give an example of a chemical reaction.
5. What is temperature? How is temperature related to heat?
6. State the three temperature scales. Which scale is most common for scientific work?
7. What is pressure? The pressure of a gas is dependent on what conditions?
8. Define vacuum. What is the most common vacuum unit and how is it defined?
9. Define condensation and vaporization. What is the difference between absorption and adsorption?
10. What is vapor pressure? Describe a volatile material.
11. Define sublimation and deposition.
12. Define and describe density. What is the specific gravity of a material?
13. What is surface tension?
14. Define the coefficient of thermal expansion (CTE) of a material.
15. Describe stress and state its formula.
16. How are stress and coefficient of thermal expansion related?
17. What is a chemical solution and what are its components
18. What is an aqueous solution?
19. What do the abbreviations *ppm*, *ppb*, and *ppt* represent?
20. What is an acid? List three common acids used in a wafer fab.
21. What is a base? List three common bases used in a wafer fab.
22. Explain the pH scale and how it measures acids and bases. Identify the regions in the pH scale attributed to weak solutions of a strong acid, a strong base, and water.
23. What is a solvent? List three common solvents used in a wafer fab.
24. Describe deionized (DI) water for the wafer fab.
25. Describe the bulk chemical distribution system.
26. What is the point-of-use chemical delivery?
27. What are the two types of gas categories?
28. Describe the purity required for gases in semiconductor manufacturing.
29. What five gases are classified as bulk? Give three benefits of bulk gas distribution.
30. What is a specialty gas?
31. What is the greatest challenge for dealing with specialty gases?
32. How are specialty gases usually transported and stored in the wafer fab?
33. What is a gas purge? What type of gas is this done with?
34. How is gas piping constructed? What is electropolishing, and why is this process beneficial?
35. What is CGA? What is DISS and why is it used?
36. Describe a gas stick.
37. Describe a specialty gas cylinder change-out. What are the safety concerns associated with this procedure?
38. State the three classes of specialty gases. Give an example of a common gas for each class.

**CHEMICAL SUPPLIERS
WEB SITES**

AERONIX Inc.

Air Products and Chemicals

Ashland Specialty Co.

ATMI Inc.

BOC Edwards

Dow Chemical/Filmtec

Dow Corning

DuPont

Eastman Chemical Co.

EKC Technology Inc.

J. T. Baker

Leybold Inficon Inc.

Linde

Matheson Gas Products

Millipore Corp.

MKS Instruments Inc.

Parker Hannifin Corp.

Praxair

<http://www.aeronex.com><http://www.airproducts.com/><http://www.ashchem.com/><http://www.atmi.com/><http://www.boc.com/edwards/><http://www.dow.com/liquidseps/><http://www.dowcorning.com/><http://www.dupont.com/semiconductor><http://www.eastman.com/><http://www.ekctech.com/><http://www.jtbaker.com/><http://www.leyboldinficon.com/><http://www.linde.de/english/Home.htm><http://www.mathesongas.com/><http://www.millipore.com/><http://www.mksinst.com/><http://www.parker.com/><http://www.praxair.com/>

PTI Advanced Filtration Inc.
Scott Specialty Gases
Solkatronic Chemicals
Swagelok Company
Union Carbide
Voltaix Inc.

<http://www.pti-afi.com/>
<http://www.scottgas.com/>
<http://www.solkatronic.com/>
<http://www.swagelok.com/>
<http://www.unioncarbide.com/>
<http://www.voltaix.com/>

REFERENCES

1. M. Silberberg, *Chemistry: The Molecular Nature of Matter and Change*, (St. Louis: Mosby, 1996), p. 23.
2. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., (New York, NY, McGraw-Hill, 1997), p. 206.
3. T. Hattori, "Trends in Wafer Cleaning Technology," *Solid State Technology* (May 1995): p. 7.
4. A. Braun, "PPT - Time for a Reality Check?" *Semiconductor International* (June 1998).
5. J. Sargent, V. Starov, and R. Werner, "Transition in the Post-Etch Wafer-Cleaning Market and Technologies," *Solid State Technology* (May 1997): p. 180.
6. R. DeJule, "Bulk Chemical Distribution Addresses Tightening Specs," *Semiconductor International* (August 1996): p. 75.
7. P. Singer, "Effective Gas Handling: A Balance of Cost and Purity," *Semiconductor International* (September 1994): p. 64.
8. P. Singer, "Trends in Gas Management and Use," *Semiconductor International* (April 1998): p. 112.
9. Ibid.
10. A. Braun, "Cleanroom Technologies Continue to Keep Contamination at Bay," *Semiconductor International* (March 1998): p. 59.
11. L. Laurin, "Bulk Silane—a Potential Hazard or a Potential Hazard Reducer," *Solid State Technology*, (January 1998): p. 104.
12. P. Singer, "Trends," p. 112.
13. N. Chowdhury and L. Mostowy, "Developing a Bulk Distribution System for High-Purity Hydrogen Chloride," *Micro Contamination Identification, Analysis and Control* (September 1995): p. 33.

CHAPTER 6

CONTAMINATION CONTROL IN WAFER fabs

Contamination-free wafer fabrication is absolutely essential in order for devices on chips to function properly. As device critical dimensions decrease, contamination requirements become more stringent. We will learn in this chapter about the different types of contamination important for wafer fabrication, their sources, and how to control contamination to effectively manufacture high-performance ICs with minimal contamination-generated defects.

To control unacceptable contamination during manufacturing, the semiconductor industry has developed

cleanrooms. A cleanroom is essentially a purified space with ultraclean air that isolates the chip manufacturing from the dirty conditions of the outside world, including chemicals, humans, and ordinary work conditions.

It is important to understand the conditions of a cleanroom because this is where the intricate microchips are manufactured. Numerous details of the work procedures in a wafer fab are defined by how we maintain the integrity of the cleanroom. One of the primary elements for successful cleanrooms is human discipline.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. State and describe the five different types of cleanroom contamination, and discuss the problems associated with each type of contamination.
2. List seven sources of contamination in a cleanroom, and describe how each one affects wafer cleanliness.
3. Interpret and use the class number for cleanroom air quality.
4. State and discuss seven appropriate actions for workers entering a cleanroom that follow acceptable protocol.
5. Describe the different aspects of an ultraclean cleanroom facility, including air filtering, electrostatic discharge, ultrapure DI water, and process gases.
6. Explain how modern workstation design and a minienvironment contribute to contamination reduction.
7. State the chemistry of the two standard wet-cleaning methods, explain the type of contamination removed by each, and discuss wet-cleaning modifications and alternatives.
8. Describe the different types of wet-cleaning equipment, and state how each cleaning process contributes to wafer cleanliness.

INTRODUCTION

A wafer has multiple microchips on its surface, and each chip has literally millions of devices and interconnection circuitry that are highly sensitive to contamination. As the feature size on a chip shrinks to accommodate higher performance and denser circuitry, the need to control surface contamination becomes more critical (see Figure 6.1 on page 114). To achieve contamination control, all wafer fabrication is done in a cleanroom where contaminants are strictly controlled.

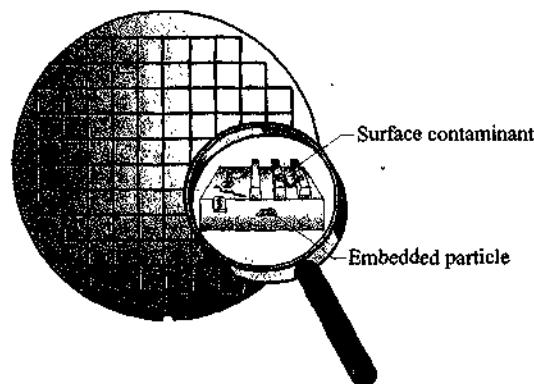
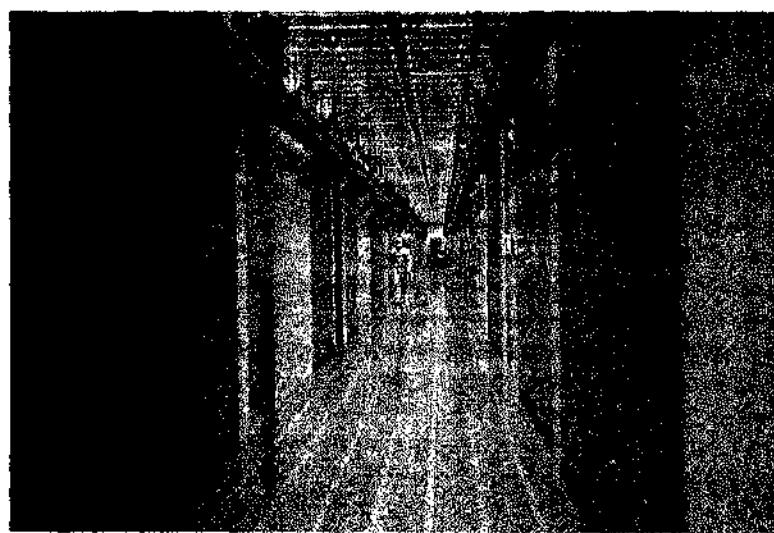


FIGURE 6.1 Wafer Contaminants

Cleanroom Background

When semiconductor manufacturing began nearly half a century ago, the need to control contamination was evident. Early cleanrooms were based around local clean zones, with clean benches used by operators wearing smocks and gloves. The introduction of the high-efficiency particulate air filter (HEPA filter) in the 1960s was a first step toward significant particulate reduction in the wafer fab. The HEPA filters delivered clean air at the workbench to efficiently move particles away from the product.

Modern semiconductor manufacturing is performed in a sophisticated facility known as a *cleanroom*. This is a wafer fabrication facility that is isolated from the outside environment and free from contaminants such as particles, metals, organic molecules, and electrostatic discharge (ESD). When we say free, that means that these contaminants are not detectable at the detection level of the most advanced test instrument. A cleanroom represents the comprehensive set of procedures and practices that are followed to ensure a wafer fabrication facility is contamination-free for semiconductor manufacturing.



Wafer Fab Cleanroom
(Photo courtesy of
Advanced Micro
Devices)

TYPES OF CONTAMINATION

Contamination in semiconductor manufacturing is any undesirable substance introduced to the semiconductor wafer that impacts the production yield or electrical performance of a microchip. Since our interest is wafer manufacturing, we will focus on the various types of surface contamination introduced during the manufacturing process.

Contamination often leads to a defective chip. *Killer defects* are those causes of failure where the chip on the wafer fails during electrical test. It is estimated that 80% of all chip failures at electrical test are due to killer defects from contamination.¹ Failure at electrical test results in a yield loss, causing the defective die on the wafer to be scrapped (thrown away) at a significant cost to the chip manufacturer.

Cleanroom contamination is grouped into five categories:

- ◆ Particles
- ◆ Metallic impurities
- ◆ Organic contamination
- ◆ Native oxides
- ◆ Electrostatic discharge (ESD)

Particles

Particles are small objects that can adhere to the surface of a wafer. Airborne particles suspended in the air are referred to as *aerosols*. The relative size distribution of various particles from pebbles down to atoms is shown in Figure 6.2.²

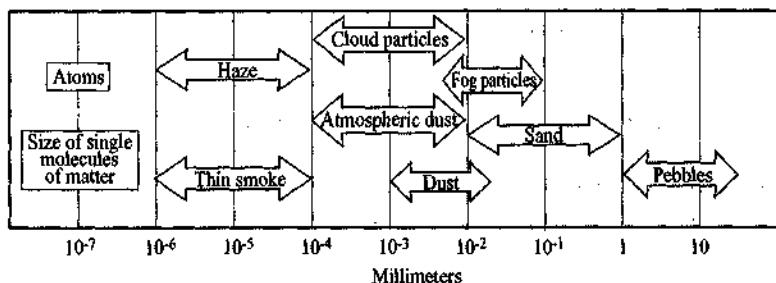
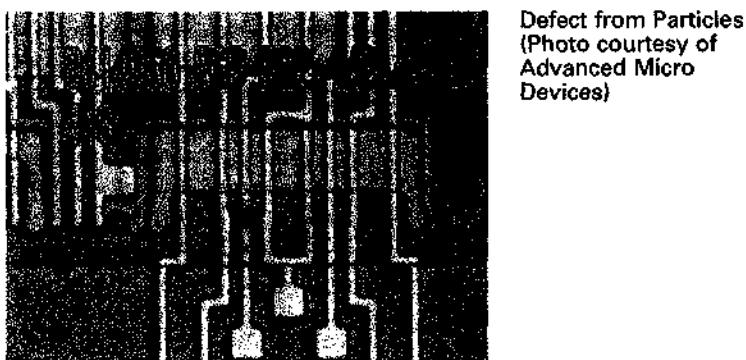


FIGURE 6.2 Relative Size of Particles

Problems From Particles ■ For semiconductor manufacturing, our objective is to control and reduce wafer exposure to particles. Particles can cause defects during the wafer fabrication process leading to open or bridged circuitry. They can create an electrical short between adjacent conductors. Particles also can be sources of other types of contamination, as discussed in the following section.



The rule of thumb for an acceptable particle size in semiconductor manufacturing is that it must be less than one-half the minimum device feature size.³ Particles larger than this size will cause killer defects. For example, a 0.18- μm feature size cannot be exposed to 0.09 μm and larger particles. To

appreciate these dimensions, consider that a human hair is about $90\text{ }\mu\text{m}$ in diameter. A dimension of $0.18\text{ }\mu\text{m}$ would be about 500 times smaller than the human hair (see Figure 6.3).

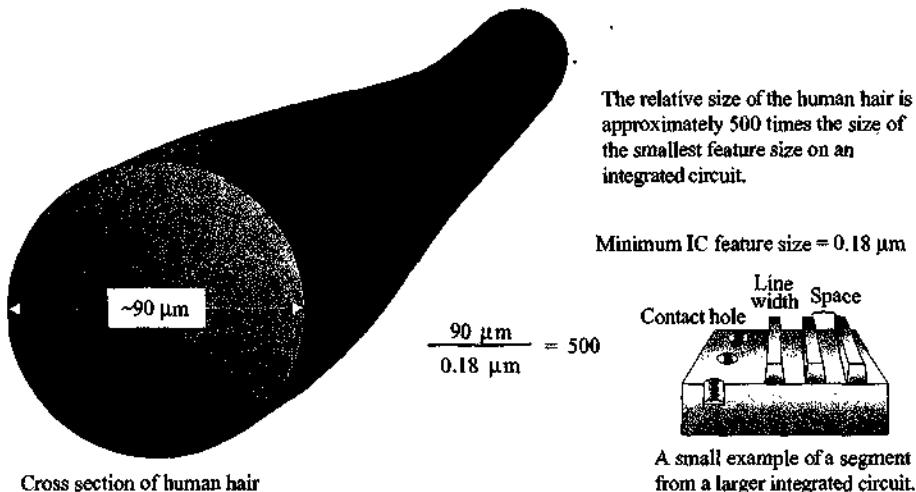


FIGURE 6.3 Relative Size of Human Hair to $0.18\text{ }\mu\text{m}$ Particle Size

The *particle density* on a wafer surface represents the number of particles in a given area. A higher particle density produces a higher chance of a killer defect. The number of particles above a certain critical size that are added to a wafer surface at an operation is termed the *particles per wafer per pass (PWP)*. In the early days of semiconductor manufacturing, skilled operators visually inspected wafers for particles using simple tools such as a microscope. However, this approach is unacceptable for the VLSI and ULSI eras. Since the mid-1980s, particle detection has been widely done by scanning the wafer surface with a laser beam and detecting the position and intensity of the scattered light caused by particles (see Figure 6.4). The smallest detectable diameter of current particle-detection equipment used in production is roughly $0.1\text{ }\mu\text{m}$.⁴

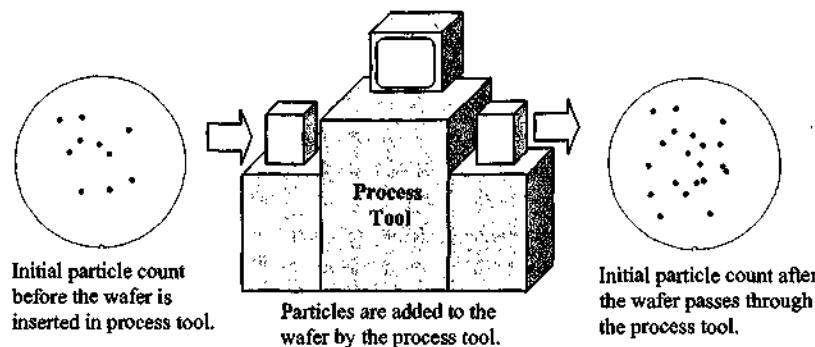


FIGURE 6.4 Particles Per Wafer Per Pass on a Wafer

Semiconductor manufacturing in the ULSI era also requires stringent control of a wide range of airborne molecular particles, including water vapor, acid vapors, hydrocarbons, and gases that add impurities to the wafer surface. The measurement of airborne molecular contamination at the ppb level is now becoming standard for advanced fabrication facilities. Molecular compounds include metals, nonmetals, and organic and inorganic contaminants, all of which are damaging for device performance.

Metallic Impurities

Contamination in a wafer fab also can come from metal compounds. The typical *metal impurities* damaging to semiconductor processing are the alkali metals that are found in many common chemicals and processes. These metals are strictly controlled in all materials used in a wafer fab (see

Table 6.1).⁵ The alkali metals are from Group IA of the periodic table and are extremely reactive elements because they give up one valence shell electron, thus forming a cation that reacts with a nonmetal anion to form an ionic compound.

TABLE 6.1 Typical Metal Impurity Elements

Heavy Metals	Alkali Metals
Iron (Fe)	Sodium (Na)
Copper (Cu)	Potassium (K)
Aluminum (Al)	Lithium (Li)
Chromium (Cr)	
Tungsten (W)	
Titanium (Ti)	

Metals come from chemical solutions or different process steps during semiconductor manufacturing. The process of ion implantation (discussed in a later chapter) exhibits the highest metal contamination, on the order of 10^{12} to 10^{13} atom/cm².⁶ Another source of metallic contamination is through reaction of the chemicals with the transport piping and containers. For instance, carbon monoxide gas, used as an additive gas to improve various wafer processes, can react with the nickel in stainless steel, gaskets, and other components of the gas delivery system.⁷ From this reaction gas-phase carbon monoxide forms nickel tetracarbonyl particles that are then distributed on the wafer surface. These particles can be redistributed into the bulk wafer and lead to increased device defects.

There are two basic ways for metallics to deposit on a wafer surface.⁸ In the first way, metal binds to the silicon surface by the charge exchange between a metallic ion and a hydrogen atom located on the wafer surface. These types of metallic impurities are difficult to remove. The second way metal is deposited on the wafer surface is when the surface oxidizes and metallic impurities are located in the oxidation layer. For this reason cleanliness is critical in oxidation processes (see Chapter 10). Metallic impurities in an oxide layer are removed only by removing the oxide from the wafer surface.

Metallic ions are highly mobile in semiconductor materials and are referred to as *mobile ionic contaminants (MICs)*. When introduced into a wafer, MICs move throughout the silicon and seriously damage the device's electrical performance and long-term reliability. Sodium is typically one of the most prevalent MICs in untreated chemicals, with people as its greatest carrier. The human body contains a high percentage of sodium in the form of fluids (e.g., saliva, tears, perspiration, and so on). Sodium contamination is rigorously controlled in wafer fabs.

It is hard to imagine how mobile an MIC material is in silicon. Consider an MIC such as sodium. A single crystal of table salt (NaCl) contains enough sodium to deposit one quadrillion (10^{12}) atoms of sodium per square centimeter on 5,000 wafers (150-mm diameter) and therefore destroy all chips on the wafers.

Problems From Metallic Impurities ■ Metallic impurities lead to reduced device yield in semiconductor manufacturing, including structural defects in the oxide-polysilicon gate structure.⁹ Additional problems include increased leakage currents at a pn junction and reduced minority carrier lifetime. MIC contamination can migrate to the oxide-silicon interface in the gate structure and alter the threshold voltage required to turn on a transistor (see Figure 6.5 on page 118). Because they are so mobile, metallic ions can move around a device long after electrical test and shipment and cause the device to fail during usage. A major goal of semiconductor manufacturing is to minimize exposure to metallic impurities and MICs.

Organic Contamination

Organic contaminants are those that contain carbon, nearly always bonded to itself and to hydrogen, and sometimes to other elements as well. Some sources of organic contamination are bacteria, lubricants, vapors, detergents, solvents and moisture. Equipment used in wafer fabs today is designed with components that require no lubricant, for example, oil-free pumps and bearings.

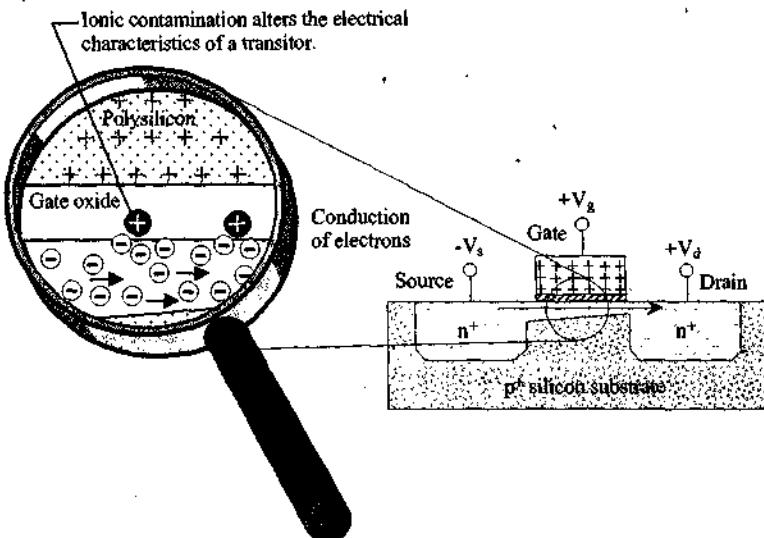


FIGURE 6.5 Mobile Ionic Contaminant Altering Threshold Voltage

Problems From Organic Contaminants ■ Trace organic contaminants degrade the integrity of the gate oxide material under certain processing conditions.¹⁰ Another problem with organic materials on the surface of semiconductor wafers during processing is incomplete cleaning of the surface. This condition permits contaminants such as metal impurities to remain intact on the wafer surface after cleaning.

Native Oxides

The surface of a silicon wafer will oxidize if exposed either to air at room temperature or to DI water that contains dissolved oxygen. This thin oxide layer is termed *native oxide*. The initial native-oxide growth on a silicon wafer occurs in the presence of moisture. When the wafer surface is exposed to air, several tens of molecular layers of moisture adsorb on the wafer within a second. Oxygen from the room is dissolved into the adsorbed moisture layer on the wafer surface and penetrates into the silicon surface. This process causes the silicon surface to oxidize even at room temperature. The thickness of the native oxide increases as exposure time lengthens.

Problems From Native Oxides ■ A silicon surface that is native oxide-free is important for semiconductor performance and reliability. Native oxide interferes with other process steps, such as the growth of single-crystal film on the wafer or the growth of the ultrathin gate oxide.¹¹ Native oxide also includes some metallic impurities, which can move into the silicon wafer and cause electrical defects.¹²

Another problem created by native-oxide growth occurs in contact regions for metal conductors. Contacts make electrical connections between interconnect wiring and the source and drain region of the semiconductor device. If there is a native-oxide layer, then this will increase the resistance of the contact and reduce and perhaps even prevent current flow (see Figure 6.6).¹³

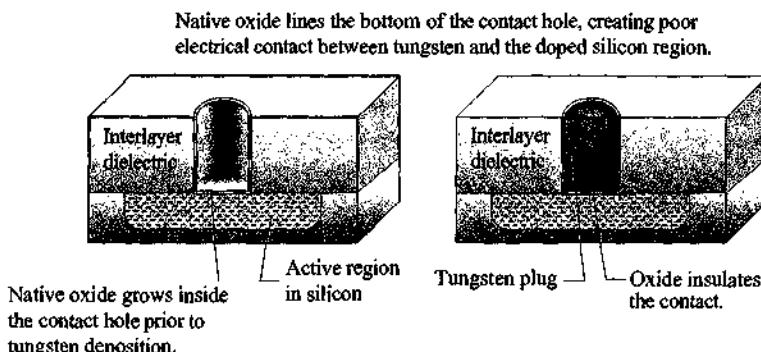


FIGURE 6.6 Native Oxide

Native oxide requires removal through a cleaning step with a mixture of HF acid (see the section on wet cleaning later in this chapter). Another approach to inhibiting native oxide is to integrate multiple process steps into a multichamber tool with an evacuated, high-vacuum chamber so that the wafers are not exposed to ambient atmosphere and moisture.

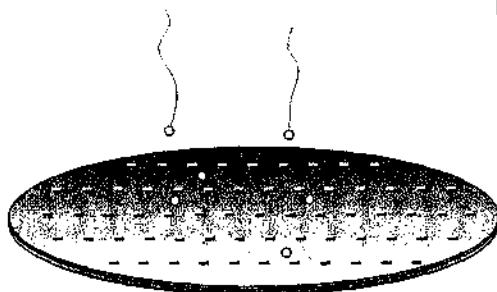
Electrostatic Discharge

Electrostatic discharge (ESD) is a form of contamination because it is an uncontrolled transfer of static charge from one object to another that potentially damages the microchip. ESD is generated when two materials with differing static-charge potential are touching or rubbing each other (this action is referred to as *triboelectricity*). Atoms with excess negative charges are attracted to adjacent atoms that have an excess of positive charges. This attraction creates a discharge of electricity that can be up to tens of thousands of volts.

Semiconductor manufacturing is especially prone to developing static electrical charges because the wafer fab is maintained at a low humidity, typically $40\% \pm 10\%$ relative humidity (RH). This condition is conducive to the generation of high levels of static charge.¹⁴ Although increasing the RH would decrease static-charge generation, it would also increase contamination through corrosion, thus making it an impractical approach.

Problems From Electrostatic Discharge ■ Although the amount of static charge transferred during ESD is usually small (on the order of nanocoulombs), the discharge deposits its energy into a very small area of the wafer. An electrostatic discharge that occurs in a few nanoseconds can generate peak currents over 1 amp, literally vaporizing metal conductor lines or punching through oxide layers.¹⁵ This discharge also can be the cause of gate oxide breakdown. Another significant problem with ESD is that once a wafer surface has a charge buildup, its resulting electric field can attract charged particles or polarize and attract neutral particles to its surface (see Figure 6.7). An example of this is how a TV screen attracts dust particles. Furthermore, the smaller the particle, the more effect the electrostatics have on particle attraction.¹⁶ As device critical dimensions are reduced, smaller particles attracted by ESD become more significant and can create killer defects. In order to minimize particle contamination, wafer charging must be controlled.

FIGURE 6.7 Particles Attracted to a Charged Wafer



SOURCES AND CONTROL OF CONTAMINATION

The wafer fab cleanroom is stringently controlled for contamination to reduce killer defects that impact microchip performance. Nearly anything that comes in contact with a wafer is a potential source of contamination. Seven sources of contamination in a wafer fabrication facility are:

- ◆ Air
- ◆ Humans
- ◆ Facility
- ◆ Water
- ◆ Process chemicals
- ◆ Process gases
- ◆ Production equipment

Air

The most fundamental concept of a cleanroom is the control of particles in the wafer fab air. The air that we normally breathe is not acceptable for semiconductor manufacturing because it contains excessive airborne contaminants. These small airborne aerosol particles float and remain in the air for a long period of time, depositing on wafers as contamination and creating killer defects.

The *class number* designates the air quality inside a cleanroom by defining the particle size and density in the cleanroom air. This figure represents how well particles are controlled to reduce particulate contamination. Class numbers originated from Federal Standard 209, first released in 1963 and revised several times until the most recent version of 209E (see Appendix B). Table 6.2 shows the number of acceptable particles per cubic foot for the different cleanroom class numbers and particle size.

TABLE 6.2 Definition of Airborne Particulate Cleanliness Classes Per Federal Standard 209E

Class	Particles/ft ³				
	0.1 µm	0.2 µm	0.3 µm	0.5 µm	5 µm
1	3.50×10	7.70	3.00	1.00	
10	3.50×10^2	7.50×10	3.00×10	1.00×10^1	
100		7.50×10^2	3.00×10^2	1.00×10^2	
1,000				1.00×10^3	7.00
10,000				1.00×10^4	7.00×10
100,000				1.00×10^5	7.00×10^2

If the cleanroom class is stated only by the number of particles, such as a cleanroom of class 1, then assume a particle size of 0.5 µm. This means there is a maximum of one particle of size 0.5 µm or larger per cubic foot. For particle sizes different from 0.5 µm, the cleanroom class should be expressed as the class number at a specific particle size. Examples are: class 10 at 0.2 µm (read from Table 6.2 as a maximum of 75 particles per cubic foot at 0.2 µm or larger) and class 10 at 0.1 µm (a maximum of 350 particles per cubic foot at 0.1 µm or larger).

Ultrafine Particles ■ There recently has been usage of a class 0.1, with particle sizes down to 0.02 to 0.03 µm. The latest clean air standard also has a provision for the number of *ultrafine particles* in a cubic meter of air, which is called the "U" descriptor. The U descriptor specifies ultrafine particles as those smaller than 0.1 µm in diameter, down to the smallest diameter detectable with a discrete particle counter. Without referring to a specific particle size, the U descriptor defines cleanliness as U(x), where (x) is the maximum allowable number of ultrafine particles per cubic meter of air.

Humans

A human is a particle generator. People continually enter and leave cleanrooms and are the greatest sources of contamination in a cleanroom.¹⁷ Particles come from hair and hair products (hair spray, gel), lint from clothes, flakes of dead skin, and so on. On the average, a person sheds over one ounce of particles per day, which can amount to an astonishing 10,000,000 particles per minute of 0.3 μm size and larger (see Table 6.3).

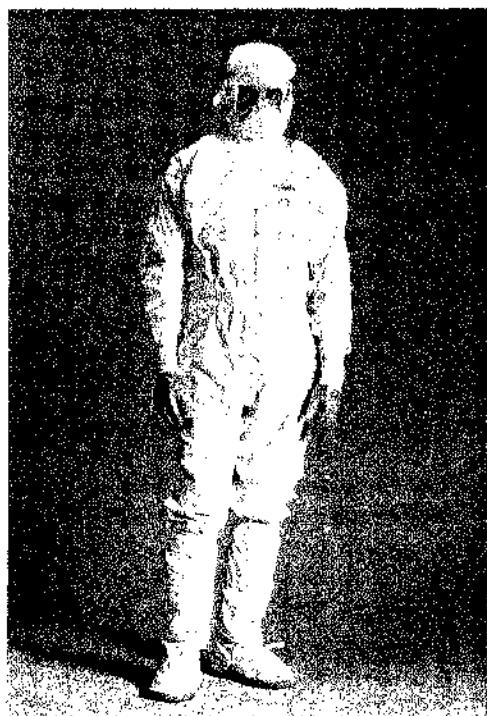
TABLE 6.3 Particles Emitted by Human Activities

Source of Particles	Average Number of Particles per Minute > 0.3 μm
Motionless (sitting or standing)	100,000
Moving hands, arms, trunk, neck, and head	500,000
Walking at 2 miles per hour	5,000,000
Walking at 3.5 miles per hour	7,500,000
Cleanest skin (per square foot)	10,000,000

Simple activities in the wafer fab such as opening and closing doors or excessive movements around process tools create particulate contamination. Normal human activities such as talking, coughing, and sneezing are damaging to semiconductors.

Cleanroom Garments ■ To attain ultraclean conditions in a cleanroom, humans must follow certain procedures, known as *cleanroom protocol*, and be covered with special cleanroom garments (also referred to as a “bunny suit”). The garment is made up of a hood with a facemask, coveralls, boots, and gloves, and will completely cover the body. The goal of the cleanroom garment system is to meet these functional criteria:

- ◆ Total containment of body-generated particles and aerosols.
- ◆ Zero particle release from the garment system.
- ◆ Zero electrical-charge buildup for ESD.
- ◆ No release of chemical or biological residues.



Technician in
Cleanroom Garment.
(Photo courtesy of
International
SEMAPTECH)

The modern cleanroom garment is a high-technology membrane fabric or densely woven polyester fabric. Advanced materials will have a 99.999% efficiency rating (stops 99.999% of all particles from passing) for 0.1 micron particles and greater. The extent to which the cleanroom garment system covers the human body can vary. Some fabs require inner clothing layers, such as polyester underwear, in addition to the outer bunny suit. People may be required to take a shower and use lotion to prevent skin flaking prior to final dressing and entering fabs. Some wafer fabs also require fab workers to put on a bubble helmet with a breathing recirculator and exhaust blower that pumps the user's breath through a filter pack strapped to the waist. This helmet prevents particles from human saliva from contaminating the fab work area.

Cleanroom Protocol ■ Each semiconductor company has a strict procedure for cleanroom protocol to minimize contamination in the cleanroom. Some standard cleanroom protocols are listed in Table 6.4.¹⁸

TABLE 6.4 Proper Cleanroom Protocol

Should Do	Should Not Do	Why?
Only authorized personnel are allowed within the cleanroom.	No people allowed that have not been properly trained in what the cleanroom expects of them. The cleanroom supervisor has the last word.	Authorized personnel are familiar with the many strict and demanding restrictions of cleanroom operations.
Take only what is necessary into the cleanroom.	No cosmetics, tobacco products, handkerchiefs, tissues, food, drink, candy, wooden/mechanical pencils or pens, perfumes, colognes, watches, jewelry, cassette players, phones, beepers, video cameras, audiocassette recorders, gum, combs, hair brushes, cardboard or noncleanroom-approved paper. No blueprints, operations manuals or instruction sheets.	Sources of unwanted contaminants.
Gown in the prescribed manner according to your company training.	No uncovered street clothes allowed within the cleanroom.	The proper cleanroom-approved garmenting protects the product from the contamination associated with humans.
Always make sure that all head and facial hair is covered.	Do not expose any facial or head hair.	Sources of unwanted contaminants.
Follow procedures for entering the cleanroom, such as an air shower (if required).	Do not open any door into the cleanroom until all procedures are complete.	Air showers may assist in removing contaminants; many firms have stopped using this procedure due to problems with airborne contamination.
Keep cleanroom garment closed at all times while in the cleanroom.	Do not expose any street clothing while in the cleanroom. Do not allow any part of your skin to touch the outer parts of the cleanroom garment.	Sources of unwanted contaminants.
Move slowly.	Do not congregate or move quickly.	This disrupts the airflow pattern.

Facility

For semiconductor manufacturing to function as an ultraclean environment, a systems approach is necessary to control all inputs and outputs to the cleanroom area. There are three basic strategies for eliminating particles from cleanrooms:¹⁹

1. Start out with a cleanroom that is free of particles.
2. Minimize the introduction of particles into the cleanroom through equipment, tools, personnel, and cleanroom supplies.
3. Continuously monitor the cleanroom for particles for timely response to cleaning maintenance.

Cleanroom Layout ■ In the early 1970s, the LSI manufacturing environment had an overall cleanliness class of 10,000 in the manufacturing space and a local class of 100 at the individual workbenches. The industry developed a *ballroom layout* approach, with one large fabrication room with a class 10,000 rating and laminar flow benches that provided the class 100 work environment (see Figure 6.8).

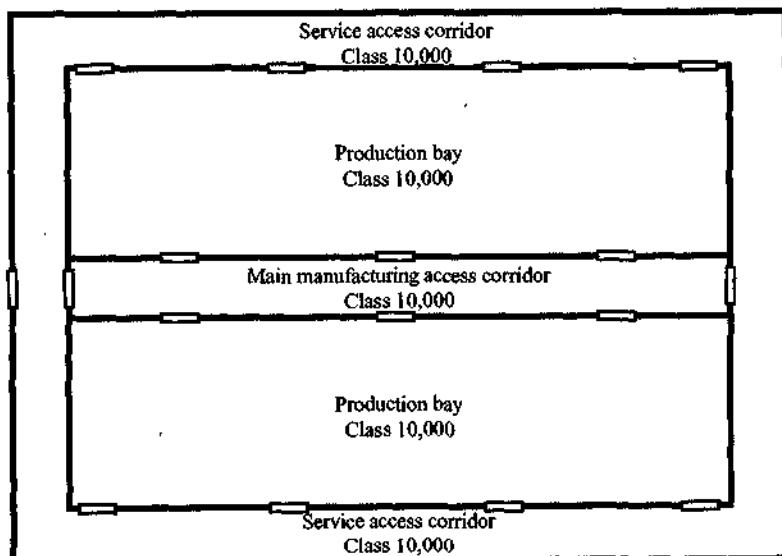


FIGURE 6.8 Early Ballroom Cleanroom Layout

With the submicron technologies of the 1980s came the introduction of the *bay and chase layout* approach to cleanrooms. In this cleanroom layout, a common corridor separates the production area (referred to as production bay or process bay) from the service area (referred to as service chase, equipment chase, or gray area), as shown in Figure 6.9 on page 124. Inside the production bay where wafers are processed, the cleanliness class is typically class 1. Most of the equipment maintenance occurs in the class 1,000 service chase.

Modern wafer fabs are built based on either type of cleanroom design. The ballroom approach is promoted as an advanced cleanroom design that encompasses automated wafer handling and localized contamination control at the process tools.²⁰ For both types of cleanrooms, there is typically a *sub-fab* area below the cleanroom that contains much of the facilities equipment (e.g., pumps, piping, ductwork, cabling, and so on).

Airflow Principles ■ To achieve ultraclean conditions in a cleanroom, the nature of the airflow is critical. For a cleanroom class of 100 and below, *laminar airflow* is necessary.²¹ Laminar airflow means that the airflow is smooth with no turbulence in the airflow pattern (see Figure 6.10 on page 124). The vertical laminar airflow has a slight positive pressure relative to the outside pressure and acts as a curtain to minimize cross-contamination from equipment or personnel to any exposed product.²²

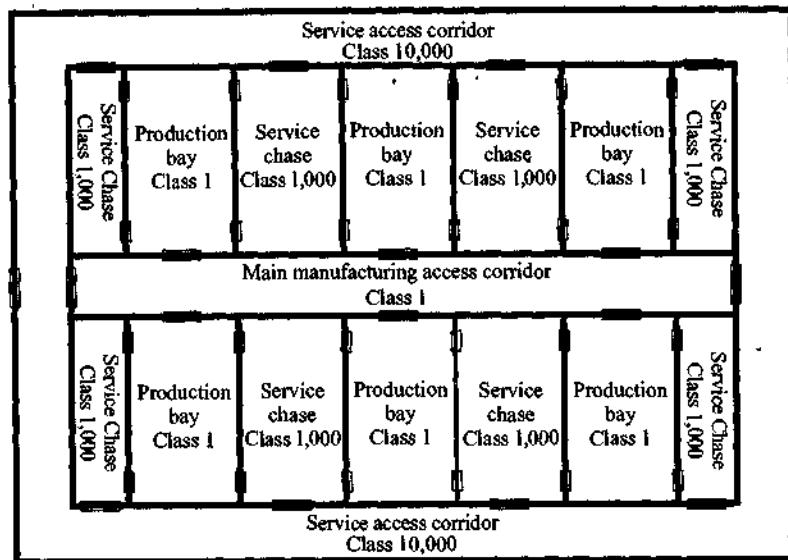


FIGURE 6.9 Bay and Chase Concept for Cleanroom

Air Filtering ■ A simplified diagram of the air handling system in a wafer fab cleanroom is shown in Figure 6.11. Air enters the cleanroom through special particulate filters in the ceiling and passes with laminar flow to the floor and into the recirculation air system to return back to the air filtering system with makeup air. In a modern fab, air may turn over every six seconds to improve recovery of ultraclean conditions during disturbances such as shift changeover. An exhaust system is used to remove undesirable heat and chemicals from the process tools and work areas.

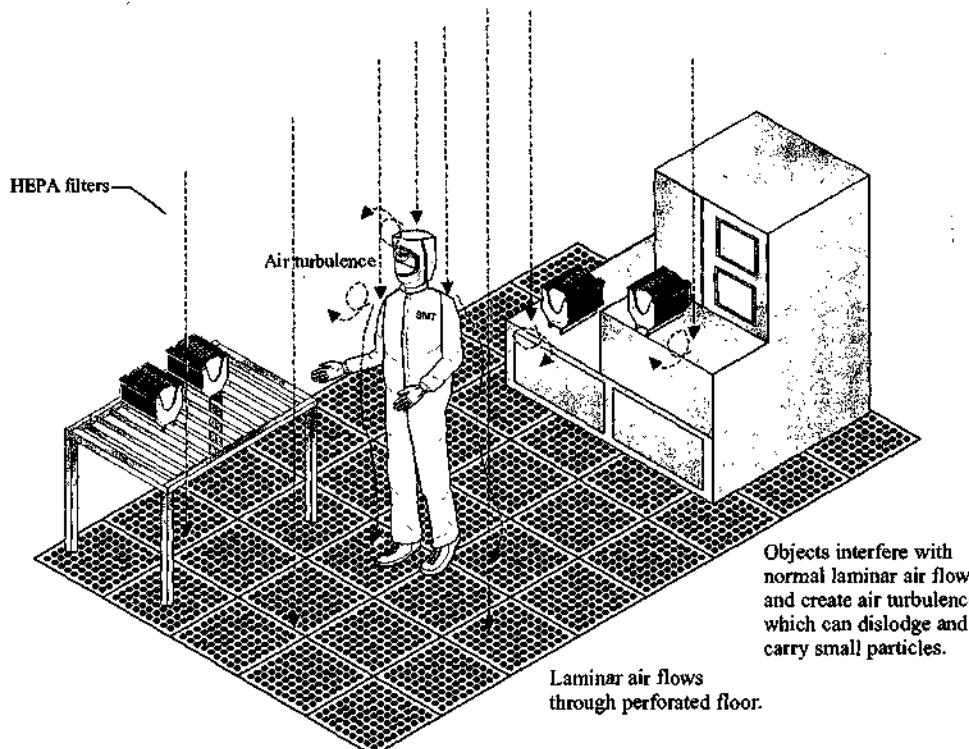


FIGURE 6.10 Laminar Airflow

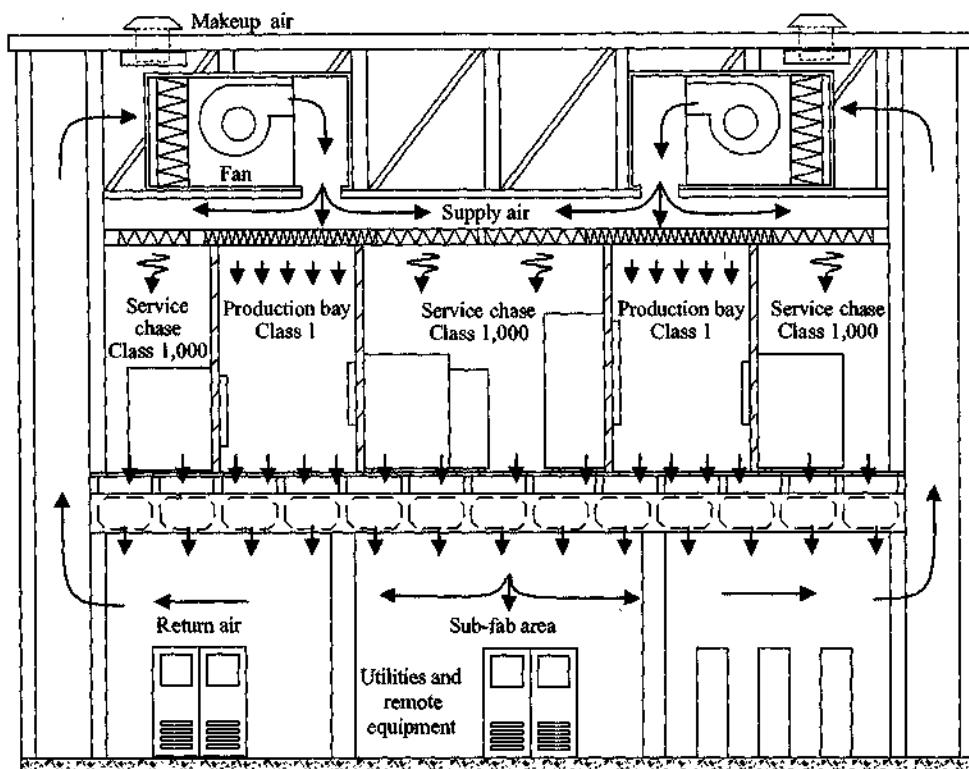


FIGURE 6.11 Wafer Fab Air Handling System

The special particulate filters located in the ceiling are either *high-efficiency particulate air (HEPA)* or *ultra-low penetration air (ULPA)*. HEPA fibrous filters are made with fiberglass fibers and constructed to create laminar airflow (see Figure 6.12). In general, ULPA represents filters that have an efficiency of 99.9995% or better for particulate diameters greater than 0.12 microns.²³

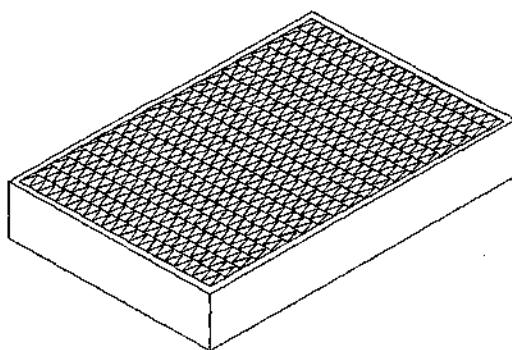


FIGURE 6.12 HEPA Filter

Temperature and Humidity ■ Temperature and humidity settings are specified for wafer fab facilities. An example of temperature control in a class 1 at 0.3 μm cleanroom is $68 \pm 0.5^\circ\text{F}$.²⁴ Relative humidity (RH) is important because of its contribution toward corrosion (higher RH promotes more corrosion, just as moisture causes steel to rust). A typical RH setting may be 40% +/- 10%.

Electrostatic Discharge ■ Most electrostatic discharge (ESD) problems are controlled through the proper use of equipment and procedures. The principal ESD control methods are:

- ◆ Static-dissipative cleanroom materials
- ◆ ESD grounding
- ◆ Air ionization

Cleanroom materials such as carts, cassettes, equipment, and so on must be *static dissipative*. This term means that the resistivity of the material is lowered through the use of conductive additives, permitting mobile electrostatic charges to flow through the material. At the same time, people and objects in the cleanroom must be continuously connected to ground. In this manner, electrostatic discharge is conducted through the human body and all cleanroom materials that the wafer comes in contact with and flows harmlessly away from the product through ground. This setup avoids potentially discharging through a device on a wafer and causing irreparable damage to the chip.

Air Ionization. The wafer has various insulating materials placed on it during processing, such as an oxide film. These materials are easily charged and will hold this charge for a long period. Since this insulation material is in intimate contact with the product, it requires a method to neutralize the charge buildup. The most common way to neutralize insulating materials on a wafer is with *air ionization*. Special ionizer emitters located on the ceiling in the cleanroom produce a high electric field that ionize the air molecules and make it conductive by gaining or losing an electron. When this conductive air contacts a charged surface, such as an insulator material on a wafer, the surface attracts ions from the opposite polarity and neutralizes the electrostatic surface charge (see Figure 6.13).

Air ionization by emitters can be limited because many ions are eliminated before reaching the wafer surface due to recombination. A recent development is air neutralization using soft-X-ray radiation.²⁵ Exposing the ambient air surrounding the charged wafer to soft X-rays will generate large ion pairs. This process effectively neutralizes the charge on the wafer surface to 0 volts after roughly two seconds.

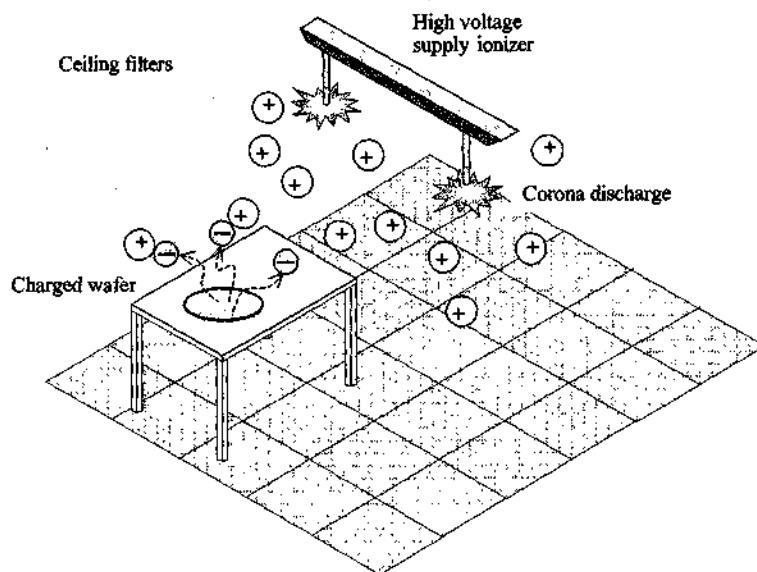


FIGURE 6.13 Neutralizing Static Charge on a Surface with Air Ionization

Water

To manufacture semiconductors, a high quality, *ultrapure deionized (DI)* water (sometimes referred to as UPW) is required in large quantities. City water has too many contaminants to be acceptable for the production of wafers. DI water is the most heavily used chemical in semiconductor manufacturing, primarily in the chemical wafer cleaning solutions and as a postclean rinse. It is estimated that current DI water consumption in a wafer fab runs up to 2,000 gallons of ultrapure deionized water for each wafer produced in a modern 200-mm process line.²⁶

Unacceptable contaminants in ultrapure DI water are:

- ◆ Dissolved ions
- ◆ Organic materials
- ◆ Particulates
- ◆ Bacteria
- ◆ Silica
- ◆ Dissolved oxygen

Figure 6.14 shows different water particles and their sizes.²⁷

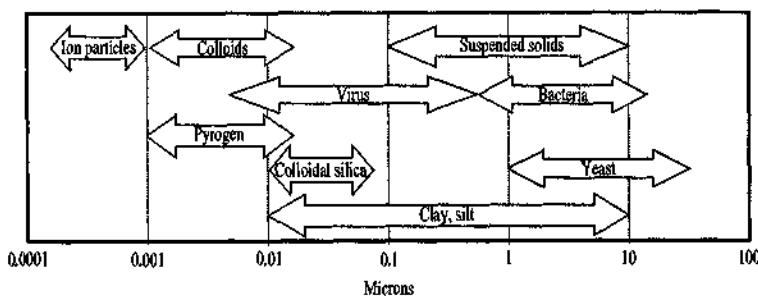


FIGURE 6.14 Size of Particles in Water

Dissolved ions in water are from minerals such as sodium and potassium that easily form ions. An example is salt (NaCl) that breaks up into Na^+ and Cl^- ions. These mobile ionic contaminants (MICs) are undesirable and create performance problems in semiconductor devices.

Organic materials, referred to as *total organic carbon (TOC)*, are the sum of all carbon-containing compounds dissolved in the water. These organic contaminants have a damaging effect on the ability to grow oxide films.

Bacteria are live, reproducing organisms in the water. Fragments shed by bacteria are known as *pyrogens*. Bacteria found in water lead to defects in oxidation, polysilicon, and metal-conducting layers. Some bacteria contain phosphorus that could cause uncontrolled doping.

Silica is found in city water as finely suspended particles. These particles range in size from ten angstroms to ten microns. A high silica content fouls filtration equipment in the water purification equipment and also decreases the reliability of thermally grown oxides.

Another type of contaminant in water is dissolved oxygen. It creates a problem because it leads to native-oxide formation on the wafer surface. Dissolved oxygen in water also creates problems if the water is pressurized as it is in some semiconductor processes. When the water is depressurized, dissolved gases may come out of solution forming bubbles that can cause incomplete water wetting on the wafer surface.²⁸

DI-Water Installation ■ A DI-water installation has two major parts for purifying water, referred to as the makeup loop and the polishing loop (see Figure 6.15 on page 128).²⁹ The *makeup loop* removes particles, total organic carbons (TOCs), bacteria, microorganisms, ionic impurities, and total dissolved minerals from the raw water. It has a prefilter that removes particles above one micron and a purifier section for removal of ionic impurities, bacteria, and dissolved gases. The *polishing loop* is the final part of the water purification system that removes the remaining contaminants.

Deionization ■ *Water deionization* to make DI water is the process of removing the electrically active salt ions using specially manufactured ion-exchange resins. This process changes the water from a conductive medium to a resistive medium with a resistivity of 18 megohm-cm (18,000,000 ohm-cm) at 25°C. DI water used in the wafer fab is referred to as 18-megohm ($\text{M}\Omega$) water. Ultrapure DI water is passed through two deionizers, once in the makeup loop and another time in the polishing loop.

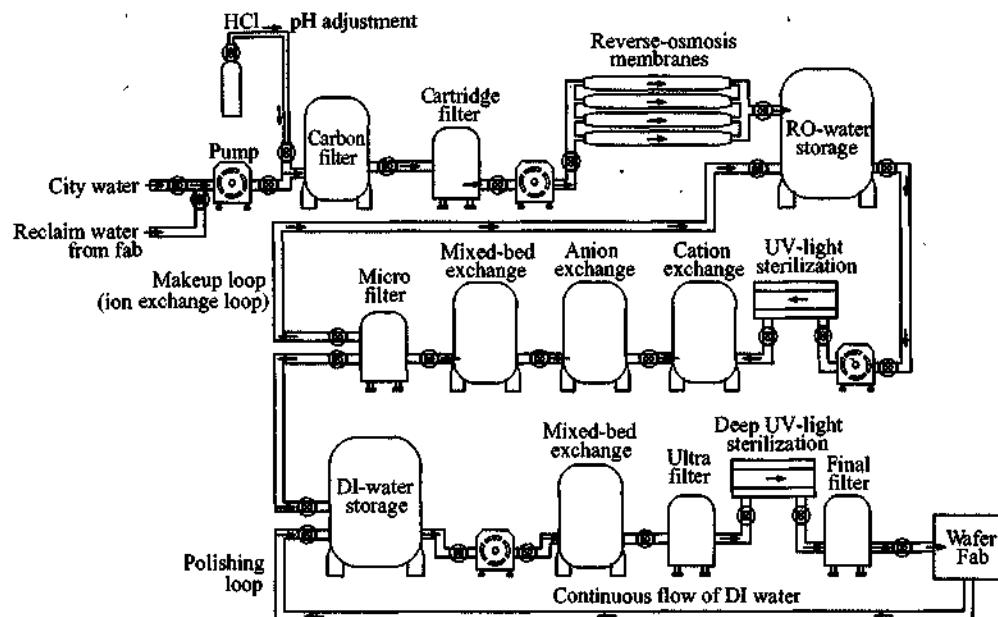


FIGURE 6.15 DI-Water Makeup and Polishing Loops

DI-Water Filtration ■ The DI-water makeup loop employs various filters with the intent of letting water pass through while trapping the particulates and colloids (extremely fine particles) in the filter media. A common filtering technique for ultrapure DI water is *reverse osmosis (RO)* to remove smaller particles and metallic ions. The operating principle for RO filters is to flow the water under pressure across a membrane filter to separate ionized salts, colloids, and organic materials down to 150 molecular weight (see Figure 6.16). RO can separate impurities as small as 0.005 microns and is also referred to as *hyperfiltration*.

Ultrafiltration is used at point-of-use (POU), meaning its filtering location is at the equipment, to remove submicron-size particles. This filter uses pressure and flow through a membrane with pore sizes ranging from 10 angstroms to 0.2 microns.

Other parts of an ultrapure DI-water system include a degasifier unit used to remove dissolved gases in DI water, such as oxygen, to reduce contamination such as native-oxide growth on the wafer surface. An improved technique for removing dissolved gases (primarily oxygen) to less than 10 ppb is a *membrane contactor* (see Figure 6.17).³⁰ Membrane contactor filters consist of

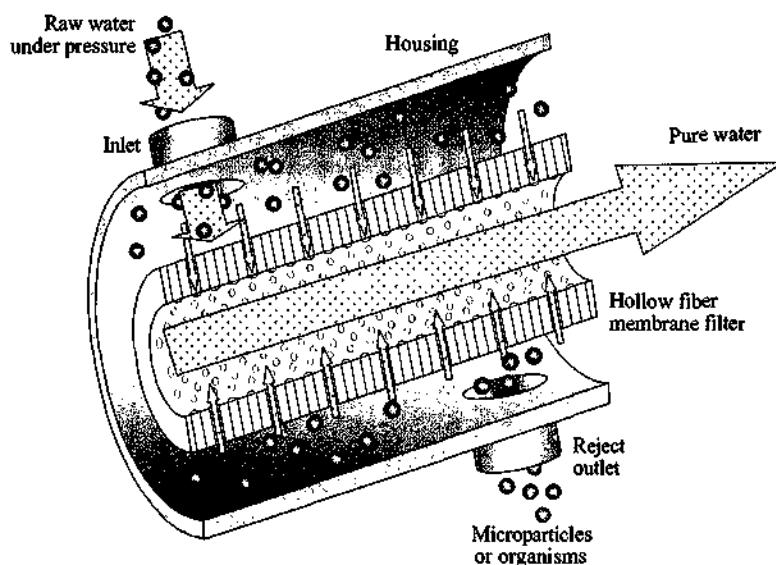


FIGURE 6.16 Principle of Reverse Osmosis Filtration

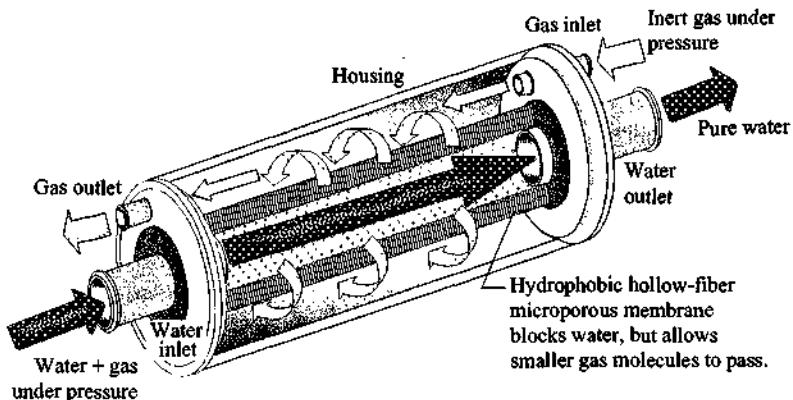


FIGURE 6.17 Membrane Contactor Filter

hydrophobic polypropylene microporous hollow fibers. The fibers have small pores in their walls to allow dissolved gases to pass through. However, the fiber's hydrophobic nature does not permit liquid water to pass through the pores. The membrane removes oxygen and all free gases.

Zeta Potential ■ The *zeta potential* represents a positive or negative electrical charge that can build up in colloids (very fine suspended particles in a liquid). Particles in water generally have a zeta potential with a negative charge, such as colloidal silica, bacteria, and pyrogens. These colloids can be filtered out of the water using a positively charged filter that traps particles that are smaller than the pore diameter of the filter.³¹ This form of particle removal uses electrostatics to attract particles with opposite zeta potential rather than relying on the pore size of the membrane (see Figure 6.18).

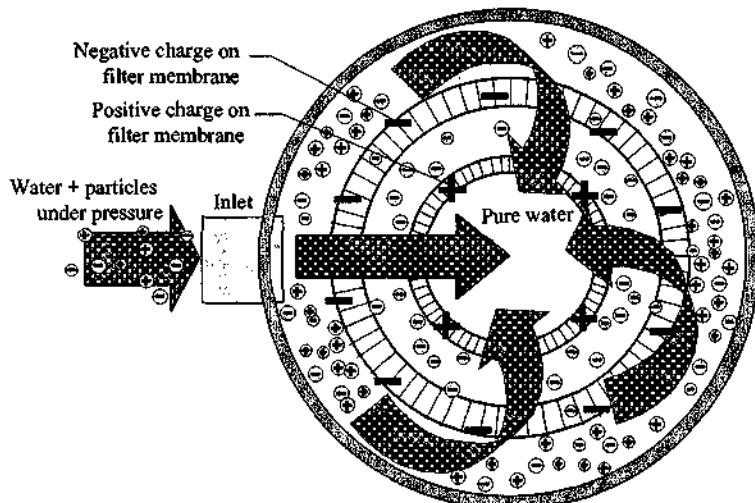


FIGURE 6.18 Electrostatic Filtration Using the Zeta Potential

Bacteria Control. Ultrapure water systems use *ultraviolet (UV) lamps* for bacterial sterilization. Water is exposed at a controlled rate to UV light waves, which deactivates certain molecules to reduce bacteria. UV systems are simple and reliable and can reduce bacteria to less than 1%.

Biological control of an ultrapure electronic-grade water system is also done by ozonating the water. Ozone (O_3) is created by discharging an electric current through dry air. The resulting ozone mixture is purified and then injected into the ultrapure water to kill the bacteria. All ozone is removed from the water by exposing the ozonated water to ultraviolet light, which breaks down the ozone to oxygen. Since most bacteria have a physical diameter in excess of 0.2 microns, point-of-use microfiltration can also be effective in removing bacteria.

Process Chemicals

Liquid chemicals used in semiconductor processing must be free of contamination for successful device yield and performance. Chemical purity is identified by an *assay number*, which describes the percentage of a particular chemical in the container (without reference to other substances present). For example, an assay of 99.99% for a bottle of hydrofluoric acid means the bottle contains 99.99% hydrofluoric acid and 0.01% other substances.

Filters are used to prevent chemical degradation during delivery or to maintain chemical purity during a recirculation step. Filters should be located at appropriate locations (e.g., near input to the gas controller) and as close to the process chamber as possible through point-of-use filtration. The different filter classes are:

Particle filtration: Depth-type filtration (see Figure 6.19) for particles from approximately 1.5 microns and larger.

Microfiltration: Membrane filtration of a liquid that removes particles in range of 0.1 to 1.5 microns.

Ultrafiltration: Pressure-driven membrane process that rejects large molecules from approximately 0.005 to 0.1 microns.

Reverse osmosis: Also called hyperfiltration. A pressure-driven solution process which transports liquid through a semipermeable membrane with the exclusion of particles and metal ions as small as approximately 0.005 microns.

A *membrane filter* uses a thin membrane of polymer or ceramic with small penetrating pores as the filter medium (see Figure 6.20). The size and distribution of the pores are controlled throughout the membrane. The membrane serves as a barrier that permits the passage of materials only up to a certain size or shape as defined by the pores. It is also called a surface filter because it removes particles from the air stream through interception at its surface. Membrane filters are used in reverse osmosis, microfiltration, and ultrafiltration. A membrane filter is often used as a point-of-use filter and is placed just before the process tool to provide final filtration.

A good filter will not create a significant pressure drop of the required flow, will not introduce secondary contamination, and will be compatible with the chemical. The *filter efficiency* is the percentage of particles of a specific size and above that are stopped in the filter. For liquid filters in a ULSI process, a typical efficiency rating for a membrane filter is 99.9999999% of 0.02 micron particles and above (referred to as *nine nines efficiency*).

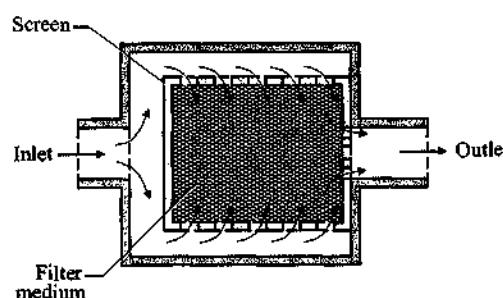


FIGURE 6.19 Depth-Type Filter
(Used with permission from International SEMATECH)

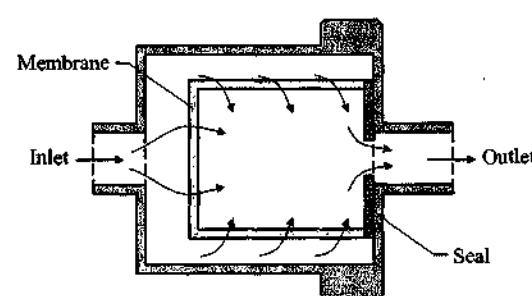


FIGURE 6.20 Membrane Filter
(Used with permission from International SEMATECH)

Process Gases ■ The delivery and use of ultraclean gas is critical for semiconductor manufacturing in the ULSI era. However, the handling and delivery system can introduce impurities that adversely affect the yield of semiconductor devices. Gases pass through purifiers and gas filters to remove impurities and particles. Some gas filters are all-metal (e.g., nickel) that do not shed particles or outgas organic contaminants. These filters have a nickel membrane that is able to withstand the corrosive gases with a proven efficiency for particles as small as 0.003 microns. Other gas filters are made of a Teflon polymer.

Production Equipment

The production equipment used to manufacture the semiconductor wafer is the most significant source of particles in a wafer fab.³² During the wafer fabrication process, silicon wafers are repeatedly loaded from cassettes into tools, processed through multiple equipment operations, unloaded back into cassettes, and then transported to the next workstation. This sequence occurs over and over for the 450 or more process steps necessary to fabricate a wafer, exposing the wafer to a multitude of mechanical and chemical operations at the different tools.

Many wafer fabrication operations occur in a vacuum, which requires special design considerations to avoid contamination. Examples of different sources of particle contamination from process tools are:

- ◆ Flaking of by-products built up on chamber walls.
- ◆ Automated wafer handling and transportation.
- ◆ Mechanical operations such as shaft rotation and valve opening or closing.
- ◆ Pumping and venting in vacuum environments.
- ◆ Cleaning and maintenance procedures.

Since equipment automation means fewer humans interacting with the product, there is less concern about particles from humans and more emphasis on reducing particles from equipment.³³ The number of particles on a wafer surface will increase during the fabrication process as the wafer is exposed to more tool operations. This condition is shown in Figure 6.21.³⁴

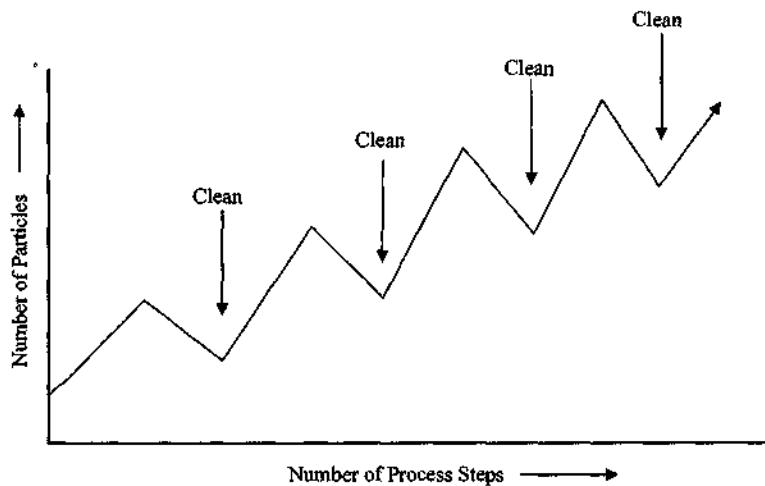


FIGURE 6.21 Wafer Surface Particles as a Function of Process Steps

Workstation Design

The right materials for *workstation design* of process tools and general workstations are necessary for achieving an ultraclean cleanroom. All materials emit some particles, but the goal is to reduce the emission to an acceptable level. Smooth, highly polished surfaces are best to reduce particulate contamination. Stainless steel has become a widely used material for work surfaces and equipment in the cleanroom. When properly processed, stainless steel has a relatively low particle emission rate. Electropolishing is a crucial finishing step.

Bulkhead Installation ■ Some fabs choose to install their equipment with a *bulkhead equipment layout*. In this approach, the major portion of the equipment is located behind the production bay fab wall in the service chase (see Figure 6.22 on page 132). Only the user-interface operator's controls and the wafer cassettes are located in the fab. This configuration isolates the equipment and its servicing in the chase, which is typically at a lower-class level of contamination.

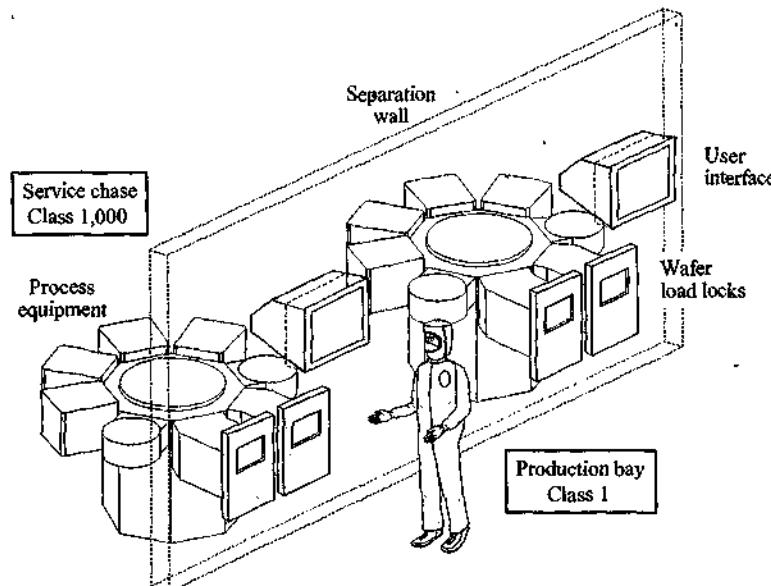


FIGURE 6.22 Bulkhead Installation

Handling ■ From the early days of semiconductor manufacturing until the 1970s, wafers were handled either manually with tweezers and vacuum wands (wafer holders that use vacuum to hold the wafer). Manual handling created particulate contamination and caused killer defects as device geometries decreased. Eventually manufacturers used *wafer cassettes* to transport wafers between tools (typically 25 wafers per cassette), with conveyor systems and elevators as a way to pick up and transfer wafers into and out of the tool (see Figure 6.23). Cassettes are designed to minimize particle generation, be electrostatic dissipative, and have minimal chemical outgassing.

Currently wafer handling within tool operations is largely done by robotic technology. *Robotic wafer handlers* do most cleanroom wafer handling to load and unload wafers from cassettes to the work area of the process tools and to manipulate wafers during the defined sequence of operations in a tool. This development has significantly reduced wafer particulate contamination over manual handling.

Once wafers arrive at a process chamber, they are placed on a chuck that holds the wafer during processing. In the early days of semiconductor manufacturing, this chuck was a mechanical clamp. It changed to a vacuum chuck to reduce particles on the topside of the wafer. However, a vacuum chuck tends to distort the shape of the wafer, which is undesirable during processing. In

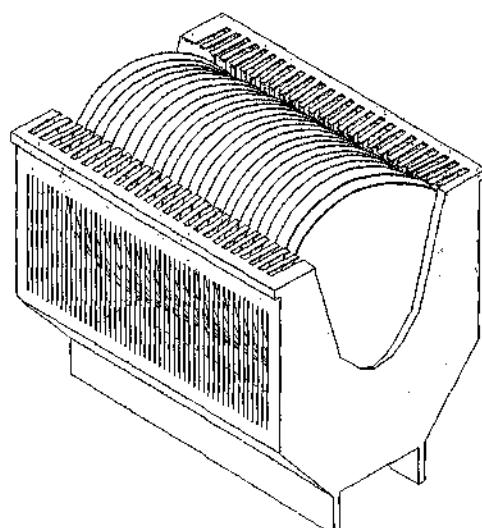


FIGURE 6.23 Wafer Cassette

order to improve process uniformity across the wafer surface, an *electrostatic chuck (ESC)* is commonly used today. It generates minimum particles and holds the wafer flat during processing (see Figure 6.24). An electrostatic chuck works by applying a voltage to an electrode on the chuck and generating a static charge. This electrode is isolated from the rear surface of the wafer by means of a dielectric material. An opposite charge is induced into the bottom side of the wafer, which pulls the wafer toward the chuck.

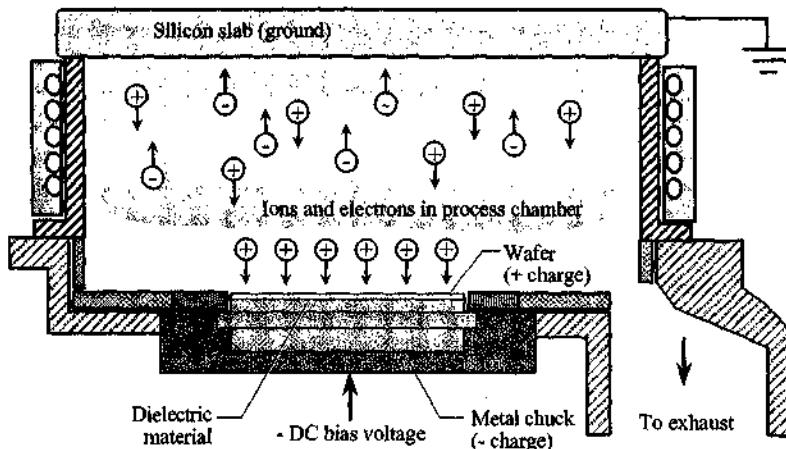


FIGURE 6.24 Electrostatic Wafer Chuck
(Used with permission from Applied Materials, Inc.)

Minienvironment ■ The concept of a cleanroom is continually being reassessed, primarily due to the need for more stringent control of contaminants and the high cost of constructing cleanrooms. There has been increased interest in controlling contamination at the specific workstation location where the wafer is processed through the use of minienvironments.

A *minienvironment* is a localized environment created by an enclosure that isolates wafers from the cleanroom environment while they are not in a process chamber (see Figure 6.25). This concept is also referred to as *wafer isolation technology*. A minienvironment clean zone can include the wafer cassettes used to hold wafers, wafer processing stations, loading ports, and storage locations.

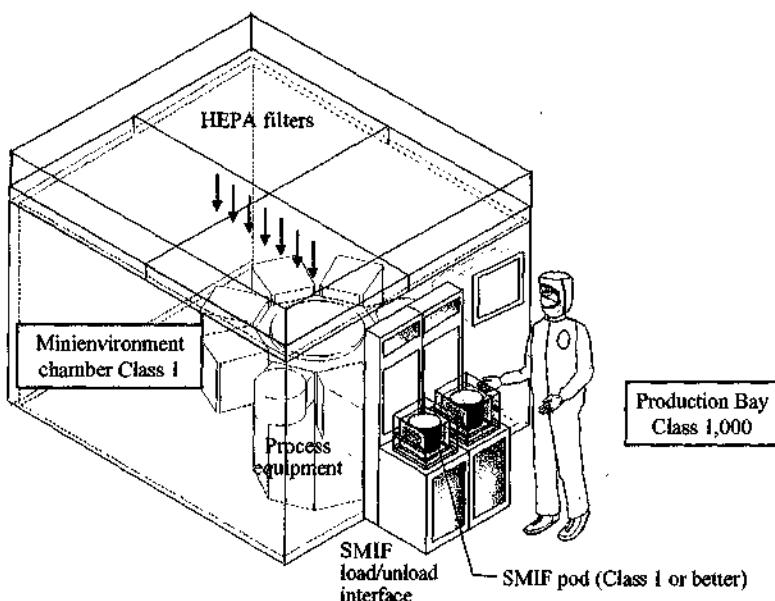


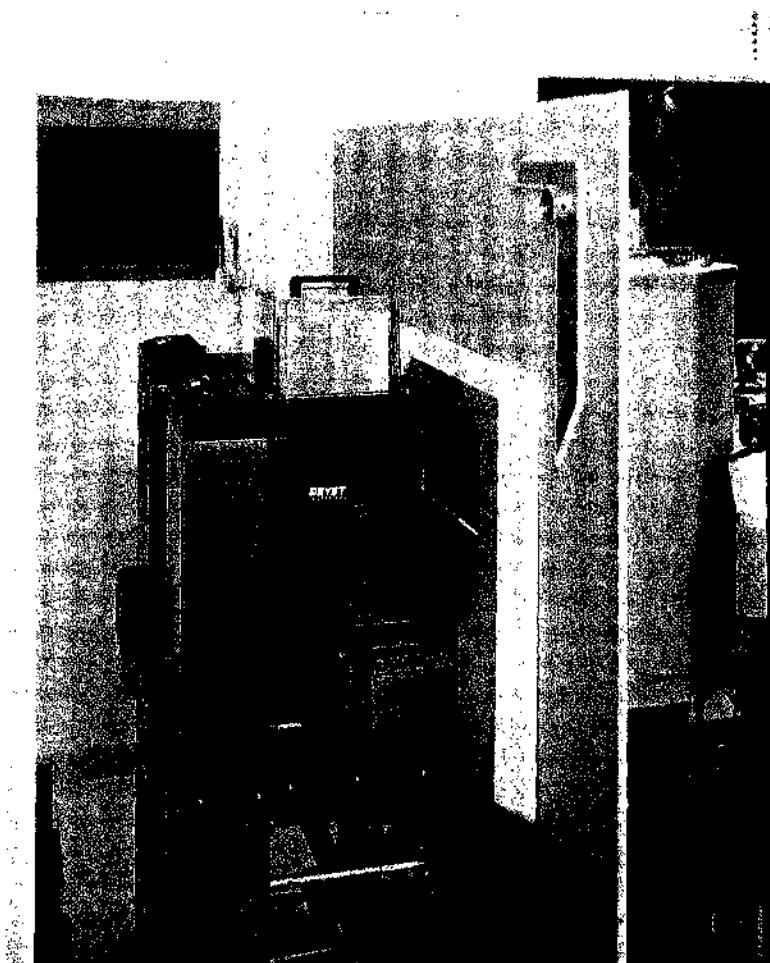
FIGURE 6.25 Minienvironment Concept

Minienvironments are controlled to an extremely pure cleanliness class (e.g., class 0.1 at 0.1 micron), while the cleanroom itself is at a higher class, such as a class 10. This condition makes it much simpler to achieve an ultraclean environment around the wafer during load/unload and processing at a workstation.

To transfer wafers between tools within a minienvironment, a standardized pod is used to enclose and transport cassettes of wafers. The pod has a *standard mechanical interface (SMIF)* to interface with the various tools, which was originally developed and patented by Hewlett-Packard. When a pod is presented to a tool, a robot on the tool automatically opens a door on the pod, removes the cassette, and presents it to the tool robot. SMIF systems can be added to existing tools or integrated into the tool enclosure.

For fabs processing 300-mm wafers, the cassette and pod will become one, permitting wafers to be handled directly from the pod. This new pod is referred to as a *front-opening unified pod (FOUP)*. Handling will be completely automated with no manual lifting of wafer pods. The SMIF interface is fully integrated into the tool.

An advantage of a minienvironment is the possibility to control molecular contamination.³⁵ Wafers are isolated from molecular contamination through inert-gas purging of the minienvironment with nitrogen during loading, unloading, and transport. This control would be much more difficult to achieve in a large cleanroom. Minienvironments also reduce wafer exposure to water vapor during the nitrogen purge, which serves to inhibit native-oxide growth.



SMIF Pod on Bulkhead Installation
(Photo courtesy of Applied Materials, Inc.)

WAFER WET CLEANING

Due to the continual reduction in critical dimensions on a wafer, the wafer surface must be clean prior to undergoing processing. The most important way to control contamination on a wafer is to prevent contaminating the wafer. However, once a wafer surface is contaminated, then the contaminant must be removed by cleaning.

The goal for wafer cleaning is to remove all surface contaminants: particles, organics, metallics, and native oxides. Every wafer process step is a potential source of contamination to the devices on the wafer. Throughout the entire ULSI fabrication process, it is estimated that the surface of an individual wafer is wet cleaned up to 100 times.³⁶

Wet-Cleaning Overview

The predominant wafer surface cleaning process is with *wet chemistry*. In the early 1980s, the consensus was that wet cleaning would be replaced by dry cleaning methods by the turn of the century. There has been a substantial effort to achieve this, but no completely successful replacement for wafer wet cleaning has been found. Wet cleaning of wafers is thriving and is being improved to attain more effective surface cleaning. The typical chemicals used in wafer wet cleaning and the contaminants they remove are shown in Table 6.5.³⁷

TABLE 6.5 Wafer Wet-Cleaning Chemicals

Contaminant	Name	Chemical Mixture Description (all cleans are followed by a DI water rinse)	Chemicals
Particles	Piranha (SPM)	Sulfuric acid/hydrogen peroxide/DI water	H ₂ SO ₄ /H ₂ O ₂ /H ₂ O
	SC-1 (APM)	Ammonium hydroxide/hydrogen peroxide/DI water	NH ₄ OH/H ₂ O ₂ /H ₂ O
Organics	SC-1 (APM)	Ammonium hydroxide/hydrogen peroxide/DI water	NH ₄ OH/H ₂ O ₂ /H ₂ O
	SC-2 (HPM)	Hydrochloric acid/hydrogen peroxide/DI water	HCl/H ₂ O ₂ /H ₂ O
Metallics (not Cu)	Piranha (SPM)	Sulfuric acid/hydrogen peroxide/DI water	H ₂ SO ₄ /H ₂ O ₂ /H ₂ O
	DHF	Hydrofluoric acid/water solution (will not remove copper)	HF/H ₂ O
Native Oxides	DHF	Hydrofluoric acid/water solution (will not remove copper)	HF/H ₂ O
	BHF	Buffered hydrofluoric acid	NH ₄ F/HF/H ₂ O

One of the most critical surface cleaning process steps during wafer fabrication occurs prior to growing a thermal oxidation layer on the wafer. Ultrathin oxide layers must start with a completely clean wafer surface.

RCA Clean ■ The industry standard wet-clean process is referred to as the *RCA clean* process, developed by W. Kern and D. Puotinen at RCA in the 1960s and first published in 1970. RCA wet clean consists of sequential immersion in two different chemical baths, Standard Clean 1 (SC-1) and Standard Clean 2 (SC-2).

The chemical mixture of *Standard Clean 1 (SC-1)* is NH₄OH/H₂O₂/H₂O (ammonium hydroxide/hydrogen peroxide/DI water). The three chemicals are mixed with a ratio range of 1:1:5 to 1:2:7. *Standard Clean 2 (SC-2)* is a composition of HCl/H₂O₂/H₂O (hydrochloric acid/hydrogen peroxide/DI water), and is mixed in a ratio range of 1:1:6 to 1:2:8.³⁸ Both of these chemical mixtures are based on hydrogen peroxide (H₂O₂) and are traditionally used at a temperature of 75 to 85°C, with a 10 to 20 minute exposure time.

Standard Clean 1 (SC-1). As seen in Table 6.5, the SC-1 clean is an alkaline solution capable of removing particles and organic materials. For particles, the SC-1 wet-chemical process

works primarily through oxidation of the particle or by electric repulsion.³⁹ To understand the oxidation mechanism, realize that hydrogen peroxide is a powerful oxidizing agent that oxidizes the wafer surface and the particle. The oxidation layer on the particle can provide a liftoff mechanism that degrades and dissolves the particle, which breaks the adhesion forces between the particle and the surface. The particle then becomes soluble in the SC-1 solution and leaves the surface. This action is shown in Figure 6.26. This oxidizing action from the hydrogen peroxide also forms a protective layer on the silicon surface that keeps the particle from reattaching to the wafer surface.

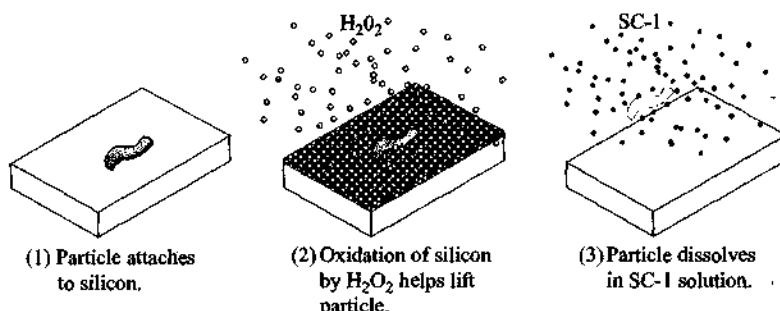


FIGURE 6.26 Oxidation and Solubility of Particle in SC-1

The SC-1 particle removal mechanism actually achieves an electrical repulsion of a particle. The hydroxide ion (OH^-) from ammonium hydroxide (NH_4OH) slightly etches the wafer surface and undercuts beneath the particle. The hydroxide ion also builds up a negative charge on the silicon surface and the particle. This negative charge on the particle and surface serves to repulse the particle from the surface and move it into the SC-1 solution (see Figure 6.27). Another benefit of the negative surface charge is that it prevents particles from redepositing.

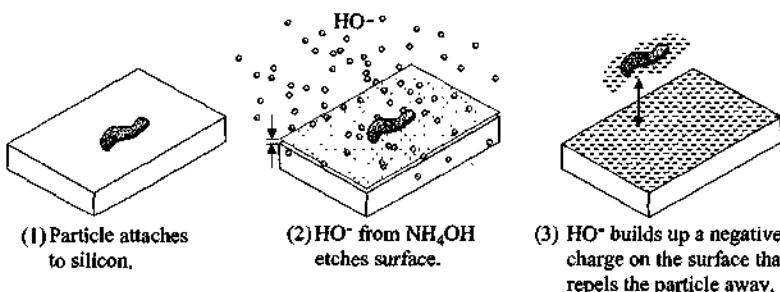


FIGURE 6.27 Particle Removal Through Negative-Charge Repulsion

Because the SC-1 step removes particles through an etching process on silicon, there is some microroughening of exposed silicon. This microroughening is a concern because it makes it difficult to grow very thin oxidation layers required in ULSI technology. Another concern regarding the use of SC-1 is a documented correlation between metal contamination on the wafer and the decomposition of the hydrogen peroxide in the mixture.⁴⁰ This contamination requires the bath to be replaced frequently, which is undesirable because it increases the use of chemicals in the fab.

It should be noted that the buildup of a charge potential on the wafer surface and the particle could be described by their zeta potential.⁴¹ For particle removal, the zeta potential of the wafer and particle should be in the same polarity direction in the solution in order to have repulsive electrical forces between the wafer and particle.

Standard Clean 2 (SC-2). The SC-2 wet-clean process is used to remove metals from the surface of the wafer. To remove metallic (and some organic) contaminants from the silicon surface, it is necessary to have a solution with a high oxidation potential and low pH. In this manner, metals become ionized and dissolve in the acid solution that has a strong oxidizing action. Now the

cleaning solution can capture electrons from the metallic and organic contaminants and oxidize them. Metals are ionized to be dissolved in the solution while organic impurities are decomposed.

Modifications to RCA Clean ■ There have been modifications to the RCA clean process, mainly because this process uses chemicals and ultrapure water in high volume and at high temperature.⁴² Very few chip manufacturers still use RCA clean in the same ratio as the original solution. Companies now will use mixtures that are up to 100 times more dilute with ultrapure H₂O, referred to as *dilute cleaning chemistries*, which achieves equal to or better cleaning effectiveness as the original solution.⁴³ For instance, a dilute SC-1 chemical ratio may be 1:4:50 for NH₄OH:H₂O₂:H₂O instead of the traditional 1:1:5 ratio. Dilute chemistries are an improvement for safety and health, plus they have the cost benefits of reduced chemical usage and disposal.

An important reason for the continued success of the RCA wet-clean is the availability of ultrapure water and chemicals. New cleaning approaches such as point-of-use chemical generation provide even higher levels of purity than ever before, which lead to more effective cleaning action. RCA clean does generate a large amount of chemical vapor, increasing the load on the cleanroom exhaust system to keep the chemical vapors from getting into the cleanroom. Another problem with bath evaporation is its effect on changing the bath composition over time.

Piranha Mixture. *Piranha* is a strong cleaning solution that combines sulfuric acid (H₂SO₄) and hydrogen peroxide (H₂O₂) to remove organic and metallic impurities on the wafer surface. Piranha is used at different steps in the process, sometimes before the SC-1 and SC-2 cleaning steps. The most common mixture is seven parts of concentrated H₂SO₄ to three parts of 30% (by volume) of H₂O₂. The usual cleaning method is to immerse the wafers in the piranha at 125°C for about 19 minutes followed by a thorough DI-water rinse. A variation of piranha is *Caro's acid*, which is prepared by mixing 380 parts of concentrated H₂SO₄ with 17 parts of 30% H₂O₂ and 1 part of ultrapure water.

HF Last Step. Many cleaning steps expose the wafer surface to a last step of hydrofluoric acid (HF) to remove native oxides on the wafer surface. Native oxide-free silicon is critical for producing a high-purity epitaxy film and ultrathin (50 angstroms and less) oxide layers for the gate region of MOS circuits. After the HF exposure, the wafer surface is completely terminated with hydrogen atoms and has a high stability against reoxidation in air.⁴⁴ A hydrogen-terminated silicon surface is maintained in the same condition as if it were bulk silicon crystal.⁴⁵

Chemical Vapor Cleaning. Another method used in a few fabs is a chemical vapor to remove residual oxide and metallic contamination from a single wafer in a process chamber. The wafer is exposed to a fine mist spray of dilute HF:H₂O, followed by a DI-water rinse and an IPA (isopropyl alcohol) vapor-dry step. This method was developed to minimize HF chemical use, but it is not widely used because the increase in cleaning performance is minimal.

Wafer Cleaning Steps ■ A typical wafer cleaning sequence is shown in Table 6.6⁴⁶ on page 138. There are variations where some of the HF/H₂O steps are omitted.

TABLE 6.6 Typical Wafer Wet-Clean Sequence

Cleaning Step	Purpose
$\text{H}_2\text{SO}_4/\text{H}_2\text{O}_2$ (Piranha)	Organics and Metals
UPW rinse (ultrapure water)	Rinse
$\text{HF}/\text{H}_2\text{O}$ (dilute HF)	Native oxides
UPW rinse	Rinse
$\text{NH}_4\text{OH}/\text{H}_2\text{O}_2/\text{H}_2\text{O}$ (SC-1)	Particles
UPW rinse	Rinse
$\text{HF}/\text{H}_2\text{O}$	Native oxides
UPW rinse	Rinse
$\text{HCl}/\text{H}_2\text{O}_2/\text{H}_2\text{O}$ (SC-2)	Metals
UPW rinse	Rinse
$\text{HF}/\text{H}_2\text{O}$	Native oxides
UPW rinse	Rinse
Drying	Dry

Wet-Clean Equipment

Because of the extensive use of wet cleaning in semiconductor manufacturing, the type of equipment used is a factor in reducing the chemical concentration and use of chemicals during wet cleaning. Traditional wet-clean processing has been done in *wet sinks*, consisting of a series of acid and rinse tanks housed in fume hoods. Self-contained cleaning equipment with microprocessor controls, robotic handling, and auto-dispensing of chemicals is common. The trend for wet cleaning and rinsing is for robotic handling of wafers with cassetteless operation within a minienvironment. Without the cassette, there is less obstruction of chemical flow to the wafer surface. This condition can improve the cleaning efficiency, reduce the amount of chemical usage, and lead to shorter rinse times.

Megasonics ■ One of the most widely used technologies with SC-1 for wet cleaning is megasonic cleaning. *Megasonic cleaning* uses ultrasonic energy with frequencies near 1 MHz during the clean process (see Figure 6.28). This process achieves much more effective particle removal at lower bath temperatures (30°C versus the original 80°C). This fact is important because of the difficulty in removing smaller particles, simply because of the difficulty in delivering the necessary force to such minuscule particle dimensions.⁴⁷

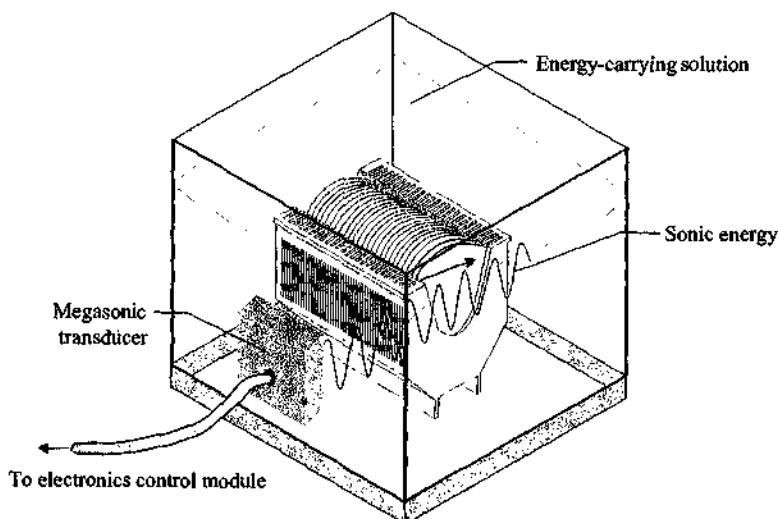


FIGURE 6.28 Megasonic Cleaning

Megasonic cleaning is accomplished when the vibrations of a megasonic transducer creates pressure waves in the liquid of the cleaning tank. The main particle removal mechanisms with megasonic cleaning are cavitation (formation of bubbles) and acoustic streaming.⁴⁸ *Cavitation* occurs when the low-pressure portion of the pressure waves create small bubbles that become filled with gas and/or vapor. These cavity bubbles oscillate through the liquid medium and violently collapse (implode) due to the sonic energy. This collapsing action is cavitation, which enhances particle removal without damage to the silicon wafer.⁴⁹ *Acoustic streaming* is the steady flow of liquid induced in the megasonic tank from the ultrasonic energy. Flowing liquid has more cleaning action than stationary water because it transports the particles away from the surface.

When the vibration frequencies are less than 100 kHz, then the process is referred to as ultrasonic. However, cavitation-induced pitting of silicon wafers occurs at ultrasonic frequencies that have not been found in the megasonic frequency range (800-1200 kHz).⁵⁰ Another reason for megasonics becoming more widespread in chemical cleaning and subsequent DI-water rinse operations is because it reduces the volume of chemicals required.

Spray Cleaning ■ With *spray cleaning* technology, the wet-cleaning chemicals are sprayed onto wafers, placed in a cassette and rotated inside a sealed chamber (see Figure 6.29). A DI-water rinse is sprayed on the wafers after each cleaning step and the DI water's resistivity is monitored to determine when all the chemicals are removed. The spray chamber is sealed during the process to isolate the chemicals and their vapors. After completion of the cleaning and rinsing cycles, the chamber is purged with hot nitrogen and the spin rate is accelerated to dry the wafers.

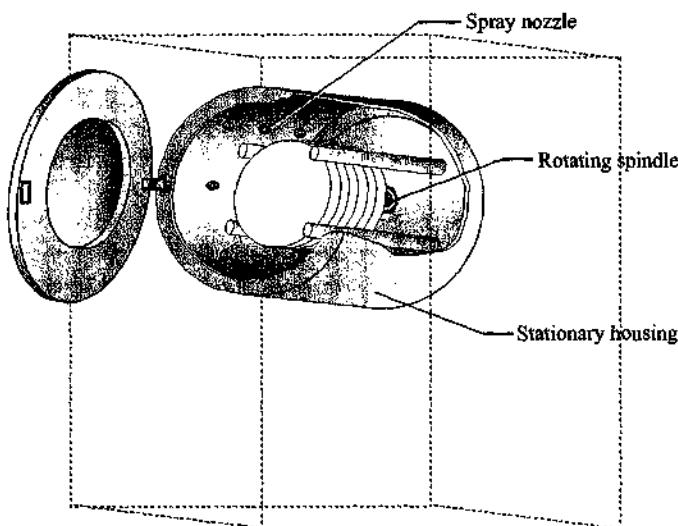


FIGURE 6.29 Spray Tool Designs for Wafer Cleaning

An advantage of spray tools is a continuous supply of premixed, freshly blended chemicals sprayed on the wafer. There is a physical force created during cleaning from the many small droplets of the spray hitting the wafer with a high turnover of chemicals and rinse water. The use of spray and centrifugal force from the rotating wafer ensures effective rinsing at a reduced chemical and water usage rate. There are also improvements in metallic and organic contamination removal when using spin-spray technology. However, spray cleaning does not provide uniformity of cleaning and rinsing because the center of the wafer is not turning at the high velocity of the outer wafer edge. This problem becomes worse with larger diameter wafers.

Scrubbers ■ *Wafer brush scrubbing* is an effective method for removing particles from the wafer surface. Brush scrubbing is widely used following wafer chemical mechanical planarization (known as CMP, which will be discussed in Chapter 18) due to the extensive particles generated during CMP. Brush scrubbing is able to remove particles one micron in diameter and smaller.

Early versions of wafer scrubbers made with nylon brushes damaged the wafer surface because of the stiffer nylon in conjunction with high pressure water sprays. Modern brushes are made with polyvinyl alcohol (PVA), which is a soft, compressible, spongelike material (see Figure 6.30). PVA brushes are effective at removing particles without wafer damage.⁵¹ Brush scrubbers are available in a double-sided version that brushes both sides of the wafer at once. Brush scrubbers are often used with room temperature solutions of nontoxic chemicals or DI water that is sprayed on the wafer while brushing is underway.

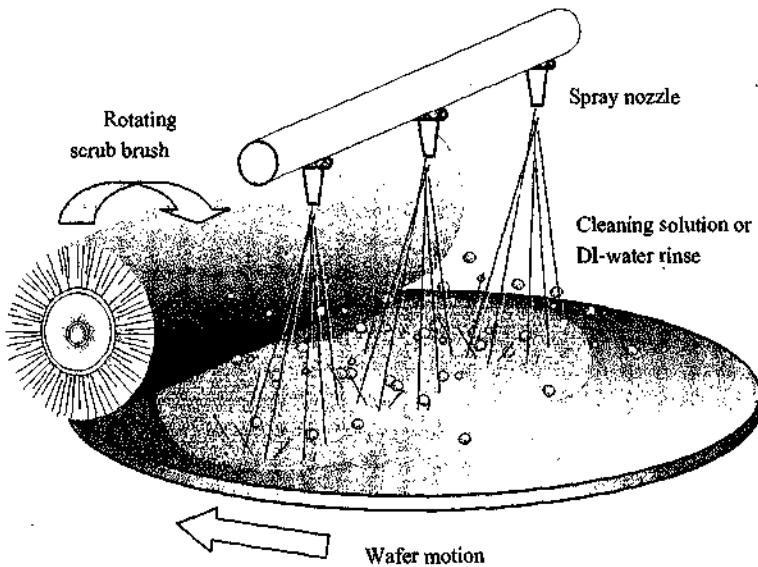


FIGURE 6.30 Wafer Brush Scrubber

Wafer Rinse ■ Chemical residues must be removed from the wafer after cleaning. Every wet-clean process step is followed by an ultrahigh purity *DI-water rinse*. This rinsing step requires a continuous supply of clean water to the wafer surface. The water rinse can also serve to end oxide-etching action from an HF cleaning step.

Overflow Rinsers. Traditionally the most common type of DI-water rinse has been the *overflow rinser* (see Figure 6.31). DI water is brought into the rinse system to flow through and around the wafers, sometimes with a nitrogen bubbler to aid in the mixing of chemicals at the wafer surface. The fluid motion in overflow rinsers serves to sweep away contaminants that have diffused into the flow stream from the surface of the wafers. High flow rates with no dead spots are goals for rinsing.

The overflow rinse process has also been applied to a cascade rinse system. In this case, the DI water cascades between two or three overflow rinsers connected to each other. The wafers start

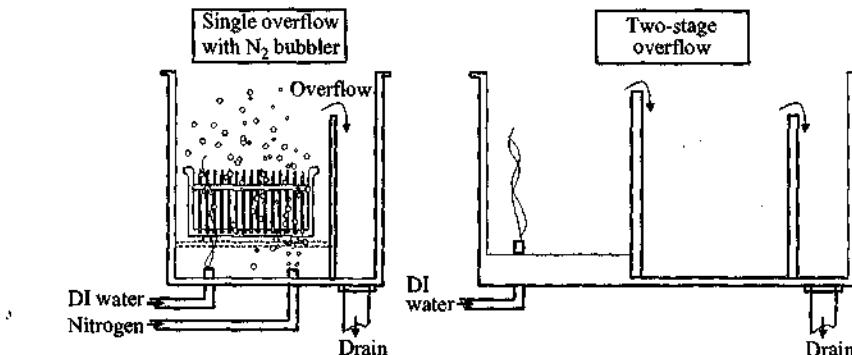


FIGURE 6.31 Overflow Rinser

the rinse process in the downstream rinser and are moved sequentially through the rinsers to the first one with the direct DI-water supply.

Overflow rinsers are the conventional system, but they consume a large amount of DI water. Given the large amount of water used in semiconductor facilities, there has been extensive investigation into alternative rinse processes that consume less water.

Dump Rinse. A common rinse method is the *dump rinser*. DI water is sprayed on the wafers while filling the rinse tank. At a certain water level, a drain in the bottom of the tank quickly opens and the water instantly dumps out (see Figure 6.32). The drain then closes and the cycle repeats for a set number of cycles. An inert gas, such as nitrogen, is often bubbled through the water to aid in removing contaminants such as particles by creating a scrubbing action.

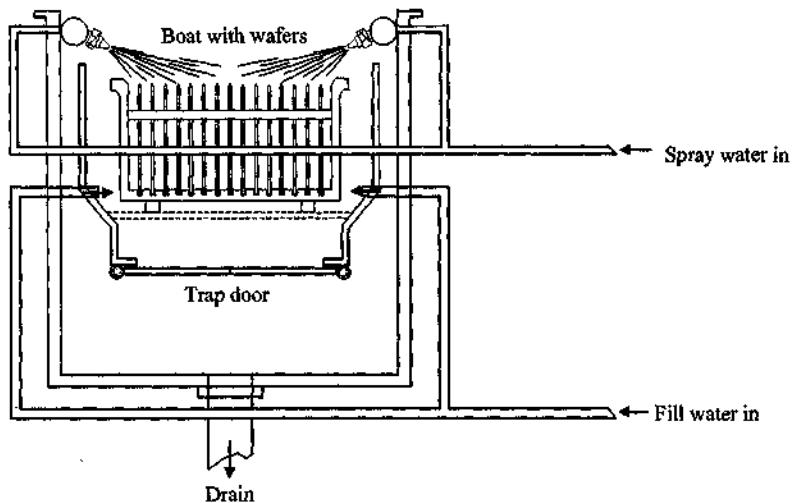


FIGURE 6.32 Dump Rinse

Spray Rinse. A *spray rinser* is typically used in conjunction with wafer cleaning and drying. It uses the physical force of the flowing water to dislodge residual chemicals on the wafer surface.

Hot DI-Water Rinsing. *Hot DI-water rinsing* (70 to 80°C) has become widely used for rinsing the wafer. The benefits of hot DI-water rinsing are that it aids in the removal of residual chemicals (especially if some type of HF solution has been used) and it improves the performance of the dry wafer. However, there has been some research that shows that the use of DI-water rinse at elevated temperatures creates a problem with etching the silicon surface, leading to surface microroughness.⁵² Surface microroughness is undesirable for the extremely thin films required in ULSI technology.

Wafer Drying ■ Given the amount of aqueous rinse in a wafer process, it is important to dry the wafer, ideally with no drying spots. How a wafer responds to water refers to its wettability. Water will either adhere to a clean hydrophilic wafer or bead up on a hydrophobic surface. A *hydrophilic surface* has an affinity toward water, allowing water to spread across it in large puddles. Oxide-coated and RCA-cleaned wafer surfaces are hydrophilic. A *hydrophobic surface* has an aversion to water and does not support large pools of water. The water generally beads up, such as water on a newly waxed car, and is called *dewetted*. Oxide-free surfaces that have been hydrofluoric acid (HF) etched are hydrophobic due to the hydrogen-terminated surface. As the wafer undergoes processing, semiconductors require a hydrophobic surface to promote adhesion during subsequent layering processes.⁵³ Because HF-cleaned wafer surfaces are susceptible to contamination, there must be an adequate effort to properly dry the wafer surface.

Spin Dryers. The *spin dryer* has been widely used in the past. A cassette of wafers is placed in the spin equipment and high-speed rotation removes moisture while the wafers are being sprayed

with heated nitrogen. However, there are problems, such as the difficulty of removing moisture in holes and the generation of particulate contamination from the mechanism. In addition, the high-speed rotation of the wafer causes a charge buildup, which attracts particles. Adding static eliminators inside the equipment controls this static electricity buildup.

IPA Vapor Dry. In this process, wafer drying occurs through the displacement of water by heated solvent vapors of isopropyl alcohol (IPA). This *IPA vapor dry* method has a lower probability of contamination from particles, but the purity level of the IPA must be controlled. The IPA is heated in a tank with the wafers suspended in the vapor above the liquid level. When the wafers are removed from the vapor dryer, the solvent vapors evaporate, leaving the wafer dry.

Alternatives to RCA Clean

There are several alternative cleaning technologies to RCA clean that have been researched and used in varying degrees.

Dry Cleaning ■ Wet-cleaning methods are still meeting most cleaning requirements in wafer fabs. Dry-surface cleaning techniques are implemented primarily in integrated cluster tools, rather than stand-alone tools. Ultimately, the cost benefits through increased yield or reduced cost of ownership will decide the frequency of dry cleaning in wafer fabrication.

Plasma-Based Dry Cleaning. *Plasma-based dry cleaning* must be considered as a serious alternative to wet clean. Today, dry-plasma technology is used for removing organic photoresist from wafers (known as *ashing*), or as an integrated preconditioning step in process tools to remove native oxides. In the plasma process, gases and plasma energy are used to cause the chemical reaction that removes the contaminant (plasma processes will be discussed in detail in later chapters).

An example of plasma cleaning is a microwave downstream-plasma process.⁵⁴ This plasma can remove film residues containing organic material plus metallic and other inorganic components. The plasma can remove the sidewall film while minimizing removal of the underlying dielectric layer. This plasma clean is followed by a DI-water rinse.

There will be further development and use of dry-plasma cleaning processes for semiconductor manufacturing since the advanced processes required for 0.18- μm critical dimensions and below are using new materials with different forms of residues (within the same classes we have defined). In many cases, the chemical reactions of wet cleaning do not possess enough activation energy (energy required to cause a reaction) to remove all the residues. This is why the higher-activation energies associated with plasma techniques and nonplasma methods are needed for wafer cleaning.

Chelating Agents ■ A *chelating agent* binds and removes metallic ions. If this agent, such as ethylenediamine-tetra-acetic acid (EDTA) is added to a cleaning solution, it can reduce the redeposition of metals in the solution. It does this by altering the reduction-oxidation potential of the metallic species. An example would be adding a chelate to the ammonium hydroxide of SC-1 as a means of preventing metals from any chemical impurities from adhering to the wafer surface.

Ozone ■ *Ozone* (O_3) injected into ultrapure water has been identified as a possible replacement for some RCA wet cleaning. It has been shown that a combination of ozonized ultrapure water followed by an SC-2 cleaning step is effective in removing metals such as copper (Cu) and silver (Ag), as well as organic contaminants.⁵⁵ Research has indicated that ozone injected into DI water could replace a piranha step used for light organic cleaning.

Cryogenic Aerosol Cleaning ■ The principle of *cryogenic aerosol cleaning* is to sufficiently cool a gas (argon) to form solid ice particles that are injected onto a wafer surface to remove particulate contamination. The cooling of the argon is done by expansion cooling (lowering of pressure) in a vacuum chamber as a mixture of argon and nitrogen flows through a jet nozzle array. The nitrogen is used to dilute the argon and control the diameter of the solid argon particles. Research data has shown that cryogenic aerosol cleaning can be superior to wet cleaning. A benefit of this process is that it is environmentally benign. This process is not currently in wide use.

SUMMARY

Wafer fabrication is done in ultraclean cleanrooms. There are five types of contamination controlled in cleanrooms: particles, metallic impurities, organic contamination, native oxides, and electrostatic discharge (ESD), all of which can affect device performance. Particles must remain less than one-half the critical dimension or else they are known as a killer defect. Air is controlled by filtering and has a class number to designate the particle size and density. Humans must follow strict cleanroom protocol to minimize contamination. Facilities have special floor layouts to minimize the introduction of contaminants, with laminar airflow and HEPA filters used to attain ultraclean air. Air ionization is done to control ESD. Ultrapure DI water controls many types of contaminants through reverse osmosis, ultrafiltration, and bacteria control. Process chemicals and gases have different levels of filtering, transportation, and handling procedures to achieve high purity.

Cleanroom equipment has special workstation designs to minimize contaminants and is becoming more controlled through use of the minienvironment.

The predominant method of wafer cleaning is a wet process using the RCA cleans of SC-1 and SC-2. Particles and organics are removed by SC-1 and metallics are removed by SC-2. Additional wet cleans are the piranha mixture and HF last step. Two common cleaning methods with RCA clean are megasonics and spray cleaning. Brush scrubbers are often used at the chemical mechanical planarization (CMP) operation to remove particles. Different types of DI-water rinse procedures are overflow rinsers, dump rinsers, spray rinsers, and hot DI water. Wafers are dried by spin dryers or IPA vapor dry. Alternatives to RCA wet clean are dry cleaning with plasma and the use of chelating agents, ozone, and cryogenic aerosol cleaning.

KEY TERMS

- | | |
|--|--------------------------------------|
| cleanroom | makeup loop |
| contamination | polishing loop |
| killer defects | water deionization |
| particles | reverse osmosis (RO) |
| aerosols | hyperfiltration |
| particle density | ultrafiltration |
| particles per wafer per pass (PWP) | membrane contactor |
| metal impurities | zeta potential |
| mobile ionic contaminants (MICs) | ultraviolet (UV) lamps |
| organic contaminants | assay number |
| native oxide | filters |
| electrostatic discharge (ESD) | membrane filter |
| class number | filter efficiency |
| ultrafine particles | workstation design |
| cleanroom protocol | bulkhead equipment layout |
| ballroom layout | wafer cassettes |
| bay and chase layout | robotic wafer handlers |
| sub-fab | electrostatic chuck (ESC) |
| laminar airflow | minienvironment |
| high-efficiency particulate air (HEPA) | standard mechanical interface (SMIF) |
| ultra-low penetration air (ULPA) | front-opening unified pod (FOUP) |
| air ionization | wet chemistry |
| ultrapure DI-water | RCA clean |
| total organic carbon (TOC) | Standard clean 1 (SC-1) |
| pyrogens | Standard clean 2 (SC-2) |
| silica | dilute cleaning chemistries |

piranha
Caro's acid
wet sinks
megasonic cleaning
cavitation
acoustic streaming
spray cleaning
brush scrubbing
DI-water rinse
overflow rinser
dump rinser

spray rinser
hot DI-water rinsing
hydrophilic surface
hydrophobic surface
spin dryer
IPA vapor dry
plasma-based dry cleaning
chelating agent
ozone (O_3)
cryogenic aerosol cleaning

REVIEW QUESTIONS

1. Give a general description of a cleanroom.
2. What is contamination in semiconductor manufacturing?
3. Define a killer defect.
4. State the five categories of cleanroom contamination.
5. What is a particle? What is an aerosol? Why are particles a problem for semiconductor manufacturing?
6. Explain the rule of thumb for acceptable particle size in semiconductor manufacturing.
7. Describe the particle density on a wafer surface. What is PWP?
8. Give an example of a typical metal impurity. Give the two basic ways for metallics to deposit on a wafer surface.
9. Describe what an MIC is.
10. Identify a problem that results from metallic impurities in semiconductor manufacturing.
11. What is an organic contaminant? Provide two possible sources of organic contamination in wafer fabrication.
12. Identify two problems that result from organic contamination in semiconductor manufacturing.
13. Explain native oxide. Identify three problems that result from native oxide.
14. Explain electrostatic discharge (ESD).
15. Give three different problems in wafer fabrication resulting from ESD.
16. List seven sources of contamination in a wafer fabrication facility.
17. Explain the class number for air quality.
18. Interpret the following: (a) class 10 at 0.3 μm and (b) class 1 at 0.5 μm .
19. Describe an ultrafine particle.
20. Explain how a human can generate particles.
21. List four criteria that cleanroom garments should meet. Describe the modern cleanroom garment.
22. List seven proper cleanroom protocols that should be followed by cleanroom personnel.
23. Give three strategies for eliminating particles from cleanrooms.
24. Describe the ballroom layout for cleanrooms.
25. What is the bay and chase approach to cleanrooms?
26. What is laminar airflow, and at what class number does laminar airflow become critical?
27. How often may air turnover in a modern wafer fab?
28. What are HEPA and ULPA filters? What is the efficiency of an ULPA filter?
29. What are typical temperature and relative humidity (RH) settings in a wafer fab? Why is RH control important?
30. List and explain three ESD control methods.
31. Briefly describe each of the six unacceptable contaminants in ultrapure DI water.
32. What are the two major parts of a DI-water installation?
33. Explain the deionization of water. At what resistivity level is water considered deionized?
34. Describe reverse osmosis (RO) filtration. What is ultrafiltration?
35. Explain how a membrane contactor filter works.
36. Explain zeta potential.
37. How is bacteria controlled in ultrapure DI water?
38. What is the assay number for chemical purity?
39. List and discuss four classes of filters.
40. What is a membrane filter?
41. Describe the filter efficiency. What is the typical filter efficiency of a membrane filter?
42. Describe the filtration of a process gas.
43. What is the most significant source of particles in a wafer fab?
44. Give four examples of particle contamination from process tools.
45. What is a bulkhead equipment layout?
46. Describe how wafer cassettes reduced contamination over manual handling.
47. How is most cleanroom wafer handling done?
48. Explain the function and purpose of an electrostatic chuck.
49. Describe a minienvironment, and explain why this situation improves contamination control in a cleanroom.
50. What is a SMIF? What is a FOUP? How do they benefit wafer fabrication?

51. What is the goal for wafer cleaning?
52. What is the predominant wafer surface cleaning process?
53. Describe the RCA cleaning process.
54. What chemical mixture is used in SC-1? What contaminants are removed by SC-1?
55. Describe how the SC-1 wet-clean process removes wafer surface particles.
56. State two concerns that arise from cleaning with the SC-1 wet-clean process.
57. What chemical mixture is used in SC-2? What contaminants are removed by SC-2?
58. Explain what are dilute cleaning chemistries.
59. What is a piranha mixture, and what contaminants are removed from the wafer?
60. Discuss the HF last cleaning step and why it is used.
61. List a typical wafer wet-clean sequence. What is a wet sink?
62. Describe megasonics cleaning and why it is used, including cavitation and acoustic streaming.
63. Discuss spray-cleaning technology. What is an advantage from this cleaning method?
64. Describe wafer brush scrubbing. What contaminant does it remove, and at what process step is this method often used?
65. When is a DI-water rinse done? Describe three different methods for wafer rinsing.
66. Explain the difference between a hydrophilic and a hydrophobic surface.
67. Describe two different methods for wafer drying.
68. Discuss plasma cleaning. Why will the usage of dry plasma cleaning increase?
69. What is a chelating agent, and how is it used in cleaning?
70. How could ozone be used for wafer surface cleaning?
71. Describe cryogenic aerosol cleaning.

CHEMICAL AND EQUIPMENT SUPPLIERS' WEB SITES

Adept Technology Inc.

AERONEX Inc.

Air Kontrol Inc.

Air Products and Chemicals

Amerimade Technology

Apex Industries

Applied Science and Technology

Aquionics Inc.

Asahi/America Inc.

Ashland Specialty Co.

ASI Technologies

AST Products

BOC Edwards

Clean Air Products

Contamination Control Products

Dow Chemical/Filmtec

Dow Corning

Dryden Engineering Co.

Eastman Chemical Co.

EKC Technology Inc.

Entegris Inc.

Environflex Inc.

Filtration Technology Inc.

FSI International

General Chemical Corp.

IN USA Inc.

Integrated Designs LP

Ion Systems

J. T. Baker

Kappler Protective Apparel

Koch Microelectronics Service Co.

Meissner Filtration Products Inc.

<http://www.adept.com>

<http://www.aeronex.com>

<http://www.airkontrol.com/>

<http://www.airproducts.com>

<http://www.amerimade.com/>

<http://www.apexind.com/>

<http://www.astex.com/>

<http://www.aquionix.com/>

<http://www.asahi-america.com/>

<http://www.ashchem.com/>

<http://www.asidoors.com/>

<http://www.astp.com>

<http://www.boc.com/edwards/>

<http://www.cleanairproducts.com/>

<http://www.ccpcleanroom.com/>

<http://www.dow.com/liquidseps/>

<http://www.dowcorning.com/>

<http://www.drydeneng.com/>

<http://www.eastman.com/>

<http://www.ekctech.com/>

<http://www.entegris.com/>

<http://www.enviroflex.com/>

<http://www.filtrationtechnology.com/>

<http://www.fsi-intl.com/>

<http://www.genchem.com/>

<http://www.inusaoozone.com>

<http://www.pumpless.com/>

<http://www.ion.com/>

<http://www.jtbaker.com/>

<http://www.kappler.com/>

<http://www.kochmicroelectronic.com/>

<http://www.meissner.com/>

MicroChem Corp.	http://www.microchem.com/
Micro Magazine	http://www.micromagazine.com/
Millipore Corp.	http://www.millipore.com/
Modutek Inc.	http://www.modutek.com
NetMotion Inc.	http://www.netmotion.com/
Pall Corp.	http://www.pall.com/
Parker Hannifin Corp.	http://www.parker.com/
Pope Scientific Inc.	http://www.popeinc.com/
PTI Advanced Filtration Inc.	http://www.pti-afi.com/
PURAC America Inc.	http://www.purac.com/
Schumacher	http://www.schumacher.com
SCP Global Technologies	http://www.scpglobal.com/
Sage Technologies Corp.	http://www.sagetech.net/
Semitool	http://www.semicon.com/
Simco Static Control	http://www.simco-static.com/
TEL, Tokyo Electron Ltd.	http://www.teainet.com
The Texwipe Co. LLC	http://www.texwipe.com/
Ultrapure & Industrial Services	http://www.ultrapure.com/
US Filter/Filterlite	http://www2.usfilter.com/
Verteq Inc.	http://www.verteq.com/

REFERENCES

1. T. Hattori, "Particle Reduction in VLSI Manufacturing," *Contamination Control and Defect Reduction in Semiconductor Manufacturing III* (Pennington, NJ: The Electrochemical Society, 1994) p. 3.
2. A. French, *Newtonian Mechanics*, (New York: W. W. Norton, 1971), p. 32.
3. C. Gross et al., "Assessing Future Technology Requirements for Rapid Isolation and Sourcing of Faults," *Micromagazine* online edition (July/August 1998).
4. T. Hattori, "Detection and Analysis of Particles in Production Lines," *Ultraclean Surface Processing of Silicon Wafers: Secrets of VLSI Manufacturing*, ed. T. Hattori, (Berlin: Springer, 1998), p. 245.
5. W. Kern, *Handbook of Semiconductor Wafer Cleaning Technology*, (Park Ridge, NJ: Noyes Publications, 1993), p. 9.
6. C. Chang and T. Chao, "Wafer-Cleaning Technology," *VLSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 67.
7. G. Cooper, "The Effect of CO Cylinder Materials on Wafer Contamination," *Semiconductor International* (July 1997): p. 301.
8. C. Chang and T. Chao, "Wafer-Cleaning Technology," p. 67.
9. S. De Gendt et al., "Silicon Surface Metal Contamination Measurements Using Grazing-Emission XRF Spectrometry," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997) p. 397.
10. K. Saga, "Influence of Surface Organic Contamination on the Incubation Time in Low-Pressure Chemical Vapor Deposition of Silicon Nitride on Silicon Substrates," *Science and Technology of Semiconductor Surface Preparation Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997) p. 379.
11. T. Ohmi, "Total Room Temperature Wet Cleaning for Si Substrate Surface," *Journal of the Electrochemical Society* (September 1996): p. 2957.
12. C. Chang and T. Chao, "Wafer-Cleaning Technology," p. 80.
13. M. Suzuki et al., "Etching Characteristics During Cleaning of Silicon Surfaces by NF₃-added Hydrogen and Water-Vapor Plasma Downstream Treatment," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997) p. 167.
14. D. Tolliver, *Handbook of Contamination Control in Microelectronics: Principles, Applications and Technology*, (Park Ridge, NJ: Noyes Publications, 1988), p. 175.
15. L. Levit and J. Menear, "Measuring and Quantifying Static Charge in Cleanrooms and Process Tools," *Solid State Technology* (February 1998): p. 85.
16. A. Braun, "Cleanroom Technologies Continue to Keep Contamination at Bay," *Semiconductor International* (March 1998): p. 58.
17. R. Kraft, "Proper Cleanroom Protocol," *Semiconductor International* (March 1998): p. 73.
18. Summarized from R. Kraft, "Proper Cleanroom Protocol," *Semiconductor International* (March 1998): p. 73.
19. R. Jarvis and L. Armentrout, "Full-Fab Surface Particle Detection Improves Yields," *Semiconductor International* (June 1997): p. 199.
20. J. Schroeder, "Automation-Centric Processing Bay Layout," *Semiconductor International* (June 1997): p. 209.

21. H. Tseng and R. Jansen, "Cleanroom Technology," *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 14.
22. S. Middleman and A. Hochberg, *Process Engineering Analysis in Semiconductor Device Fabrication*, (New York: McGraw-Hill, 1993), p. 125.
23. H. Tseng and R. Jansen, "Cleanroom Technology," p. 28.
24. D. Mahoney, "The Construction of a Class 1 Cleanroom: A Case Study," *Semiconductor International* (October 1997): p. 141.
25. H. Inaba, "Effect of Electrostatic Charge on Particle Adhesion on Wafer Surfaces," *Ultraclean Surface Processing of Silicon Wafers: Secrets of VLSI Manufacturing*, ed. T. Hattori (Berlin, Springer, 1998): p. 143.
26. T. Roche, T. Peterson, and E. Hanson, "Water Use Efficiency in Immersion Wafer Rinsing," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477*, (Warrendale, PA: Materials Research Society, 1997): p. 527.
27. R. Mohindra and W. Kern, "New Process for Producing Particle-Free Deionized Water," *Semiconductor International* (July 1997): p. 191.
28. M. Dax, "Membrane Contactor Technology Gives PPB Dissolved Oxygen in Water," *Semiconductor International* (December 1996): p. 36.
29. H. Tseng and R. Jansen, "Cleanroom Technology," *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 39.
30. M. Dax, "Membrane Contactor," p. 36.
31. R. Mohindra and W. Kern, "New Process," p. 194.
32. T. Hattori, "Detection and Analysis of Particles in Production Lines," *Ultraclean Surface Processing of Silicon Wafers: Secrets of VLSI Manufacturing*, ed. T. Hattori (Berlin: Springer, 1998), p. 245.
33. A. Braun, "Cleanroom Technologies Continue to Keep Contamination at Bay," *Semiconductor International* (March 1998): p. 58.
34. T. Hattori, "Trends in Wafer Cleaning Technology," *Ultraclean Surface Processing of Silicon Wafers: Secrets of VLSI Manufacturing*, ed. T. Hattori (Berlin: Springer, 1998), p. 438.
35. R. McIlvaine, "Cleanroom Demands for the Next Generation," *Semiconductor International*, (March 1998): p. 68.
36. T. Ohmi, "Revolution of Silicon Substrate Surface Cleaning," *ULSI Science and Technology Proceedings 97-3*, (Pennington, NJ: The Electrochemical Society): p. 197.
37. T. Hattori, "Trends in Wafer Cleaning Technology," Supplement to *Solid State Technology* (May 1995): p. S8.
38. C. Chang and T. Chao, "Wafer-Cleaning Technology," p. 61.
39. Ibid., p. 64.
40. M. Meuris et al., "The IMEC Clean: A New Concept for Particle and Metal Removal on Si Surfaces," *Solid State Technology* (July 1995): p. 109.
41. T. Ohmi, "Total Room Temperature," p. 2957.
42. Ibid., p. 2958.
43. M. Heyns et al., "Advanced Wet and Dry Cleaning Coming Together for the Next Generation," *Solid State Technology* (March 1999): p. 40.
44. J. Rosamilia et al., "Hot Water Etching of Silicon Surfaces: New Insights of Mechanistic Understanding and Implications to Device Fabrication," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997): p. 181.
45. T. Ohmi, "Total Room Temperature," p. 2958.
46. T. Ohmi, "Total Room Temperature," p. 2961.
47. R. DeJule, "Trends in Wafer Cleaning," *Semiconductor International* (August 1998): p. 64.
48. J. Liu et al., "Si₃N₄ Particle Removal Efficiency Study," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997), p. 27.
49. G. Gale, A. Dai, and I. Kashkoush, "How to Accomplish Effective Megasonic Particle Removal," *Semiconductor International* (August 1996): p. 133.
50. Ibid.
51. R. DeJule, "Trends in Wafer Cleaning," p. 65.
52. J. Rosamilia, "Hot Water Etching," p. 181.
53. J. Park et al., "The Formation of Water Marks on Both Hydrophylic and Hydrophobic Wafers," *Science and Technology of Semiconductor Surface Preparation, Symposium Proceedings 477* (Warrendale, PA: Materials Research Society, 1997): p. 513.
54. K. Lao and W. Wu, "Microwave Downstream Plasma Removes Metal Etch Residue," *Semiconductor International* (July 1997): p. 231.
55. T. Ohmi, "Total Room Temperature," p. 2959.

METROLOGY AND DEFECT INSPECTION

Inspection has been done since the beginning of wafer fabrication. Semiconductor manufacturing veterans can recall when simply observing the color of the oxide film on the wafer surface was acceptable to estimate film thickness. Whichever color tint the oxide film appeared to be was compared to a color chart made up of pieces of wafers with different film thickness associated with each color tint (see Appendix D).

Wafer process inspection technology has undergone major changes. Feature sizes continue to shrink below the quarter micron era and chip density on wafers is continually increasing. Each process step has critical issues that define success or failure: contamination, junction depth, and film quality, to name a few. Furthermore, the introduction of new materials and processes will also bring new ways chips can fail. Measurements

are critical to characterize the wafer materials and verify their acceptability.

To maintain good process yield and improve device performance, wafer fabs have improved control over process parameters and reduced sources of defects during fabrication. Some of these improvements have come from areas such as equipment automation, robotic handling, contamination reduction, and more consistent flow of wafers through the fab to avoid long waiting periods. Other improvements could not have occurred without the ability to measure wafers and evaluate the process performance. This evaluation is done using highly accurate measurement equipment that provides real-time data about the wafer fabrication performance, giving key information to the engineer and technician for the decision-making process.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain why IC metrology is performed, and discuss equipment, yield and data collection issues associated with metrology.
2. Identify twelve different quality measures used in wafer fabrication, and identify the fabrication processes where each is used.
3. Describe the various metrology methods and equipment associated with the different quality measures.
4. List and discuss the purpose of seven different types of analytical equipment used to support IC fabrication.

INTRODUCTION

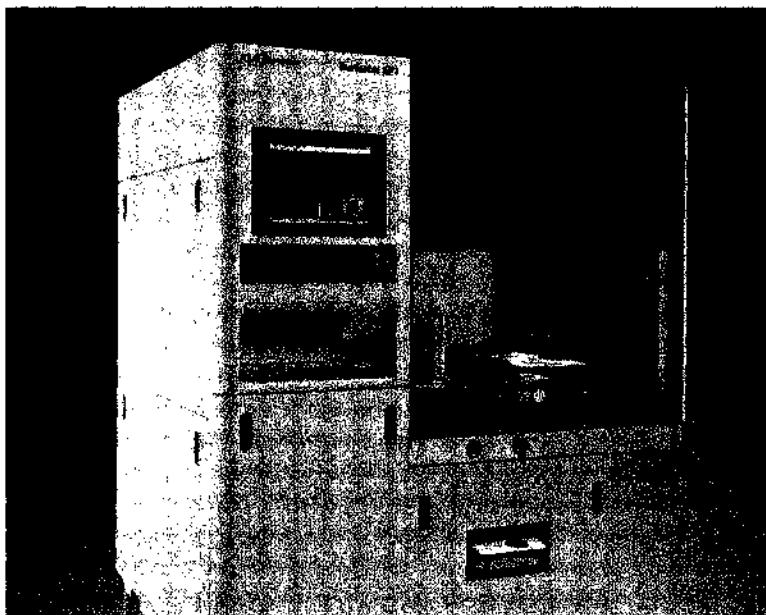
Metrology is the science of measurement to determine dimensions, quantity, or capacity. For IC fabrication, *metrology* refers to the techniques and procedures for determining physical and electrical properties of the wafer during the fabrication process. Metrology used during fabrication employs measurement equipment and sensors to collect and analyze data about wafer parameters and defects. *Defects* are the characteristics of the wafer or the results of the wafer fabrication process that cause nonconformance to the specified wafer requirement. The wafer *defect density* is the number of defects per unit area of wafer surface, usually cm^2 . Wafer defects are specified by type and dimension. Metrology measurements are used by manufacturing personnel (e.g., technicians,

engineers, and managers) to ensure product performance and to make meaningful decisions regarding changes to improve process performance.

It is essential that accurate evaluation of wafers be performed throughout the fabrication process to verify that the product meets the specified requirements. To achieve this, every process step in wafer fabrication has stringent quality measures that define the requirements needed at each step in order for the die to pass electrical test and meet reliability specifications during usage. Quality measures require extensive data collection on either test wafers or production wafers to demonstrate that the chip product's process requirements have been met.

IC METROLOGY

IC metrology is a necessary means for measuring the performance of fabrication processes to ensure the specific quality standards are attained. To accomplish this requires a sample wafer, a measurement tool, and the means to analyze the data. Traditionally, most in-process data has been collected on *monitor wafers* (also referred to as *test wafers*) which are blank (or unpatterned) wafers included during process runs specifically to characterize the process. These blank wafers are recirculated for repeated usage (after undergoing appropriate refurbishing steps, such as surface stripping and cleaning). For instance, monitor wafers are included during thermal oxidation in a vertical diffusion furnace to measure oxide thickness and to check for particles inside the process chamber. Depending on the process step, blank wafers may include a predeposited film layer on the wafer surface.



Unpatterned Surface Inspection System
(Photo courtesy of KLA-Tencor Corporation)

Many semiconductor manufacturers have begun using production wafers, sometimes patterned wafers, for in-line tool monitoring (see Figure 7.1). Using actual production wafers more closely simulates what is happening during the process run, providing the manufacturing team members better information for making decisions.

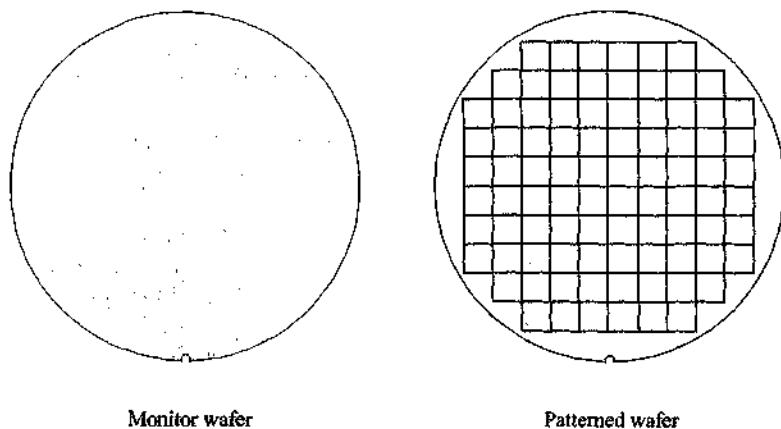


FIGURE 7.1 Monitor Wafer vs. Patterned Wafer

Measurement Equipment

There are different types of metrology equipment used to perform measurements in wafer fabrication. A major way to distinguish these tools is by how the measurement is done—with a stand-alone tool independent of the process or with metrology equipment integrated with the processing equipment. The two dominant categories for measurement equipment are shown in Table 7.1.¹ A *stand-alone measurement tool* functions independent of the process to perform metrology measurements. An *integrated measurement tool* has sensors that permit it to function as a part of the process and deliver *in situ* (real-time) measurement data.

Yield

Yield is an important indicator of a wafer fab's ability to produce high quality die. *Yield* is defined as the percent of good parts produced out of the total group of parts started. For example, if there are 200 die on a wafer and 190 die are acceptable, then the wafer yield is:

$$\text{Yield (\%)} = \frac{190}{200} \times 100 = 95\%$$

Different methods are used to calculate yield. One method is to measure yield relative to the type of part produced for a period of time (e.g., the yield of all wafers produced within a week). Yield can also be measured with respect to a particular location in the process, such as the die yield at the etch process. An important yield measurement for semiconductor manufacturing is wafer sort yield, which indicates the percentage of die that are acceptable after functional test (see Chapter 19).

Yield is widely used in semiconductor manufacturing to describe the health of the fabrication process. A high yield means that the process is producing good parts and performing as intended. A low yield is a strong indicator that there are quality problems in the product design or production process that must be resolved through improvement.

TABLE 7.1 Measurement Tool Classification

Stand-Alone Tools	
Off-line	Available only outside fab (usually destructive or contaminating).
At-line	Available in the fab (measures monitor wafers that are destructive, contaminated or unpatterned).
In-line	Used during production (can measure patterned wafers).
Integrated Tools	
On-line	Available at the process workstation to measure patterned wafers but not able to measure during wafer processing.
In situ	Measures wafer, process, or equipment during processing (real-time measurement).

In situ measurements are a recent development that were not very common even in the mid-1990s. As equipment efficiency improvements occur, there will be more integration of in situ metrology with fabrication equipment.

Data Management

Detecting defects is a challenge in IC fabrication because of shrinking feature sizes. The semiconductor manufacturer must have a method of categorizing the defects, sorting between false and real defects, and determining the root cause of each defect for determining the appropriate corrective action. Defect analysis should differentiate between random or nonrandom causes and correlate defect data with electrical and other tests to track how different defects affect wafer yield.

To a large degree, defect analysis is supported by an integrated software system. Large software programs connect major measurement tools and provide the ability to set up sample plans, analyze defect data, and detect trends amongst defects. Defects must be categorized in order to conduct failure analysis. The most advanced metrology software has *automatic defect classification (ADC)* to identify and classify defects based on software recognition patterns. Advanced defect management software also catalogs wafers and images based on user-defined criteria and includes statistical process control (SPC) capability for process monitoring and control.

QUALITY MEASURES

The breadth of metrology for IC fabrication is demonstrated by the many quality measures used throughout wafer processing. Semiconductor *quality measures* define the requirements for specific aspects of wafer fabrication to ensure acceptable device performance as measured by electrical tests and device reliability. Major quality measures used in the wafer fab are shown in Table 7.2, including the process areas where each measure is applicable. Semiconductor manufacturers specify their particular wafer quality measures by stipulating the exact requirements for their products at each process step. The manufacturer may specify other quality requirements beyond those listed in Table 7.2.

TABLE 7.2 Quality Measures in Wafer Fabrication by Production Areas

Quality Measure	Implant	Diffusion	Thin Films		Polish	Etch	Photo
			Metals	Dielectric			
1 Film Thickness			✓	✓	✓	✓	✓
2 Sheet Resistance	✓	✓	✓				
3 Film Stress			✓	✓	✓		
4 Refractive Index			✓		✓		
5 Dopant Concentration	✓	✓					
6 Unpatterned Surface Defects	✓	✓	✓	✓	✓	✓	
7 Patterned Surface Defects						✓	✓
8 Critical Dimensions (CDs)						✓	✓
9 Step Coverage					✓	✓	
10 Overlay Registration							✓
11 Capacitance-Voltage			✓				
12 Contact Angle							✓

*The diffusion bay processes include: oxidation, deposition, diffusion, anneal, and alloy.

Film Thickness

Since wafer fabrication is a layering process, there are many different types of films found on the wafer surface throughout the fabrication process. Some of these different types of films are metal, dielectric, photoresist, and polysilicon. The quality of these films is essential to a high-yield fabrication process to produce reliable chips.

A critical quality parameter for films are their thickness. *Film-thickness metrologies* can be divided into two general types: whether they measure opaque (i.e., light-blocking, such as metal) films or transparent films. In some instances, such as for the gate oxide dielectric, the film thickness must be measured with an accuracy of an angstrom (\AA) or less. Additional film quality parameters are surface roughness, reflectivity, density, and lack of pinholes and voids.

Resistivity and Sheet Resistance ■ One of the most practical methods for evaluating thickness for conductive films is to measure the *sheet resistance*, R_s . To discuss R_s , we need to understand how resistivity is related to resistance in a thin film layer. Consider current flow through a square sheet of conductive material, with the thickness, length, and width shown in Figure 7.2. Recall from Chapter 2 that resistance, R , of a conductor is given as:

$$R = \frac{\rho(l)}{a} \quad (\text{ohms})$$

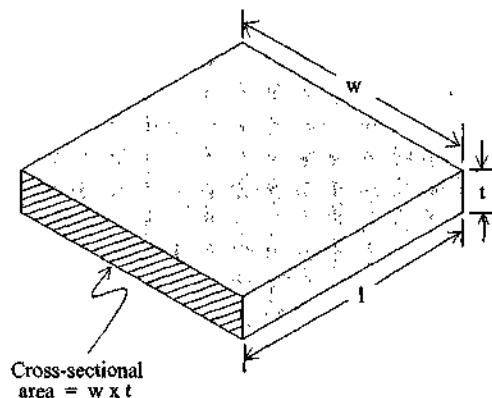


FIGURE 7.2 Illustration of Square Thin Film

The cross-sectional area is given by the product of width and thickness ($w \times t$) and can be substituted in the equation for the area. Based on the assumption that the length (l) is the same dimension as width (w), then l can be substituted for w . Making these substitutions yields:

$$\begin{aligned} R_s &= \frac{\rho(l)}{w \times t} \\ R_s &= \frac{\rho(l)}{w \times t} \\ R_s &= \frac{\rho}{t} \quad (\text{ohms/square, or } \Omega/\square) \end{aligned}$$

Where,

R_s = Sheet resistance in ohms/square

ρ = Film resistivity in ohm-cm

t = Thickness of film

Sheet resistance, R_s , can be interpreted as the end-to-end resistance of a square sample of a thin film on a wafer. It depends on the film resistivity and thickness. Sheet resistance is independent of the size of the square film sheet. Measuring the sheet resistance between two points at the same distance

apart on any size square of the same thickness will produce the same resistance. For this reason, the units for R_s are given as ohms/square (Ω/\square). Dimensionally the units for R_s are the same as ohms, but ohms/square is typically used as a reminder that this is actually the resistance of a sheet of film. If the thickness and sheet resistance are known, then the *sheet resistivity*, ρ_s , can be calculated:

$$\rho_s = R_s(t) \quad (\text{ohms-cm})$$

The terms sheet resistance and sheet resistivity are sometimes used interchangeably since they only differ by the thickness of the film. A tool used for measuring sheet resistance is the four-point probe.

Four-Point Probe. Resistance cannot be practically measured on thin films with two simple probes, as usually done with multimeters commonly used by electronic technicians. This is because of excessive contact resistance at the contact interface between the probes and the wafer material. In the semiconductor industry, a widely used method to measure sheet resistance is the *four-point probe*. This method has four in-line probes equally spaced that touch the wafer surface in a single file (see Figure 7.3). A known value of current (I) is passed between the two outer probes and the potential difference (V) developed across the two inner probes is measured. This approach avoids dealing with the effect of contact resistance.

The spacing between probes, s , should be less than the wafer diameter and less than the film thickness. The resistivity of the film sheet is related to the four-point probe current and voltage by:

$$\rho_s = \frac{V}{I} \times 2\pi s \quad (\text{ohms-cm})$$

Where,

ρ_s = Sheet resistivity in ohm-cm

V = DC voltage across the voltage probes (in volts)

I = Constant DC current passing through the current probes (in amperes)

S = Spacing between the probes

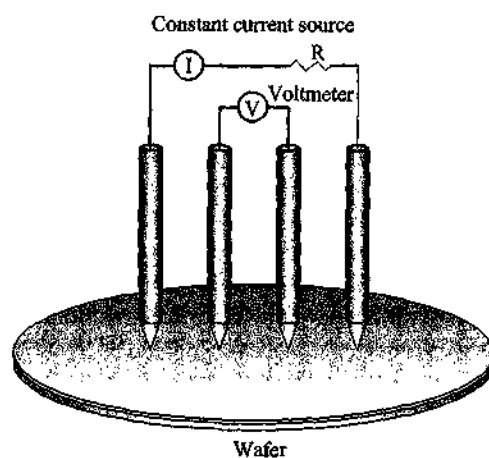


FIGURE 7.3 Four-Point Probe

Resistivity measurements with a four-point probe can be reproduced to within $\pm 2\%$ if care is taken in selecting instrumentation, and in determining probe pressure and current levels.² The four-point probe is a nondestructive technique (e.g., the wafer does not have to undergo permanent damage to perform the test). However, due to the potential for damage to the wafer from the probe contacts, the semiconductor industry has moved toward noncontacting probes.³ One contactless method is the use of eddy-current probes. A high-frequency ac current flowing through a coil induces eddy currents in conducting films placed under the coil. The eddy currents are a loss in energy that can be attributed to the loading effects produced by the resistance of the conducting film. The resulting change in electrical energy can be measured and used in computing the value of the sheet resistance of the film being measured.

Sheet Resistance (Opaque Films) ■ Sheet resistance is used to indirectly measure the thickness of opaque conductive films (e.g., metal, silicide, or semiconductor films) deposited on an insulator substrate such as a silicon wafer. As long as the film layer is large and the probe spacing is small, the sheet resistance, R_s , is given by:⁴

$$R_s = 4.53 \frac{V}{I} \quad (\text{ohms/square, or } \Omega/\square)$$

This formula is actually derived from the equation for sheet resistivity listed on page 154. The constant 4.53 results from a correction factor based on the assumption of an infinitely large sheet of film with small probe spacing. The correction factor can be modified to a different value if the film sheet is not infinitely large.

Van der Pauw. A modification of the four-point probe is the *Van der Pauw Method* that measures sheet resistivity with four probes on the periphery of an arbitrarily shaped sample (see Figure 7.4). The concept is based on measuring electrical current on two corners of a square pattern and measuring voltage on the other two. The results are the same as those achieved with the four-point probe.

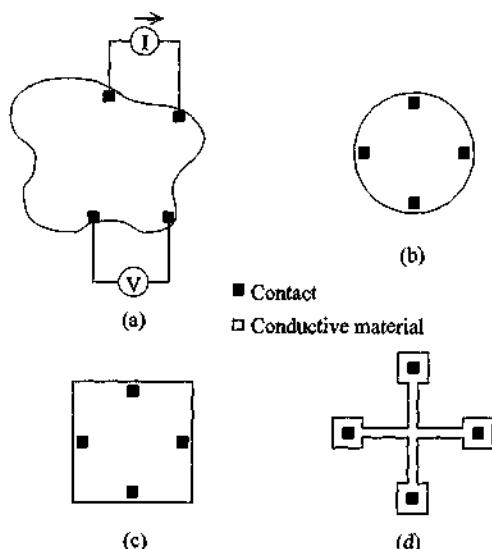


FIGURE 7.4 Van der Pauw Sheet Resistivity

Contour Maps. A practical benefit of sheet resistance measurements with the four-point probe is that the technique can be done at many sites across a test wafer to provide *contour maps* (see Figure 7.5). A contour map will typically present the data with a contour line showing the nominal Ω/\square and then deviations above and below this nominal value. The sheet resistance data results from a predetermined number of measurement sites across the wafer.

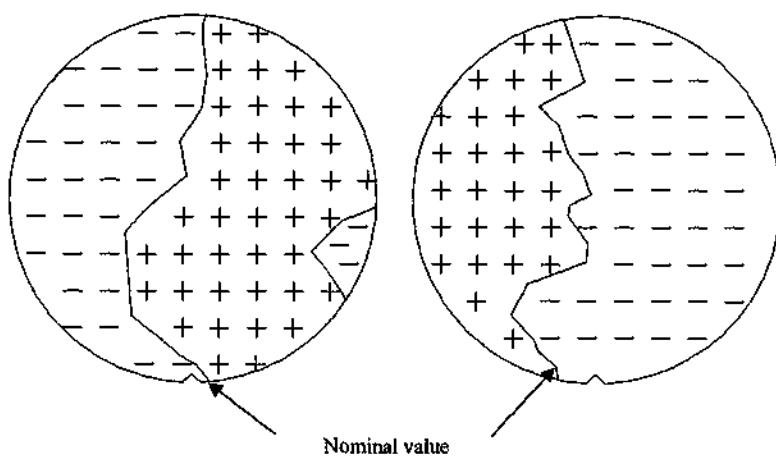


FIGURE 7.5 Sheet Resistance Contour Map

Ellipsometry (Transparent Films) ■ *Ellipsometry* is a nondestructive, noncontact optical film-thickness measurement technique primarily used to measure thin, transparent films. It is the leading method for film thickness metrology in the wafer fab.⁵ The basic principle of ellipsometry is to use a linearly polarized laser light source that, when reflected from the sample, becomes elliptically polarized (see Figure 7.6). Polarized light consists of all light rays traveling in one plane. Ellipsometry measures the shape of this reflected ellipse, and, based on known inputs such as the angle of reflection, it accurately determines the film thickness. It is common in ellipsometry tools for the angle of the incident light to be varied to provide more measurement samples that are optimized for the film material. This practice is referred to as *variable-angle spectroscopic ellipsometry* (VASE). VASE has improved ellipsometry for measuring multilayer stack structures, which is common in ULSI (see Chapter 12 for a description of multilayer stacks).⁶

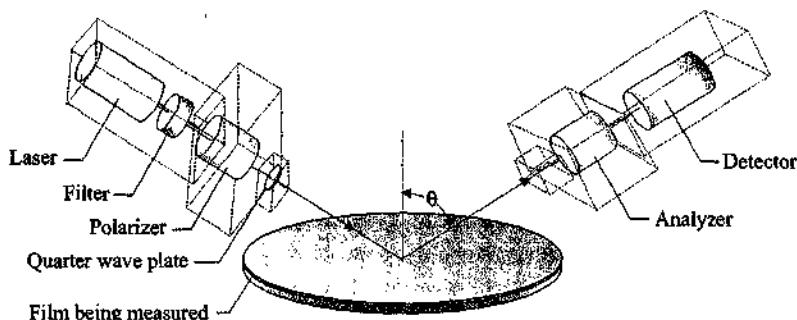


FIGURE 7.6 Basic Principle of Ellipsometry

Spectroscopic ellipsometry analyzes the polarization change of light over a wide spectral range (spectroscopy uses the characteristic light wavelengths emitted by different materials). Information about how the different wavelengths vary relative to one another after being reflected from the film is very useful in a production environment because it makes ellipsometry sensitive to different physical properties of the thin film. Ellipsometry is capable of measuring different types of very thin films on the order of tens of angstroms thick.⁷ The types of films include dielectrics, metals, and polymer coatings. The main requirement is that the film be transparent or semitransparent. Fully automated thin film ellipsometry tools routinely measure gate oxide films < 40 Å thick with a repeatability better than 0.1%. Another example of ellipsometry measurements is sub-100 Å stacked oxide/nitride/oxide (ONO) film structures used in MOS capacitors with measurement repeatability in the 1% range. Thin metal layers (< 500 Å) are considered semitransparent and can be measured with ellipsometry. An example of a semitransparent metal layer is the thin copper seed layer used in copper interconnects. Metal layers (> 1000 Å) are generally considered opaque and cannot be measured with ellipsometry.

Ellipsometry measurement-tool features such as small measurement spots, pattern-recognition software, and high-precision wafer positioning hardware have made it possible to do many transparent film process-control measurements directly on production wafers. This reduction in unpatterned monitor wafers has reduced process-control costs and increased ellipsometry efficiency.⁸ Ellipsometer tools are also being directly integrated into process tools for *in situ* (real time) measurement applications in areas such as etch and planarization. Real-time thickness measurement is desirable because it permits the process to define a precise endpoint of the film thickness during runs for better accuracy and repeatability in film thickness.

Reflection Spectroscopy ■ When light reflects off a surface, *reflection spectroscopy* (also referred to as *reflectometry*) is one of the three general types of optical measurement techniques, with the other two being optical microscopy (discussed later in this chapter) and ellipsometry (see above). Reflection of a structure is often used to characterize layer thickness for light-absorbing dielectric layers on a nonabsorbing wafer substrate (see Figure 7.7). Based on the relationship of how light reflects off the top and bottom surface of the film layer, reflectometry can be used to compute the film thickness. An advanced reflection-spectroscopy technique utilizes a dual-beam

spectrometer in order to give a clearer light signal for imaging. One light beam is used for reflection measurements from the film layer, while the second light source provides a reference source to correct for real-time lamp deviation or noise.

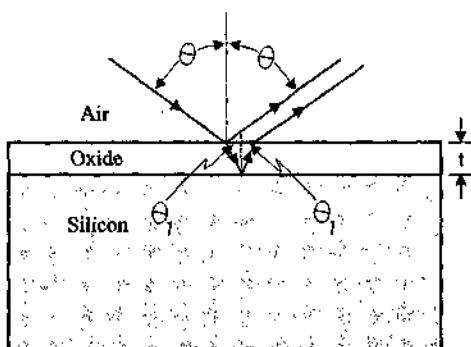


FIGURE 7.7 Light Reflection from a Thin Layer

X-Ray Film Thickness ■ X-ray beams can be focused on a surface and used for making film thickness measurements through a lesser-used technique known as *X-ray fluorescence (XRF)*. When X-rays strike a film, the absorption of radiation excites electrons in the film. The excited electrons drop into a lower energy state, emitting X-ray photons (known as fluorescence) whose energy represents the identity of the film atom. By measuring these X-ray photons, film thickness can be determined (see Figure 7.8). A modification is to use *total-reflection XRF (TRXRF)*, which uses a small angle and reduces the amount of X-ray scattering to improve the measurement sensitivity.⁹

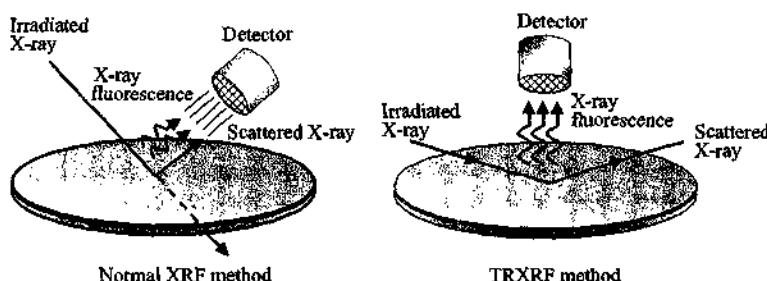


FIGURE 7.8 Film Thickness Measured with X-Ray Fluorescence (XRF)

A thin film's X-ray reflectivity will vary depending on the angle of incidence and wavelength. The reflectivity can be used to calculate the thickness and density of the film layer. It is difficult to use X-rays to measure films with mixed compositions or film stacks with two or more layers. X-ray fluorescence applies primarily to single-layer films.

Photoacoustic Technology ■ A recent development in metal thin film measurement is *photo-acoustics*. This noncontact technology is based on light-induced sound pulses that generate an acoustic pulse that is directed toward the film stack. When the acoustic pulse strikes the surface and underlying film interface, an echo is created that bounces back toward the surface. This echo causes a slight change in reflectivity that is detected at the wafer surface (see Figure 7.9 on page 158). The time it takes for the pulse echoes to bounce back is used to calculate the film thickness.

This technique has a spot size of $< 8 \mu\text{m}$, which because of its small size, makes it capable of probing structures on patterned wafers. It can measure film stacks with an individual layer thickness down to $< 20 \text{ \AA}$, which is critical as device scaling requires smaller structures for increased performance.¹⁰

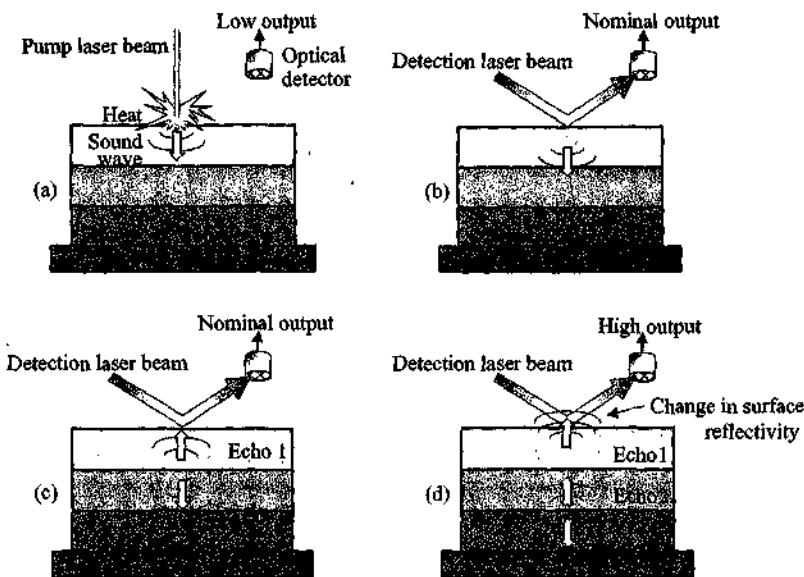


FIGURE 7.9 Photocoustic Film-Thickness Measurement
Redrawn from *Solid State Technology*, (June 1997), p. 86.

Film Stress

Highly localized *film stress* can be introduced during routine manufacturing processes into thin films that may deform the substrate and create reliability concerns. The amount of this deformation is measured with thin film stress measurement tools. Stress measurement is done by analyzing the changes in the radius of curvature of the substrate resulting from the film deposition and applies to all types of standard thin films, including metals, dielectrics, and polymers. The wafer radius is measured before and after film deposition using either a scanning laser-beam technique or split-beam laser technique to create a stress profile map across the wafer (see Figure 7.10). Automated stress measurement tools have SMIF handling capability.

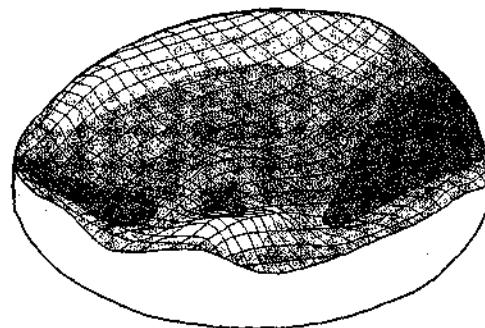


FIGURE 7.10 Detailed Stress Map of Wafer

Refractive Index

Refraction is the property of a transparent substance that addresses how much a light beam bends as it travels through it (see Figure 7.11). Variations in the refractive index can indicate contamination in the film and lead to incorrect thickness measurements. The index of refraction for pure oxide is 1.46. The index of refraction for thin films is measured by interference and ellipsometry techniques with the same ellipsometer measurement tool used to determine thin film thickness.

- Index of Refraction, $n = \frac{\sin \theta_i}{\sin \theta_r}$
- Examples of n:
 - air = 1.00
 - SiO_2 = 1.46
 - diamond = 2.12

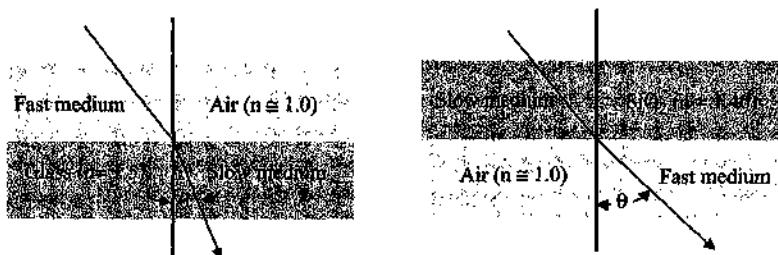


FIGURE 7.11 Index of Refraction

Dopant Concentration

The performance of semiconductor devices is directly affected by distribution of dopant atoms in the silicon in areas such as the formation of pn junctions, the epilayer, and the doping of polysilicon (see Figure 7.12). Modern processes employ dopant concentrations ranging from about 10^{10} atoms/cm² to about 10^{18} atoms/cm² (see Chapter 17).¹¹ There are several techniques for measuring the dopant concentration, or dose, of atoms in silicon. A common in-line method is the four-point probe measurement, typically used for high-dopant concentration. The thermal-wave system is also used in-line and is acceptable for low-dose readings. For off-line measurements, secondary-ion mass spectrometry (SIMS, described later in this chapter) with whole-wafer positioning has recently been used as an alternative method for doping-concentration process control.¹² The capacitance-voltage test (described later in this chapter) can also be used to characterize dopant concentration.

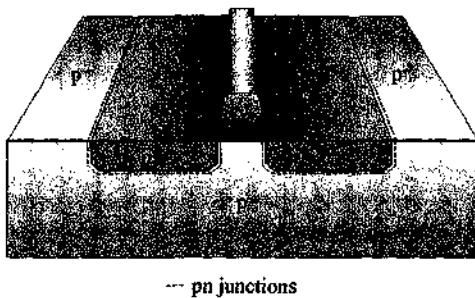


FIGURE 7.12 pn Junction

The four-point probe method has been previously discussed in this chapter, and is used in the same manner for dopant concentration as for sheet resistance (since sheet resistivity of the implanted silicon layer is related to the dopant concentration). The contour mapping technique possible with the four-point probe method is useful for routinely monitoring the implantation dose concentration.

Thermal-Wave System ■ A widely used method for monitoring ion-implant dose concentration is the *thermal-wave system*. This method measures the lattice damage in the implanted wafer due to ion implantation.¹³ This is done by measuring changes in the wafer surface reflectivity from two lasers focused on the same localized spot on the wafer (see Figure 7.13 on page 160). One laser heats the wafer with a modulated Ar laser to create waves of heat (thus the term thermal wave). The thermal waves cause a change in the reflectance of the other HeNe probe laser that is proportional to the number of crystal defect sites in the wafer. A thermal-wave signal detector with calibration data is used to characterize the implant process by correlating the amount of crystal damage to the dopant concentration and other implant parameters.

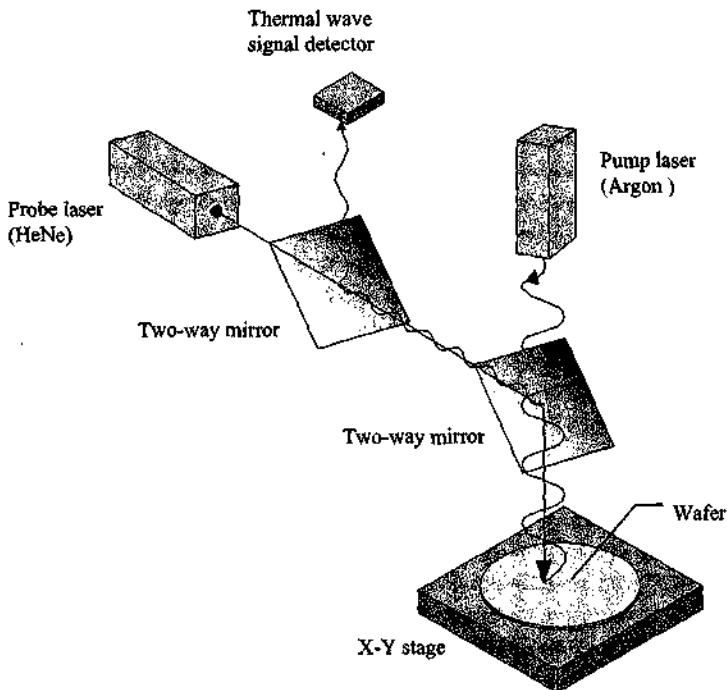


FIGURE 7.13 Thermal-Wave System for Measuring Dopant Concentration

The thermal-wave system has become common for ion-implant monitoring because of its ease of use and ability to be used on both patterned and unpatterned wafers. Its main disadvantage is that it measures damage and therefore needs calibration curves to indirectly assess dopant concentration.

Spreading Resistance Probe ■ The *spreading resistance probe (SRP)* has been a metrology tool in wafer fabrication since the 1960s and is used to measure both dopant concentration depth profiles and resistivity.¹⁴ It is capable of profiling very shallow pn junction depths. The spreading resistance probe has two carefully aligned probes that are moved in steps along a beveled wafer surface, with the resistance between the probes measured at each step (see Figure 7.14). As the probes pass through the junction, the probes sense the change in conductivity type (n or p). The sample must be carefully prepared with a bevel angle, usually < 1°, which makes the SRP a destructive test. As the probes step through the wafer bevel, the spreading resistance, R_{sp} , of a flat circular contact of radius r on a planar surface of semiinfinite material of resistivity ρ is given by,

$$R_{sp} = \frac{\rho}{4r} \quad (\text{ohms})$$

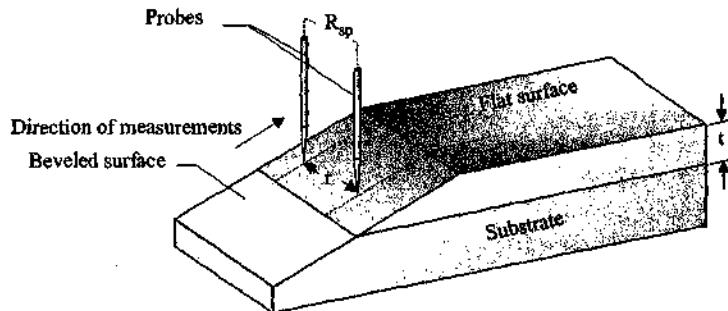


FIGURE 7.14 Spreading Resistance Probe (SRP)

The SRP employs calibration curves specific for a manufacturer's instrument to overcome a number of error sources associated with this measurement (e.g., silicon surface finish, contact resistance, tip deformation, bevel surface preparation, and so on). Computer algorithms are used to correlate the depth and resistance values to the dopant concentration at each level. The weak points

of SRP are the need for skilled operators, the sample preparation, and the destructive nature of the test. Its strong point is its ability to provide accurate dopant concentration measurements with no junction-depth limitation.

Unpatterned Surface Defects

Unpatterned wafers are bare silicon wafers or wafers with various blanket films used as test wafers and incorporated in process runs to provide a source of characterization information about process conditions. The unpatterned wafer may be polished to a mirror finish or have a film that has a rougher surface. After use in a process run, the unpatterned wafer is typically cleaned and reused, which adds cost to the manufacturing process and reduces a company's profit margin. Typical defects inspected for on unpatterned wafers used for process monitoring include particles, scratches, slip lines, and other material defects.

Defect-detection equipment for wafer-surface defects falls into two general categories: darkfield and brightfield optical detection.¹⁵ *Brightfield detection* is the traditional light source for microscope equipment—it examines the wafer surface for defects with directly reflected visible light. With bright-field detection, horizontal surfaces reflect most of the light while slanted or vertical surfaces reflect less. *Darkfield detection* examines light scattered off defects located on the wafer surface. This is done by directing light to the wafer surface at a shallow angle through the outside of the optic's objective body (see Figure 7.15). This light impinges on the wafer surface and passes back up through the center of the optics. This action renders all flat surfaces black, while irregularities appear as bright lines. This fact makes darkfield detection useful for bringing out small defects on the wafer surface that might be difficult to see with brightfield detection. An example of darkfield detection is seeing dust particles in a ray of sunshine in a dark room. Both systems usually use some form of signal or image processing to locate defects based on the light signal received from the wafer surface.

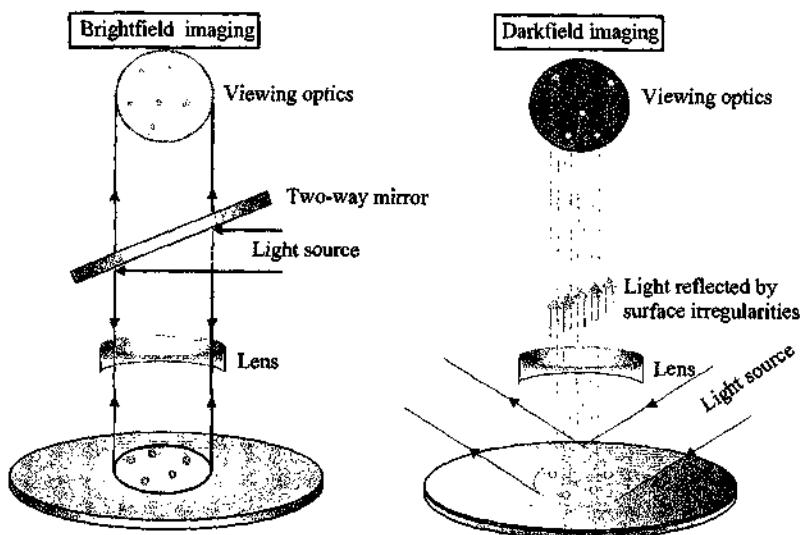
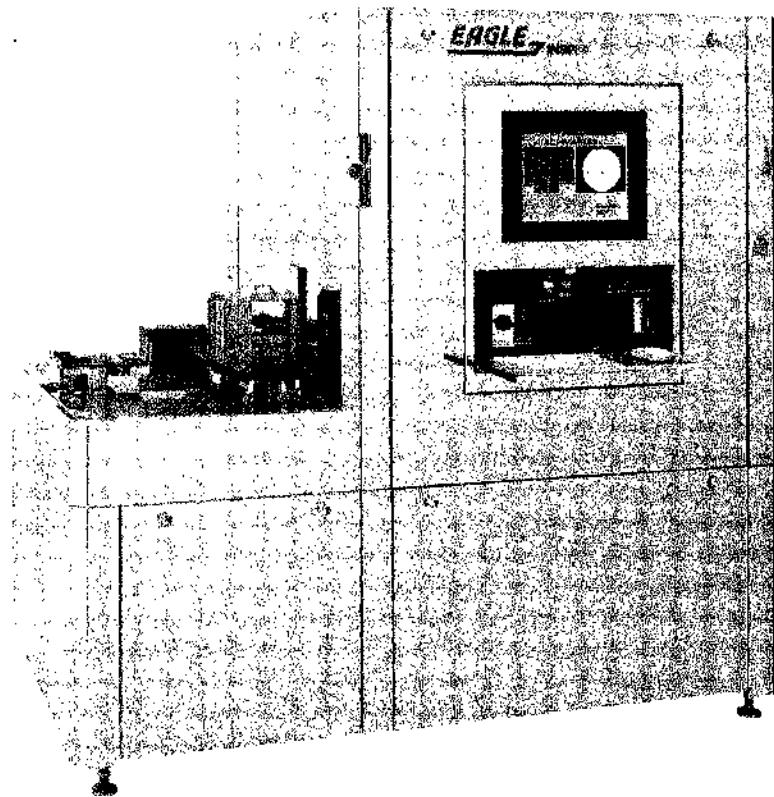


FIGURE 7.15 Darkfield and Brightfield Detection

Optical Microscopy ■ The *optical microscope*, or *light microscope*, has been used in science since the seventeenth century. For semiconductor manufacturing, it has traditionally been one of the most widely used methods of inspecting the wafer surface for defects such as particles and scratches. The optical microscope provides low-magnification views of the wafer, typically less than 1,000X magnification. Previously, optical microscopy was not capable of performing in the submicron regime. With shrinking device geometries, the size of killer defects to be detected has decreased, which has required the optical microscope to also improve over the years. Depending on the type of optical system used, particulate defects down to a size of 0.1 μm are currently able to be detected.¹⁶ Not all layers on a chip are critical in terms of dimensions, and in many instances it is necessary to detect only gross defects. In this case, the optical microscope is quick and cost-effective.



Wafer Inspection System
(Photo courtesy of Inspek)

Optical System. An optical microscope uses light reflection to detect surface defects. The modern optical microscope is integrated into a wafer inspection station that includes robotic wafer handling and software interface with video and defect classification. A typical optical system is shown in Figure 7.16. Microscope manufacturers have achieved optical systems capable of detecting smaller objects by improving the optics and using new light sources with shorter wavelengths (the ability to detect smaller objects improves with shorter light wavelengths). The technology of optics and light are discussed in detail in Chapter 14.

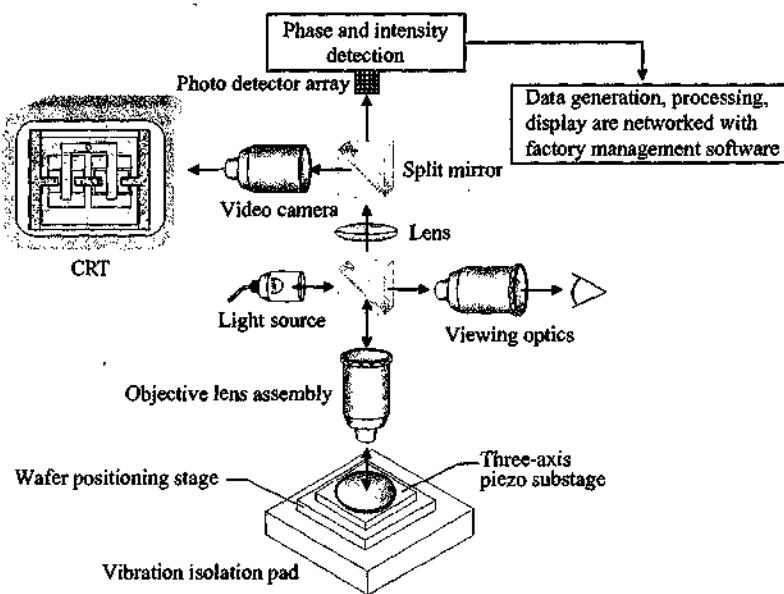


FIGURE 7.16 Schematic of Optical System

An important parameter for optical microscopy is *contrast*, which is the ability to distinguish between parts of an object. Techniques used in modern optical microscopes to enhance contrast are brightfield and darkfield detection (described on page 161), confocal contrast, and color interference contrast. A *confocal contrast microscope* uses a scanning technique to view a single point of the object at a time, thus providing better image contrast and therefore better visualization of the object (see Figure 7.17). A confocal microscope uses either visible light or laser scanning. *Color interference contrast* splits a beam of light into a direct and a reference beam. The direct beam is altered by the sample and then recombined with the reference beam, creating an image based on interference.

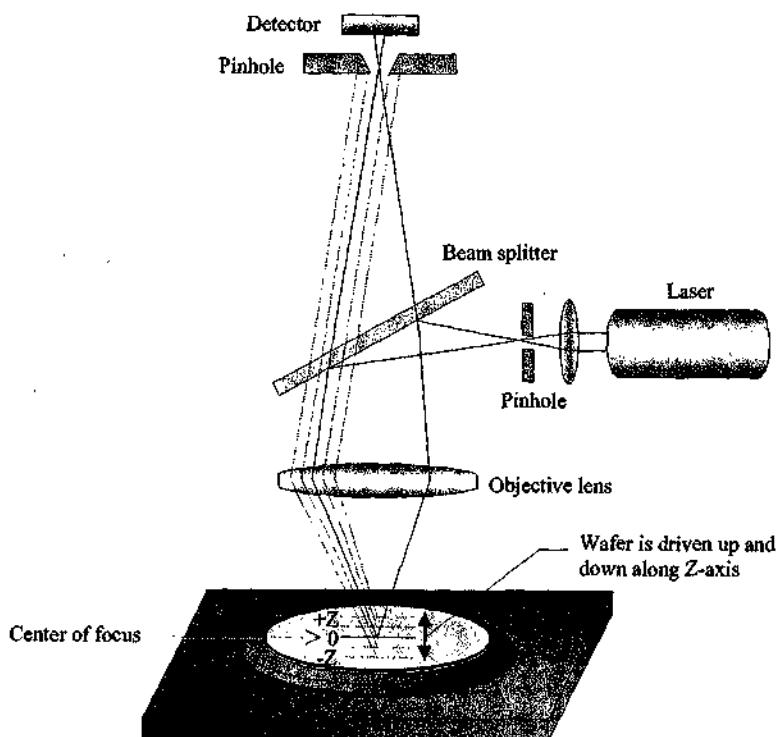


FIGURE 7.17 Principle of Confocal Microscopy

Light Scattering Defect Detection ■ In the beginning of semiconductor manufacturing, particles and surface defects were inspected by an operator using a light source and a microscope. This method was unreliable, leaving much to the subjectivity of the operator. A technique for identifying particles based on *light scattering* (also referred to as *laser scattering* or *scatterometry*) came into widespread use in the mid-1980s. This darkfield-detection method identifies surface particles and other defects by illuminating the surface with laser light and then using optical imaging to detect any light scattered by particles (see Figure 7.18 on page 164). The strength of the scattered light signal received from an individual particle depends on factors such as size, shape, composition, surrounding wafer surface (e.g., roughness and haze), and the type of equipment used. Particle diameters down to about 0.1 μm are currently capable of being detected with light-scattering techniques.¹⁷

Initially, when light-scattering particle detection became widespread in the mid-1980s, it could only detect particles on polished wafer surfaces. As the technology has advanced, it is now possible to detect particles on patterned product wafers and wafers with a thin film surface. As device geometries decrease, the critical particle size that creates killer defects also decreases. It was thought that the need to detect smaller particles would cause light-scattering technology to disappear, mainly because the particle size detection limit of about 1- μm was on the order of the light wavelength. When the light wavelength is the same as the particle size, it is difficult to detect particles because the scattered light intensity is not clearly defined.¹⁸ Improvements in light scattering have been attained as sensitivity factors such as the choice of wavelength of the incident light beam are reduced. In this manner, light-scattering surface measurement has become a dominant

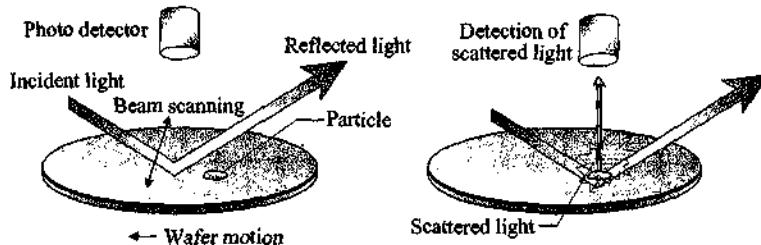


FIGURE 7.18 Particle Detection by Light Scattering

technology for tool monitoring of surface defects with both unpatterned and patterned wafers, primarily because its high-speed measurement provides rapid results during manufacturing process runs.¹⁹

Particles Per Wafer Per Pass. A practical aspect of measurement tools that detect wafer surface defects is the ability to create particle maps to count the particles on the wafer, showing where they occur as well as the distribution of particle diameters. A major effort in the fab is to perform *particles per wafer per pass* (PWP) measurements with unpatterned wafers to check the defect counts from process runs (see Figure 7.19). Examples of processes analyzed for PWP are photoresist spin tracks, dielectric deposition chambers, and chemical mechanical planarization (CMP) equipment. To perform a PWP procedure, a technician uses a wafer inspection tool to count the number of defects on a test wafer, runs the test wafer through the production equipment, and then recounts the number of defects. Corrective action is taken to reduce the number of defects and thereby increase the number of functional die on a wafer. It is often difficult to identify the smaller particles. In this case, the light-scattering detection equipment may work in conjunction with the scanning electron microscope (SEM) to identify particle composition for corrective action.

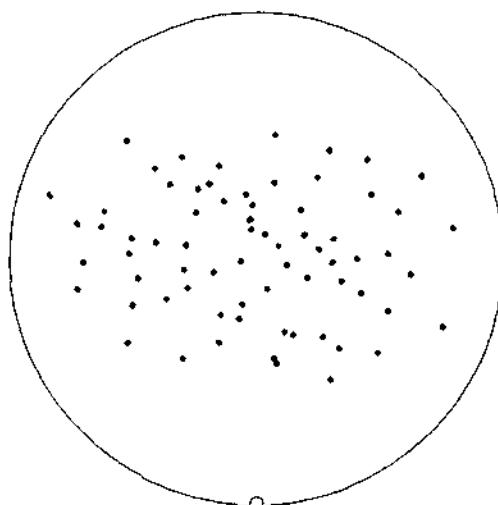


FIGURE 7.19 Particle Map

Patterned Surface Defects

Patterned wafers for fabrication metrology are regular production wafers used for in-line monitoring of surface defects. Production wafers common in IC fabrication may consist of multilayer metals with up to six or more layers that have been grown or deposited in furnaces and cluster tools. These multilayers make it challenging to distinguish between surface defects and the circuit pattern on the wafer. The majority of defects on patterned wafers are detected using light-scattering techniques. Optical microscopy (see previous section) is commonly used to detect surface defects on patterned wafers. There is also a less-common method of digital comparison, which correlates the wafer surface to a defect-free reference wafer.

Light Scattering on Patterned Wafers ■ Patterned wafers are becoming common for tool monitoring during process runs. The major defects on product wafers are particles, scratches, and pattern defects. With reduced geometries and increased process complexity, pattern defects occur more frequently. This condition makes defect detection on patterned wafers all the more critical.

Defect detection on patterned wafers with light scattering is similar to unpatterned wafers. However, the light-scattering process must be modified so that the measurement equipment can distinguish between scattered light from a particle and scattered light from a pattern edge. One technique is to use the regularity of the repeated circuit patterns to identify this particular scattered light and block it using a filter. Other patterned-wafer measurement equipment has a function that electrically cancels the cyclical signal pulses from the edge of the repeated patterns.

Critical Dimension (CD)

An important reason for critical dimension (CD) measurements is to achieve precise control over all line widths of the product being built. In CMOS technology, the transistor's gate structure is very critical. The gate width determines the channel length and the length of the channel affects speed. CD variation often indicates some instability in a critical part of the semiconductor manufacturing process. It is projected that the critical dimension (CD) for the 16-Gbit DRAM will be 0.1 μm in the year 2006.²⁰ To achieve this CD control, the metrology tool will need a precision and accuracy of better than 2 nm—the size of four silicon atoms side by side.²¹ The measurement tool that is capable of attaining this level of metrology measurement is the scanning electron microscope (SEM).

Scanning Electron Microscope (SEM) ■ The *scanning electron microscope (SEM)* has been the dominant measurement tool to verify acceptable CD control in all submicron generations since the early 1990s. The first commercial SEMs were produced in the 1960s. The SEM can achieve magnifications up to 100,000 to 300,000X, which is significantly higher than optical microscopes. The resolution (smallest feature detectable) for SEMs is on the order of 40 to 50 Å. Cross sections of wafers viewed with the SEM can provide defect information. Because of the concern for controlling submicron linewidths, SEM technology was developed specifically for CD measurement in the 1980s, and is sometimes referred to as CD-SEM. The SEM measurement tool is often coupled with other analytical techniques like EDX or FIB (see the following section).

SEM Basics. The SEM is a sophisticated microscope that functions by creating a highly focused beam of electrons that scans an object while detectors measure the resulting scattered electrons.²² It is a nondestructive and noncontact metrology tool. The SEM has an electron gun, focusing elements for shaping the electrons into a beam, and a final electrostatic-magnetic focusing system that makes the electrons strike the sample within a small, 2 to 6 nm spot (see Figure 7.20 on page 166). Since electrons have a very short wavelength (for instance, 1.22 Å for 100 eV electrons, versus about 5,500 Å for visible light), atomic-scale objects can be viewed with SEM electron imaging.

The electron gun produces a beam of electrons in a vacuum chamber of about 10^{-6} torr. It is desirable to have a high current of stable electrons with a narrow energy spread. Near the wafer, the electrons are focused into a narrow beam with a cylindrical magnetic objective lens often combined with electrostatic focusing elements to create a high-energy beam. The beam undergoes x-y deflection with an electrostatic deflector to scan the wafer. When the incident beam strikes the wafer, secondary and backscattered electrons, along with other electrons, X-rays, and photons, are emitted or transmitted due to interaction between the beam and the sample surface. The Everhart-Thornly (E-T) detector is used to collect the secondary electrons and create an electronic image that represents the sample surface. Backscattered electrons are also collected and offer superior compositional contrast between different materials.

The energy of the electron beam is directly related to the image needed. A low-energy beam (<2 keV) has a low accelerating voltage needed for nondestructive, in-line CD measurement. A high-energy electron beam (100 – 200 keV) is used to image underlying or deep structures such as contact holes. High-energy beams make it possible to nondestructively image below the surface of the wafer (say ~ 20 μm for a high-energy beam). Nevertheless, the problems with SEMs are most severe when imaging dense or deep structures, such as very fine resist lines or deep and narrow contact holes on insulator layers. This problem is mainly due to the difficulty of getting a satisfactory signal of

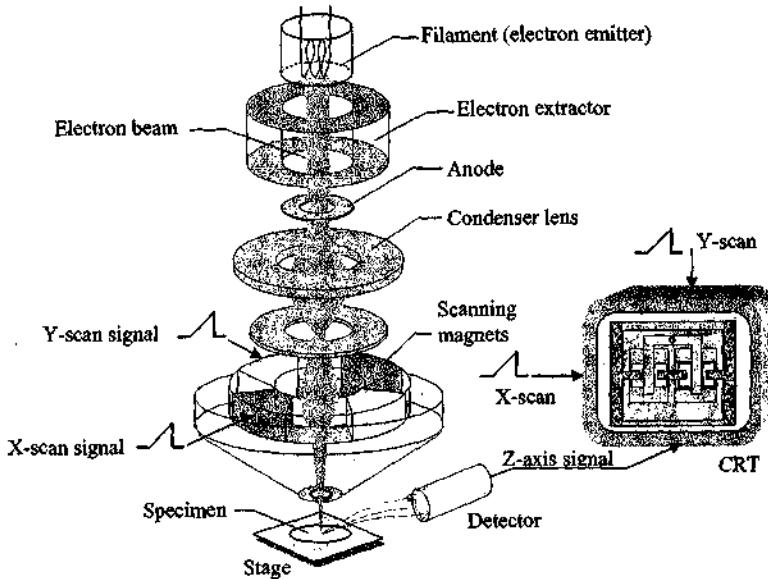
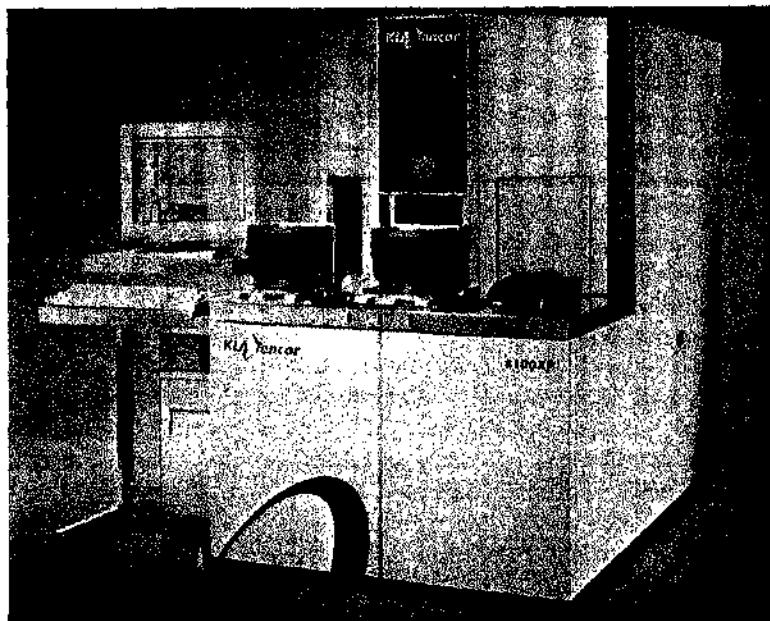


FIGURE 7.20 Simple Schematic of CD-SEM

secondary electrons from such structures. The limitations of the SEM are that it requires a high vacuum and it is necessary to coat insulating samples with conductive films prior to imaging.

CD-SEM. The *CD-SEM* has high resolution imaging with automated equipment control that rapidly evacuates the chamber to the desired vacuum, automatically positions the wafer, and uses a preprogrammed process recipe to select certain measurement sites. It is ideal for the CD-SEM to have a high wafer throughput to support production (up to 70 wafers/hour at five measured sites per wafer). The instrument can examine and measure wafers from all orientations and tilt angles of up to 60° (critical for measuring fine process patterns, sidewalls, and holes). A CD-SEM also can be used to perform defect review and analysis. Additional benefits of SEM imaging include the integration of other metrology tools, such as X-ray compositional analysis, and focused ion beam milling (see the following section).



CD-SEM
(Photo courtesy of KLA-Tencor)

Step Coverage

Because of surface topography during the fabrication of wafers, the ability to achieve conformal step coverage is a desirable material property (see Figure 7.21). *Conformal step coverage* has uniform material thickness in all regions of the step, including the sidewalls and corners.

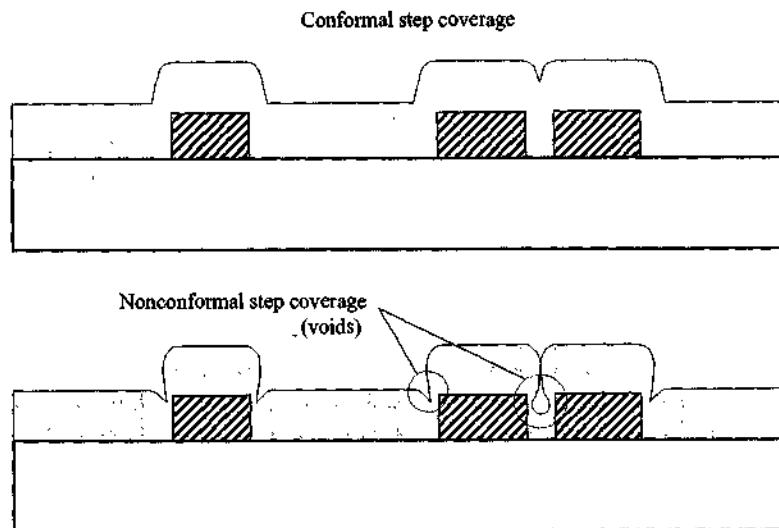


FIGURE 7.21 Step Coverage

A high resolution, nondestructive *surface profiler* with a stylus tip is often used to measure step coverage and other features on the wafer surface. This automated surface metrology tool uses a stylus that comes in contact with the wafer surface with a force as low as 0.05 mg, capable of profiling soft production-wafer films without damage to the wafer surface (see Figure 7.22). The stylus usually has a diamond tip with a radius of 0.1 μm , with older stylus having radii up to 12.5 μm .²³ Current profilers can measure wafer features as small as 0.1 μm with a 7.5 Å step height repeatability.²⁴

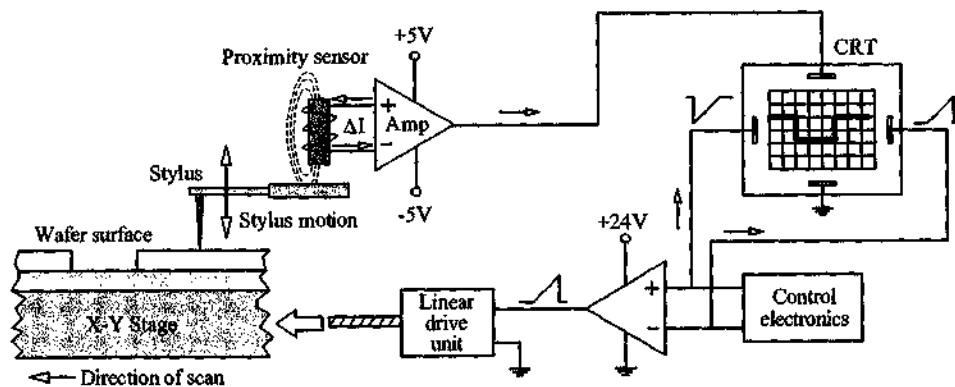


FIGURE 7.22 Surface Profiler

Overlay Registration

Overlay registration is used following the photolithography process to measure the ability of the tools and processes to accurately print photoresist patterns over a previously-etched pattern on a wafer. Due to shrinking feature sizes, the registration of the mask pattern to the wafer is a challenge because of reduced tolerances available for overlay registration. Furthermore, the increased use of chemical mechanical planarization (CMP) creates very low-contrast images on the wafer that are difficult to distinguish. This condition makes alignment of the wafer to the mask more complicated.

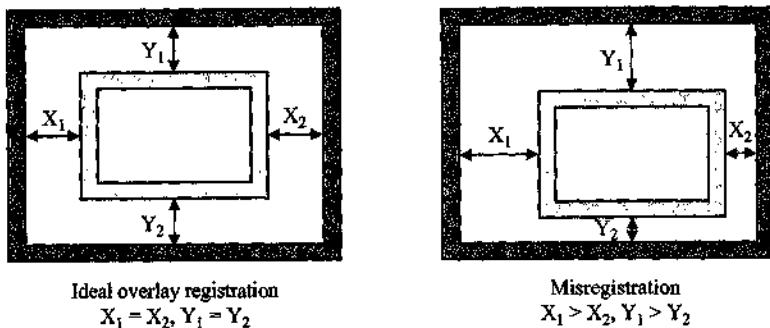


FIGURE 7.23 Overlay Registration Inspection Patterns

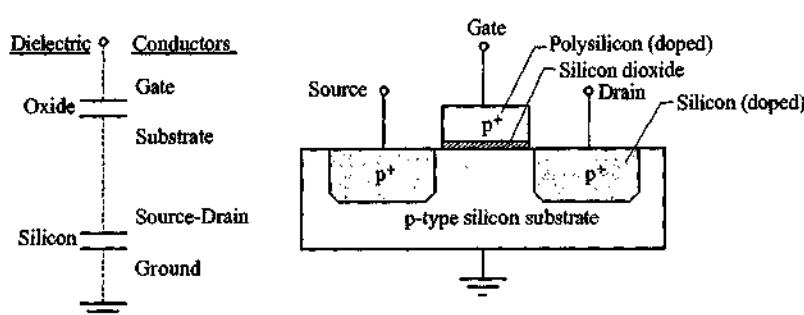
An automated-overlay metrology tool is used in photolithography to compare the registration of the special overlay mask pattern transferred to the photoresist with the etched overlay features on the wafer. It is important for this tool to be capable of measuring alignment targets on process layers that have low-contrast and grainy patterns in the wafer. Using brightfield reflected light is not sufficient to focus and measure images of low-contrast alignment targets on the wafer surface.

The primary method used today to measure registration of overlay targets is *coherence probe microscopy (CPM)*. It is also referred to as a *correlation microscope*. Coherent light has waves with a definite phase relationship to one another, meaning that the individual light waves do not move relative to one another. Using coherent light permits CPM to image not only the amount of light scattered off a surface but also its phase relationship. This method is able to capture wafer surface information in the Z axis along with the wafer surface plane, which improves focusing of the wafer target and enhances the low-contrast image of a polished overlay target. Out-of-focus and diffracted signals are ignored.²⁵

Capacitance-Voltage (C-V) Test

MOS device reliability is highly dependent on the thin layer of high-quality oxide in the gate structure. Contamination in the gate oxide region can lead to a shift in the threshold voltage requirement, causing device failure. Mobile ionic contaminants (MICs) and other undesirable charge conditions are detected at the SiO_2 layer after oxidation steps with a test known as *capacitance-voltage test*, or *C-V test*. The C-V test is commonly done to detect ionic contamination after oxidation steps. In addition, the C-V test provides information about gate oxide integrity (GOI), including the dielectric thickness, the dielectric constant, k , the resistivity of the silicon between the electrodes (to characterize majority carrier concentration), and the flatband voltage (a voltage level where there is no potential difference across the oxide structure).

The appropriate model for understanding gate oxide performance is the parallel-plate capacitor, which was explained in Chapter 2. For a MOS device, there are two capacitors in series that are functioning when the threshold voltage is applied. The first is the gate oxide, sandwiched between the doped polysilicon gate and the channel region below the gate structure. The second capacitor is formed in the silicon substrate material. This results from the attraction of charges into the gate region to form the conducting channel (known as inversion) during application of the threshold voltage. The oxide and silicon substrate are modeled as series capacitors during the C-V test (see Figure 7.24).

FIGURE 7.24
MOS Model of Two Capacitors at Gate Region

Since two capacitors in series have a combined lower capacitance than any individual capacitor, there is a drop in the capacitance of the gate structure when the threshold voltage is applied. This expected drop in capacitance is used in the C-V test to verify that no undesirable charges (e.g., mobile ionic contaminants) exist.

Steps for C-V Test ■ During a C-V contamination test, the two series capacitors in the gate region are modeled by using a special wafer. A variable voltage is applied between a metallized area on an oxide layer and the lightly-doped silicon beneath the oxide (see Figure 7.25).

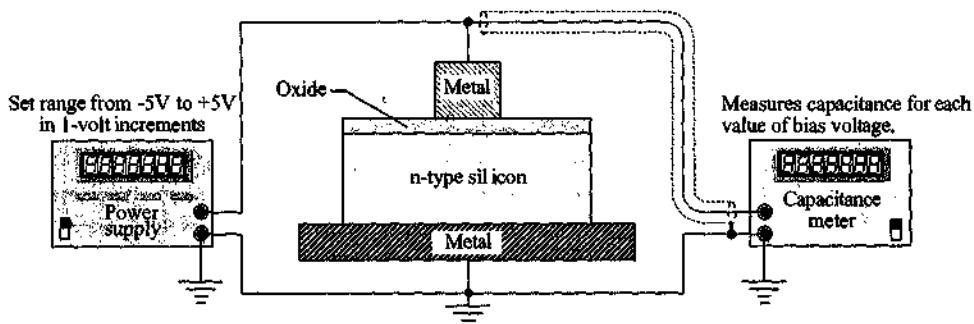


FIGURE 7.25 C-V Test Setup and Plotting

The first step for a C-V test is to apply a variable-voltage bias to both the metal contact of the oxide film under test and the lightly-doped silicon beneath the oxide. The bias is varied from positive to negative for a p-type silicon and negative to positive for n-type silicon. The objective of this first step is to deplete the silicon directly beneath the metallized area of majority carriers. In this manner, the silicon functions as a dielectric, which serves to reduce the total capacitance of the test structure because the oxide and silicon substrate are capacitors in series. The capacitance versus voltage during this test is plotted (see Figure 7.26).

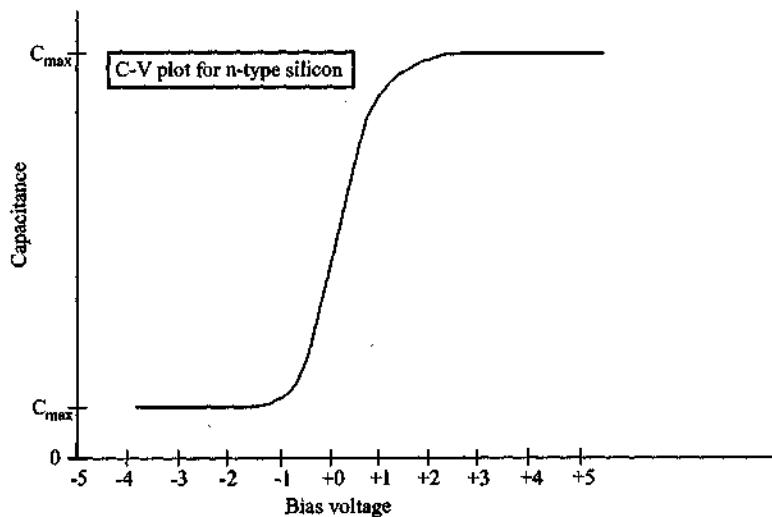


FIGURE 7.26 Capacitance versus Voltage for n-Type Silicon (First Step of C-V Test)

The second step is to apply a constant positive DC voltage to the metallized area (its magnitude depends on the oxide thickness) while heating the wafer to 300°C for five minutes followed by a cooldown prior to removing the bias (see Figure 7.27 on page 170). The elevated temperature serves to increase the mobility of the ionic contaminants. The positive voltage bias repels the positive ionic contaminants and drives them toward the oxide-silicon interface.

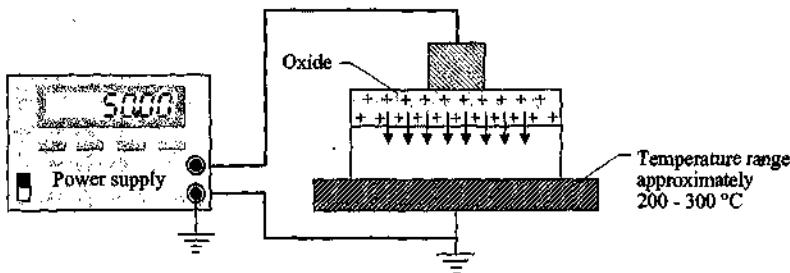


FIGURE 7.27 Ionic Charge Collection in C-V Test

The third step is to repeat the first C-V plot. However, now if there are positive ionic contaminants collected at the Si/SiO_2 interface, then a more negative voltage is required to get an equivalent charge in capacitance. This is a voltage shift, V_s , that is a measure of the amount of contamination in the oxide (see Figure 7.28). The magnitude of the voltage shift is proportional to the ionic contamination in the oxide, the oxide thickness, and the wafer doping. The actual amount of ionic contamination can be calculated from this plot.

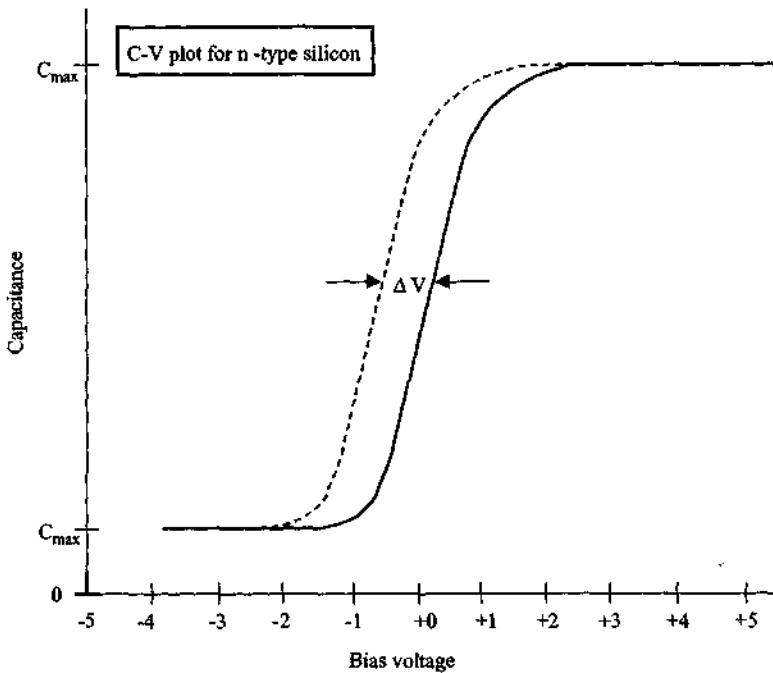


FIGURE 7.28 Voltage Shift in n-Type Silicon

The final step is to verify that this shift is due to contamination and not oxide charging. A negative voltage is applied while heating the substrate to move the ionic contaminants away from the Si/SiO_2 interface. Performing the C-V test again should produce the original plot, verifying that it was actually the ionic contaminants that produced the voltage shift.

Benefits of C-V Test ■ The C-V analysis is used to monitor the wafer to ensure that the cleaning process removes ionic contaminants. This analysis is done either on fabricated test wafers or production wafers with a special test structure (test structures are described in Chapter 19). The analysis cannot tell you where the contamination comes from, such as the wafer surface, cleaning step, equipment maintenance, or furnace process. The C-V test is usually a part of any evaluation for process changes. In this case, wafers are tested in two groups. One group undergoes normal processing prior to C-V testing while the other group follows the new process and then is C-V tested. The results are compared to verify that the proposed process is acceptable. There is a small amount of voltage shift that is acceptable, depending on the oxide thickness and the sensitivity of the measurement equipment.

Contact Angle

Contact-angle meters are used to measure adhesion of liquids to the wafer surface and to calculate surface energies or adhesion tension. This measurement characterizes wafer surface parameters such as wettability, cleanliness, finish, and adhesion (see Figure 7.29). The contact (tangent) angle formed between a liquid drop and its supporting surface is relative to the forces at the liquid/solid or liquid/liquid interface and can be used as a wafer-test specification or as a quality characteristic. Both direct-angle measurements and indirect dimensional measurements methods are employed to obtain highly accurate and repeatable contact angle measurements.



FIGURE 7.29 Contact Angle

ANALYTICAL EQUIPMENT

This section provides an overview of the major analytical equipment used to support wafer fabrication. These analytical tools provide highly accurate wafer measurements and are typically located in an off-line lab to support production problems. Figure 7.30 shows when some of these instruments were first used and how important each one is or is projected to become in either process development or manufacturing. The analytical equipment reviewed includes:

- ◆ Secondary-ion mass spectrometry (SIMS)
- ◆ Time of flight secondary-ion mass spectrometry (TOF-SIMS)
- ◆ Atomic force microscope (AFM)
- ◆ Auger electron spectroscopy (AES)
- ◆ X-ray photoelectron spectroscopy (XPS)
- ◆ Transmission electron microscope (TEM)
- ◆ Energy- and wavelength-dispersive spectrometer (EDX and WDX)
- ◆ Focused ion beam (FIB)

	Year							Importance to Manufacturing
	1950s	1960s	1970s	1980s	1990s	2000s	2010s	
AES			Development	Manufacturing				Useful
AFM					Research, Development			Useful
FIB					Research, Development,	Manufacturing		Critical
SEM		Research, Development, Manufacturing						Critical
SIMS		Research, Development, Manufacturing						Critical
TEM		Research, Development, Manufacturing						Critical
TOF SIMS				Research, Development				Useful
XPS			Development					Useful

FIGURE 7.30 Relative Importance of Analytical Equipment

Secondary-Ion Mass Spectrometry (SIMS)

Secondary-ion mass spectrometry (SIMS) is a method of eroding a wafer surface with ions accelerated in a magnetic field to analyze the surface material composition. These ions strike the surface of the wafer and dislodge, or sputter, other ions, some of which are known as secondary ions (see Figure 7.31 on page 172). Secondary ions contain the wafer material and any impurities such as dopants. They are collected and analyzed in a vacuum chamber with a mass spectrometer that can

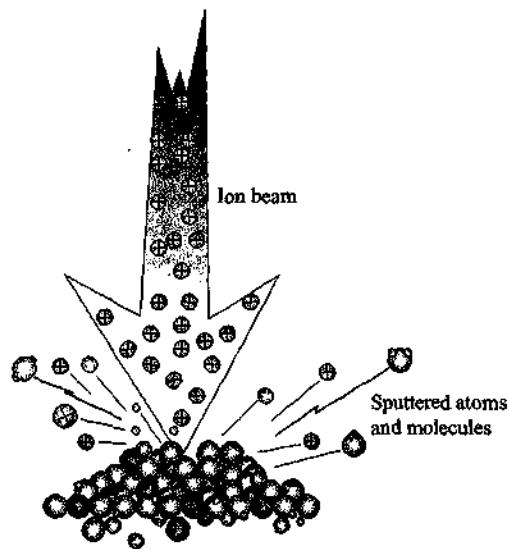


FIGURE 7.31 Ion Beam Sputtering of Surface Material

identify the type of dopant and its concentration in the silicon.²⁶ SIMS is essentially a destructive measurement technique and requires an ultrahigh vacuum environment (around 10^{-10} to 10^{-12} torr).

Recall that ions are charged particles, able to be accelerated or deflected in a magnetic field. When accelerated ions strike a wafer surface, this ion-beam sputtering process involves violent collisions at the atomic level. This action could be compared to a sandblaster, which propels sand with such energy that it removes surface pieces but leaves behind craters in the base material. However, with an ion beam, there are also chemical reactions that take place during the collision, creating many complex molecules in the sputtered material. The particular incident ion could be nearly any element, but is typically chosen as Cs^+ , O_2^- , O_2^+ or Ar^+ and is selected based on the collision chemistry that will best produce secondary ions.

SIMS Tool Description ■ A common method to produce ions used in SIMS is the duoplasmatron (see Figure 7.32). A gas, commonly O_2 or Ar , is converted into a plasma by a low-pressure arc formed between a hot filament and an anode. An extraction electrode is used to draw charged atoms from the plasma.

Cs^+ ions are produced by a different method. A pellet of cesium chromate is vaporized in a heated reservoir. The vapor feeds into a hot ($\sim 1100^\circ\text{C}$) and porous tungsten plug in an ionization chamber. The vapor diffuses through the pores of the plug and becomes ionized, followed by an extractor that collects the ions and forms them into an ion current.

After the ions are produced, they go through an analyzer magnet to select the ions of interest (e.g., Cs^+ or O^-). These ions are then either focused into a small spot with a magnet, known as an ion microprobe, or used to flood the surface of the sample. The latter is referred to as an ion microscope and is the most common technique today. With an ion microscope, the sputtered secondary ions are collected simultaneously from many points on the surface and then identified by analyzing their mass-to-charge ratio in the mass spectrometer. A common mass spectrometer, the quadrupole, is based on an oscillating electric field that separates ions into defined oscillations based on their mass, permitting certain ions to pass through an aperture and be identified.

The rate of material removal using SIMS can range from a high sputter rate, known as *dynamic SIMS* and often used for ion implanter characterization, to a slow sputter rate of a single monolayer of atoms every few hours. The latter is referred to as *static SIMS*, and is frequently used to analyze contamination in thin oxide and nitride films. The SIMS measurement tool is sensitive enough to measure ppba (parts per billion-atomic) impurities in a junction or contact that is only 1 to $10 \mu\text{m}$ across. It has become the primary measurement tool used to verify the performance of ion implanters because it can characterize the dose and depth of the junction plus reveal any undesirable metal impurities at the junction.²⁷

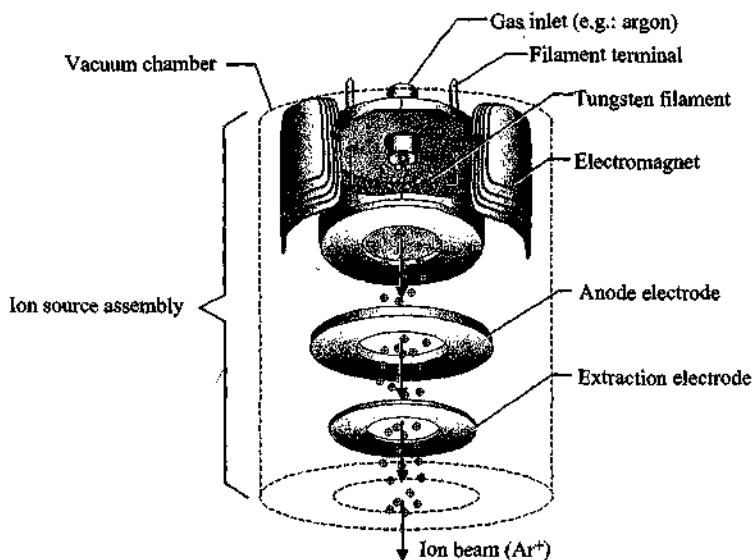


FIGURE 7.32 Ion Production in a Duoplasmatron

Time of Flight SIMS (TOF-SIMS) ■ A limiting factor for SIMS is that the mass spectrometer only detects around 0.001% of all the ions emitted from the sample. Another approach, referred to as *time of flight SIMS (TOF-SIMS)*, can detect as many as 10 to 50% of the ions emitted by the sample. The TOF-SIMS identifies ions of different mass by measuring the time it takes for an ion to travel the length of a fixed path, since the charged particle velocity is a function of mass (see Figure 7.33). This fact makes it possible to reduce the incident beam current by a factor of up to 10^5 relative to a SIMS with a quadrupole spectrometer. The rate of material removal is so low that only a fraction of a single-surface monolayer is removed in an hour. For this reason TOF-SIMS is essentially nondestructive and is ideal for very thin films on the wafer surface.²⁸

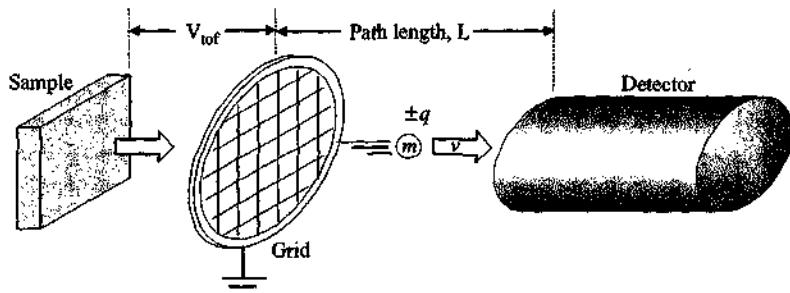


FIGURE 7.33 TOF-SIMS Mass Spectrometer Principle

Atomic Force Microscope (AFM)

The *atomic force microscope (AFM)* is a surface profiler that scans a small, counterbalanced tip probe over the wafer to create a 3-D surface map.²⁹ It was first demonstrated in 1986 and uses optical technology, direct tip contact, and a laser to sense the position of the tip relative to the surface (see Figure 7.34 on page 174). Probe and surface separation is so small (on the order of 2 Å) that atomic forces affect the probe between the surface and the tip. The tip geometry is critical and must be characterized to provide accurate measurements. AFM measurements are exceedingly slow, which is a problem for in-line measurements in a production environment.

There are several variations of AFM technology.³⁰ In the simplest system, a laser beam reflects off the top surface of the probe tip and is directed to a photodiode. When the surface topography moves the probe, it changes the position of the laser in the photodiode, creating an electronic map of the surface. A more advanced method has the probe oscillate very near the surface. Atomic van der Waals forces interact with the vibrating tip, inducing phase shifts in the vibration that are detected by an imaging system. The benefit of this approach is that it can measure vertical

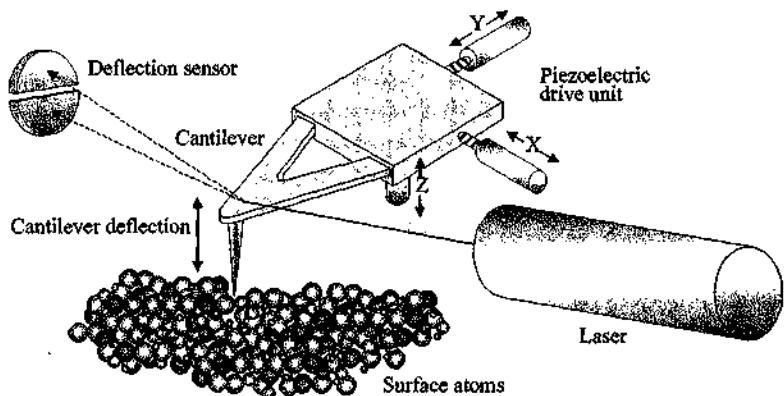


FIGURE 7.34 Schematic of Atomic Force Microscope

photoresist profiles. AFM is a recent measurement technology and is considered a promising alternative to SEMs.

Auger Electron Spectroscopy (AES)

Auger electron spectroscopy (AES) measures the energy of auger electrons emitted by the surface of a sample when struck by an incident electron beam. It is very sensitive to the surface, with a depth of only 10 to 50 Å. Auger (pronounced *ōzhā'*) electrons were discovered in 1923 by Pierre Auger in France, and make up a very small percent (< 0.1%) of the total electrons produced in a sample (secondary electrons are the dominant species). The energy associated with Auger electrons provides a distinct link back to the parent atom that is used for identification of the sample elements.

Because Auger electrons are easily absorbed by the sample, only those Auger electrons on the outer monolayers of the surface escape and are detected in AES. This makes Auger technology especially suited for analyzing the surface of materials, usually to a depth of about 2 nm. Oxides of metals, silicides, and wafer surfaces are readily sensed by Auger spectroscopy. Furthermore, the Auger spectrometer uses a highly focused electron beam, with the smallest beam spots on the order of 12 nm, which is beneficial in microcircuit analysis. AES requires an ultrahigh vacuum environment in order to reduce contaminant formation on the sample.

X-Ray Photoelectron Spectroscopy (XPS)

X-ray photoelectron spectroscopy (XPS) is primarily used to identify chemical species on a sample surface, analyzing to a sample depth of about 2 nm (equivalent to AES). All elements are detectable except hydrogen and helium (their detection requires a good spectrometer). In XPS, X-ray photons are directed toward the wafer surface and interact with certain core-level electrons referred to as XPS electrons (see Figure 7.35). If the X-ray energies exceed the XPS electrons' binding energy, then the electrons are emitted from the sample. Because the binding energy of an electron is influenced by its chemical surroundings, XPS identifies both the element and its chemistry.³¹

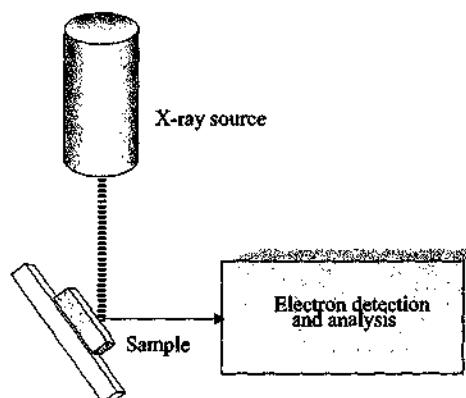


FIGURE 7.35 Schematic of XPS Measurement

Transmission Electron Microscope (TEM)

The principle of operation for the *transmission electron microscope (TEM)* is similar to an SEM, with the major difference that the beam of electrons is transmitted through an ultrathin slice of the sample (on the order of 10 to 100 nm thick). Based on factors such as the electron wavelength, accelerating voltage and specimen thickness, an image is formed and magnified on a screen with a resolution of about 2 Å (see Figure 7.36).

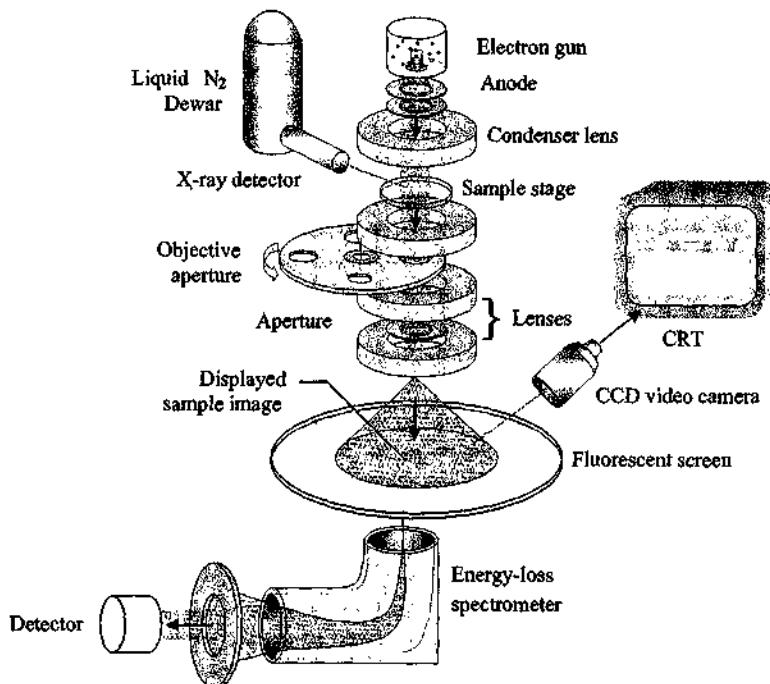


FIGURE 7.36 Schematic of TEM

TEM is the only metrology tool capable of quantifying some very small features on a wafer. For instance, silicon crystal point defects, such as single dislocations introduced into an active junction by ion implant (which leads to junction leakage), can be imaged on an atomic scale by TEM. Table 7.3 outlines some process areas of wafer fabrication that benefit from TEM analysis.³² The most difficult aspect of TEM is sample preparation, with a wide array of techniques used, such as mechanical polishing, chemical etching, and ion-beam milling.

TABLE 7.3 Sample TEM Applications in Semiconductor Manufacturing

TEM Measurement	
Silicon Material	Dislocation and stacking fault densities in Si shallow junctions
Patterning	Precise sidewall profile in polysilicon and metal structures
Metallization	Characterization of metal silicides and alloys
Implantation	Surface and buried implant damage
Contamination	Thin organic and oxide films at submicron contact interfaces

Energy- and Wavelength-Dispersive Spectrometer (EDX and WDX)

A useful signal generated by the incident electron beam of an SEM is the characteristic X-ray produced by the energetic primary electrons emitted by the sample material. These X-rays carry information related to the atom species present in the sample, permitting the identification of a variety of thin films, particles, and defects created in semiconductor manufacturing. X-ray detectors are found on nearly all electron microscopes today.

The *energy-dispersive spectrometer (EDX)* is the most widely used X-ray detection method for identifying elements and is complementary to SEM. This is primarily because the EDX can quickly detect X-rays of all energies simultaneously, which means it is a relatively fast measurement. However, because it penetrates well beyond the sample surface, it should not be considered a surface analysis. EDX operation is based on a large diode made from a high-quality doped Si crystal and isolated from the SEM vacuum chamber by a thin ($\sim 25 \mu\text{m}$) beryllium window (see Figure 7.37). An X-ray passing through the window produces a series of electron-hole pairs, which are electronically detected and identified back to the energy level of the X-ray. EDX can acquire all energy peaks and make a spectrum plot within a few minutes.³³

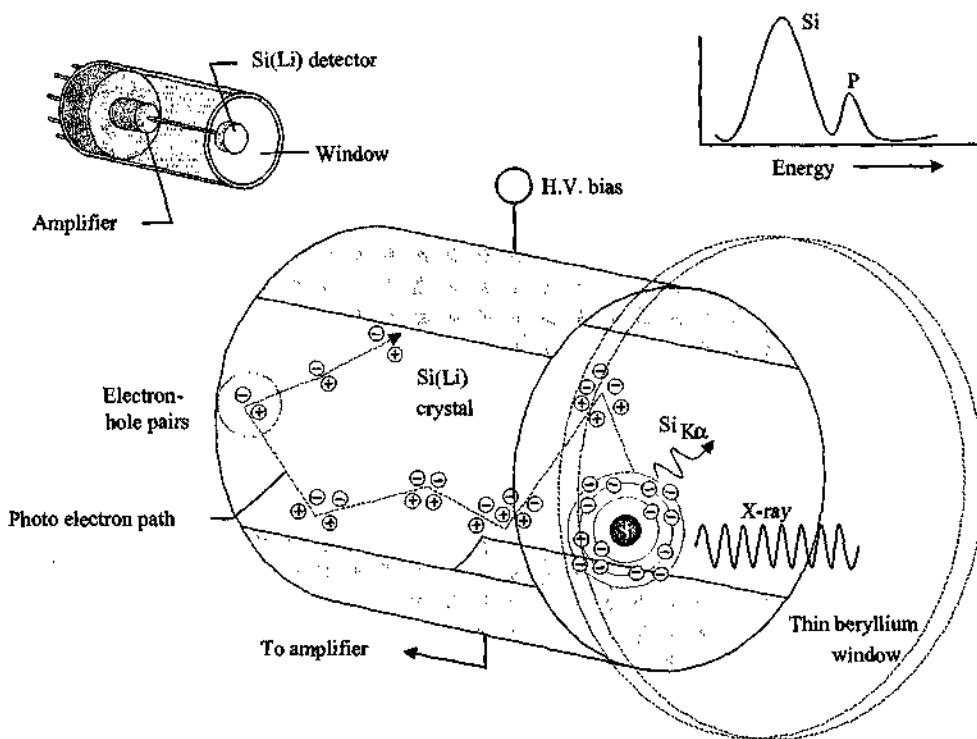


FIGURE 7.37 Energy-Dispersive Spectrometer (EDX)

The *wavelength-dispersive spectrometer (WDX)* operates based on a diffracting crystal and photo counter. The crystal separates and disperses the incoming X-rays by wavelength, which are then collected by a photo counter detector. The appropriate crystal is selected and placed in the path of the X-rays. Critical parameters are X-ray wavelength, lattice spacing of the crystal, incident angle, and order of reflection. WDX is a slow measurement but yields superior measurements with high accuracy.

Focused Ion Beam (FIB)

Traditionally, and often still today, a wafer is analyzed by removing it from the cleanroom to a lab where it is cross-sectioned and examined on an SEM. This process started changing when *focused ion beam (FIB)* systems became prevalent in the early 1990s because of their convenient cross-sectioning capabilities. The FIB is a destructive technique similar to an SEM in design and operation,

except that the primary beam is made of Ga^+ ions instead of electrons. These ions are focused through a set of lenses into a small spot. At the location where they impact the wafer, atoms are ejected into the vacuum, creating a small void of precisely controlled shape and depth in the sample material. This precision gives the FIB the possibility of making cross sections in specific locations of a wafer. The FIB technique is, however, a time-consuming process.

Focused ion beam (FIB) milling is capable of carving a 10 to 100-nm thick cross section from any area on a wafer inside the cleanroom (see Figure 7.38). It can cut through metal, polysilicon, oxide, and nitride layers with little or no damage to adjacent structures. Typically a high-current beam is used for the initial cut and is followed by a low-current beam with a tighter focus for final sample polishing. FIB can make the thin samples necessary for TEM. For this reason, equipment suppliers make custom FIB designs specifically for this application.

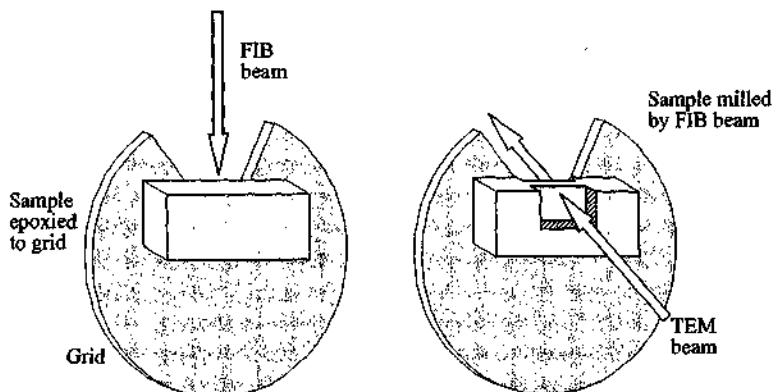


FIGURE 7.38 Focused Ion Beam (FIB) Milling

SUMMARY

IC metrology measures the performance of fabrication processes, typically with monitor or production wafers. Metrology tools are either stand-alone or integrated into the process tool, and they often use advanced defect-analysis software. Specified quality measures are used at different process locations to verify acceptable performance. Film thickness is most often measured with the four-point probe sheet resistance (opaque films), ellipsometry (transparent films), and reflection spectroscopy. Lesser-used methods are X-ray and photoacoustic. Film stress on a wafer is determined by measuring the wafer curvature. The refractive index of a transparent substance is measured using ellipsometry. The dopant concentration is measured using four-point probe, infrared interference, and spreading resistance probe. Surface defects are measured with unpatterned

or patterned wafers, with the latter becoming more cost-effective. The optical microscope is used, with advanced features to improve wafer-surface contrast. Light scattering is often used to detect particles on the wafer surface, with contour maps generated to highlight the defects. The recent trend is for more surface defect analysis to be done with patterned wafers. Critical dimensions are measured using the scanning electron microscope (SEM). Enhanced techniques are used to measure overlay registration. The capacitance-voltage test provides information on the quality of the gate oxide structure, primarily to ensure the oxide is contamination-free. The contact angle verifies acceptable wafer surface quality. Sophisticated analytical tools are used to characterize the wafer and verify acceptable process performance.

KEY TERMS

metrology
 defects
 defect density
 monitor wafers (test wafers)
 yield
 stand-alone measurement tool
 integrated measurement tool
 automatic defect classification (ADC)
 quality measures
 film-thickness metrologies
 sheet resistance, R_s
 sheet resistivity, ρ_s
 four-point probe
 Van der Pauw method
 contour maps
 ellipsometry
 variable-angle spectroscopic ellipsometry (VASE)
 spectroscopic ellipsometry
 reflection spectroscopy (reflectometry)
 X-ray fluorescence (XRF)
 Total-reflection XRF (TRXRF)
 photoacoustics
 film stress
 refraction
 dopant concentration (dose)
 thermal-wave system
 spreading resistance probe
 unpatterned wafers
 brightfield detection

darkfield detection
 optical microscope (light microscope)
 contrast
 confocal contrast microscope
 color interference contrast
 light scattering (laser scattering or scatterometry)
 particles per wafer per pass (PWP)
 patterned wafers
 scanning electron microscope (SEM)
 CD-SEM
 conformal step coverage
 surface profiler
 overlay registration
 coherence probe microscopy (CPM)
 correlation microscope
 capacitance-voltage test (C-V test)
 contact angle meters
 secondary ion mass spectrometry (SIMS)
 dynamic SIMS
 static SIMS
 time of flight SIMS (TOF-SIMS)
 atomic force microscope (AFM)
 Auger electron spectroscopy (AES)
 X-ray photoelectron spectroscopy (XPS)
 transmission electron microscope (TEM)
 energy-dispersive spectrometer (EDX)
 wavelength-dispersive spectrometer (WDX)
 focused ion beam (FIB)

REVIEW QUESTIONS

- What is metrology? What is the purpose of metrology in IC fabrication?
- Define a defect. What is meant by wafer defect density?
- Discuss the difference between monitor wafers and production wafers for metrology measurements.
- What is the difference between a stand-alone measurement tool and an integrated measurement tool?
- Describe automatic defect classification (ADC), and explain how this analysis can improve metrology.
- Define semiconductor quality measures. List twelve different quality measures used in IC fabrication. State the process areas where the different quality measures are used.
- List the two general areas of film-thickness metrology.
- State the formula for sheet resistance for a thin film. Explain how sheet resistance is interpreted for a square sample. What are the units for sheet resistance?
- What is the formula for the sheet resistivity of a thin film?
- Describe the four-point probe and give its benefits for measuring sheet resistance.
- What is the Van der Pauw method?
- Explain what a contour map is.
- Explain the basic principles of ellipsometry. What are the benefits for using ellipsometry to measure thin film thickness?
- Explain what VASE is. What applications does VASE offer improved ellipsometry?

15. Describe spectroscopic ellipsometry.
16. Describe reflection spectroscopy. What is a dual-beam spectrometer?
17. How is film thickness measured using X-rays? What does XRF stand for? What is total reflection XRF?
18. Describe how photoacoustic technology is used for film-thickness measurement.
19. How is thin film stress measured on a wafer?
20. How is refraction measured on a wafer? What does a variation in the refractive index indicate?
21. Can the four-point probe be used to characterize dopant concentration in a film?
22. Explain the principle of thermal wave for measuring dopant concentration.
23. Describe the spreading resistance probe method for dopant concentration.
24. What is brightfield detection? What is darkfield detection?
25. Describe the general state of optical microscopy for wafer-surface defect detection.
26. What is contrast? Discuss two methods to improve inspection contrast on the wafer surface.
27. Explain how light scattering can be used to detect surface defects.
28. Explain particles per wafer per pass (PWP).
29. What is the dominant measurement tool for wafer critical dimension?
30. Explain the basic operation of an SEM.
31. What is a CD-SEM?
32. How is step coverage often measured?
33. What is overlay registration? State and discuss the primary technique for measuring overlay registration.
34. Why is the capacitance-voltage test done? Describe four steps necessary to perform this test.
35. State the purpose of the contact-angle metrology measurement.
36. Describe secondary-ion mass spectrometry (SIMS).
37. What is TOF-SIMS and in what condition is it desirable to use?
38. Explain atomic force microscopy (AFM).
39. Describe the technique for using the Auger electron spectroscope.
40. What is X-ray photoelectron spectroscopy?
41. Explain transmission electron microscopy.
42. Explain the difference between EDX and WDX.
43. Describe focused ion beam (FIB) milling and explain its benefits.

METROLOGY EQUIPMENT SUPPLIERS' WEB SITES

Applied Materials	http://www.appliedmaterials.com/products/
Carl Zeiss Microelectronics	http://www.zeiss.de/
Cerprobe Corp.	http://www.cerprobe.com
FEI Company	http://www.feic.com
Gaertner Scientific Corp.	http://www.gaertnerscientific.com/
Hitachi	http://www.hitachi.com/semequipment/products.html
Inspek	http://www.inspek.com/
International SEMATECH	http://www.sematech.org/public/index.htm
JA Woollam Co. Inc.	http://www.jawoollam.com/
JEOL	http://www.jeol.com/
Kaman Instrumentation	http://www.kamaninstrumentation.com/
Keithley Instruments	http://www.keithley.com/
Kernco Instruments Co.	http://www.kerncoinstr.com/cam.htm
Kevex Spectrace	http://www.kevexspectrace.com/
KLA-Tencor	http://www.kla-tencor.com/splash.html
Leica	http://www.leica.com/
Leybold Inficon Inc.	http://www.leyboldinficon.com/
The Micromanipulator Co. Inc.	http://www.micromanipulator.com/
Nanometrics	http://www.nanometrics.com/
National Institute of Standards	http://www.nist.gov/
Nicolet Instruments	http://www.nicolet.com/
Nikon	http://www.nikonusa.com/
Olympus America Inc.	http://www.olympus.com/
Perkin-Elmer	http://www.perkinelmer.com/
Rudolph	http://www.rudolphtech.com/home/
Schlumberger	http://www.1.slb.com/ate/diagsys
SEMI	http://www.semi.org/

Sonoscan Inc.
 Therma-Wave
 Veeco Instruments

<http://www.sonoscan.com/>
<http://www.thermawave.com/index.htm>
<http://www.veeco.com/>

REFERENCES

1. S. Butler, "Process Control Through Integrated Metrology," *Solid State Technology* (January 1999); p. 34.
2. C. W. Pearce, "Crystal Growth and Wafer Preparation," *VLSI Technology*, 2nd ed., ed. S. Sze, (Boston: McGraw Hill, 1988), p. 300.
3. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., (New York: McGraw-Hill, 1998), p. 117.
4. J. C. C. Tsai, "Diffusion," *VLSI Technology*, 2nd ed., ed. S. Sze, (Boston: McGraw-Hill, 1988), p. 300.
5. Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors: Technology Needs*, (San Jose, CA: SIA 1997), p. 80.
6. P. Burggraaf, "Thin Film Metrology: Headed for a New Plateau," *Semiconductor International* (March 1994); p. 58.
7. R. DeJule, "Advances in Thin Film Measurements," *Semiconductor International* (May 1998); p. 64.
8. C. Morath et al., "Ultrasonic Multilayer Metal Film Metrology," *Solid State Technology* (June 1997), p. 85.
9. M. Dax, "X-Ray Film Thickness Measurements," *Semiconductor International* (August 1996), p. 98.
10. R. DeJule, "Advances in Thin Film Measurements," *Semiconductor International* (May 1998); p. 56.
11. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, (Boston: Kluwer Academic, 1995), p. 371.
12. Semiconductor Industry Association, *National Technology Roadmap*, p. 80.
13. E. Rimini, *Ion Implantation: Basics to Device Fabrication* (Boston: Kluwer Academic, 1995), p. 70.
14. D. Schroder, *Semiconductor Material and Device Characterization*, 2nd ed., (New York: Wiley, 1998), pp. 31-35.
15. J. Baliga, "Defect Detection on Patterned Wafers," *Semiconductor International* (May 1997); p. 64.
16. A. Braun, "Defect Detection Overcomes Limitations," *Semiconductor International* (February 1999); p. 52.
17. A. Braun, "Defect Detection and Review Enter New Era," *Semiconductor International* (May 1998); p. 61.
18. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., p. 227.
19. A. Braun, "Defect Detection and Review Enter New Era," *Semiconductor International* (May 1998); p. 66.
20. Semiconductor Industry Association, *National Technology Roadmap*, p. 14.
21. M. Davidson and A. Vladar, "The Physics of Metrology Instruments," *Solid State Technology* (June 1998); p. 135.
22. Material for description of CD-SEM taken from M. Davidson and A. Vladar, "Physics of Metrology Instruments," pp. 136-142.
23. A. Braun, "Analytical Techniques for Process Problem Solving," *Semiconductor International* (October 1997); p. 112.
24. KLA-Tencor, Product literature for *High Resolution Profiler*, *KLA-Tencor HRP-220*, (August 1998).
25. KLA-Tencor, Product literature for *Automated Overlay Metrology Tool*, *KLA-Tencor 5200 XP*, (May 1998).
26. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., p. 88.
27. Ibid., p. 402.
28. Ibid., p. 418.
29. K. Wilder, B. Singh, and W. Arnold, "Novel In-Line Applications of Atomic Force Microscopy," *Solid State Technology* (May 1996), p. 109.
30. M. Davidson and A. Vladar, "Physics of Metrology Instruments," p. 144.
31. D. Schroder, *Semiconductor Material and Device Characterization*, 2nd ed., p. 701.
32. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation*, 2nd ed., p. 332.
33. Ibid.

GAS CONTROL IN PROCESS CHAMBERS

Semiconductor manufacturing is a cyclic repetition of several major process steps. Many fabrication processes involve chemical reactions that take place in process chambers. The driving force for these chemical reaction processes is to optimize the required chemical reactions by introducing the correct chemicals in the proper environment (e.g., vacuum) while providing energy to drive the reaction. At the same time, detrimental aspects of

the reaction are minimized, such as exposure to moisture, the ambient environment, and contaminants. This optimum condition is reached by carefully introducing the necessary mix of precursor chemicals into the process chamber, often as a gas, and then monitoring the chemical reaction to achieve the desired conditions on the wafer surface.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain why process chambers are used in semiconductor manufacturing.
2. Describe the benefits of a vacuum, the vacuum ranges, and appropriate pumps.

3. Explain the need for gas flow in process chambers, and describe how it is controlled.
4. Explain what an RGA is and why it is beneficial in process chambers.
5. Describe what plasma is and how it is obtained.
6. Discuss the effects of contamination in process chambers and explain how to minimize it.

INTRODUCTION

During the beginning of the semiconductor industry, only two wafer fabrication steps required a vacuum chamber for processing. A vacuum was used for the evaporation of aluminum for the one and only metal layer and for the evaporation of gold on the back of the silicon wafer so that the circuit die could be mounted in its transistor package.¹ Vacuum processes in those early days occurred in a chamber known as a bell jar (see Figure 8.1 on page 182).

Present-day wafer processing often requires chemical reactions that take place in process chambers. The *process chamber* is a controlled vacuum environment where intended chemical reactions occur under controlled conditions. For this reason, a process chamber for chemical reactions is sometimes referred to as a *reactor*. Process chambers have many functions:

- ◆ Controlling how gas chemicals flow into and react in the chamber in close proximity to the wafer.
- ◆ Maintaining a prescribed pressure inside the vacuum environment.
- ◆ Removing undesirable moisture, air, and reaction by-products.
- ◆ Creating an environment for chemical reactions such as plasma to occur.
- ◆ Controlling the heating and cooling of the wafer.

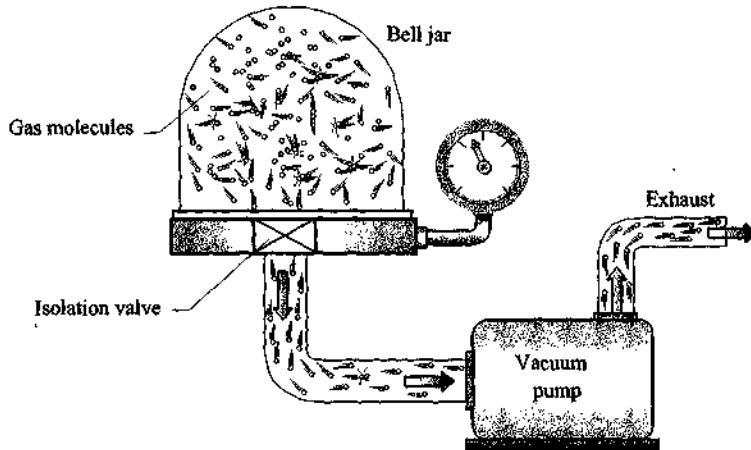


FIGURE 8.1 Early 1960s Vacuum Bell Jar

Gas is typically the material state used to transport the various chemicals needed in the process chambers. Gas flows into the process chamber as a result of the pressure differential between the supply system and the vacuum conditions in the chamber. Sometimes solid materials are used in the process chamber. An example is the solid metal target used as a source of material during sputtering (see Chapter 12).

Since the late 1980s, process chambers have been configured in the form of a *cluster tool*. The multiple process chambers are clustered around a central transfer chamber with a wafer transport system that is typically a robot arm (see Figure 8.2). This equipment design permits integration of multiple process steps. Wafers are transported from process chamber to process chamber under vacuum, eliminating native-oxide formation and reducing contamination on the wafer. Cluster tools also improve wafer manufacturing throughput (defined as the number of wafers processed per unit time) because there is no need to vent the chamber during wafer transfer steps.

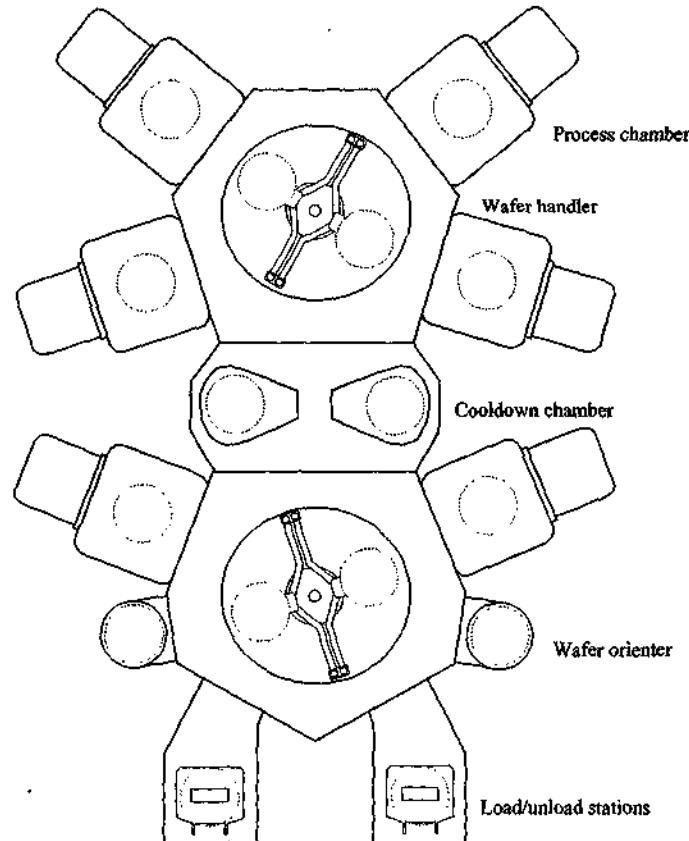


FIGURE 8.2 Integrated Cluster Tool
(Used with permission from Applied Materials, Inc.)



Cluster Tool with Integrated Process Chambers
(Photo courtesy of Applied Materials, Inc.)

VACUUM

Many chemical reactions in wafer fabrication are performed in a vacuum environment. A *vacuum* exists when there is less pressure in an enclosed volume than in the surrounding atmospheric pressure. The benefits of a vacuum in semiconductor manufacturing are shown in Table 8.1.²

TABLE 8.1 Benefits of Vacuum in Semiconductor Manufacturing

Vacuum Condition	Benefit
1. Create clean environment	Removes particles, unwanted gases, moisture, and contaminants.
2. Low molecular density	Reduces the number of molecules in the system to reduce contamination and to move a gas out of the way (lower the molecular interference).
3. Extend distance between collisions of molecules (Mean Free Path)	Provides the necessary condition for creating the plasma needed in semiconductor processes such as sputtering and etch.
4. Accelerate reactions	Helps accelerate processes by lowering the vapor pressure of materials so they can react faster with other chemicals.
5. Create a force	Creates a force, such as a vacuum pickup on a robot arm for wafer handling.

Vacuum Ranges

The following terms describe the different vacuum ranges: low vacuum, medium vacuum, high vacuum (also called high vac), and ultrahigh vacuum (UHV). These ranges are shown in Table 8.2 on page 184.³ As a reference, the vacuum of deep space is about 10^{-16} torr.

Low vacuum (also referred to as *rough vacuum*) has two important characteristics:⁴ gas flow is primarily by collisions between molecules (also known as *viscous flow*) and the pressure is high enough that true mechanical pressure gauges can be used. Low vacuum is commonly used in fabrication processes that depend on gas-phase chemical reactions, momentum transfer between

molecules and/or a high rate of interactions between the gas and surfaces. *Medium vacuum* is from about 1 torr to 10^{-3} torr and is a transition between low and high vacuum. *High vacuum* is characterized by having few collisions between gas molecules (molecular flow). This condition results in very clean wafer surfaces. *Ultrahigh vacuum* is a continuation of high vacuum with stringent control of the vacuum chamber design and materials to minimize undesirable gas contaminants.

TABLE 8.2 Vacuum Ranges

Wafer Fab Processes	Vacuum Ranges in Torr				Chapter in book
	Rough $759 \cdot 10^9$	Medium $10^0 \cdot 10^{-3}$	High $10^{-3} \cdot 10^{-6}$	Ultrahigh $10^{-6} \cdot 10^{-9}$	
Oxidation	Wafer handlers Atmospheric tools				10
Photo	Vacuum chucks Wafer handlers				13 - 15
Polish	Wafer handlers Slurry removal				18
Etch	Plasma resist strippers Plasma etchers				16
Deposition	Batch deposition tools Single wafer deposition tools				11
Metallization	Metal evaporators Metal sputtering tools				12
Ion Implant	Batch ion implanters Single wafer ion implanters				17
Metrology	Wafer inspection tools for quality checks and diagnostics Analytical tools for research and failure analysis				7

Mean Free Path

The average distance a gas molecule moves before it strikes another molecule is known as the *mean free path (MFP)*. When the pressure is lowered in a vacuum, the space between the gas molecules increases, which is an important factor for how gases flow through the system and for creating a plasma in the process chamber. The mean free path for air at different pressure regimes and standard temperature is given in Table 8.3.

TABLE 8.3 Mean Free Path and Molecular Density versus Pressure

	760 Torr (atmosphere)	1×10^{-3} Torr	1×10^{-9} Torr
# of molecules/cm ³	3×10^{19} (30 million trillion)	4×10^{13} (40 trillion)	4×10^7 (40 million)
Mean Free Path	2×10^{-6} inches	2 inches	30 miles

VACUUM PUMPS

There are many different vacuum pumps used in semiconductor manufacturing. For our purposes, they can be categorized into two general types: roughing pumps and high vacuum pumps. *Roughing pumps* have several purposes: to achieve a rough to medium vacuum (pressure down to 10^{-3} torr) in a chamber, to evacuate the entry area for wafers into a cluster tool (known as the loadlock), and to exhaust a high vacuum pump (see Figure 8.3). *High vacuum pumps* are used for achieving high and ultrahigh vacuum from 10^{-3} torr to 10^{-9} torr. Modern vacuum pumps for new wafer fabs are dry, meaning they contain no oils or lubricants that can backstream into the process chamber and contaminate wafers as well as the process chamber. This chapter addresses only dry pumps as the intent is to focus on current technologies needed to support sub-quarter micron processes.

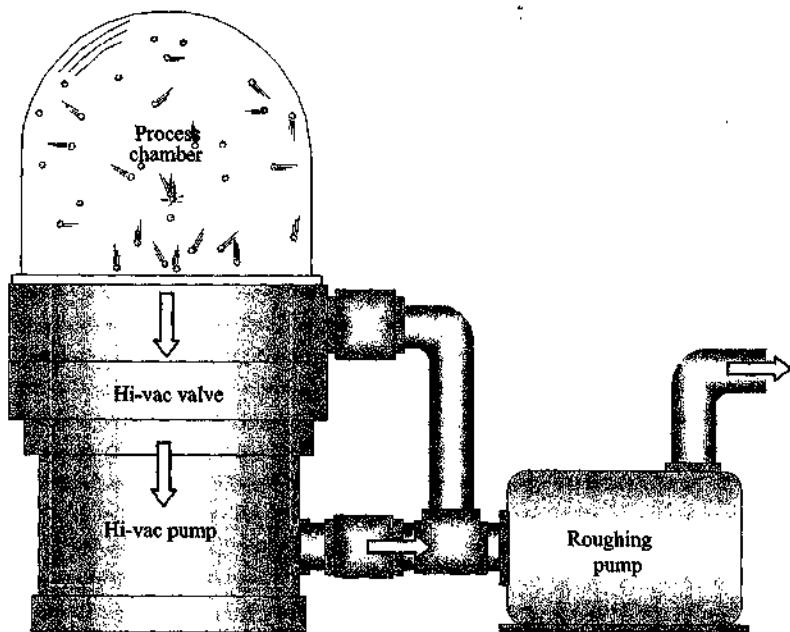


FIGURE 8.3 Roughing Pump Exhausting a High Vacuum Pump

Roughing Pump

When chamber pumpdown starts, roughing pumps remove more than 99.99% of the initial air or contaminants in the chamber. Removing this material with the roughing pump permits the high-vacuum pump to be optimized for removal of the remaining water vapor and gas molecules attached to the walls of the chamber. There are various types of roughing pumps, each with characteristics that work for a specific application. Two types of roughing pumps are:

- ◆ Dry mechanical pump
- ◆ Blower/booster pump

Dry Mechanical Pump ■ A *dry mechanical pump* employs mechanical devices to remove gases, such as the rotary claw dry pump (see Figure 8.4). The pump principle is typically based on increasing the chamber volume, thus lowering the pressure (i.e., Boyle's law). A mechanical pump often uses nonmetallic materials on the moving surfaces to avoid the use of sealing or lubricating fluids.

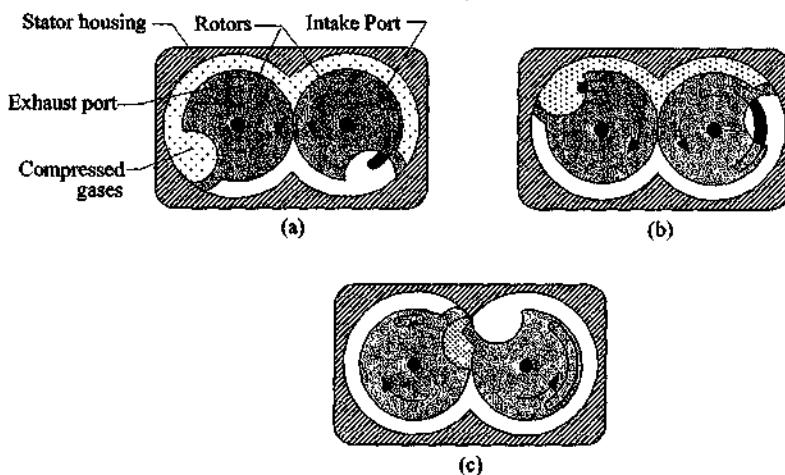


FIGURE 8.4 Rotary Claw Dry Mechanical Pump
(Used with permission from International SEMATECH)

Blower/Booster Pump ■ The *blower* or *booster pump* (also referred to as a *lobe pump*) is a widely used mechanical pump because it provides for high throughput of gas and requires no lubricants. It is desirable for systems where a high volume of gas must be pumped at a rough vacuum range. A blower is also referred to as a *Roots blower* or a *Roots-type blower*. The principle for how a blower works with the mating lobes is shown in Figure 8.5. A blower is often exhausted into a roughing pump because it will not pump at atmosphere under viscous flow. Newer blower pumps are being designed to exhaust directly to the atmosphere without the use of a roughing pump.

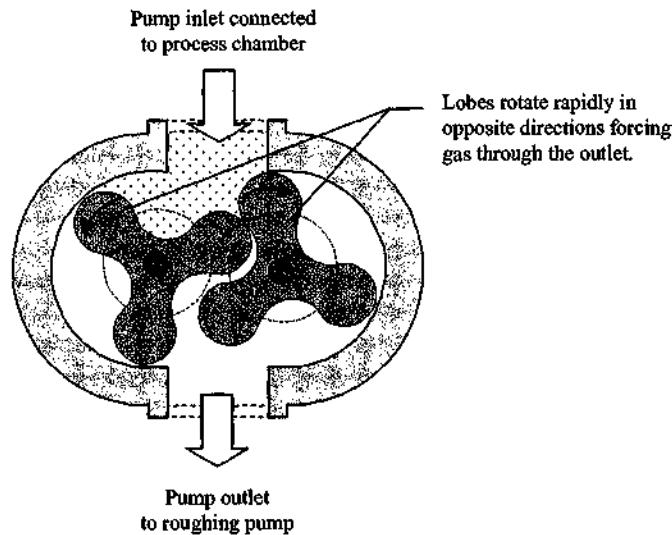


FIGURE 8.5 Roots Blower Pump

High Vacuum Pump

Two common high vacuum pumps are:

- ◆ Turbomolecular pump
- ◆ Cryopump

Turbomolecular Pump ■ The *turbomolecular pump* (usually called a *turbo pump*), is a versatile, reliable, and clean pump that is widely used in wafer fab equipment. The turbo pump is capable of reaching a pressure of 10^{-10} torr if a bakeout is done to drive out moisture in the process chamber. This pump also gives a fast start-up to full pumping action once turned on.

A turbo pump works on the principle of mechanical compression. There are a number of high-speed rotating blades (resembling blades in a jet engine) positioned between fixed blades that impart momentum and direction to the gas molecules (see Figure 8.6). Each set of rotating and fixed blades is a compression stage—a pump may have from ten to forty of these stages. Turbo pump blades will rotate at speeds up to 90,000 RPM. A turbo pump exhausts into a roughing pump because it cannot pump with viscous flow at atmospheric pressure.

Turbo pumps are specially designed for semiconductor applications. Magnetically levitated bearings (referred to as mag-lev) are common to eliminate the need for lubrication and special antivibration designs. The turbine blades can be designed to produce high gas throughput and high compression ratio for fast pumpdown of light gases.

The most common causes for failure of a turbo pump are exposing the pump suddenly to atmospheric pressure (called dumping the pump), particulate entering the pump, or physical shock. When a turbo pump is suddenly dumped to atmosphere, the turbine blades flex and come in contact with each other, causing catastrophic failure. Because the rotors on a turbo pump are precisely balanced, the pump should never be moved or bumped while it is in use. Extensive maintenance is not required on turbo pumps; most fabs run pumps until they crash and then change them.⁵

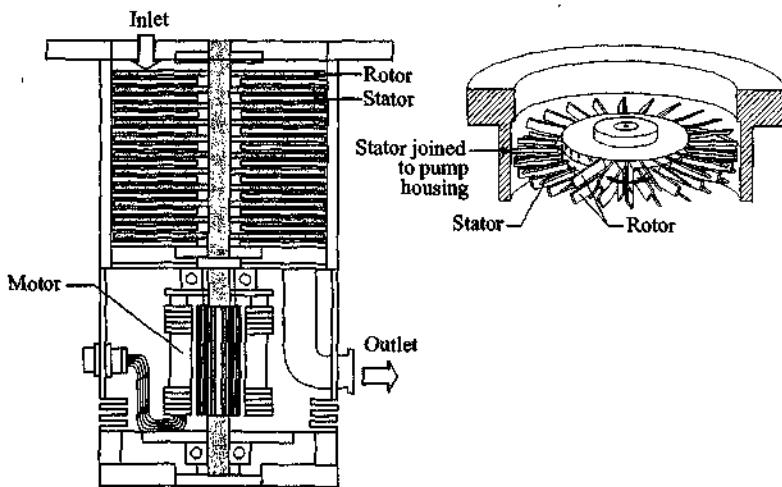


FIGURE 8.6 Turbo Pump Blades
(Used with permission from Varian Vacuum Systems)

Cryopump ■ A *cryopump* is a capture pump that removes gases from the process chamber by making them so cold that they are frozen and captured in the pump, thus providing the pumping action. The cryopump has become the industry standard in semiconductor equipment for pumping to the high and ultrahigh vacuum range. It has high gas throughput while pumping, plus has a very high water vapor pumping speed. This feature is useful when pumping down from atmosphere to remove the moisture in the chamber. The cryopump is extremely clean, with no oils or moving parts exposed to the vacuum. This attribute is desirable for wafer fabrication, which is a reason why this pump is common on new production equipment. There are two main components to a cryopump: a gaseous helium compressor and a pump module with a cold head, baffle, and body (see Figure 8.7).

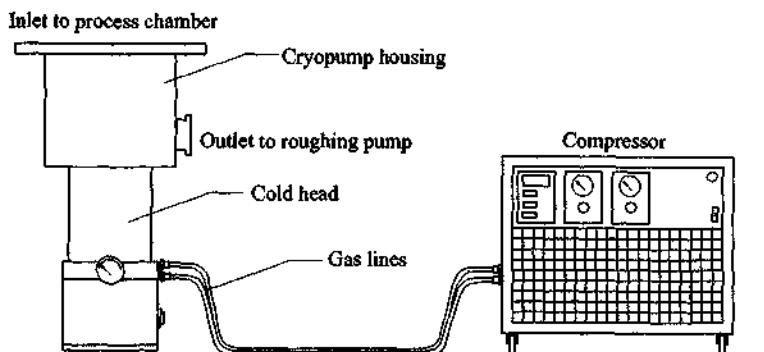


FIGURE 8.7 Cryopump Compressor and Pump Module
(Used with permission from Varian Vacuum Systems)

The helium compressor works similar to a regular refrigeration compressor (e.g., in the air conditioning system at home), except that it supplies high-pressure, high-purity, room-temperature helium to the expander module in the pump. When the gas expands in the pump module from a high-pressure to a low-pressure condition, the helium takes in heat and refrigeration occurs. This process produces cryogenic temperatures from 80 K down to 50 K. The pump has a second stage that expands the helium again and reaches approximately from 20 K to 10 K. Once the helium expands, it comes in contact with and cools multiple surfaces called *cryoarrays* (see Figure 8.8 on page 188). It is on these surfaces that the gases from the vacuum chamber are cooled and condensed or adsorbed. The condensed gases are immediately frozen on the cold cryoarray surfaces and trapped, which is essentially the pumping action.

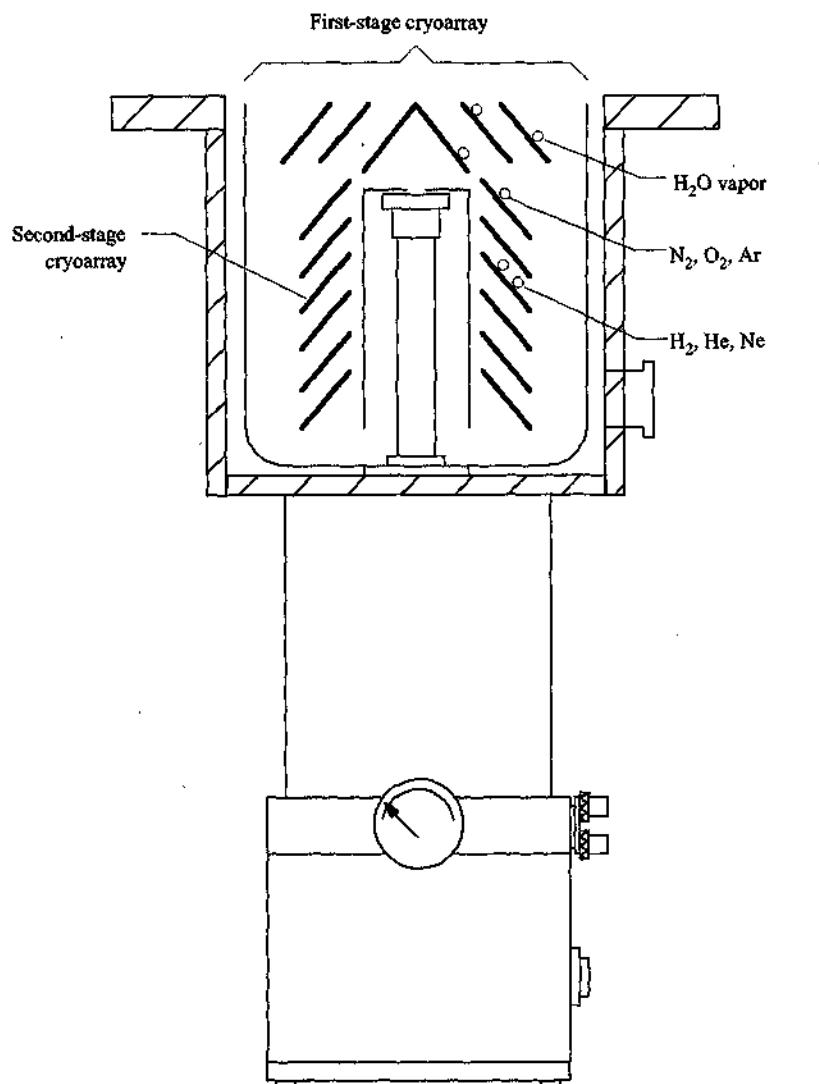


FIGURE 8.8 Cryoarray Surfaces in Pump Module
(Used with permission from Varian Vacuum Systems)

Since cryopumps capture gases by freezing the molecules rather than compressing and expelling them, the gases actually accumulate as frozen solids on the cryoarray surfaces. This process is similar to ice (frozen water) that forms on the inside of a freezer. These captured gases are removed periodically through a process called *regeneration*, where the pump is warmed to room temperature or above and the gases are vented through an appropriate vent line. Cryopumps require a roughing pump to remove the air from the pump and vacuum system.

Vacuum in Integrated Tools

The vacuum environments of an integrated cluster tool depend on the requirements of each individual process chamber (see Figure 8.9).⁶ Chambers are isolated from one another, with a progressively better level of vacuum from the loadlock to the process chamber. The *loadlock* is where wafers enter the cluster tool, isolating the inner regions of the tool from the workplace environment. The system is designed to provide a well-controlled, low-contamination environment for wafer preparation and processing.

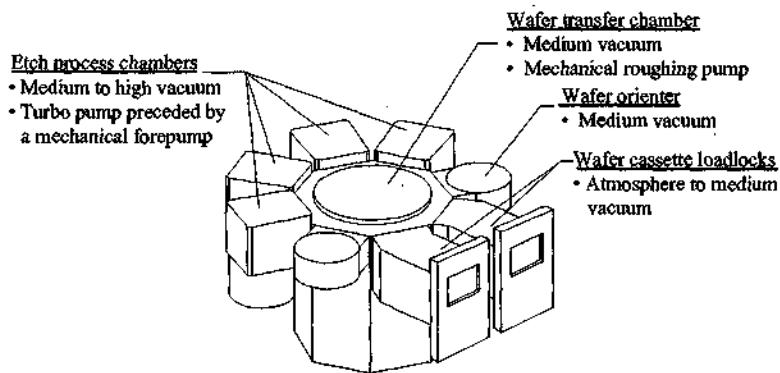


FIGURE 8.9 Cluster Tool Layout with Vacuum Environment

PROCESS CHAMBER GAS FLOW

The flow of gases into the process chamber is critical to attain the desired chemical reaction. The basic process chamber requirements for gas flow are:⁷

- ◆ Ability to handle a wide variety of bulk and specialty gases, many of which are corrosive and toxic.
- ◆ The control of gas flow into the process chamber is accurate and repeatable.
- ◆ The gas mix proportions are able to be controlled during the process run.
- ◆ Materials used in the chamber are not affected by the process gases and do not introduce contaminants into the gas stream.

When discussing gas delivery systems or vacuum, the mass quantity of gas flow at standard conditions is called *throughput* (Q). Throughput is the net number of gas molecules that pass through a point of the vacuum system in a specified period of time. Throughput defines the volumetric gas flow in a system at standard conditions. The most common units for throughput are torr-liters per second, standard cubic centimeters per minute (sccm), or standard liters per minute (slm).

Pumps are specified for a certain *pump speed*, which indicates how effectively the pump can remove gases. Pump speed is typically expressed in units of volume per unit time (such as liters/second or cubic feet per minute).

Throughput and pump speed are important in many semiconductor process steps. The chemical reactions that take place at the surface of the wafer may require high gas flow rates (which means high throughput). Factors such as the type of pump, its pumping speed, its location in the system, and the gas throughput are important variables that can determine whether an acceptable chemical reaction occurs at the wafer surface.

Mass Flow Controllers

Chemical reactions involve physical processes where molecular quantities are important for proper control of the reactions. From the ideal gas law we know that the number of gas molecules in a given volume changes in proportion to the absolute pressure and temperature. Thus, controlling gas flow into a chamber only by volume will not always yield the same number of gas molecules, which is undesirable for controlling chemical reactions.

To overcome this problem, gas flow into process chambers is controlled by use of a *mass flow controller* (MFC), shown in Figure 8.10 on page 190. MFCs use the heat-transfer property of the gas to directly measure the mass flow rate into the chamber. It employs a thermal sensor to detect changes in the mass flow of the gas. Integrated tools will typically use many MFCs to control the flow of the various gases into the process chambers. A pressure regulator is required ahead of the MFC to ensure that a constant specified pressure is delivered to the MFC.

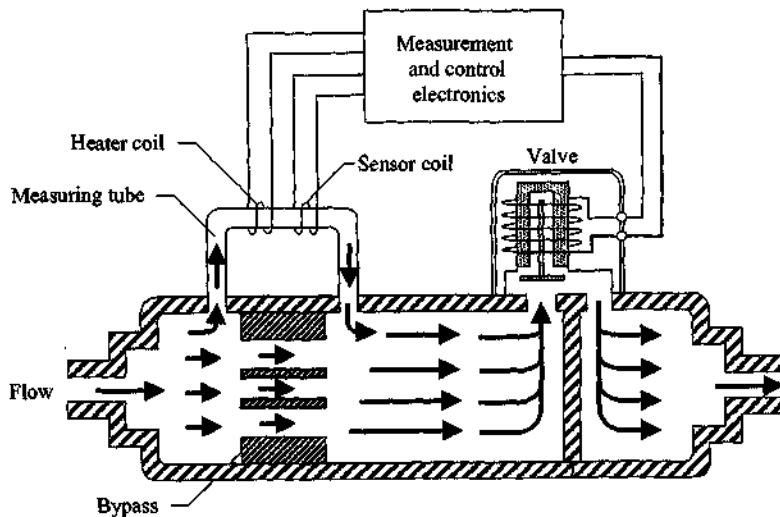


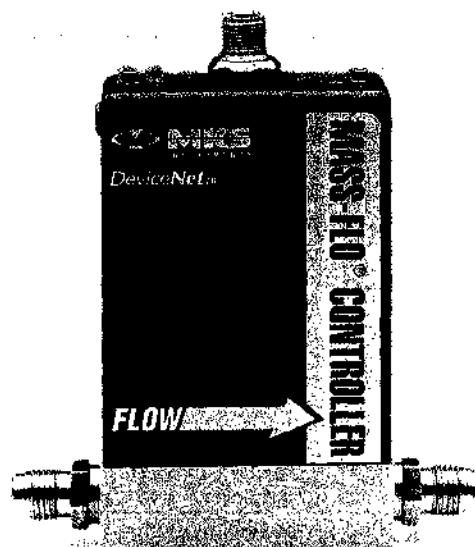
FIGURE 8.10 Thermal Mass Flow Controller
(Used with permission from International SEMATECH)

RESIDUAL GAS ANALYZER (RGA)

The *residual gas analyzer (RGA)* is an important process chamber instrument to identify the types of gas molecules remaining in an evacuated system. In this manner, it can be used for leak detection, and analysis of contamination in process chambers, and as a troubleshooting tool to solve vacuum-based problems in chambers.⁸ Its most common applications are leak detection and process troubleshooting.

RGA Basics

The principle of the RGA is to separate, identify, and measure the quantity of all gas molecules in the chamber. An RGA measures the partial pressure contribution of each gas present in the vacuum system, as well as the total pressure from all gas molecules. It typically operates only in the high and ultrahigh vacuum ranges, but it is acceptable to use an RGA for vacuum conditions up to 10 millitorr.



Mass Flow Controller
(Photo courtesy of MKS Instruments, Inc.)

There are four basic parts to an RGA: an ionizer, an aperture, an analyzer, and a detector (see Figure 8.11). These are essentially the parts of a mass spectrometer, but an RGA is much smaller and is designed for attaching to process tools. These four parts are located in the sensing head of the RGA behind a process-specific valve inlet attached directly to the process chamber.

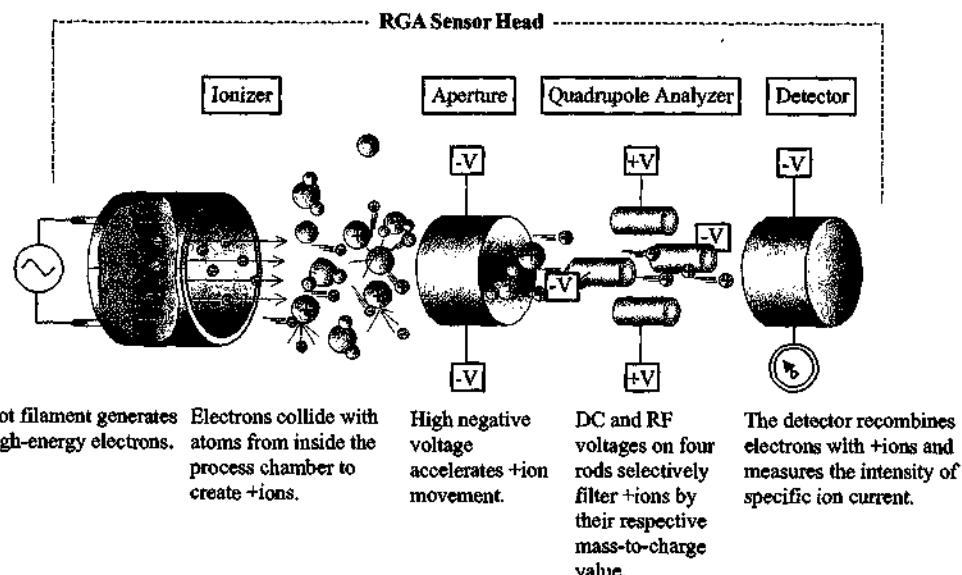


FIGURE 8.11 Basic Parts of Residual Gas Analyzer (RGA)

The ionizer creates ions from the gas molecules present in the system by bombarding the molecules in the chamber with electrons generated in the ionizer. The bombardment removes an electron from the gas molecules and creates positive ions. These ions are directed toward the analyzer by applying a voltage potential or magnetic field in the aperture. The analyzer is where the ions are separated by mass. There are several different types of analyzers; a common type is the *quadrupole mass analyzer (QMA)*. It consists of four cylindrical rods that have both a constant DC potential and a high-frequency RF component (see Figure 8.12). For a given voltage level applied to the cylinders, only ions of a given atomic mass-to-charge pass through the filter. All other ions are grounded on the cylinder rods. The detector on modern RGAs will have better than 1 atomic mass unit (amu) resolution to differentiate between the various ions. As voltages are progressively changed on the filter, different ions are passed through the cylinders. In this manner, various gases of different masses are separated and identified.

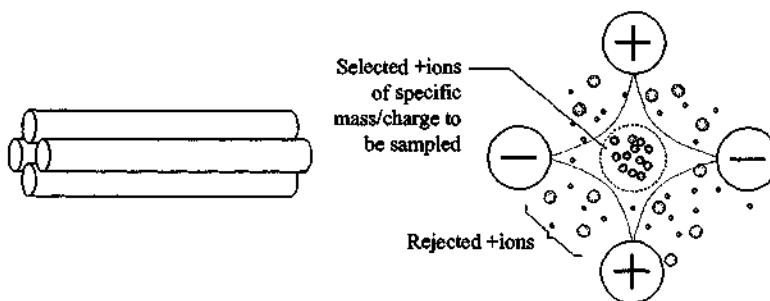
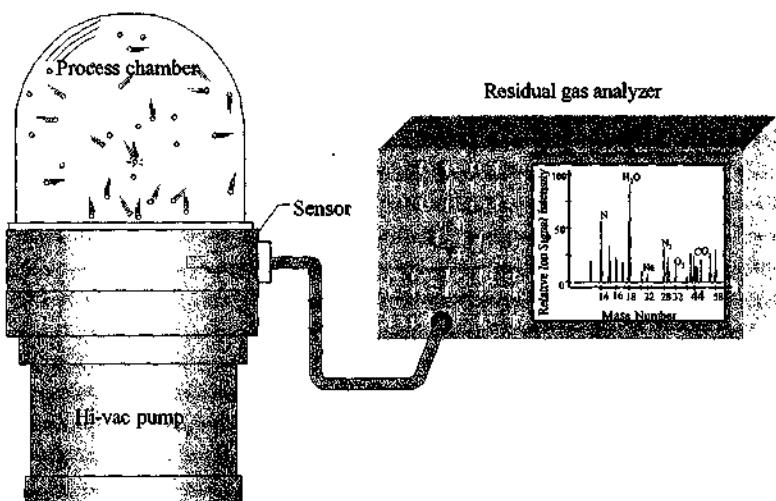


FIGURE 8.12 Quadrupole Mass Filter

RGA as Real-Time Monitor

RGA data can be an important part of verifying that a process chamber is ready for processing, especially for large-diameter wafers (see Figure 8.13 on page 192). This is because RGA data can provide real-time information about the cleanliness and stability of the process chamber during pumpdown.

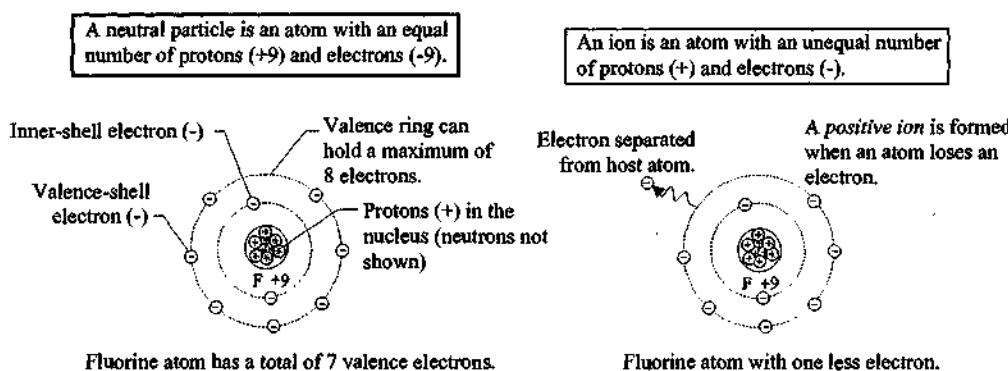
**FIGURE 8.13** RGA Data Output

Using the RGA as an in situ (real-time) process monitor is a recent development. While the wafer is undergoing processing, the RGA will directly monitor the chemical constituents in the chamber to verify that the process is proceeding correctly.⁹ The RGA can be used to replay conditions in the event of a wafer quality problem. This replay ability allows for more rapid diagnosis of process problems.

There is a growing need for RGAs to monitor chemical processes, such as plasma etching and plasma-enhanced chemical vapor deposition (PECVD).¹⁰ For plasma applications, an RGA can follow the different process chemicals as the reaction occurs in the chamber and give insight into plasma behavior and how the chemical species change over time.

PLASMA

Plasma is a neutral, highly energized, ionized gas consisting of neutral atoms or molecules, positive ions, and free electrons. Positive ions and free electrons are formed when a valence electron is removed from a neutral atom. For instance, fluorine is neutral when there are an equal number of protons and electrons in its atomic structure. Fluorine is ionized when an electron is separated from its host atom (see Figure 8.14). Ionization of gas atoms in a confined process chamber can occur by subjecting the gas to strong DC or AC electromagnetic fields or by bombarding the gas atoms with some sort of electron source. These methods will be covered later in greater detail.

**FIGURE 8.14** Creation of an Ion

Plasma is used at various process steps during wafer fabrication because it supplies much of the energy needed to support a gas reaction near the wafer surface in a process chamber. For example, plasma is used in lieu of thermal energy to ionize and excite a source gas to deposit thin films in high-density plasma chemical vapor deposition (HDP-CVD), discussed in Chapter 11. Another plasma application is to selectively remove metal through plasma etching (see Chapter 16). The most common indication that a plasma exists in a process chamber is the characteristic observable light referred to as a glow discharge (see Figure 8.15).

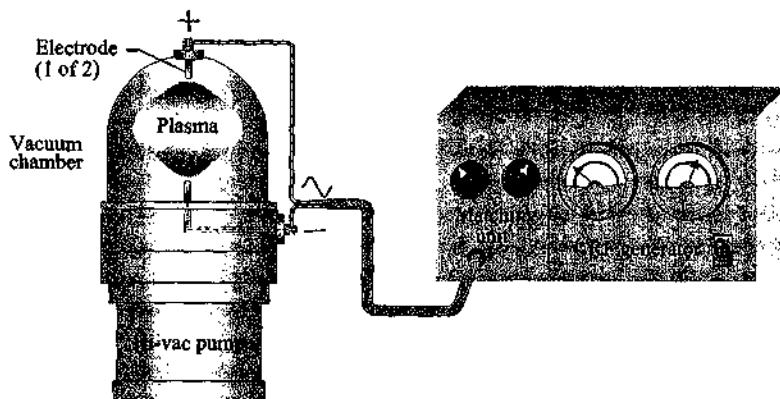


FIGURE 8.15 Plasma Glow Discharge

Glow Discharge

Although a glow discharge across a gas mixture can be created by applying DC power, the most common methods use AC power in the radio frequency (RF) range. When the electrical power is first applied, free electrons in the gas mixture are greatly influenced by the presence of electric fields. Electrons are accelerated through the gas mixture and collide with atoms and molecules, releasing additional electrons during the collisions.

In a weakly ionized plasma, or glow discharge, such as that commonly used in wafer fabrication, the high-energy electrons collide with neutral atoms or molecules and excite them. These excited atoms or molecules are short-lived, with lifetimes measured in nanoseconds. When an excited atom or molecule returns to its lowest energy state, energy is released during this relaxation in the form of a radiated photon (or light). This release causes the characteristic glow in the glow discharge (see Figure 8.16), with different glow colors for different species (e.g., oxygen, nitrogen, fluorine, and so on). The high-energy electrons in a plasma transfer energy to the neutral atoms and molecules by impact and initiate a reaction which could not have occurred in a high pressure environment. Typical parameters sustaining a glow discharge include the RF power and frequency, pressure, gas mixture and flow rate, vacuum pumping speed, and surface temperature.

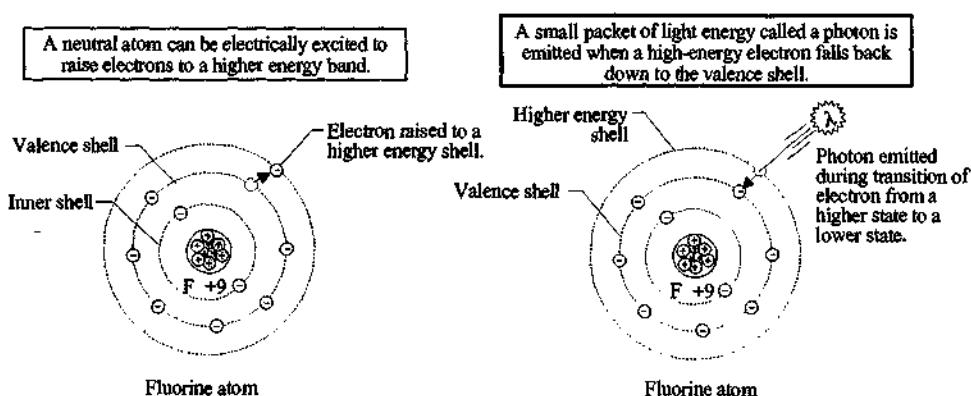


FIGURE 8.16 Electrically Exciting and Relaxing an Atom

Radicals ■ Radicals are highly reactive chemical species. As electrons collide with atoms or molecules, different kinds of products are created. The dissociation (or fragmentation) of a molecule occurs when a large molecule is broken into smaller particles (see Figure 8.17). Recall that the number of electrons in the valence shell are related to the chemical properties of an atom. A *radical* is created when a neutral molecule is bombarded by an energetic electron. This action cleaves a bond without adding or removing an electron, creating a highly reactive species. This radical is an uncharged atom or molecular fragment that has incomplete bonding (or unpaired electrons) in the valence shell.

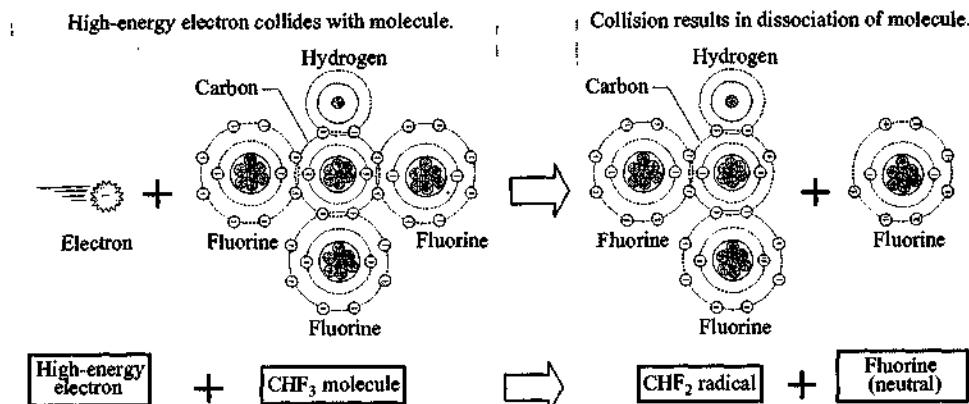


FIGURE 8.17 Dissociation of a Molecule

RF Energy ■ The energy of plasma is sustained through the absorption of RF radiation in an applied alternating current (AC) electric field of several hundred volts rms. The use of RF power, typically at 13.56 MHz (chosen for industrial purposes by the FCC, or Federal Communication Commission), creates a high-efficiency plasma. In recent years, different frequencies have started being used for plasma generation, such as 400 KHz, 2 MHz, 4 MHz, and 2.45 GHz. The frequency has a direct influence on the ion mobility, thus affecting process uniformity and the rate of the process (e.g., the metal removal rate for etching). The RF field is applied between a negative cathode and a positive anode, referred to as electrodes. The wafer typically will be held against the grounded electrode, with the RF power applied to a parallel electrode (see Figure 8.18).

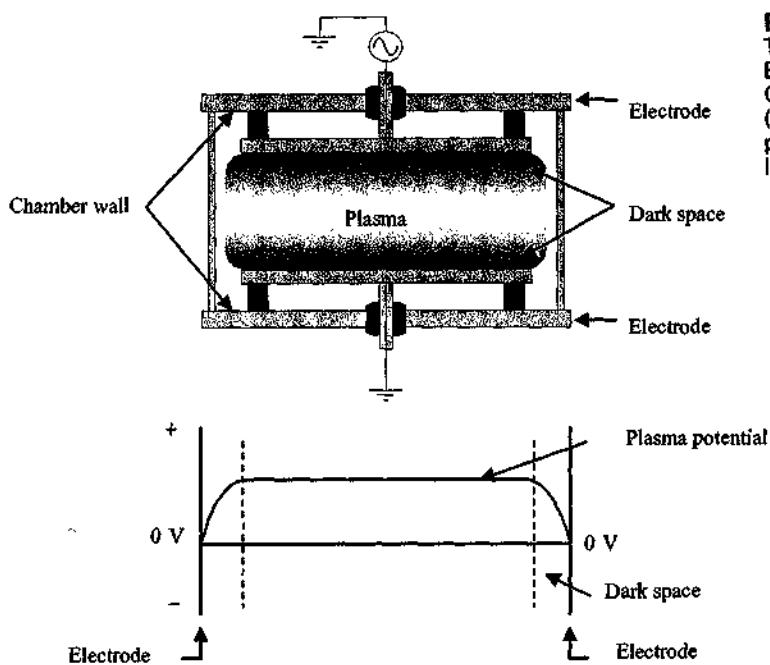


FIGURE 8.18
Typical Plasma
Electrode
Configuration
(Used with
permission from
International SEMATECH)

The electrons, ions, and other species generated in a glow discharge will move toward the electrodes. Electrons in the glow discharge can move faster toward the positive electrode when compared to the slow moving massive positive ions. This movement produces a positive potential in the glow discharge region. The positive ions move toward the cathode and cross through a *dark space*, or ion sheath, adjacent to the cathode. The dark space results from the lack of electrons in this region. With respect to the glow discharge region, the dark space has a large voltage drop and strong electric field. This condition accelerates the positive ions toward the cathode, which then ejects secondary electrons. These secondary electrons are repelled from the cathode and cross back through the dark space where they maintain the glow discharge.

PROCESS CHAMBER CONTAMINATION

Wafer fabrication equipment with vacuum can be viewed as two general types: tools that process large batches of wafers (e.g., 150 or 200 wafers) or single-wafer, multichamber tools with loadlocks and several integrated process chambers. In general, chamber contamination on wafers is lower in single-wafer reactors than in large batch tools. The improvement in single-wafer cluster tools is attributed mostly to maintaining the processing chambers under relatively constant temperature and pressure conditions, which minimizes factors that contribute to particle generation.

Trace amounts of water are probably the most significant source of contamination in process chambers with vacuum.¹¹ Water is a problem in vacuum systems since it is reduced to ions as well as H₂ or O₂ gas molecules because of the reactor chemistry and vacuum. It sticks tenaciously to surfaces, outgasses slowly, and is a chemical poison in reactors. For example, the tenacity of water sticking to chamber walls is so high that it is estimated that there are 10,000 times more water molecules on the chamber walls than in the vacuum space of one cubic meter of processing chamber. It is also believed that water creates particle contamination during pumpdown by the formation of particles.¹²

In order to minimize water adsorption contamination in process chambers, it is desirable to reduce the need to open or disassemble process reactors for cleaning purposes. This goal is often accomplished by using *in situ* cleaning techniques (cleaning while the wafers are processed). Nevertheless, process equipment still requires maintenance actions, which necessitate equipment shutdown and repair. The specific steps followed by the maintenance technician during servicing have a significant effect on the control of contamination in the equipment. Recommendations for minimizing contamination during the servicing of wafer fabrication equipment are listed in Table 8.4.¹³

TABLE 8.4 Recommendations to Minimize Contamination During Equipment Servicing

Recommendations
<ol style="list-style-type: none"> 1. Maintain good temperature and humidity control in the cleanroom environment where the equipment is located. 2. Control the equipment's pump and vent cycles to minimize turbulence and prevent particle generation when processing wafers. 3. Avoid abrasive cleaning materials. 4. Use exact replacement parts and materials to avoid subtle sources of equipment contamination. 5. Use low particle-generating gas-handling components, such as regulators and automatic valves that have a tendency to generate particles.

SUMMARY

Process chambers have many functions in wafer fabrication, such as controlling gas flow. A vacuum in a process chamber creates a clean environment with low molecular density and high mean free path. Vacuum ranges are low, medium, high, and ultrahigh. The most common vacuum pumps in the wafer fab generally can be categorized as roughing (dry mechanical and blower/booster) and high vacuum (turbo pump and cryopump). The mechanical pump is a low vacuum pump. The blower is a low to medium vacuum that has high throughput. The turbo pump is a common high vacuum pump that uses turbine blades, while the cryopump is a capture pump that removes gases by freezing them. Gas

flow in the process chamber is controlled using a mass flow controller (MFC). The residual gas analyzer (RGA) is used to detect leaks or to analyze residual gas in a chamber after vacuum pumpdown. It also can do real-time monitoring of the reaction in a chamber during a process run. Plasma is a highly energized gas that is commonly used in process chambers to excite source gases during a reaction. It involves a glow discharge that results from collisions between energetic electrons that produce radicals. Plasma energy is sustained by the absorption of RF radiation. Cleanliness in the process chamber is critical, with water moisture being the most serious contaminant.

KEY TERMS

process chamber
reactor
cluster tool
vacuum
torr
low vacuum (rough vacuum)
medium vacuum
high vacuum
ultrahigh vacuum
mean free path (MFP)
roughing pump
high vacuum pump
dry mechanical pump
blower (booster pump or lobe pump)

turbomolecular pump (turbo pump)
cryopump
cryoarrays
regeneration
loadlock
throughput
pump speed
mass flow controller (MFC)
residual gas analyzer (RGA)
quadrupole mass analyzer (QMA)
plasma
glow discharge
radical
dark space

REVIEW QUESTIONS

1. What is a process chamber? What are its five functions?
2. Describe a cluster tool, and explain why it is beneficial in IC fabrication.
3. What is a vacuum?
4. What are the benefits of vacuum in semiconductor manufacturing?
5. State the most common vacuum unit, and describe how to interpret it.
6. List and describe four vacuum ranges.
7. What is mean free path? Why is this condition important?
8. Give reasons for using a roughing pump and a high vacuum pump.
9. Describe two types of roughing pumps. Which one has the higher gas throughput?
10. Describe two types of high vacuum pumps.
11. What are the most common reasons for failure of a turbo pump?
12. Describe the purpose of the cryoarray, and explain how it creates pumping action.
13. What is regeneration?

14. Draw a picture of a cluster tool and show chambers with at least three different vacuum levels.
15. What is the purpose of the loadlock in a cluster tool?
16. State the four basic process chamber requirements for gas flow.
17. Describe gas throughput.
18. What is pump speed?
19. What is the purpose of a mass flow controller?
20. What does a residual gas analyzer (RGA) do?
21. List and describe the three basic parts to an RGA.
22. How does the quadrupole mass analyzer work on an RGA?
23. What is plasma? How is plasma beneficial in a process chamber?
24. Describe the plasma glow discharge region.
25. What is a radical in a plasma?
26. Why is RF energy used in plasma?
27. Why is moisture a problem in process chambers?
28. List the recommended steps to minimize contamination during equipment servicing.

VACUUM EQUIPMENT SUPPLIERS' WEB SITES

Alberta University Vacuum Page

Alcatel Vacuum Products

Apiezon Products

AVS, American Vacuum Society

BOC Edwards

CTI Cryogenics

Ebara Technologies

Granville-Phillips

Inficon Inc.

Leybold Vacuum

Millipore Corp.

MKS Instruments

Omega Engineering Inc.

Osaka Vacuum Ltd.

Parker Hanniflin Corp.

Pfeiffer Vacuum Tech., Inc.

SEMI

Unit Instruments

Varian Inc.

Varian Vacuum Technologies

VAT Valve

Veeco Instruments Inc.

<http://nyquist.ee.ualberta.ca/~schmaus/vacf/>

<http://www.alcatel.com/>

<http://www.apiezon.com/>

<http://www.vacuum.org/>

<http://www.boc.com/edwards/>

<http://www.ctivacuum.com/>

<http://www.ebaratech.com/>

<http://www.helixtechnology.com/>

<http://www.leyboldinficon.com/>

<http://www.leyboldvac.de/>

<http://www.millipore.com/>

<http://www.mksinst.com/>

<http://www.omega.com/>

<http://www.osakavacuum.com/>

<http://www.veriflo.com>

<http://www.pfeiffer-vacuum.com/>

<http://www.semi.org/>

<http://www.unit.com/>

<http://www.varianinc.com/>

<http://www.varianinc.com/vacuum/>

<http://www.vatvalve.com/>

<http://www.veeco.com/>

REFERENCES

1. R. Waits, "Semiconductor Manufacturing and Vacuum Technology: A Memoir," *Solid State Technology* (May 1997): p. 105.
2. H. Tompkins, *The Fundamentals of Vacuum Technology*, (New York: American Vacuum Society, 1997), p. 6.
3. C. Tilford and J. P. Looney, "Vacuum Measurement: The Basics," *Semiconductor International* (May 1994): p. 73.
4. Ibid.
5. J. Baliga, "Vacuum Pump Designs Adjust to Harsher Conditions," *Semiconductor International* (October 1997): p. 88.
6. S. Hansen, *Introduction to the Creation and Control of the Vacuum Process Environment*, (Andover: MKS Instruments, 1995), p. 118.

7. Ibid., p. 110.
8. L. Peters, "Residual Gas Analysis: A Technology at a Crossroads," *Semiconductor International* (October 1997): p. 95.
9. T. Banks, G. Diamond, and S. Ruck, "Integrating Mass Spectrometry Data into the Fab Environment," *Semiconductor International* (June 1997): p. 138.
10. Ibid., p. 98.
11. A. Rapa and A. Bross, "Contamination Control in Multilevel Interconnection Manufacturing," *Handbook of Semiconductor Interconnection Technology*, ed. G. Schwartz, K. Srikrishnan, and A. Bross, (New York: Marcel Dekker, 1998), p. 552.
12. Ibid.
13. Ibid., p. 553.

IC FABRICATION PROCESS OVERVIEW

The typical IC wafer fab process may take as many as six to eight weeks and involve 350 or more process steps to complete the full manufacturing flow. The complexity of this process can be overwhelming.

Recall that most semiconductor processing occurs in the top few microns of a silicon wafer. This activity corresponds to the first part (front end) of the manufacturing

process flow. All materials on top of the silicon are part of the layering strategy needed to interconnect the many devices on the chip. To add multiple metal and insulating layers, the process flow requires the wafer to cycle through the different process steps. Once you understand the process, you will realize that only a few process areas are used many times over to fabricate a high-performance microchip.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Draw a diagram showing how a typical wafer flows in a sub-micron CMOS IC fab.
2. Give an overview of the six major process areas and the sort/test area in the wafer fab.
3. Describe the primary purpose of each of the 14 CMOS manufacturing steps.
4. Discuss the key process and equipment used in each CMOS manufacturing step.

INTRODUCTION

This chapter describes a simple manufacturing process for producing 0.18- μm CMOS integrated circuits on a silicon wafer. This is presented to help the reader gain a better understanding and appreciation for semiconductor manufacturing. Specific details about each process step is provided later in this text in the process chapters.

An overview of the entire wafer fab process is first presented, with a model that depicts how wafers repeatedly cycle through a few major process areas. This explanation simplifies the concept of keeping wafer flow to a manageable level in a fab. Recognize that when changes are made to the process, such as parameter or tool changes, the results of the change may not be known until many days or weeks later during wafer test at the end of the process. This fabrication complexity makes it important for each process area to flawlessly perform its tasks by meeting the defined quality measures at each step.

CMOS PROCESS FLOW

IC manufacturing is a complicated sequence of chemical and physical operations that are performed on a silicon wafer. Simply stated, the operations fall into four basic categories: layering, patterning, etching, and doping. Figure 9.1 on page 200 illustrates how complex the process can be even for manufacturing a single MOS transistor.

Since CMOS technology is the most popular of the process families, we have chosen it as an example to describe wafer process flow. The specific example is that of a $0.18\text{ }\mu\text{m}$ CMOS integrated circuit process flow. And, since this is an overview of IC manufacturing, you will be introduced to a variety of terms and concepts, which will be further explained in subsequent chapters. As you study this chapter, keep in mind the manufacturing area where specific process operations are being performed. Note the purpose of each operation, the type of equipment and materials that are used, and the quality measures that are followed to determine the integrity of the process at each step.

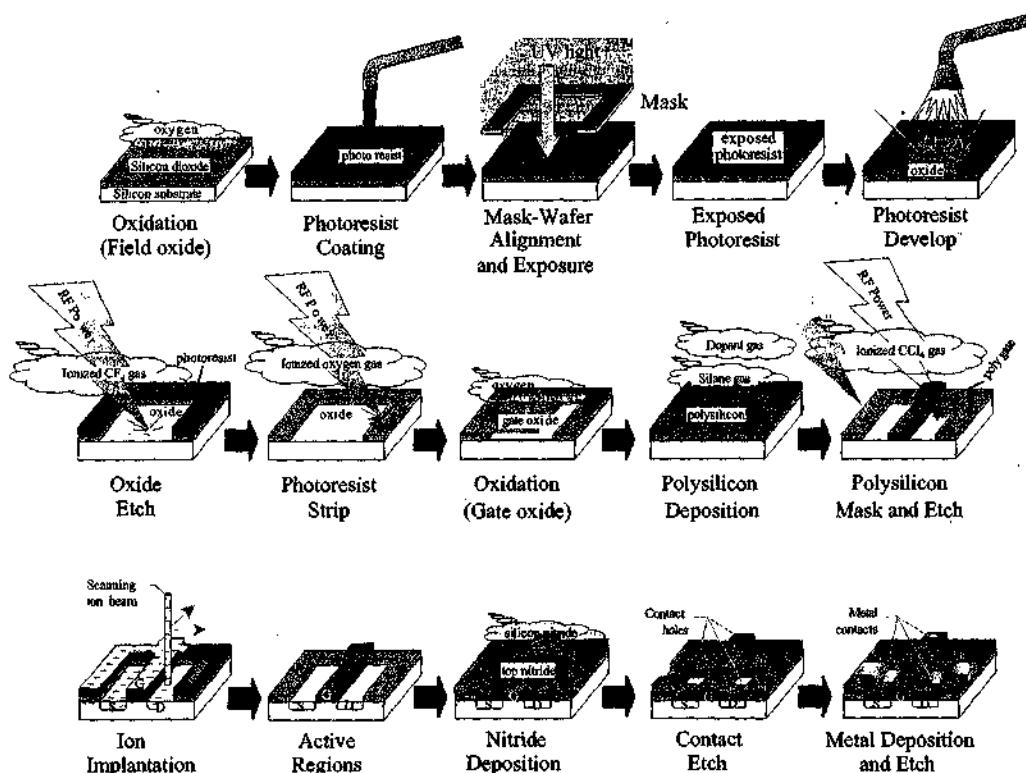


FIGURE 9.1 Major Fabrication Steps in MOS Process Flow
(Used with permission from Advanced Micro Devices)

Overview of Areas in a Wafer Fab

ICs are manufactured in a wafer fab. As shown in Figure 9.2, wafer fabs are generally divided into six distinct production areas: diffusion (which includes oxidation, film deposition, and doping processes), photolithography, etch, thin films, ion implant, and polish. These six major production areas and their related process and metrology tools are all housed in the ultraclean cleanrooms of the wafer fab. The polish area is a new addition to semiconductor manufacturing for high-performance ICs and is gaining popularity in the industry. Although located near the wafer fab, the test/sort area for testing individual die on the wafer is not housed in the same cleanroom environment as the other areas of the fab. Assembly and packaging plants are generally located in other facilities, perhaps even in other countries.

Diffusion ■ The *diffusion* bay is recognized as the area where high-temperature processing and film depositions are performed. The primary tools in the diffusion area are a *high-temperature diffusion furnace* and a *wet cleaning station*. High-temperature diffusion furnaces (see Figure 9.3) can operate at temperatures near 1200°C and are configured to run a variety of processes, including oxidation, diffusion, deposition, anneals, and alloys. These processes will be covered in detail in later chapters. Wet cleaning stations are the secondary tools used in the diffusion area. Wafers must be cleaned thoroughly to remove contamination and native-grown oxide on the surface before inserting the wafers into the furnaces.

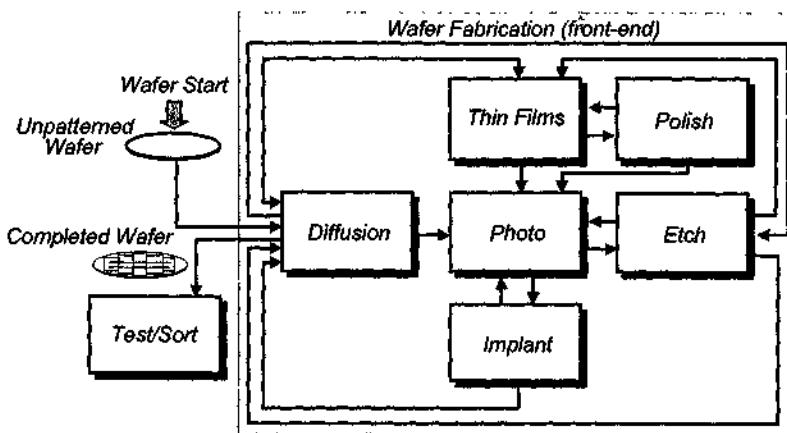


FIGURE 9.2 Model of Typical Wafer Flow in a Sub-Micron CMOS IC Fab
(Used with permission from Advanced Micro Devices)

Photolithography ■ The *photolithography* bay is recognizably different than the other areas in the fab due to the yellow fluorescent tubes that light up this production bay. The purpose of photolithography is to photograph the image of a circuit pattern onto the photoresist that coats the wafer surface. Photoresist is a light-sensitive chemical that captures the image of a mask pattern resulting from exposure to ultraviolet (UV) light. Photoresist is sensitive to certain wavelengths of light, such as UV and white light, but the wavelengths of yellow light do not affect it.

The *coater/developer track* is a cluster tool used to perform many of the operations in photolithography. This tool primes the wafer, coats it with photoresist, spins the wafer to smooth out the photoresist, bakes the wafer, and uses robotics to transfer the resist-coated wafer to the alignment and exposure tool. The purpose of the *stepper* is to align the wafer to an array of die patterns etched on a chrome-coated quartz reticle. When properly aligned and focused, the stepper exposes a small area of the wafer, then steps to the next field and repeats the process until the entire wafer surface has been exposed to the die patterns on the reticle (see Figure 9.4 on page 202). When completed, the wafer returns to the coater/developer tracks where the resist is developed, then the wafer is rinsed and baked again.

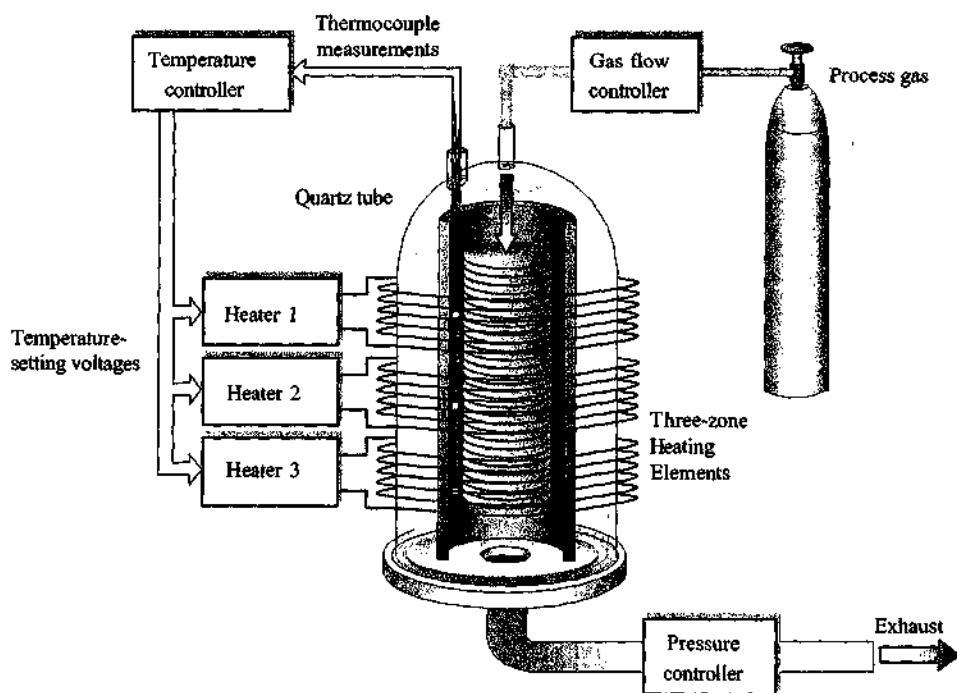
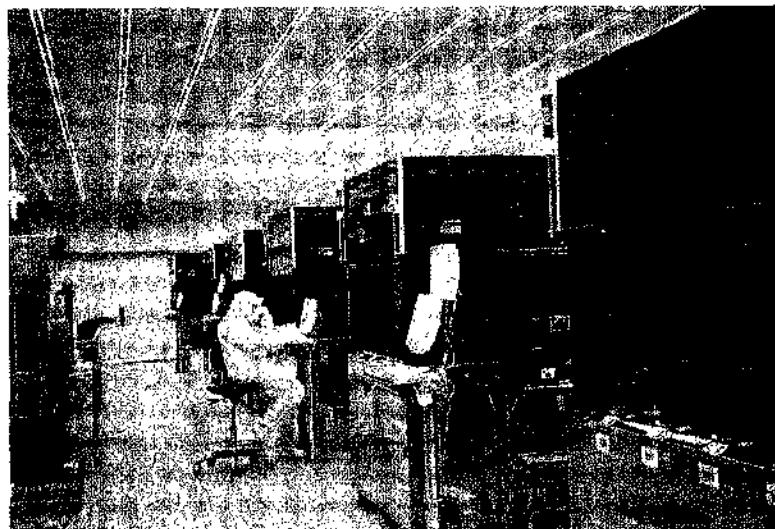


FIGURE 9.3 Simplified Schematic of a High-Temperature Furnace



Photolithography Bay in a Sub-Micron Fab
(Photo courtesy of Advanced Micro Devices)

Referring once again to Figure 9.2 on page 201, photolithography is shown to be at the center of the wafer fab. This position is due to the fact that wafers flow into photolithography from all other areas of the fab. Contamination control is especially important here because particles or defects can become imbedded in or on the resist films during the photographic process. A defect in the photographic mask or reticle or a particle on the stepper can be photographed onto all wafers that are processed with these tools.

To reduce contamination, open containers of chemicals are prohibited from being used in this area. Thus, cleaning stations and photoresist strippers are usually located in other areas of the fab other than photolithography. Note that in Figure 9.2 on page 201, wafers flow from photolithography into only two other areas: etch and ion implant. Consequently, there are only three production areas where photoresist-coated wafers can be found—photo, etch, and ion implant.

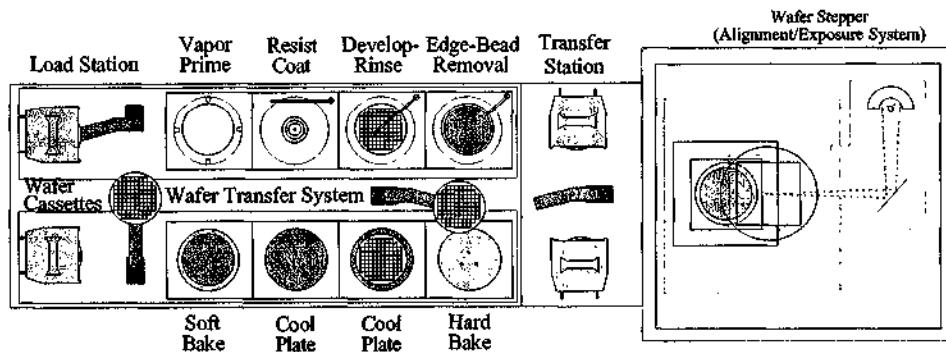


FIGURE 9.4 Simplified Schematic of a Photolithography Processing Module

Etch ■ The *etch* process creates a permanent pattern on the wafer in areas not protected by the photoresist pattern. The most common tools in the etch bay are the *plasma etcher*, the *plasma resist stripper*, and the wet cleaning station. Today most etching steps are done with dry plasma etchers (see Figure 9.5), although some wet etch processes are still in use. The plasma etchers are tools that use radio frequency (RF) energy to ionize gas molecules inside a vacuum chamber. The plasma is the glow given off by the electrically-excited gases. The gases react with the top layer of material on the wafer. After the etch process, another plasma system, called the plasma stripper, uses ionized oxygen to remove the photoresist from the wafer. This step is followed by thoroughly cleaning the wafers with a combination of chemicals.

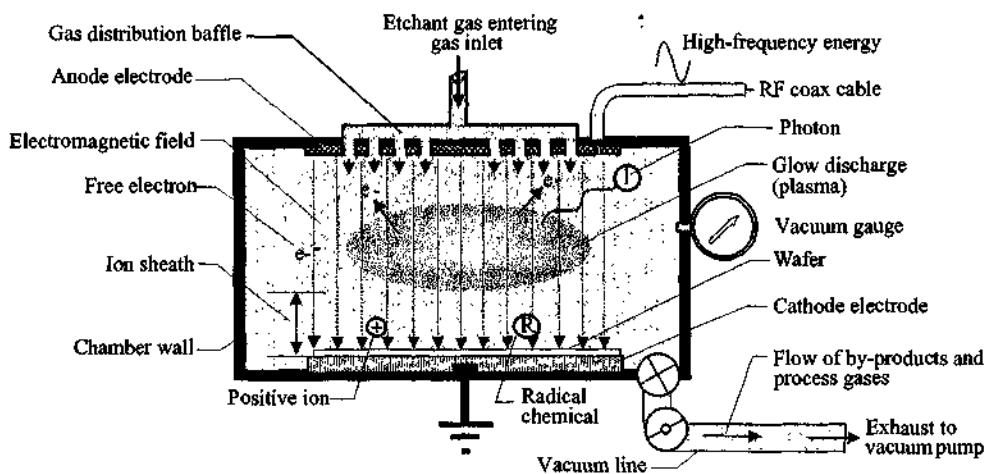


FIGURE 9.5 Simplified Schematic of a Dry Plasma Etcher

Ion Implant ■ The *ion implanter* is the most popular tool for doping wafers in a sub-micron process. Gases carrying the desired dopant, e.g., arsenic (As), phosphorus (P), and boron (B), are ionized inside the implanter (see Figure 9.6). High voltages and magnetic fields are used to control and accelerate the ions. The high energy of the dopants penetrates through the surface of the resist-coated wafer. After implantation is completed, the photoresist is stripped off and the wafer is thoroughly cleaned.

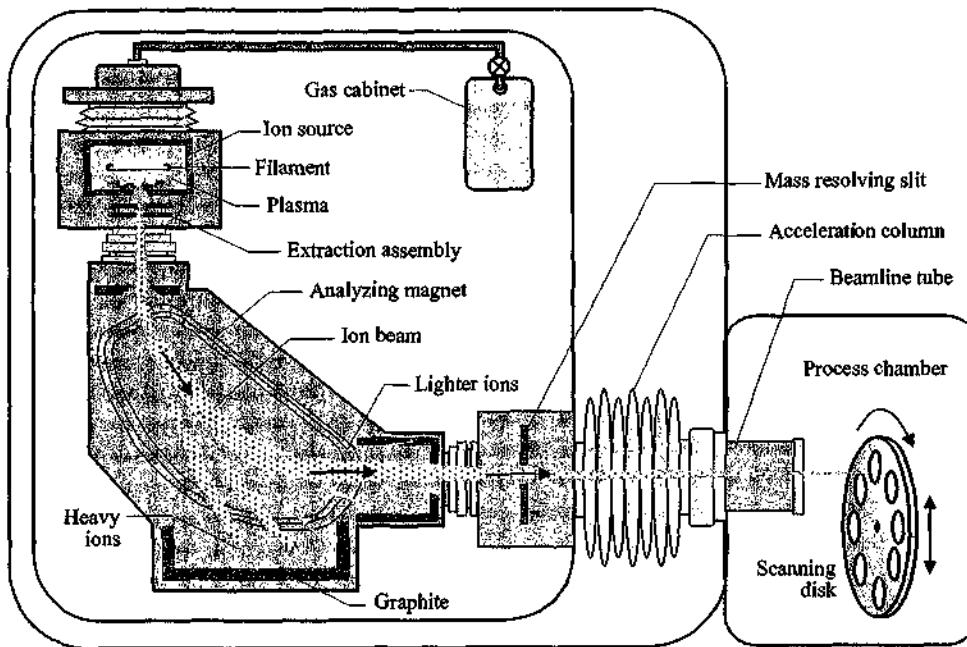


FIGURE 9.6 Simplified Schematic of an Ion Implanter

Thin Films ■ The *thin films* bay is primarily responsible for depositing dielectric and metal layers during different steps of the manufacturing process. The processes in thin films operate at lower temperatures than the furnaces in the diffusion area. There are many diverse tools in this area. All thin film deposition equipment operates under low to medium vacuum conditions (see Figure 9.7 on page 204), including chemical vapor deposition (CVD) and metal sputtering tools (PVD, or physical vapor deposition). Other tools used in this area may include the spin-on-glass (SOG) system,

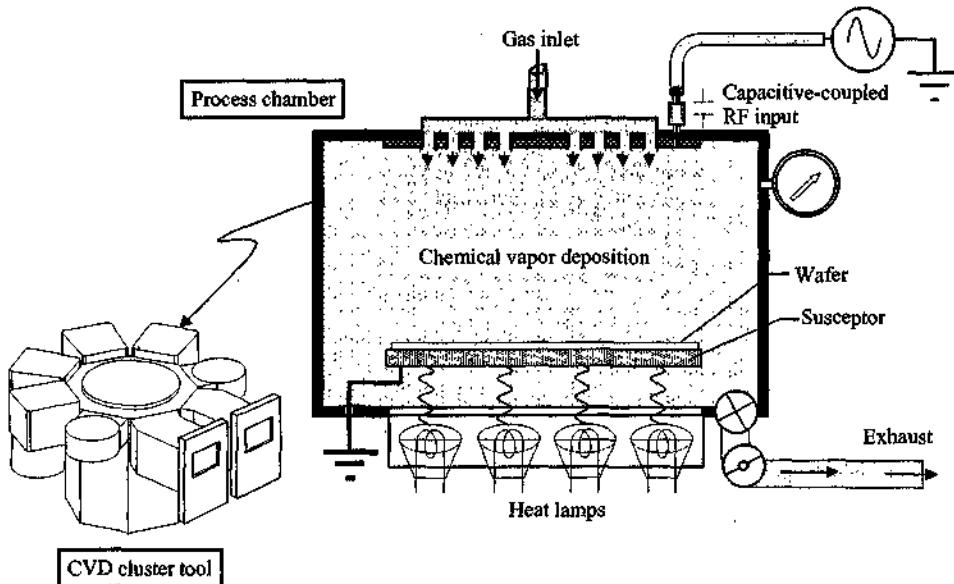


FIGURE 9.7 Simplified Schematics of a CVD Cluster Tool and Process Chamber

rapid thermal processor (RTP) system, and wet cleaning station. The SOG is used to fill in low areas of the wafer for the purpose of planarizing (smoothing) the surface of the wafer. RTAs are used to anneal ion implant damage of the silicon substrate and for metal alloy processing steps.

Polish ■ The purpose of the CMP (*chemical mechanical planarization*) process is to planarize the top surface of the wafer by lowering the high topography to be level with the lower surface areas of the wafer. CMP minimizes uneven wafer surfaces that make further processing difficult. The polisher is the primary tool in the CMP area, and this process is also referred to as *polish*. CMP uses a combination of chemical etching and mechanical abrading to remove a desired amount of the upper layer of the wafer. Other tools that support CMP include *wafer scrubbers*, cleaning stations, and metrology tools.



Polish Area in a Sub-Micron Wafer Fab
(Photo courtesy of Advanced Micro Devices)



Thin Film Metallization Work Bay
(Photo courtesy of Advanced Micro Devices)

CMOS MANUFACTURING STEPS

The remainder of this chapter focuses on the manufacturing steps that a wafer follows in a typical CMOS process. Although there are some operations in the process flow that can be performed in batches of wafers at one time, this process flow describes the manufacturing of a single wafer. For further simplification, this process describes the production of only a microscopic area of a single CMOS inverter consisting of two transistors—nMOS and pMOS. Cross-sectional views of the product will be provided for each major operation.

The CMOS manufacturing steps to be described are:

1. Twin Well Process
2. Shallow Trench Isolation Process
3. Poly Gate Structural Process
4. Lightly Doped Drain (LDD) Implant Processes
5. Sidewall Spacer Formation
6. Source/Drain (S/D) Implant Processes
7. Contact Formation
8. Local Interconnect Processes
9. Via-1 and Plug-1 Formation
10. Metal-1 Interconnect Formation
11. Via-2 and Plug-1 Formation
12. Metal-2 Interconnect Formation
13. Metal-3 to Pad Etch and Alloy
14. Parametric Testing

1. Twin Well Process

The first step in some CMOS wafer fabrication processes is to define the active regions of the MOSFETs. Many of today's sub-quarter micron processes use a *twin-well* (also referred to as *twin-tub*) approach to define the active regions of the nMOS and pMOS transistors. A twin-well consists of a p-well and an n-well, with each well requiring at least three to five major processing steps to fabricate. A *retrograde implant* technique is used to obtain optimum electrical characteristics for the FETs. This technique starts with a high-energy, high-dose implant that penetrates approximately 1 μm into the epilayer. Subsequent well implants are done at the same active site with a progressive

decrease in energy, junction depth, and dose of dopant material such as phosphorous or boron (measured in atoms/cm²). The results help set the threshold operating voltage for the FETs as well as prevent common CMOS problems, such as *latchup* and other reliability issues. Figure 9.8 and Figure 9.9 illustrate, respectively, the major steps required in the formation of the n-well and p-well.

n-well Formation ■ The five major steps in n-well formation are described below in the following process and illustrated in Figure 9.8.

Description of Five Major Steps in n-well Formation

Process Step	Description
1. Epitaxial Growth	The silicon wafer, having already a thin layer of <i>epitaxial</i> silicon (epilayer), arrives in the diffusion bay. The epilayer has the exact crystal structure as the substrate, except with improved purity and fewer crystal defects. The epilayer has already been lightly doped with p-type dopant (boron).
2. Initial Oxide Growth	In diffusion the wafer is cleaned in a series of chemical baths to remove particles, organic and inorganic contamination, and <i>native oxide</i> growth (natural silicon dioxide growth) from the wafers. After the wafer is rinsed and dried, the wafer is placed in a high temperature (~1000 °C) furnace process chamber. Oxygen is flowed into the process chamber to react with silicon to grow approximately 150 Å of oxide. This initial oxide serves several functions: (1) it protects the top surface of the epi silicon from contamination, (2) it prevents excessive damage to the silicon during implantation, and (3) as a screen oxide, it helps control the depth of the dopants during implantation.
3. 1st Mask, n-well Implant	In photolithography the wafer undergoes a series of process steps in the coater/developer track tool. The "tracks," as commonly referred to, prime the top surface of the wafer, coat the wafer with photoresist (liquid photographic film), spin the wafer, and bake the wafer. An internal automatic wafer-handling system transfers the wafer between process stations within the tracks. Another wafer handler removes resist-coated wafers, one at a time, and transfers them to the <i>alignment-and-exposure system</i> (an extremely complex and precise camera). The alignment-and-exposure system photographs the image of a specific mask layer directly onto the resist-coated wafer. In this case, the mask layer defines the areas of the product that are to be implanted in order to form the n-wells for the pMOS transistors. The exposed wafer is then transferred back to the tracks, where the image first appears when the wafer is sprayed with the chemical developer. The developed wafer undergoes another bake step and is inspected before being transferred to the ion implant bay. The printed pattern is inspected for proper <i>linewidhts</i> , called <i>critical dimensions (CD)</i> . In the event of a major defect, the photoresist can be stripped off and the wafer can then be reworked. Photolithography is the only area in the fab where a wafer can be easily reworked.
4. n-well Implant (High energy)	The patterned wafer arrives in the ion implant bay. The photoresist pattern covers specific areas of the wafer that are to be protected from ion implantation. Windows, or openings, in the photoresist allow high-energy positive dopant ions to penetrate into the upper surface of the epilayer (~1 μm <i>junction depth</i>). In this case, phosphorus is the desired dopant. The <i>ion implanter</i> is the main process tool in this bay. Its purpose is to ionize dopant atoms, accelerate them with high voltage (~200 KeV), select the most appropriate dopant species to implant, focus the ions into a narrow beam, and, finally, scan the wafer to provide uniform doping across all unprotected areas of the wafer. At this point the dopant ions penetrate the crystal lattice of the silicon causing damage to the covalent atomic structure. This damage will be repaired later in a diffusion and annealing step.
<i>Note:</i> After each ion implant operation, an oxygen-plasma reactor tool strips the photoresist off each implanted wafer. The wafer is then cleaned by a wet chemical process to remove residual photoresist and polymers created by the plasma process. This note applies to all subsequent implant operations but will be listed only this one time.	
5. Anneal	The implanted wafer is transferred to diffusion where the wafer is cleaned before being inserted into an <i>anneal</i> furnace. Four things occur as a result of this anneal process. (1) A new barrier oxide layer is grown over the bare silicon. (2) The higher temperatures cause the movement of dopants further into the silicon (called diffusion). (3) The implant damage is repaired. (4) The bonds between the dopant atoms and the silicon atoms are activated, making the dopant atoms a part of the crystal lattice structure (electrical activation).

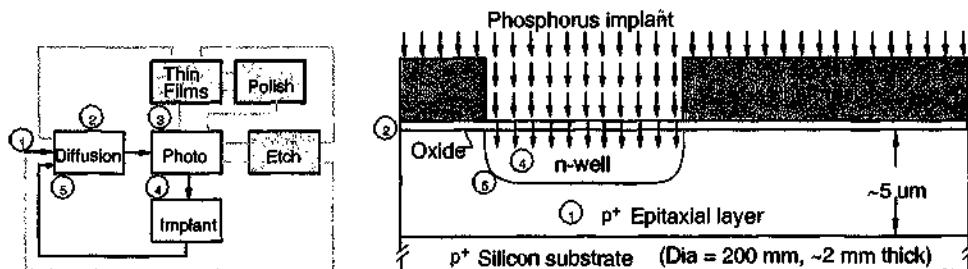


FIGURE 9.8 n-well Formation

p-well Formation ■ The three major steps in p-well formation are described in the following process and illustrated in Figure 9.9.

Description of Three Major Steps in p-well Formation

Process Step	Description
1. 2nd Mask, p-well Implant	The photolithography steps for the p-well implant mask are exactly the same as for the first mask. The only difference is the mask is the direct opposite of the n-well implant mask. Compare Figure 9.8 and Figure 9.9 to see the differences.
2. p-well Implant (High energy)	The retrograde p-well implant energy levels are considerably lower than the n-well. This condition is due to the difference in the masses of the elements being implanted. Comparing the atomic mass unit (amu) of boron (~11) to phosphorus (~31), boron is about one-third the mass of phosphorus. Thus, the energy needed to implant boron should only require about one-third of that used to implant phosphorus to the same junction depth.
3. Anneal	This anneal step is basically the same as the first annealing step.

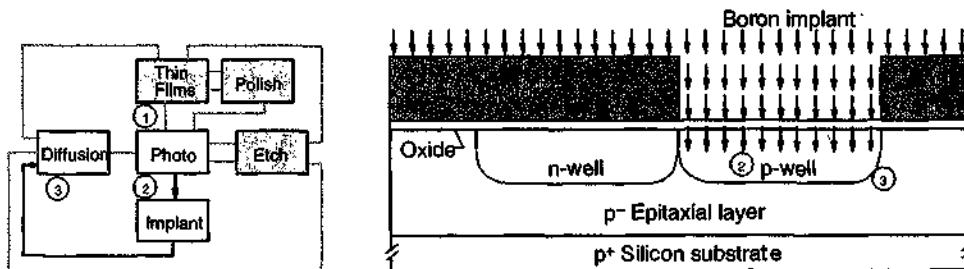


FIGURE 9.9 P-well Formation

2. Shallow Trench Isolation Process

Shallow trench isolation (STI) is an alternative method for creating isolation regions between active transistor areas on a substrate.¹ This method is especially useful in the manufacturing of quarter-micron devices. The preferred isolation technique up until the late 1990s had been the LOCOS (local oxidation of silicon) technique that was instituted in the early 1970s.² Despite its complexity, STI is gaining in popularity with ULSI chip manufacturing. STI formation is explained as follows in three major steps: trench etch (see Figure 9.10 on page 208), oxide fill (see Figure 9.11 on page 209), and oxide polish (see Figure 9.12 on page 209).

STI Trench Etch ■ The four major steps in STI trench etch are described in the following process and illustrated in Figure 9.10.

Description of Four Major Steps in STI Trench Etch

Process Step	Description
1. Barrier Oxide	The wafer arrives in the diffusion bay where the wafer is again cleaned to remove contamination and oxide. After rinsing and drying, the wafer is inserted into a high-temperature oxidation furnace. Approximately 150 Å of new oxide is grown on the wafer. This layer will form a barrier to protect active regions from chemical contamination during nitride strip.
2. Nitride Deposition	The wafer is inserted into a high-temperature (~750 °C) <i>LPCVD (low-pressure chemical vapor deposition)</i> furnace. Inside the process chamber ammonia and dichlorosilane gases react to form a thin layer of silicon nitride (nitride, Si ₃ N ₄) on the wafer. The nitride will serve two functions throughout the STI formation process steps. (1) Nitride is a durable masking material, which protects active regions during the STI oxide deposition process and (2) nitride serves as a polish-stop material during the chemical mechanical planarization (CMP) step.
3. 3rd Mask, STI	The wafer arrives in photolithography from diffusion. This photolithography process step is similar to the previous masking step, except a different mask is used. This masking step is much more critical than the first mask due to the smaller dimensions. Inspection of the wafer includes CD measurements as well as checking for <i>defect inspection (DI)</i> , and optical <i>visual inspection (VI)</i> . These measurements require checking the accuracy of the alignment-exposure system relative to the written specifications.
4. STI Trench Etch	The photoresist pattern is designed to protect areas of the silicon that are not to be etched. The windows in the photoresist allow ions and highly reactive <i>radical chemicals</i> to etch nitride, oxide, and silicon in unprotected areas. The preferred tool for etching these deep trenches is the <i>dry plasma etcher</i> . The etcher uses high-powered radio frequency (RF) energy to ionize a fluorine- or chlorine-based gas inside a vacuum process chamber. The RF energy dissociates molecules and ionizes atoms to create a chamber filled with a variety of plasma components. These plasma components provide the chemical and physical etching that results in the removal of silicon in locations designated as <i>isolation regions</i> . The slanted profile and the rounded bottom of the trenches improve the filling process and the electrical characteristics of the isolation structure.

Note: Following completion of each etch operation, the wafer is stripped of photoresist and wet cleaned in a series of chemical baths. Key inspection procedures include measurements that verify proper step height (Å), etch rate (Å /min), CD, and DI. This note applies to all subsequent etch operations but will be listed only this one time.

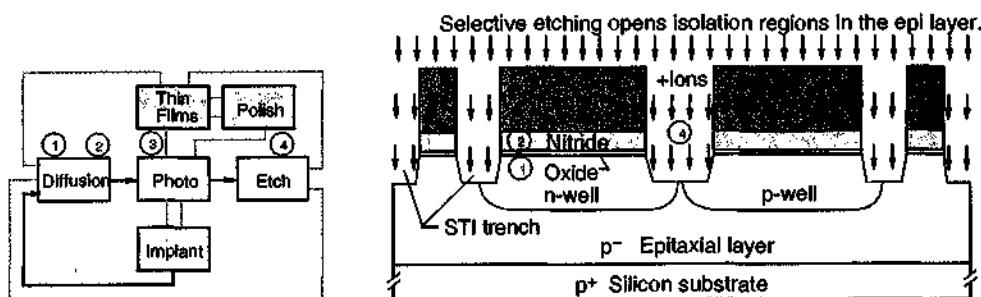


FIGURE 9.10 STI Trench Etch

STI Oxide Fill ■ The basic steps in STI oxide fill are described in the following process and illustrated in Figure 9.11.

Description of Basic Steps in STI Oxide Fill

Process Step	Description
1. Trench Liner Oxide	The wafer is again cleaned in the diffusion bay to remove contamination and oxide. After rinsing and drying, the wafer is inserted into a high-temperature oxidation furnace. Approximately 150 Å of oxide grows in the exposed walls of the isolation trenches. The nitride mask prevents oxygen diffusion into the active regions. The purpose of the liner oxide is to improve the interface between the silicon and the trench CVD oxide.
2. Trench Fill with CVD Oxide	This deposition process is performed either in diffusion using an LPCVD furnace or in thin films using a variety of oxide CVD systems. Higher throughput and deposition rates are obtained with the furnace.

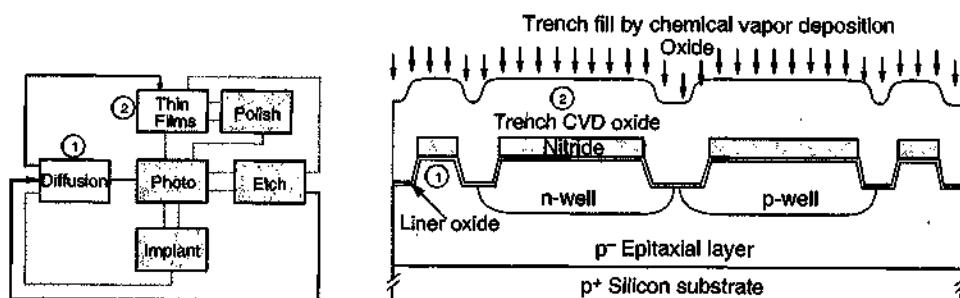


FIGURE 9.11 STI Oxide Fill

STI Oxide Polish-Nitride Strip ■ Planarization of the wafer surface is accomplished by several methods. In the past, filling the gap to planarize the wafer could be done using *spin-on-glass (SOG)*, consisting of 80% solvent and 20% silicon dioxide. Baking the SOG after the deposition evaporated the solvent, leaving oxide inside the gaps. A complete surface *etchback* also could be performed to reduce the overall thickness. However, no method has been as effective a planarization technique as CMP, chemical mechanical planarization (also referred to as *polish*). The following table and Figure 9.12 on page 210 describe and illustrate the process.

Description of Basic Steps in STI Oxide Polish and Nitride Strip

Process Step	Description
1. Trench Oxide Polish (CMP)	In the CMP area the wafer is turned upside down and held by vacuum to a wafer carrier. The wafer rotates in the same direction as a polishing pad attached to a rotating polishing table. A chemical slurry keeps the pad and wafer wet during the polishing process and provides reactants to help abrade the oxide surface. The nitride, being a harder material than the oxide, serves as the polish-stop material to prevent over-polishing the isolation structures. <i>Note:</i> The following statement applies to all polishing operations and will not be repeated again. The polishing process creates excessive particles and chemical contaminants that must be thoroughly removed before transferring the wafer to other areas in the wafer fab. Special scrub stations clean the top and bottom of the wafer and ensure it is rinsed and dried. Inspections for film thickness, particles, and defects are also conducted before wafer is transferred.
2. Nitride Strip	The diffusion bay is traditionally equipped with hot phosphoric acid stations for stripping the nitride from the wafer. In the diffusion bay the wafer is stripped, cleaned, rinsed, and dried, then inspected to check the thickness of the isolation oxide.

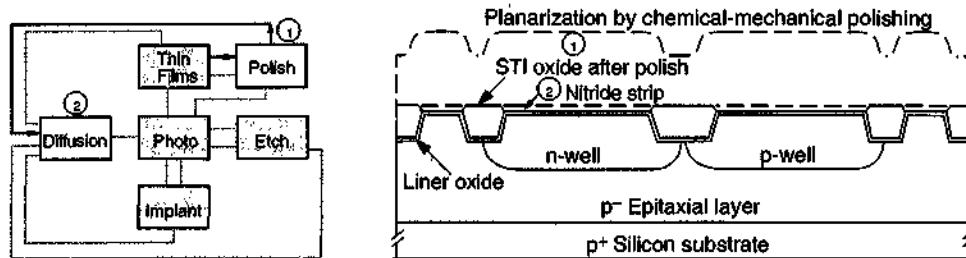


FIGURE 9.12 STI Formation

3. Poly Gate Structural Process

Forming the structure of the transistor gate is a critical step in the process because it includes thermally growing the thinnest *gate oxide* and patterning and etching the *polysilicon gate*, one of the smallest physical structures in the entire IC process (see Figure 9.13). The width of the polysilicon gate is often the most critical CD linewidth of the entire wafer.

Description of Basic Steps in Forming the Poly Gate Structure

Process Step	Description
1. Gate Oxide Growth	The wafer is cleaned to remove contamination and oxide. There is only a matter of a few hours before the wafer must be inserted in the oxidation furnace. Oxide will form as long as the silicon is exposed to oxygen in the atmosphere. The wafer is inserted in the oxidation furnace and a thin oxide of approximately 20 to 50 Å is grown.
2. Polysilicon Deposition	The wafer is immediately loaded into an LPCVD furnace where silane (SiH_4) is introduced into the process chamber. The silane dissociates and polysilicon is deposited on the wafer. Approximately 5000 Å of polysilicon (also referred to as poly) is deposited. Some processes may call for doping the polysilicon immediately following the deposition step. This action can be done in the same process chamber or in a separate furnace.
3. 4th Mask, Poly Gate	In photolithography, deep UV lithography (assuming sub-0.25 μm technology) techniques are used to pattern the fine structures of the polysilicon gates. An <i>antireflective coating</i> (ARC) is commonly applied between the poly and the photoresist to reduce undesirable reflections. The resist linewidths for the gates are the narrowest structures on the IC; therefore, various quality measurements are required, such as CD and overlay registration (OL), and defect inspection (DI).
4. Poly Gate Etch	By far one of the most critical etch steps in the IC process requires the use of the best <i>anisotropic plasma etchers</i> available in the fab. These single-direction etchers etch the polysilicon to provide the vertical profile as shown in Figure 9.13.

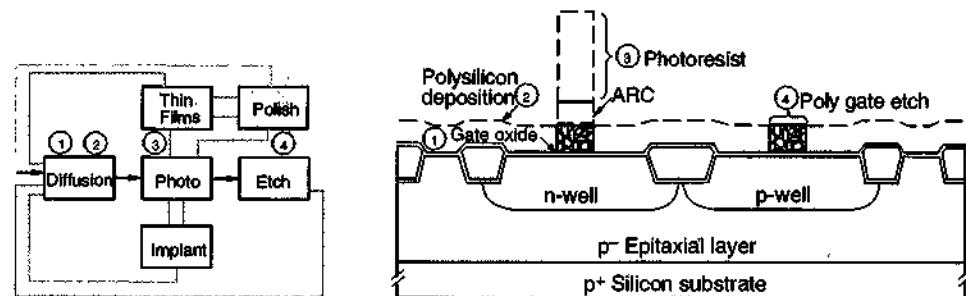


FIGURE 9.13 Poly Gate Structural Process

4. Lightly Doped Drain (LDD) Implants Process

As gate widths get smaller, the *channel length* below the gate structure (the silicon region between the source and drain) also decreases. This reduction increases the potential for charges to *punchthrough* the transistor source and drain and cause undesirable channel leakage current. There are techniques that are used to reduce the occurrence of this leakage in the channel.

The next series of ion implant steps begins to define the source and drain regions of transistors. Each pMOS and nMOS transistor will be implanted twice—once with a shallow implant called the *lightly doped drain (LDD) implant*, then followed by a medium or high dose *source/drain implant*. Lightly doped drain (LDD) implants (see Figure 9.14 below and Figure 9.15 on page 212) use the larger mass of arsenic and BF_2 dopant materials to create an *amorphous* (disorganized with no long-range order or crystal structure) upper layer of silicon. The combination of large mass and amorphous surface conditions helps maintain a shallow junction, which also helps reduce channel current leakage effects between the source and drain.

n⁻ LDD Implant ■ The steps for creating n⁻ lightly doped drain implants are described in the following process and illustrated in Figure 9.14.

Description of n- Lightly Doped Drain Implant

Process Step	Description
1. 5th Mask, n ⁻ LDD Implant	The purpose of this masking step is to pattern the wafer in a manner that opens windows in the photoresist where n-channel transistors can be implanted. All other areas are protected by the remaining photoresist.
2. n ⁻ LDD Implant (Low energy, shallowjunction depth)	Arsenic ions are selectively implanted through the windows in the patterned resist. The energy, dose and depth are significantly lower than in the previous n-well implant steps. Arsenic is preferred over phosphorus because its larger mass amorphizes the silicon surface to create a more uniform dopant depth during implant.

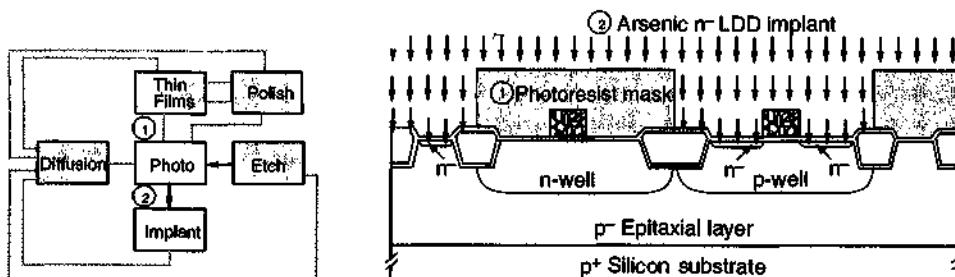


FIGURE 9.14 n⁻ LDD Implant

p⁻ LDD Implant ■ The steps for creating p⁻ lightly doped drain implants are described in the following process and illustrated in Figure 9.15.

Description of p⁻ Lightly Doped Drain Implant

Process Step	Description
1. 6th Mask, p ⁻ LDD Implant	The purpose of this masking step is to pattern the wafer in a manner that opens windows in the photoresist where p-channel transistors can be implanted. All other areas are protected by the remaining photoresist.
2. p ⁻ LDD Implant (Low energy, shallow junction depth)	Boron difluoride, BF ₂ , is preferred over boron for this implant step. BF ₂ is a heavier substance than boron. The heavier BF ₂ helps amorphize the silicon surface.

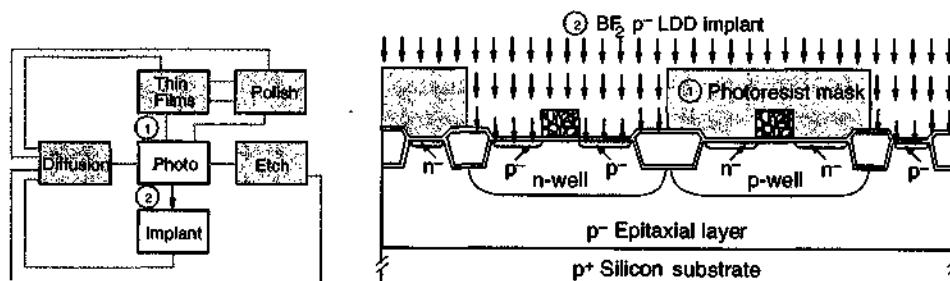


FIGURE 9.15 p⁻ LDD Implant

5. Sidewall Spacer Formation

Sidewall spacers will be used alongside the poly gates to prevent the higher source/drain (S/D) implant from penetrating too close to the channel where S/D punchthrough could occur. Sidewall spacer formation requires two major process steps (see Figure 9.16). First, an oxide layer is deposited across the surface of the wafer, then the oxide is etched back using a dry etch process. There is no requirement for a masking step because the anisotropic (single-direction) etch tool uses ions to sputter away most of the oxide. The etch process ends when the polysilicon is exposed. However, not all of the oxide is removed. Some oxide remains on the sidewalls of the polysilicon gates.

Description of Two Major Steps in Sidewall Spacer Formation

Process Step	Description
1. Spacer Oxide Deposition	This procedure is performed in the thin film bay. Approximately 1000 Å of oxide is deposited with a chemical vapor deposition (CVD) process. This layer will be used to form spacers on the sides of the polysilicon gates.
2. Spacer Oxide Etchback	The dry plasma etcher removes most of the CVD oxide leaving behind the thicker oxide on the sidewalls of the polysilicon gates.

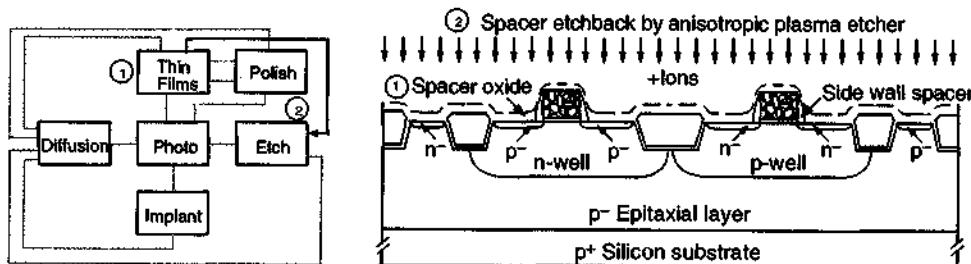


FIGURE 9.16 Sidewall Spacer Formation

6. Source/Drain (S/D) Implant Processes

To complete the retrograde implant technique, medium-dose implants are made to penetrate the silicon slightly beyond the LDD junction depth, but not as deep as the original twin-well implants (see Figure 9.17 and Figure 9.18). The spacer oxide from the previous step will protect the channel from the dopant atoms during the implant process.

n⁺ S/D Implant ■ The steps for creating n⁺ source/drain implants are described in the following process and illustrated in Figure 9.17.

Description of n⁺ Source/Drain Implant

Process Step	Description
1. 7th Mask, n ⁺ S/D Implant	This masking step defines areas of the nMOS transistors that are to be implanted.
2. n ⁺ S/D Implant (Medium energy)	This is a medium-energy implant step that penetrates the silicon deeper than the LDD junction depth. Spacer oxide prevents the arsenic dopant from encroaching into the narrow channel.

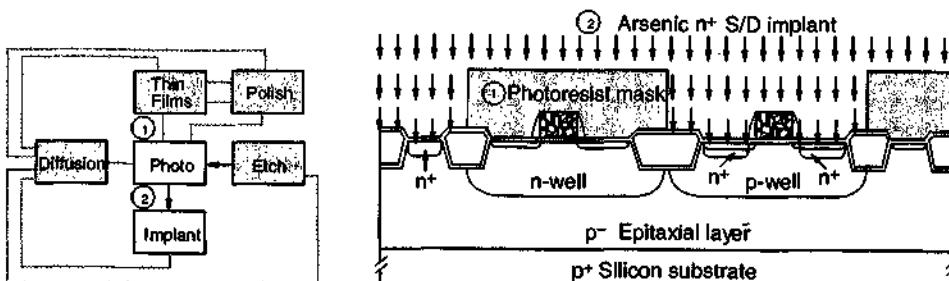


FIGURE 9.17 n⁺ S/D Implant

p⁺ S/D Implant ■ The steps for creating the p⁺ source/drain regions are similar to n⁺ S/D formation. The steps for p⁺ S/D are described in the following process and illustrated in Figure 9.18.

Description of p⁺ Source/Drain Implant

Process Step	Description
1. 8th Mask, p ⁺ S/D Implant	This masking step defines areas of the pMOS transistors that are to be implanted.
2. p ⁺ S/D Implant (Medium energy)	This implant step penetrates the silicon slightly beyond the LDD junction depth. Spacer oxide prevents boron dopant from encroaching into the narrow channel.
3. Anneal	The implanted wafer is annealed in a <i>rapid thermal process (RTP)</i> tool. RTP tools have the ability to quickly reach temperatures of ~1000 °C and maintain the setpoint for several seconds. This condition is important to prevent structures from spreading and to control the diffusion of dopants in the S/D regions.

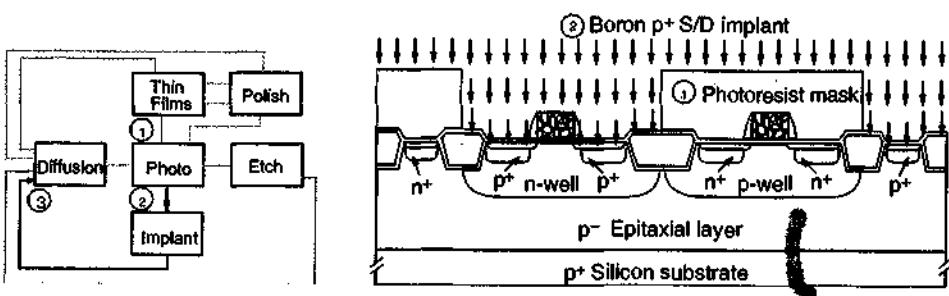


FIGURE 9.18 p⁺ S/D Implant

7. Contact Formation

The purpose of the contact formation process steps is to form *metal contacts* on all active areas of silicon. These metal contacts serve to promote adhesion between the silicon and the metal conductor material that will be deposited later (see Figure 9.19). Titanium is a good choice for the metal contact material, although cobalt can also be used. Titanium has low resistivity characteristics and reacts very well with silicon. When raised to a temperature $>700\text{ }^{\circ}\text{C}$, titanium bonds with silicon to form a *titanium silicide* (TiSi_2) compound, or *tisilicide*. Titanium and silicon dioxide do not react; therefore, no chemical bonding or physical connection is made between these two materials. Thus, titanium can be easily etched off the oxide without the need for a masking step. Tisilicide remains at all locations where active silicon exists (e.g. source, drain, and gate).

Description of Major Steps in Titanium Contact Formation

Process Step	Description
1. Titanium Deposition	The wafer is cleaned thoroughly to remove contaminants and oxides from the silicon. Titanium (Ti) is deposited on wafers using a sputtering process. <i>Sputtering</i> is a <i>physical vapor deposition (PVD)</i> process that is performed in a plasma process chamber where energetic argon ions bombard a metal target to release metal atoms, which deposit onto a wafer.
2. Anneal	The wafer is inserted in the RTP tool. The high temperature triggers a chemical reaction between titanium and silicon that forms TiSi_2 (tisilicide).
3. Titanium Etch	Chemicals etch away the unreacted titanium, leaving behind tisilicide over the active silicon areas.

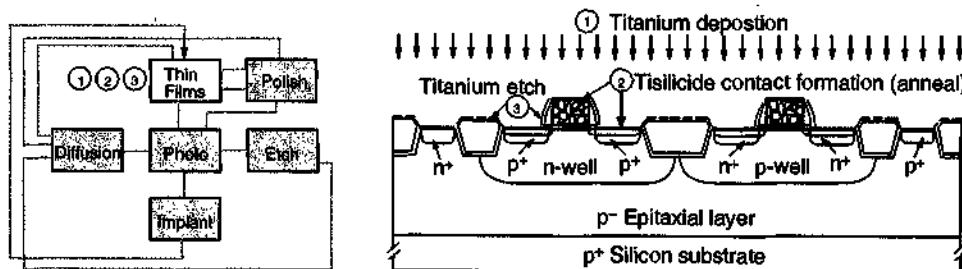


FIGURE 9.19 Contact Formation

8. Local Interconnect (LI) Process

The next step in the IC manufacturing process is to form metal connecting lines between transistors and other tisilicide contacts. The method used in this process flow is referred to as *local interconnect (LI)*. The steps leading up to the formation of LI are as complicated as the formation of the STI. The process begins with the deposition of dielectric films, followed by CMP, patterning, etch, and tungsten metal deposition and finished off with a metal polish (see Figure 9.20 and Figure 9.21). This process is referred to as *damascene*—a name adopted from a practice that began thousands of years ago by artisans in Damascus, Syria.³ The result of these steps produces a top surface that resembles intricate inlaid jewelry or artwork. The graphic in Figure 9.22 illustrates how the metal lines are formed within the walls of the etched oxide.

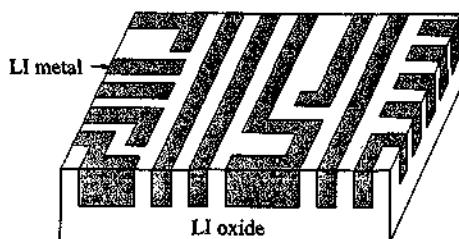
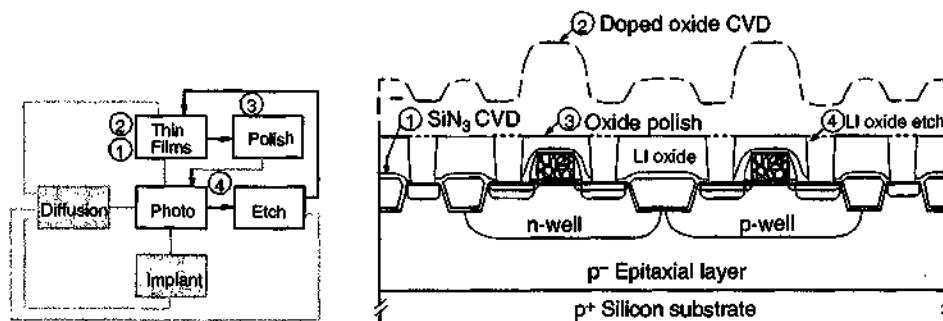


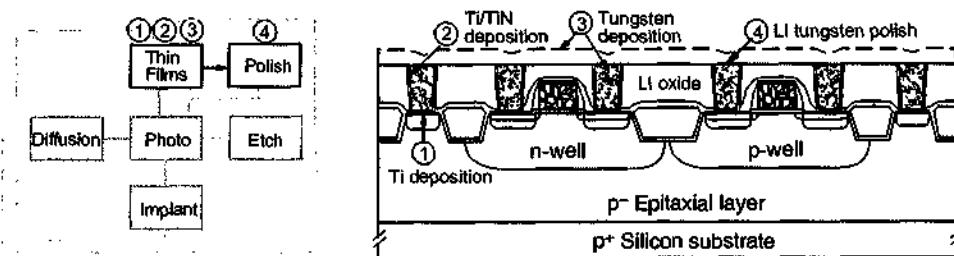
FIGURE 9.20 LI Oxide as a Dielectric for Inlaid LI Metal

Description of LI Oxide Dielectric Formation

Process Step	Description
1. Nitride (SiN_3) CVD	A barrier layer of silicon nitride is deposited using a CVD process. The nitride provides protection for the active regions from the dopants in the next film layer.
2. Doped Oxide CVD	The LI dielectric component of the local interconnect structure is formed from CVD oxide (SiO_2). This oxide is lightly doped with phosphorus or boron. The addition of dopants to the oxide improves the dielectric qualities of the glass. An additional RTP step allows the glass to flow and smooth the surface.
3. Oxide Polish (CMP)	CMP is used to planarize the LI oxide. The resulting thickness of the oxide after polish is $\sim 8000 \text{ \AA}$.
4. 9th Mask, LI Etch	The wafer is patterned in photolithography and then etched in the etch bay. Narrow trenches are created in the LI oxide that will serve as the forms for defining the paths for the LI metal.

**FIGURE 9.21** LI Oxide Dielectric Formation**Description of LI Metal Formation**

Process Step	Description
1. Titanium Deposition (PVD process)	A thin barrier layer of titanium (Ti) lines the bottom and the inside of the LI walls. The Ti serves as a “double-adhesive tape” to hold the tungsten (W) to the SiO_2 .
2. Titanium Nitride Deposition	Titanium nitride (TiN) is deposited immediately over the Ti to serve as a diffusion barrier for the tungsten metal.
3. Tungsten Deposition (CVD process polish)	Tungsten fills the LI trenches and coats the entire wafer. Tungsten is preferred over aluminum for LI metal due to its ability to fill holes without leaving voids in the formed metal plug. Another reason is for its good polishing characteristics.
4. Tungsten polish	The tungsten is polished down to the upper surface of the LI oxide.

**FIGURE 9.22** LI Metal Formation

9. Via-1 and Plug-1 Formation

The *interlayer dielectric (ILD)* serves as an insulator material between each metal layer or between the first metal layer and silicon. The ILD will have many small *vias*, which are small openings in the ILD that provide an electrical pathway from one metal layer to adjacent metal layers. Vias are filled with a conductive metal (usually tungsten, and referred to as a *tungsten plug*) and are placed at the appropriate locations to form the electrical circuit between metal layers (see Figure 9.23 and Figure 9.24). ILD-1 is the first of a series of interconnection process to be described.

Description of Major Steps in Via-1 Formation

Process Step	Description
1. ILD-1 Oxide Deposition (CVD)	Oxide is deposited across the surface of the wafer with a CVD tool in thin films. This layer (ILD-1) will serve as the dielectric material from which vias are to be formed.
2. Oxide Polish	CMP is used to planarize the ILD-1 oxide. The resulting thickness of the oxide after polish is approximately 8000 Å. The wafer is scrubbed and cleaned to remove particles generated during the polish step.
3. 10th Mask, ILD-1 Etch	The wafer is patterned in photolithography and then etched in the etch bay. Small via holes, less than a quarter micron in diameter, are etched into the ILD-1 oxide. This step is critical in terms of CD, OL, and defects.

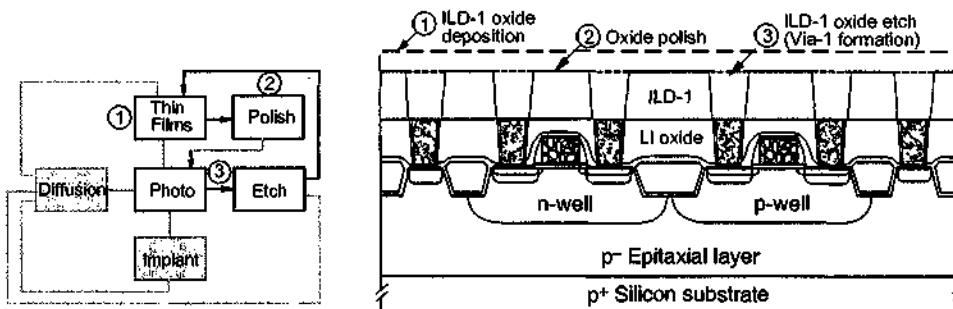


FIGURE 9.23 Via-1 Formation

Description of Major Steps in Plug-1 Formation

Process Step	Description
1. Titanium Barrier Metal Deposition (PVD)	A thin layer of titanium is deposited across the surface of the wafer with a PVD tool in thin films. Titanium lines the bottom and the walls of the via holes. The Ti serves as the glue that holds the tungsten plug inside the hole.
2. Titanium Nitride Deposition (CVD)	TiN is thinly deposited on top of the Ti layer. The TiN serves as a diffusion barrier for the tungsten in the next deposition step.
3. Tungsten Deposition (CVD)	Another CVD tool is used to deposit tungsten on the wafer. Tungsten fills the small openings to form the plug.
4. Tungsten Polish	The tungsten-coated wafer is polished down to the upper surface of the ILD-1 oxide.

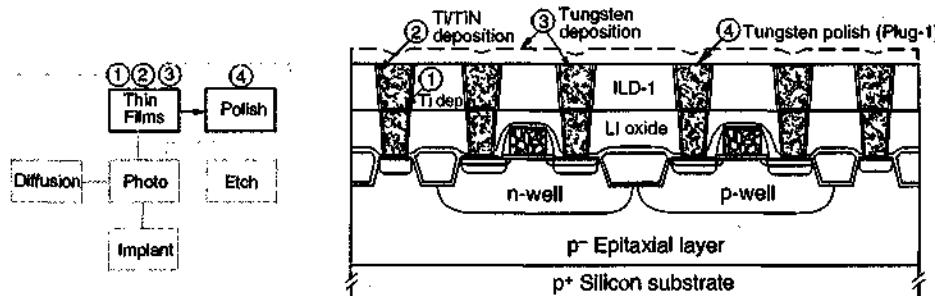
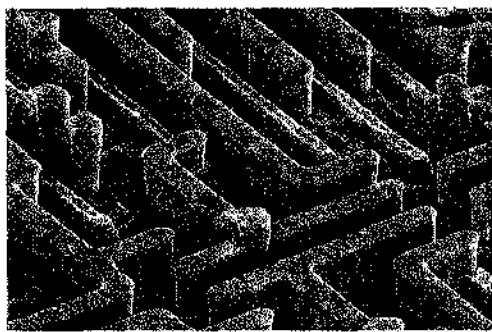


FIGURE 9.24 Plug-1 Formation



SEM Micrograph of Polysilicon, Tungsten LI and Tungsten Plugs (Polysilicon is represented by the thin lines between the taller LI lines). (Micrograph courtesy of Integrated Circuit Engineering)

10. Metal-1 Interconnect Formation

The next series of operations has to do with depositing a three-layer metal film, referred to as a *metal stack* or *sandwich*, on the wafer, followed by a masking and etching step (see Figure 9.25). The multilayer metal stack consists of different refractory metals—titanium, aluminum/copper, and titanium nitride. When finished, the wafer will have the first of five metal stacks to be constructed on the device. The number of metal layers on a device varies depending on die complexity, with advanced die currently having around eight metal layers. Including LI metal, the device covered in this process flow has a total of six metal layers.

Description of Steps in Metal-1 Interconnect Formation

Process Step	Description
1. Titanium Barrier Metal Deposition (PVD)	As with other metal processes, titanium is the first metal to be deposited on the entire wafer. It provides a good bond between the tungsten plugs and the next metal, aluminum. It also bonds well to the interlayer dielectric material and improves the reliability of the metal stack.
2. Aluminum-Copper Deposition (PVD)	The aluminum-copper (99% Al, 1% Cu) metal is sputtered onto the titanium-coated wafer by a PVD tool in thin films. The 1% Cu is added to the aluminum to improve the reliability of the aluminum.
3. Titanium Nitride Deposition (PVD)	TiN is thinly deposited on top of the aluminum-copper layer to serve as an antireflective coating for the next photomasking step.
4. 11th Mask, Metal Etch	The wafer is patterned with photoresist, then the three-layer metal stack is etched using a plasma etcher.

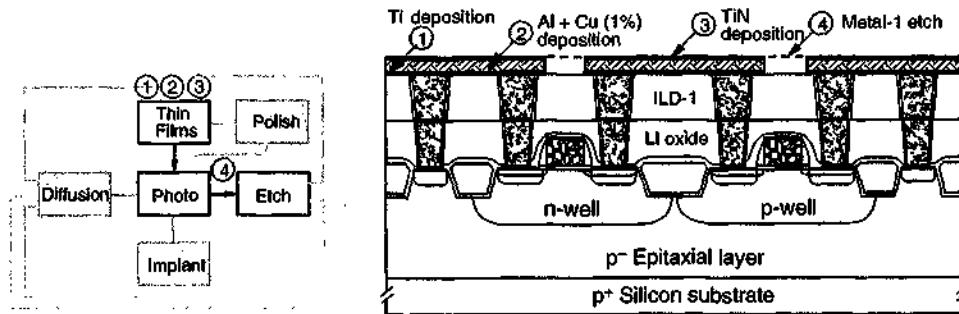


FIGURE 9.25 Metal-1 Interconnect Formation

11. Via-2 and Plug-1 Formation

The next four operations will result in the formation of the second interlayer dielectric (ILD-2) and the via openings (see Figure 9.26 and Figure 9.27). ILD-2 is similar to ILD-1 except for the very first step of filling small and large gaps that were etched into Metal-1. *Gap fill* is the filling of these gaps with a dielectric material that can be placed into the narrow spaces without creating voids or other defects that could affect electrical performance. There are two common methods for filling gaps—spin-on-glass (SOG) with etchback and *high-density plasma chemical vapor deposition* (HDPCVD). HDPCVD is the preferred method for sub-quarter micron processes. Once the gaps are filled, a plasma-enhanced CVD (PECVD) system finishes depositing the remainder of the ILD-2 oxide. Following the deposition of the ILD-2, the oxide is polished, patterned, and then etched to open the vias for the tungsten plugs to be formed.

Description of Major Steps in Via-2 Formation

Process Step	Description
1. ILD-2 Gap Fill	The latest gap-fill method for ULSI devices utilizes a high-density plasma process to alternately deposit and etch the interlayer oxide as its being deposited. The result is a dense oxide with few or no voids in the metal gaps.
2. ILD-2 Oxide Deposition	A PECVD system is used to deposit the remainder of the ILD-2 oxide layer.
3. ILD-2 Oxide Polish	The wafer is polished to planarize the surface prior to the next patterning step.
4. 12th Mask, ILD-2 Etch	The wafer is patterned with photoresist, then the ILD-2 oxide is etched using a plasma etcher.

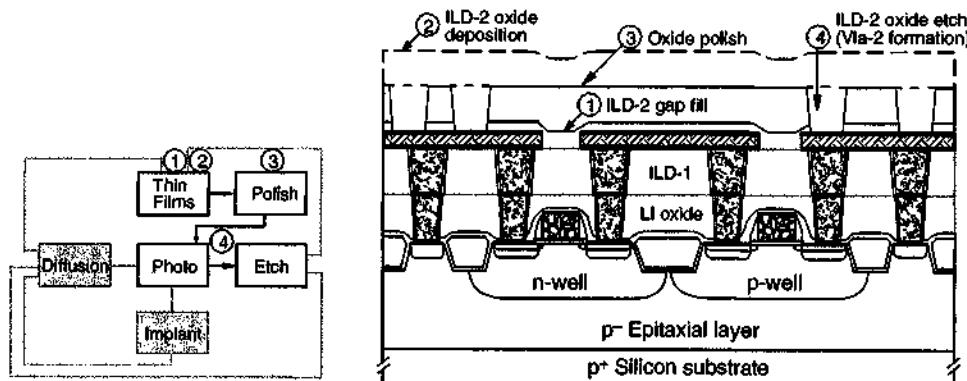


FIGURE 9.26 Via-2 Formation

Description of Major Steps in Plug-2 Formation

Process Step	Description
1. Titanium Barrier Metal Deposition (PVD)	As with other metal processes, titanium is the first metal to be deposited on the wafer. It provides a good bond between the tungsten plugs and the next metal, aluminum. It also bonds well to the interlayer dielectric material and improves the reliability of the metal stack.
2. Titanium Nitride Deposition (CVD)	TiN is thinly deposited on top of the Ti layer. The TiN serves as a diffusion barrier for the tungsten in the deposition step.
3. Tungsten Deposition (CVD)	Another CVD tool is used to deposit tungsten on the wafer. Tungsten fills the small via openings to form the plug.
4. Tungsten Polish	The tungsten-coated wafer is polished down to the upper surface of the ILD-2 oxide, leaving the tungsten plugs in the vias.

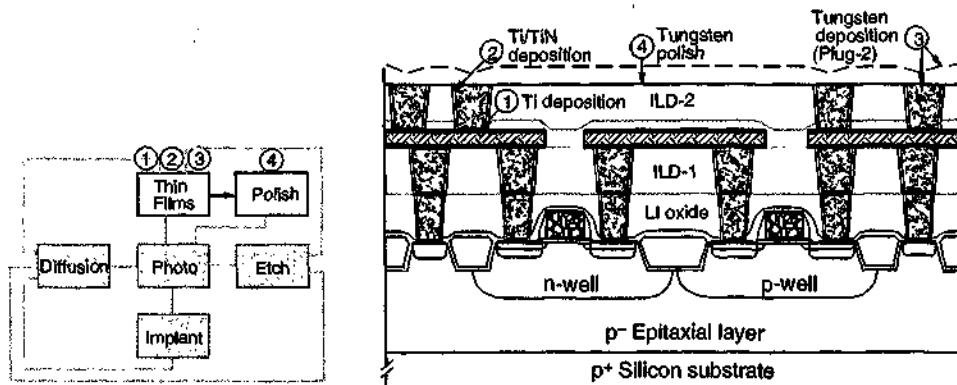
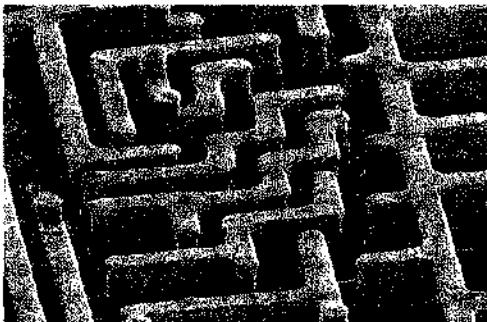


FIGURE 9.27 Plug-2 Formation



SEM Micrograph of First Metal Layer over First Set of Tungsten Vias (Note TiN Cap over Aluminum)
(Micrograph courtesy of Integrated Circuit Engineering)

12. Metal-2 Interconnect Formation

The next series of operations describes the process of forming the interconnects between layers (see Figure 9.28 on page 220). The process is repeated in the same manner for all remaining metal stacks.

Description of Major Steps in Metal-2 Interconnect Formation

Process Step	Description
1. Metal-2 Deposition to Etch	The second metal stack is deposited exactly as the first metal stack. The stack is a composite of three layers—Ti, Al/Cu, and TiN. A plasma etcher is used to etch the Metal-2 lines through windows in the patterned resist.
2. ILD-3 Gap Fill	Following the Metal-2 etch, HDPCVD is used to fill the metal gaps with dense oxide.
3. ILD-3 Oxide Deposition to Polish	ILD-3 oxide is deposited using a PECVD method. This action is followed by polishing the oxide to attain surface planarity.
4. Via-3 Etch, Ti/TiN Deposition, Tungsten Deposition, Polish	Via-3 openings are etched, then these are filled with barrier metal (Ti/TiN). Tungsten is deposited across the wafer surface. The tungsten is polished off the wafer until the ILD-3 oxide layer is reached. Tungsten plugs are left in the vias to provide the interconnect between Metal-2 and Metal-3.

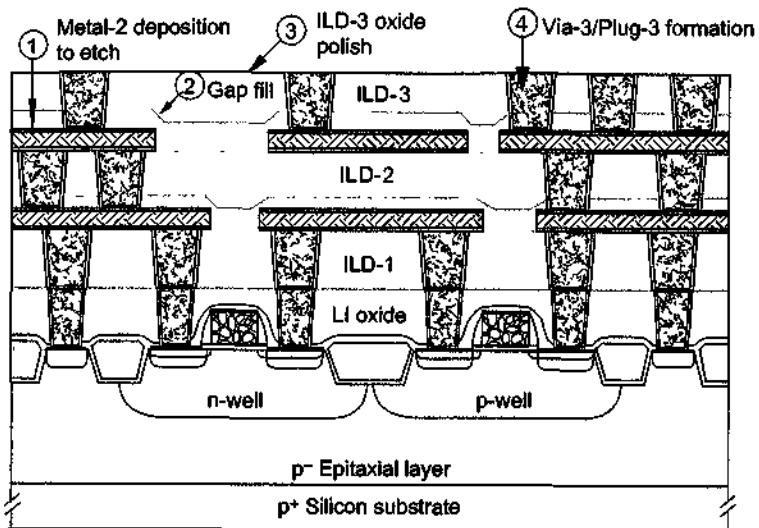
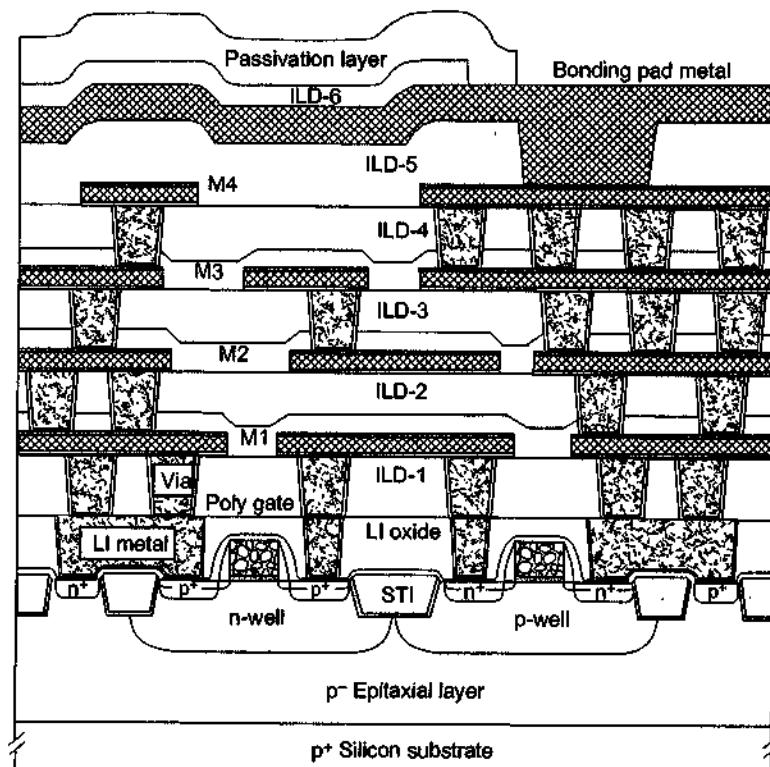


FIGURE 9.28 Metal-2 Interconnect Formation

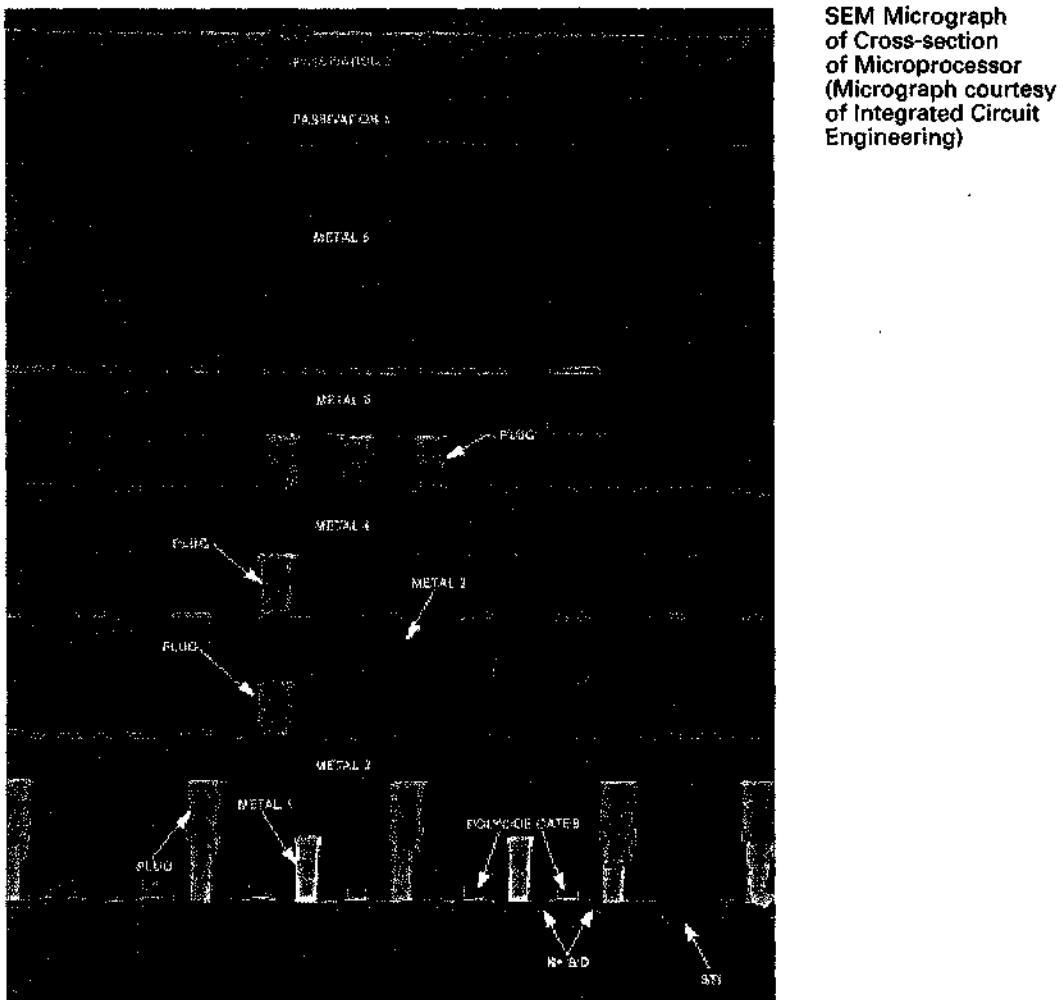
13. Metal-3 to Pad Etch and Alloy

After repeating the layering process for layers 3 and 4 and at the completion of the Metal-4 etch, ILD-5 oxide is deposited using the thin film process (Figure 9.29). CMP is not required for this dielectric layer because the structures to be patterned are much larger than the quarter micron dimensions formed earlier in the process. The ILD-5 layer is etched to allow metal filling by the Metal-5 deposition. The Metal-5 layer is deposited thicker than previous metal stacks. This metal layer is etched as necessary to form bonding pads and to remove metal from areas where it is not needed.

FIGURE 9.29 Full 0.18 μm CMOS Cross Section

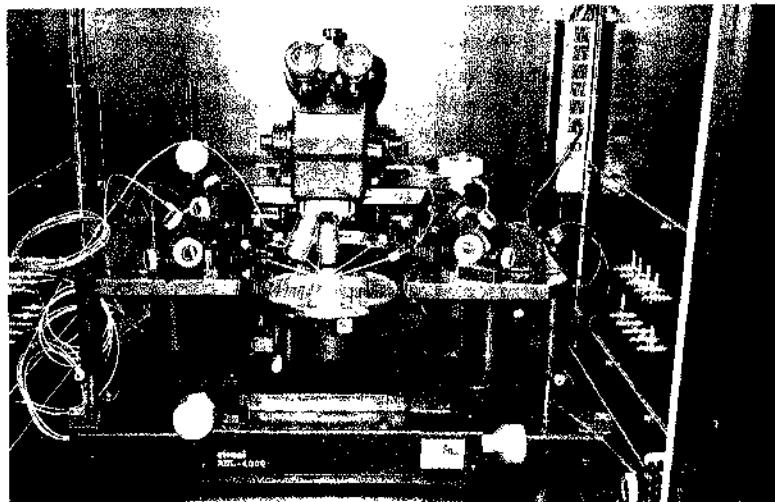
This product incorporates 0.18 μm processing technology, including twin-tub technology, retrograde implants, shallow trench isolation, shallow S/D diffused regions, sidewall spacers, titanium silicide contacts, titanium barrier metal, tungsten local interconnect, tungsten plugs, three-layer metal stacks, HDP oxide gap fill, and PECVD oxide as the interlayer dielectric.

The final steps of the process include one more oxide layer (ILD-6) followed by a top layer of silicon nitride ($\sim 2000 \text{ \AA}$). This layer is referred to as the *passivation layer*. Its purpose is to protect the product from moisture, scratches, and contamination. A final low-temperature alloy step in a diffusion furnace is performed. This heat treatment helps improve the metallurgical bonds between metal connections, thus improving electrical performance and reliability. Care must be taken with this alloy step to prevent overheating the product, which can cause permanent structural defects.



14. Parametric Testing

The wafer is tested twice to determine its product-worthiness—once when the wafer completes the first metal etch and again after the completion of the last wafer fab process step. Following metal etch, special microprobes connected to electronic test instruments are used to measure certain electrical parameters of specific device test structures on the wafer. This in-line parametric testing procedure is also referred to as *wafer electrical test*, or WET.



Wafer Electrical Test Using a Micro-manipulator Prober
 (Photo courtesy of Advanced Micro Devices)

The last process step for the wafer is at wafer test/sort just outside the wafer fab. The wafer is probed and tested automatically in the electrical test and sort department (also referred to as *E-sort*). Each die on the wafer is tested for electrical functionality. Those die that fail are tracked through software using x-y position coordinates (previously marked with an ink spot) so that the locations of good and bad die are known for later processing. This data is used to determine die yield; that is, the percentage of good die versus bad die on a wafer. The wafer is now ready to be shipped to the assembly department. In some companies, the assembly plants are located in other parts of the world.

If the wafer passes the wafer sort, the wafer is sent to the backgrind department. Here the backgrinder grinds the backside of the wafer. This process makes the wafer thinner and easier to slice for eventual separation into individual die.

SUMMARY

An overview of the wafer fab process defines six major process areas: diffusion, photo, etch, ion implant, thin films, and polish, plus wafer test/sort at the end of the process. Diffusion is for high-temperature processing; photolithography patterns the wafer with photoresist; etch replicates the photoresist pattern in the wafer; ion implant dopes the wafer; thin films deposit dielectric and metal layers; and polish planarizes the top surface of the wafer.

A simplified CMOS process is defined with 14 manufacturing steps. (1) Twin-well implants create the n-well and p-well in the silicon. (2) Shallow-trench isolation isolates the silicon-active areas for sub-0.25 μm technology. (3) The gate structure is formed through gate oxide growth, polysilicon deposition, and patterning. (4) Lightly doped drain (LDD) implant forms a shallow implant in the source and drain. (5) Sidewall spacer formation protects the channel from the

subsequent source/drain implant. (6) Source/drain implant is a medium-energy implant that forms a junction depth deeper than the LDD implant. (7) Contact formation forms a silicide contact for tungsten metal to alloy with the silicon. (8) Local interconnect (LI) forms the first metal connecting lines between transistors and contacts. (9) First interlayer dielectric (ILD) to Via-1 deposits the dielectric and creates vias between LI and first metal layer. (10) First metal layer to first metal etch deposits the metal sandwich and patterns the metal. (11) Second ILD to Via-2 deposits the second interlayer dielectric and creates the via openings. (12) Second metal layer to Via-3 deposits the second metal stack and deposits and etches ILD-3. (13) Metal-3 to pad etch and alloy repeats the layering process until the metal-5 bonding pad is deposited, followed by ILD-6 and a passivation layer. (14) The final process step is parametric testing to verify acceptability of each die on the wafer.

KEY TERMS

diffusion	dry plasma etcher
high-temperature diffusion furace	isolation regions
wet cleaning station	spin-on-glass (SOG)
photolithography	etchback
coater/developer track	polish
stepper	gate oxide
etch	polysilicon gate
plasma etcher	anti-reflective coating (ARC)
plasma resist stripper	silane
ion implanter	anisotropic plasma etcher
thin films	channel length
CMP (chemical mechanical planarization)	punchthrough
wafer scrubbers	lightly doped drain (LDD) implant
twin-well (twin-tub)	source/drain implant
retrograde implant	amorphous
latchup	sidewall spacers
epitaxial silicon	rapid thermal process (RTP)
native oxide	metal contacts
alignment-and-exposure system	titanium silicide (tisilicide)
linewidths	local interconnect (LI)
critical dimension (CD)	damascene
junction depth	interlayer dielectric (ILD)
ion implanter	vias
anneal	tungsten plug
shallow trench isolation (STI)	metal stack
LPCVD (low-pressure chemical vapor deposition)	sandwich
sidewall liner	gap fill
overlay registration (OL)	high-density plasma chemical vapor deposition (HDPCVD)
defect inspection (DI)	plasma enhanced CVD (PECVD)
visual inspection (VI)	passivation layer
radical chemicals	wafer electrical test (WET)

REVIEW QUESTIONS

1. List the six distinct production areas in a wafer fab and give a short description of each area.
2. What activity is performed in the diffusion area?
3. List five processes that are done in a high-temperature furnace.
4. What is the purpose of photolithography?
5. What is a coater/developer track used for? List five operations performed using this tool.
6. What is the major concern for particulate contamination in photolithography?
7. Identify the three production areas where photoresist-coated wafers can be found.
8. What is the purpose of the etch process? Name the most common tools used in this area?
9. What is the ion implanter used for?
10. What is the purpose of the thin films bay?
11. List four different tools used or processes done in the thin films area.
12. What does CMP stand for, and what is its purpose? What is another name for CMP?
13. List the 14 manufacturing steps outlined for a typical CMOS process.
14. Describe the epitaxial layer on the silicon wafer.
15. What is a twin-well?

16. Explain the retrograde implant technique. What problem does it help resolve?
17. What are the reasons for the thermal anneal process after ion implantation?
18. Why are the retrograde p-well implant energy levels lower than the n-well implant energy levels?
19. What is shallow trench isolation (STI)? What process did it replace?
20. What tool is used to etch silicon for STI, and why?
21. What tools are used to deposit the oxide in STI oxide fill?
22. Describe the most effective wafer planarization technique.
23. Why is the formation of the transistor gate structure a critical process step?
24. What problem can occur due to smaller gate widths?
25. How do lightly doped drain (LDD) implants reduce the channel-current leakage effects?
26. Explain the purpose of sidewall spacers.
27. What are medium-dose implants after the LDDs used for?
28. What is the purpose of the metal contact?
29. What is a local interconnect (LI)?
30. What is the purpose of the ILD?
31. What is a via? What is a tungsten plug?
32. Describe the materials used in a metal stack.
33. What is gap fill?
34. Why is CMP not required for the final ILD layer?
35. What is the purpose of the passivation layer?
36. Describe what happens in the wafer test/sort area.

REFERENCES

1. S. Ghandhi, Chap. 11 in *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: Wiley, 1994), pp. 704–800.
2. S. Wolf, *Process Integration*, vol. 2, *Silicon Processing for the VLSI Era*, (Sunset Beach: Lattice Press, 1990), pp. 17–19.
3. P. Singer, “Making the Move to Dual-Damascene Processing,” *Semiconductor International* (January 1999): pp. 70–82.

CHAPTER 10

OXIDATION

A cornerstone of silicon IC wafer fabrication is the ability to thermally grow an oxide layer on the surface of a wafer. A major development from the 1950s was *oxide masking*, which patterned and etched openings in a thermally grown oxide layer to diffuse dopants into the silicon substrate. This development was a key factor in developing a process to fabricate large quantities of transistors.¹ In this way, oxidation played a major role in the development of silicon planar technology and remains viable today, which explains the widespread usage of silicon in wafer fabrication.

With the proper manufacturing control, oxide layers have high quality, stability, and desirable dielectric properties.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe an oxide film for semiconductor manufacturing, including its atomic structure, its various uses, and its benefits.
2. State the chemical reaction for oxidation, and describe how oxide grows on silicon.
3. Explain selective oxidation and give two examples.
4. Identify the three types of thermal processing equipment, describe the five parts of a vertical furnace, and discuss the attributes of a fast ramp vertical furnace.
5. Explain what is a rapid thermal processor, its usage and design.
6. Describe the critical aspects of the oxidation process, including its quality measures and some common troubleshooting problems.

INTRODUCTION

Oxide on a silicon wafer is created by either the grown or deposited method. A *grown oxide layer* occurs on a wafer by providing externally supplied high-purity oxygen in an elevated-temperature environment to react with the silicon substrate. The high-temperature oxidation process occurs in the diffusion area of the wafer fab. This is the first process area when a silicon wafer enters the fabrication process (see Figure 10.1 on page 226). A *deposited oxide layer* is generated by using an external silicon source and O₂ and reacting these materials in a chamber to form a thin film on the wafer surface. This chapter will cover the basic technology of oxidation and explain the high-temperature process of thermal oxidation. Chapter 11 will cover the deposition of oxide and other materials.

Since a silicon wafer is a flat surface (planar), oxide is grown or deposited essentially as a layer. Because of this, it is referred to as a *thin film*. Once a layer is on the wafer surface, it is then modified during subsequent processing to attain the necessary three-dimensional shape used in the formation of circuit components, such as a trench capacitor or interconnect conductor line. This

Because of these qualities, oxidation is critical, especially for the thin oxide that is an essential part of the gate structure for MOS technology. Thermal oxide is used as a dielectric material, as well as for device isolation, oxide screens for implants, stress-relief oxides (pad oxides), and reoxidizing nitride and polysilicon surfaces for photoresist adhesion and stress reduction.

Oxides can be either deposited or grown. The focus in this chapter will be on thermally grown oxides. We will review the nature of oxide films and how they are grown, including detailed information on the growth mechanism and the high-temperature chamber where growth occurs.

modification produces three-dimensional shapes on the wafer surface referred to as *topography* or *surface topology*.

Oxide growth is a natural phenomenon that occurs by exposing the silicon wafer to oxygen at an elevated temperature. The ability to grow an oxide film on silicon is often listed as a significant reason for the use of silicon as the most common semiconductor substrate material (another major reason is silicon's relatively high melting temperature). We use the term "grow" to indicate that temperature is used to cause the oxide to grow out of the silicon semiconductor material, actually consuming silicon in the process.

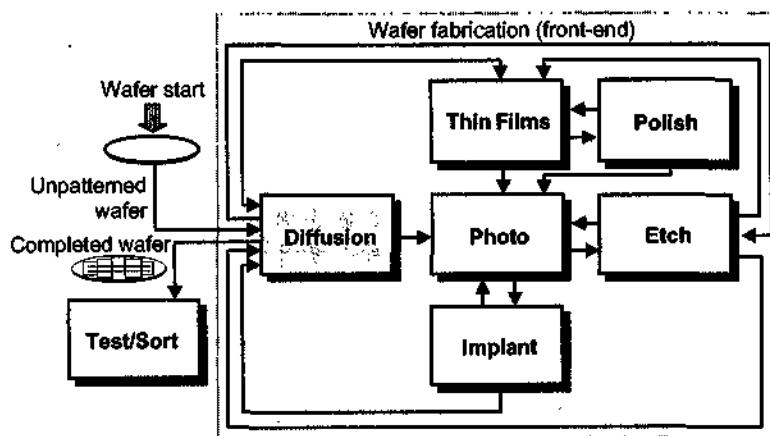


FIGURE 10.1 Diffusion Area of Wafer Fabrication
(Used with permission from Advanced Micro Devices)

The amount of thermal exposure (i.e., temperature multiplied by time) for a wafer during processing is referred to as the *thermal budget*. The thermal budget requirement for wafer fabrication is rapidly dropping.² A goal throughout semiconductor processing is to minimize the amount of thermal (heat) exposure to the wafer. As the critical dimension on device structures decreases to 0.18 μm and beyond, scaling requirements dictate shallower junction depths. To minimize the unacceptable diffusion of dopants out of a shallow junction region, the thermal budget must decrease accordingly. Another problem caused by excessive thermal budget is the increase in ohmic contact resistance of metal-layer interconnects (see Chapter 12), which increases the overall resistance of the conductive paths. A factor in determining the process conditions for many wafer fabrication process steps is the ability to minimize the thermal budget, either by reduced temperature or by minimizing the time at temperature.

OXIDE FILM

An oxide layer is grown on silicon typically at thermal oxidation temperatures between 750°C to 1100°C. An oxide layer grown on silicon is referred to as *thermal oxide* or *thermal silicon dioxide* (SiO_2). Since silicon dioxide is an oxide material, the two terms are used interchangeably. Another term for silicon dioxide is *glass*. Silicon dioxide is a dielectric material and will not conduct electricity.

Nature of Oxide Film

When a silicon surface is exposed to oxygen, an amorphous silicon dioxide film grows immediately. Although undoped silicon is a semiconductor material, SiO_2 is an insulator. The atomic structure of this SiO_2 film consists of a silicon atom surrounded by four oxygen atoms (see Figure 10.2). We can refer to this atomic structure as a silicon dioxide *tetrahedron cell*. Amorphous SiO_2 has no long-range periodic crystal order at the atomic level. Long-range order is absent because the tetrahedra are not arranged in a regular, three-dimensional array as in a crystal.³

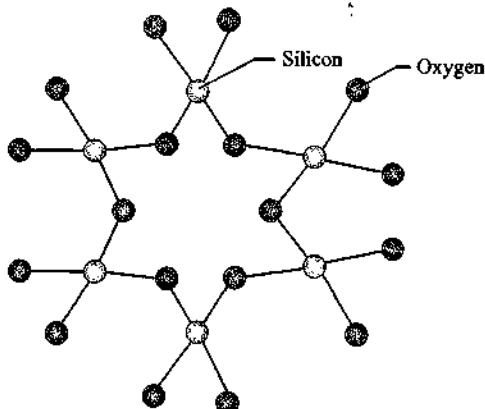


FIGURE 10.2 Atomic Structure of Silicon Dioxide
(Used with permission from International SEMATECH)

Silicon dioxide is a form of intrinsic (pure) glass with a melting temperature of 1732°C.⁴ Thermally grown SiO₂ has strong adhesion to silicon and exhibits excellent dielectric properties. A layer of SiO₂ always covers the wafer surface because silicon oxidizes immediately upon exposure to air with a few monolayers of native oxide (a monolayer is one layer of atoms on a surface). Over time, the native oxide layer will thicken to an upper limit of about 40 Å, even at 25°C room temperature.⁵ This type of oxide is nonuniform and usually considered a contaminant, with little use in semiconductor fabrication. There are some applications for use of native oxide in wafer fabrication, such as part of a dielectric film stack in memory cells. Native oxide grown from air is contaminated, but native oxide grown in chemical baths using high-purity chemicals is very clean.

Uses of Oxide Film

Oxide is important for silicon semiconductor fabrication because of its ease of formation and its excellent interface with the underlying silicon material. It is the most common layer used in semiconductor fabrication. Different ways in which an oxide layer is used to fabricate a microchip are:

- ◆ Device scratch protection and contaminant isolation
- ◆ Field isolation to confine charged carriers (termed surface passivation)
- ◆ Dielectric material in the gate oxide or memory cell structures
- ◆ Impurity-mask barrier during doping
- ◆ Dielectric layer between metal conductor layers

Device Protection and Isolation ■ Silicon dioxide grown on the surface of the wafer serves as an effective barrier to isolate and protect sensitive devices in the silicon. SiO₂ physically protects devices because it is a very hard and nonporous (dense) material that effectively insulates active devices in the silicon surface. The hard SiO₂ layer will protect the silicon from scratches and processing damage that might occur during fabrication. Transistors have traditionally been electrically isolated by thermally growing thick SiO₂ in the region between them using the LOCOS process (discussed later in this chapter). However, this process is unacceptable for sub-0.25 µm technology and has been replaced by shallow trench isolation (STI). The STI process uses deposited oxide as the main dielectric (see Chapters 9 and 11).

Surface Passivation ■ A major benefit from thermally grown SiO₂ is the reduction in the surface-state density of silicon that occurs by tying up dangling silicon bonds. This action is referred to as *surface passivation*, which prevents the deterioration of the electronic properties and reduces current-leakage paths of the semiconductor caused by moisture, ions or other external contaminants.⁶ The hard SiO₂ layer will protect the silicon from scratches and processing damage that might occur during backend final fabrication. SiO₂ growth on the silicon surface can serve to tie up electrically active contaminants (mobile ionic contaminants) in the oxide layer away from the active surface of the silicon. Passivation is also important for controlling the leakage current of

junction devices and growing a stable gate oxide.⁷ For the oxide layer to function as a good passivation layer, it has a quality requirement to be a uniform thickness and not have pinholes or voids in the material.

Another factor in using the oxide layer to passify the wafer surface is the oxide thickness. Adequate oxide thickness is necessary to prevent electrical charging in a metal layer caused by a charge buildup in the silicon wafer surface, much like charge storage and breakdown characteristics of ordinary capacitors. This charging would lead to electrical shorting and other undesirable electrical effects. A thick layer of oxide that inhibits charge buildup from metal layers is the *field oxide layer*, which typically is between 2,500 Å to 15,000 Å thick (see Figure 10.3).

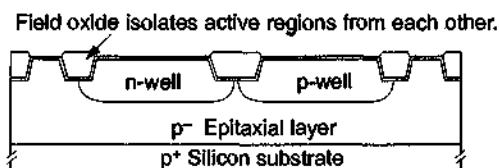


FIGURE 10.3 Field Oxide Layer

Silicon dioxide also has a coefficient of thermal expansion very similar to that of silicon. During high-temperature processing, the wafer will expand and then contract during cooling. The SiO₂ will expand and contract at close to the same rate as silicon, thus minimizing wafer warpage during the thermal excursions. This action also prevents film stress from separating the oxide film from the silicon substrate.

Gate Oxide Dielectric ■ An extremely thin layer of oxide is used as the dielectric material for the important *gate oxide structure* common in MOS technology (see Figure 10.4). The gate oxide is thermally grown because of its high quality and stability with the underlying semiconductor silicon. SiO₂ has a high dielectric strength of 10⁷ volts/cm and high resistivity of approximately 10¹⁷ ohms·cm.⁸

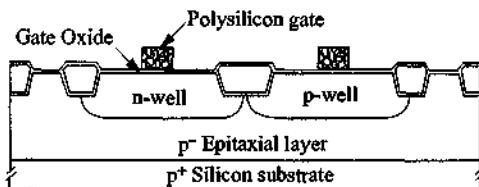


FIGURE 10.4 Gate Oxide Dielectric

The thickness of the gate oxide is chosen specifically for the scaling requirements of the device technology. For 0.18-μm generation technology, a typical gate oxide thickness is 20 ± 1.5 Å. The gate oxide has a specific thickness so that it scales properly with the entire gate structure to permit induction of a charge in the silicon wafer under the oxide. The widespread use of MOS technology in the ULSI era has made the formation of the gate oxide a primary concern in process development. A critical aspect of device reliability is *gate oxide integrity*. The gate structure in a MOS device is the part that controls the flow of electrical current through the device. Because this oxide is fundamental to the proper function of microchips based on field-effect technology, it is essential that it have high quality, with excellent film-thickness uniformity and no contaminants. Any contamination that degrades the proper functioning (integrity) of the gate oxide structure must be scrupulously controlled.

Dopant Barrier ■ Silicon dioxide can be used as an effective barrier to selectively introduce dopants into the silicon surface (see Figure 10.5). Once an oxide layer is grown on the silicon surface, then mask openings are etched into the SiO₂ to create windows where dopant materials can enter the wafer. The oxide protects the surface of silicon from impurity diffusion where there are

no openings and therefore allows selective impurity doping. Dopants have a slow rate of movement through SiO_2 when compared to silicon, requiring only a thin oxide layer to block dopants (note that this rate is temperature dependent).

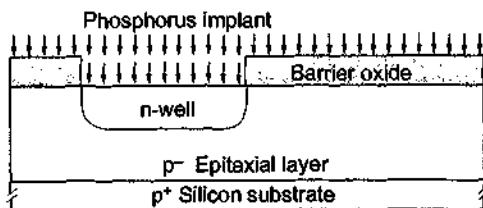


FIGURE 10.5 Oxide Layer Dopant Barrier

A thin oxide layer (e.g., 150 Å) is also grown in areas where ion implant occurs. This oxide screen serves to reduce damage to the silicon wafer and obtain better control over the depth that the dopant is implanted into the silicon by reducing channeling effects (see Chapter 17). Once implantation is done, HF (hydrofluoric acid) can be used to selectively remove the oxide, resulting in a planar silicon surface again.

Dielectric Between Metal Layers ■ Under normal conditions, silicon dioxide will not conduct electricity. For this reason, SiO_2 is an effective insulator between metal layers on the microchip. Silicon dioxide prevents shorting of an upper metal layer to the underlying metal layer, just as the insulator material on an electrical cord prevents an electrical short. The oxide quality is critical with no holes or voids allowed. This oxide is often doped to produce more effective flow properties and to better minimize the diffusion of contaminants (i.e., it acts as a getter). It is typically deposited using chemical vapor deposition (not thermally grown) and will be covered in more detail in Chapter 11.

A summary of different types of oxides and their uses in semiconductor fabrication is shown in Table 10.1.⁹

TABLE 10.1 Oxide Applications

Application	Purpose	Structure	Comments
Native Oxide	This oxide is a contaminant and is generally undesirable. Sometimes used in memory storage or film passivation.	Silicon dioxide (oxide) p ⁺ Silicon substrate	Growth rate at room temperature is 15 Å per hour up to a maximum thickness of about 40 Å.
Gate Oxide	Serves as a dielectric between the gate and source-drain parts of MOS transistor.	Gate oxide Source Transistor site p ⁺ Silicon substrate	Common gate oxide film thickness range from about 20 Å to several hundred Å. Dry thermal oxidation is the preferred growth method.

Application	Purpose	Structure	Comments
Field Oxide	Serves as an isolation barrier between individual transistors to isolate them from each other.	<p>Field oxide Transistor site p^+ Silicon substrate</p>	Common field oxide film thickness ranges from 2,500 Å to 15,000 Å. Wet oxidation is the preferred growth method.
Barrier Oxide	Protects active devices and silicon from follow-on processing.	<p>Barrier oxide Metal Diffused resistors p^+ Silicon substrate</p>	Thermally grown to several hundred angstroms thickness.
Dopant Barrier	Masking material for depositing or implanting dopants into wafer.	<p>Dopant barrier spacer oxide Ion implantation Gate Spacer oxide protects narrow channel from high-energy implant</p>	Dopants diffuse into unmasked areas of silicon by selective diffusion.
Pad Oxide	Provides stress reduction for silicon nitride (Si_3N_4).	<p>Passivation Layer Nitride Pad oxide Bonding pad metal M-4 ILD-5 M-3 ILD-4 M-3</p>	Thermally grown and very thin.
Implant Screen Oxide	Used to reduce implant channeling and damage.	<p>Ion implantation Screen oxide p^+ Silicon substrate High damage to upper Si surface + more channeling Low damage to upper Si surface + less channeling</p>	Thermally grown.
Insulating Barrier Between Metal Layers	Serves as a protective layer between metal lines.	<p>Bonding pad metal Interlayer oxide Passivation layer M-4 ILD-5 ILD-4 M-3</p>	This oxide is not thermally grown, but is deposited.

THERMAL OXIDATION GROWTH

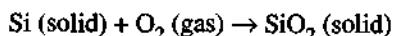
The various applications for thermally grown oxide have different thickness requirements. Silicon dioxide thickness ranges for different requirements are summarized in Table 10.2. The color chart in Appendix D shows silicon dioxide color as a function of the film thickness, which was used in the early days of wafer fabrication to estimate the oxide thickness. Other critical quality parameters for oxide layers in semiconductor fabrication are thickness, uniformity, pinholes and voids.

TABLE 10.2 Oxide Thickness Ranges for Various Requirements

Semiconductor Application	Typical Oxide Thickness, Å
Gate Oxide (0.18 μm generation)	20–60
Capacitor Dielectrics	5–100
Dopant Masking Oxide	400–1,200 (Varies depending on dopant, implant energy, time, and temperature)
STI Barrier Oxide	150
LOCOS Pad Oxide	200–500
Field Oxide	2,500–15,000

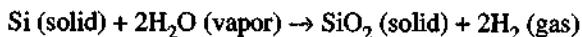
Chemical Reaction for Oxidation

Thermal oxide is grown by a chemical reaction between silicon and oxygen. Uniform oxide growth is achieved by exposing silicon to elevated temperature in the presence of high-purity oxygen. If this growth occurs with *dry oxygen*, meaning no moisture, then the following chemical equation describes the reaction:



The time and quality of this reaction varies and is affected by the purity of oxygen gas supplied to the silicon wafer surface and the reaction temperature. It occurs naturally when silicon is exposed to air at room temperature. The reaction rate is increased with an increase in temperature. The typical temperature for the oxidation of silicon during wafer fabrication is normally between 750°C to 1100°C and can vary for different oxidation process steps. The furnace temperature at any one operation is precisely controlled. The rate of oxide thickness for dry oxidation versus temperature and time is shown in Figure 10.6 on page 232.¹⁰

Wet Oxidation ■ When water vapor is introduced into the reaction, known as *wet oxidation*, the rate of the oxidation reaction is increased further. The chemical reaction for wet oxidation is:



For wet oxidation, oxygen saturated with water vapor is used instead of dry oxygen as the oxidizing gas. The water vapor could also be supplied as steam, referred to as *pyrogenic steam* (see Figure 10.7 on page 232). A wet oxidation reaction produces a silicon dioxide film and hydrogen gas during oxidation growth. The faster growth rate in a wet atmosphere is due to the faster diffusion and higher solubility of water vapor than oxygen in silicon dioxide.¹¹ However, the hydrogen molecules produced in the reaction are trapped in the solid silicon dioxide layer, making the layer less dense than oxide grown in dry oxygen. This condition can be improved to achieve a similar oxide in structure and properties such as that produced in dry oxidation by heating the oxide in an inert atmosphere.

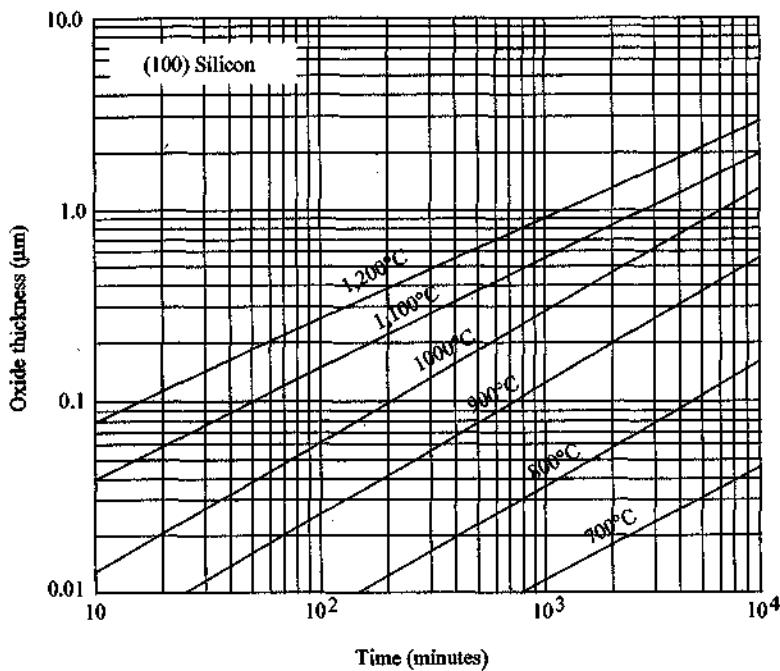


FIGURE 10.6 Dry Oxidation Time (Minutes)

Oxidation Growth Model

Silicon dioxide grows by consuming silicon, which is the case for either the dry or wet oxidation process. This is shown in Figure 10.8. The thickness of the silicon consumed is 0.46 of the total oxide thickness.¹² This means that for every 1,000 Å of oxide thickness, 460 Å of Si is consumed.

The growth of the oxidation layer is controlled and limited by the movement of oxygen through the oxide at the oxide-silicon interface. For the oxide layer to continue growing, the oxygen must come in contact with the silicon. However, the SiO_2 separates the oxygen gas molecules from the silicon wafer. Oxide growth occurs when the oxygen gas molecules move through the existing silicon dioxide layer to the silicon wafer. This movement is referred to as diffusion (more precisely, gas diffusivity through a solid barrier). *Diffusion* is the movement of one material through another.

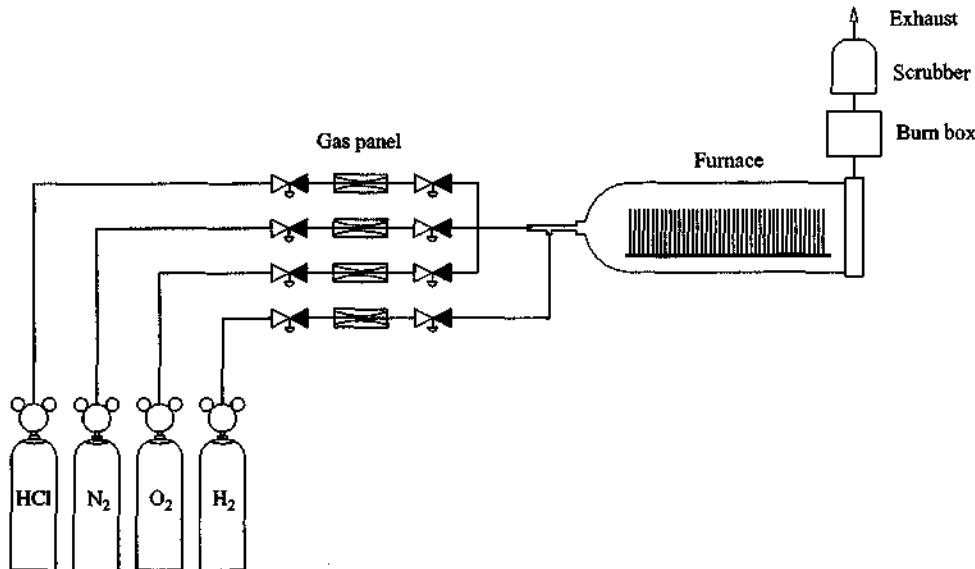


FIGURE 10.7 Wet Oxidation

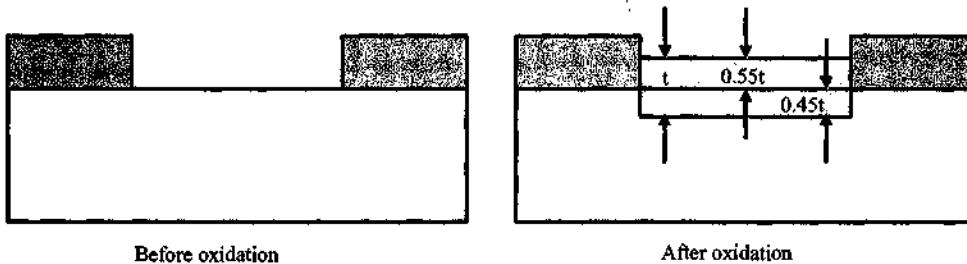


FIGURE 10.8 Consumption of Silicon During Oxidation

Atoms diffuse from regions of high concentration to regions of low concentration for solid, liquid, or gas states, and this diffusion is accelerated by thermal energy. It is a natural physical process that occurs regularly in everyday life. An example of liquid-state diffusion is putting a drop of red food colorant into a glass of water (see Figure 10.9). The colorant is initially concentrated when it enters the water, and then it will diffuse into the water until it is uniformly red. If the water is heated, then the diffusion will spread throughout the water more quickly.

The workbay in the wafer fab where oxidation occurs is still referred to as diffusion or the diffusion bay. In the early days of wafer fabrication, diffusion was an important process for creating pn junctions. Dopants were supplied to the silicon by chemical sources and then diffused to the desired junction depth by subjecting the wafers to elevated temperatures. Diffusion of dopants to create pn junctions is no longer done in wafer fabrication, although the use of the term diffusion bay is still common today. It has been replaced by ion implantation (see Chapter 17). Nevertheless, diffusion of materials (in this case oxygen) occurs in processes such as oxidation. Oxidation is well defined and governed by certain laws of diffusion that are based on a set of mathematical relationships referred to as Fick's laws. *Fick's laws* describe the rate of movement of diffusing materials based on temperature, concentration, and the energy necessary to drive diffusion. The detailed study of Fick's laws is not covered in this text.

Oxide-Silicon Interface ■ There is an abrupt transition at the oxide-silicon (Si/SiO_2) interface between single crystal silicon to amorphous SiO_2 . Recall that for the SiO_2 molecule, each silicon is bonded to four oxygen atoms and each oxygen atom is bonded to two silicon atoms. At the Si/SiO_2

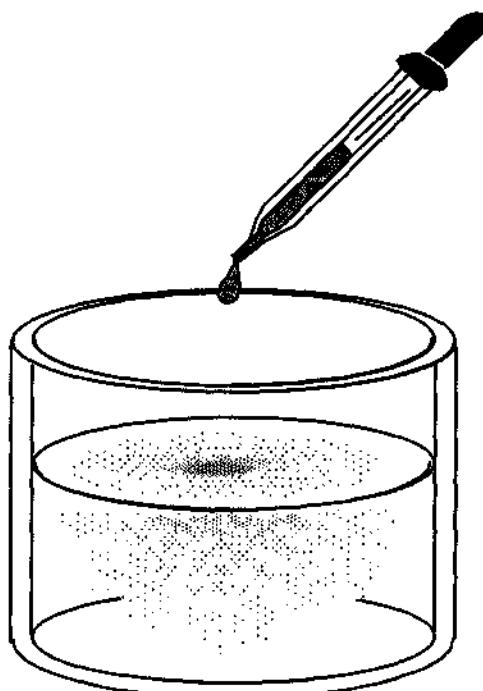


FIGURE 10.9 Liquid-State Diffusion

interface, some silicon atoms in the structure remain unbonded (see Figure 10.10). This incomplete oxidation of silicon less than 2 nm away from the Si/SiO₂ interface is the origin of a positive *fixed oxide charge*. Other charges built up at the interface include an *interface-trapped charge* consisting of positive or negative charges that result from structural defects, oxidation-induced defects, or metal impurities, and a *mobile oxide charge* due to mobile ionic contaminants (MICs). There is also the possibility of an *oxide-trapped charge* that is positive or negative and trapped in the bulk of the oxide away from the interface.¹³ This accumulation of charge at the Si/SiO₂ interface is undesirable for normal device operation and can cause the threshold voltage of a MOS device to shift to unacceptable values.¹⁴ Some of these undesirable charges can be minimized through a low-temperature (~ 450°C) anneal step in hydrogen or forming gas (hydrogen-nitrogen mixture).¹⁵

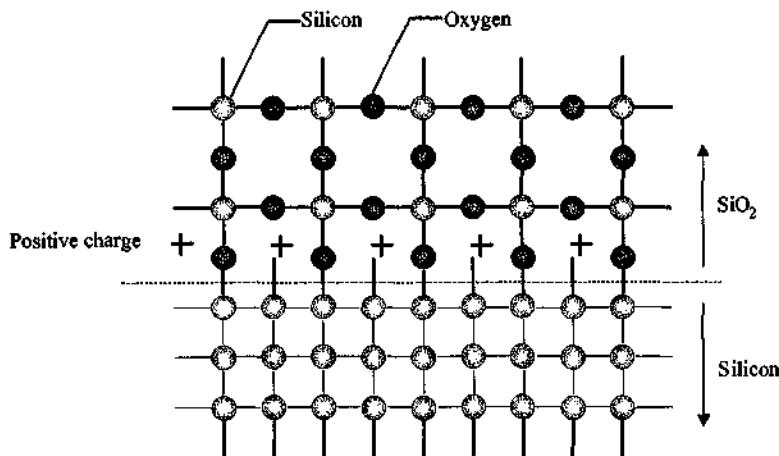


FIGURE 10.10 Charge Buildup at Si/SiO₂ Interface
(Used with permission from International SEMATECH)

Use of Chlorinated Agents in Oxidation. Using a chlorine-containing gas during the oxidation process can serve to neutralize the charge accumulation at the interface. The chlorine ions will diffuse to the positively charged layer and form a neutral layer. The chloride concentration is kept below 3% because excessive chloride ions will cause device instability. Another important advantage of chloride ions added to the thermal oxidation process is that they tend to increase the oxidation growth rate by about 10–15%. Furthermore, the presence of chlorine can actually immobilize (referred to as gettering) mobile ionic contaminants (MICs) that come from the furnace equipment, processing materials, and handling.

Early work used chlorine or vapor-phase hydrogen chloride (HCl). Chlorine is extremely toxic and HCl vapor in the presence of water vapor is corrosive. A common chlorine compound often used to transport the chlorine into the furnace is dichloroethylene (DCE) or variants of this compound.¹⁶ Other less-corrosive sources of chlorine are trichloroethylene (TCE) and 1,1,1-trichloroethane (TCA).¹⁷ TCE is carcinogenic and not used anymore, while TCA has fallen out of favor because it is an ozone-depleting chemical. Also, some semiconductor firms have switched back to HCl because of significant improvements in high-purity piping and fittings used to transport this chemical.

Rate of Oxide Growth ■ The *rate of oxide growth* describes how fast the oxide grows on the wafer. It depends on parameters such as temperature, pressure, oxidizing condition (dry or wet), silicon crystal orientation, and doping levels. The growth rate is of interest because the wafer processing time can be reduced if diffusion occurs quickly, which serves to reduce the thermal budget. The model for oxide growth on silicon is referred to as a linear-parabolic model developed by Deal and Grove and accurately represents oxide growth over a wide range of thicknesses (most optimally from 300 to 20,000 Å).¹⁸ Oxide growth is described by two growth stages: the linear stage and the parabolic stage.

The initial growth of silicon dioxide is referred to as the *linear stage* and consumes silicon on the wafer surface as a linear function of time. This means the oxide layer is growing into the wafer

at a linear rate over time. The linear stage of oxide growth is valid up to about 150 Å of oxide thickness.¹⁹ It is described by the linear equation:²⁰

$$X = \left(\frac{B}{A \text{ hr}} \right) t$$

Where, X = the thickness of the growing oxide
 (B/A) = the linear rate constant
 t = the time it takes to grow the oxide

In the linear stage, oxidation varies linearly with time. Oxidation is *reaction-rate controlled* in the linear region because the limiting factor for oxidation growth is the reaction occurring at the Si/SiO₂ interface. You will note that the linear rate constant, B/A, is the slope of this linear relationship, and therefore it controls the rate of the reaction. As temperature increases, the values for B/A increase, which means that the rate of oxidation also increases.

The *parabolic stage* of oxidation growth is the second phase of oxidation growth and starts after about 150 Å of oxide thickness. The equation that describes the parabolic stage is:²¹

$$X = (Bt)^{\frac{1}{2}}$$

Where, X = the thickness of the growing oxide
 B = parabolic rate constant
 t = the time it takes to grow the oxide

Note that this equation generates the shape of a parabola. Oxide growth in the parabolic stage is much slower than the linear stage. This is because as the oxide layer becomes thicker, the oxygen must diffuse through a larger distance to arrive at the Si/SiO₂ interface (see Figure 10.11). The reaction is thus limited by the rate at which the oxygen diffuses through the oxide. For this reason, the parabolic stage of oxide growth is said to be *diffusion controlled*. As the parabolic rate constant increases, the rate of oxide growth will increase. For instance, the parabolic rate constant, B, for wet oxidation is found to be much larger than that for dry oxidation.²² Thus the rate of oxidation increases with wet oxidation.

A general curve shows both the linear and parabolic stage, as illustrated in Figure 10.12 on page 236.²³ This is a simplified curve from the original curve published by Deal and Grove in 1965 used to describe the thermal oxidation of silicon.²⁴

Factors Affecting Oxide Growth ■ There are other factors that can affect the rate of oxide growth besides temperature and the presence of H₂O. We will review some of these other factors.

Dopant Effects. Heavily doped silicon oxidizes at a faster rate than lightly doped material. In the parabolic stage, boron doping will oxidize faster than phosphorus. Boron tends to become incorporated in the oxide film, which weakens its bond structure and leads to a subsequent increase in the diffusivity of the oxygen through it.²⁵ There is little difference in the linear rate constant between boron and phosphorus doping.

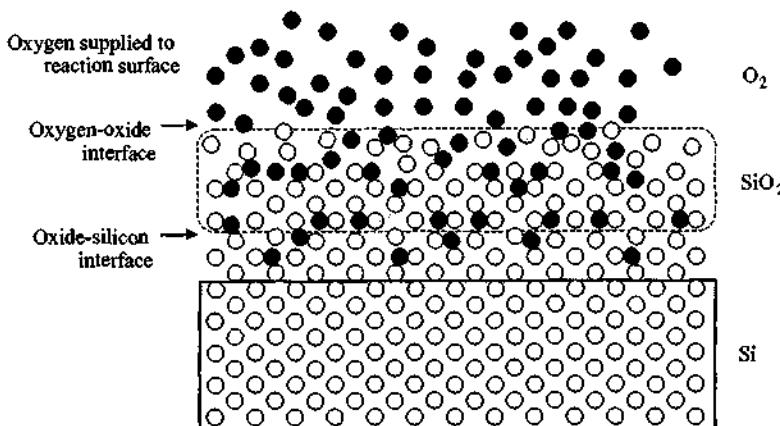


FIGURE 10.11 Diffusion of Oxygen Through Oxide Layer
 (Used with permission from International SEMATECH)

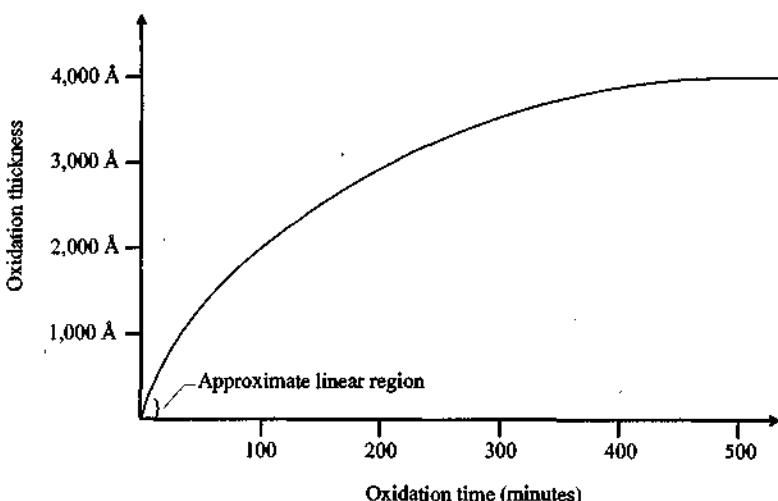


FIGURE 10.12 Linear and Parabolic Stages for Dry Oxidation Growth at 1100°C
(Used with permission from International SEMATECH)

Crystal Orientation. The linear oxidation rate is dependent on the orientation of the crystal because the density of silicon atoms on the (111) plane is greater than on the (100) plane. Therefore, (111) silicon crystal will oxidize in the linear stage at a faster rate than (100) silicon crystal, but (111) also has a greater charge buildup.

During the parabolic stage, the parabolic rate constant, B , is independent of the crystal orientation of the silicon substrate. There is no difference in oxidation growth rate for (111) vs. (100) in the parabolic stage. This is logical since the parabolic rate of oxidation growth is based on the diffusivity of the oxygen through the existing SiO_2 and is not affected by the reaction at the Si/SiO_2 interface.²⁶ However, charge buildup remains high because this factor is a function of atom density at the surface.

Pressure Effect. Since the growth rate of an oxide layer depends on the movement of the oxidizer from the gas phase to the silicon interface, the growth rate will increase with pressure. Higher pressure significantly increases the linear and parabolic rate constants, forcing the oxygen atoms to penetrate through the growing oxide more rapidly. This condition allows the temperature to be reduced to achieve equivalent growth rates or a more rapid oxide growth at the same temperature. A rule of thumb for the increased oxide growth versus temperature reduction is that for each 1-atmosphere increase in pressure, the furnace temperature can be decreased 30°C.²⁷ This ratio can be used to reduce the thermal budget. A high-pressure oxidation process could be used, for example, to grow a thick field oxide layer.

Plasma Enhancement. Plasma-enhanced oxidation is another method that increases the oxide growth rate at low temperature, thus reducing the thermal budget.²⁸ The technique is usually carried out in an oxygen-plasma discharge generated by an RF source. The silicon is biased to be at a potential less than the plasma potential, which collects the charged oxygen in the plasma on the wafer. This activity results in rapid oxidation of the silicon and allows oxides to be grown at temperatures less than 600°C. The drawbacks to this technique are problems associated with particle generation, higher film stress that is different from thermally grown oxides, and inferior film quality compared to thermally grown oxides.²⁹ For these reasons, this method is not widely used in wafer fabrication.

Initial Growth Phase ■ The Deal and Grove linear-parabolic model accurately predicts oxidation growth for thickness above 300 Å. However, below this thickness it is found that dry oxidation is faster than predicted. This is an important oxidation growth range because the gate oxide thickness for subquarter micron MOS technology is currently around 20 to 60 Å. The manufacturing process must be capable of producing these oxide layers with high yield and long-term reliability.

Because the gate oxide is becoming so thin, there is not one model that accurately predicts the rate of oxidation. One particular model has detected the presence of pores, about 10 Å in diameter, in very thin oxides grown in dry oxygen. These pores allow the oxidant to remain in direct

contact with the silicon in the early phases of growth, thus causing the rapid initial growth of oxide.³⁰ The investigation into thin gate oxides is an area of ongoing research.

Selective Oxidation ■ The oxidation of selective areas on a wafer uses SiO_2 to electrically isolate adjacent devices on the silicon surface. The traditional method for isolating devices with greater than $0.25 \mu\text{m}$ feature sizes has been *local oxidation of silicon (LOCOS)*. A deposited layer of silicon nitride (Si_3N_4) serves as an oxidation barrier. It is etched to allow selective thermal oxide growth, since oxide will not grow where the nitride covers the silicon (see Figure 10.13). After thermal oxidation, the nitride and any underlying barrier oxide are removed to expose bare silicon surface regions ready for device formation.

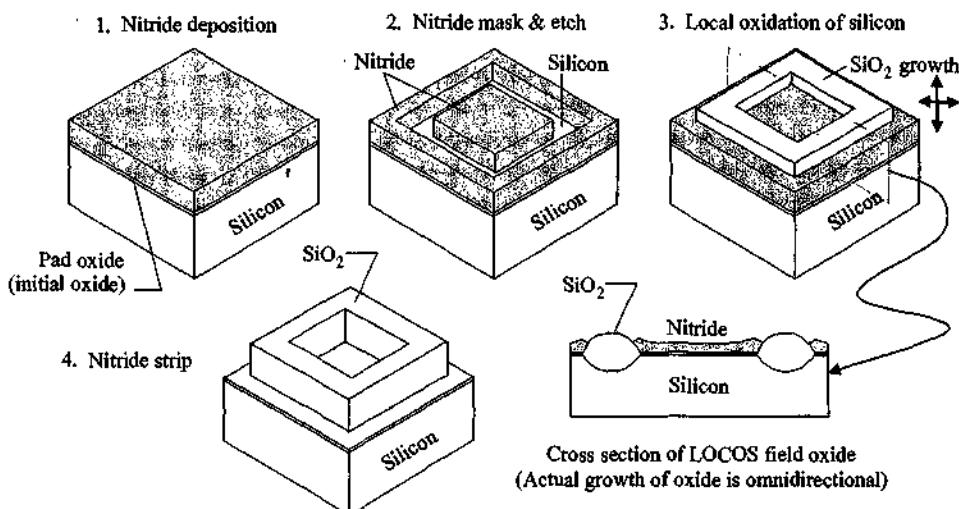


FIGURE 10.13 LOCOS Process

As the oxygen diffuses through the grown oxide, it moves in all directions. Some of the oxygen moves down into the silicon and other oxygen atoms move sideways. This means there is a slight lateral growth of the oxide under the nitride mask. Since the oxide is thicker than any silicon consumed, growth under the nitride mask serves to push up the nitride edges. This action is referred to as the *bird's beak effect*. This phenomenon is an undesirable by-product of LOCOS type oxidation processes (see Figure 10.14). The bird's beak effect is most pronounced when the oxide is relatively thick. To reduce stress between the nitride mask and silicon, a thin barrier layer of thermal oxide is used between them, which is termed a *pad oxide*.

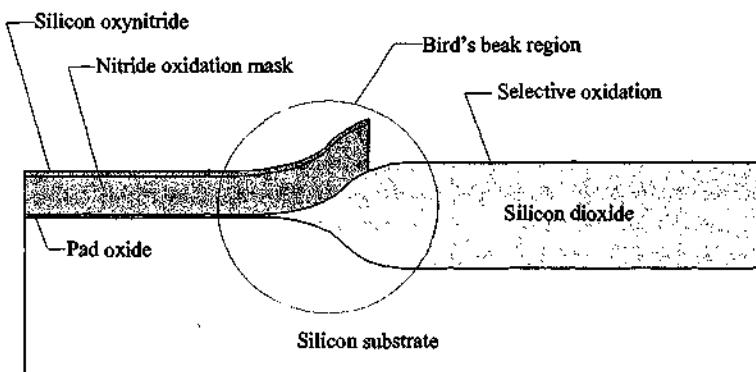


FIGURE 10.14 Selective Oxidation and Bird's Beak Effect
(Used with permission from International SEMATECH)

Shallow Trench Isolation (STI). The primary form of selective oxidation used for sub- $0.25\text{-}\mu\text{m}$ technology is *shallow trench isolation (STI)*. The main dielectric material in STI is a

deposited oxide (see Chapter 11). Selective oxidation is accomplished by creating a mask, usually of silicon nitride (Si_3N_4). The mask is deposited and patterned, followed by etching of the silicon to form a trench. The etching of silicon trenches is followed by thermal oxidation of a 150–200 Å thick oxide layer in areas exposed through the mask (see Figure 10.15). This thermally grown oxide passivates the silicon surface and serves as a barrier between the silicon and deposited trench-fill oxide. It is also an effective barrier that prevents sidewall current leakage in finished devices.

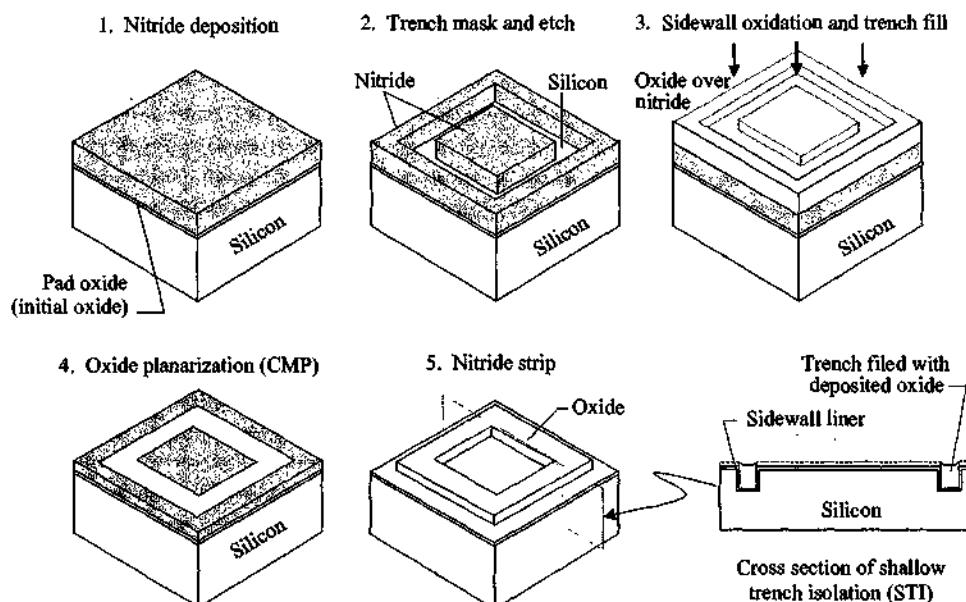


FIGURE 10.15 STI Oxide Liner

Silicon nitride is usually deposited by chemical vapor deposition (see Chapter 11), followed by etching to open up the selective areas for thermal oxidation. The silicon surfaces on the wafer covered by the silicon nitride layer do not oxidize as long as the nitride mask is sufficiently thick (the nitride mask covering the silicon will oxidize slightly). Silicon nitride can be grown on silicon by reacting nitrogen with surface silicon atoms at elevated temperatures, usually above 900°C. However, nitride grown using this method can only be grown to a very thin layer (about 5 nm), which limits its use to only special applications that take advantage of its high dielectric constant.³¹

Stress in Silicon Dioxide ■ Silicon dioxide stress is undesirable because it can contribute to wafer warpage and possibly to defect generation in the form of a slip in the silicon wafer. Measured stress in a thermally grown SiO_2 film is found to be compressive and have a relatively small magnitude. The stress in an oxide film results from the difference between the coefficients of thermal expansion of Si and SiO_2 .³² Stress from an oxide film layer can cause the wafer to bow so that the oxidized surface becomes convex. Optical measurements of the curvature can be used to quantify stress.

Oxidation-Induced Stacking Faults ■ Wet or dry thermal oxidation can cause *oxidation-induced stacking faults (OISFs)* at the interface between Si and SiO_2 . Recall from Chapter 4 that stacking faults are a form of unit-cell dislocation due to layer stacking errors. It is believed that OISFs are caused by incomplete oxidation at the Si/ SiO_2 interface, which leads to excess interstitial silicon in this region.³³ Stacking faults can cause increased leakage current if the fault is in the vicinity of a pn junction. OISF formation is greatly reduced by performing thermal oxidation with chlorine because chlorine promotes vacancy formation at the silicon surface, providing a means to remove the excess interstitial silicon atoms.



Horizontal Diffusion Furnace
(Photo courtesy of International SEMATECH)

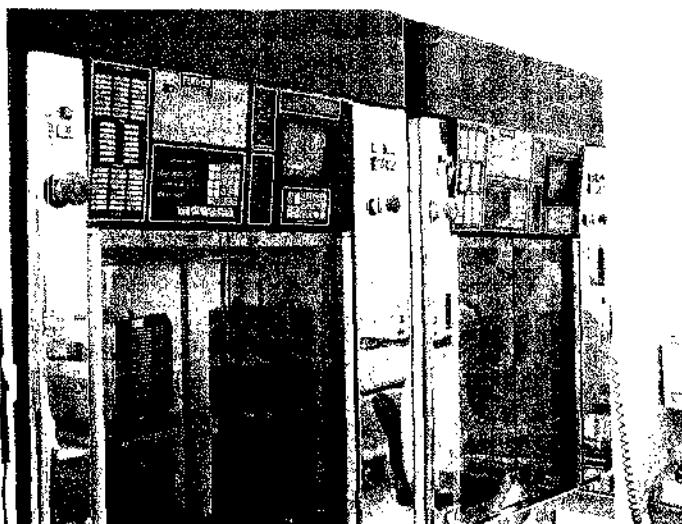
FURNACE EQUIPMENT

In this section we will review basic furnace equipment technology. Furnace equipment serves a variety of purposes during the wafer fabrication process. Thermal oxide growth, including the formation of the critical gate oxide, is a major reason for the use of furnaces. Other applications include thermal anneal of the wafer surface after ion implantation (Chapter 17); the deposition of a variety of films, such as doped and undoped polysilicon, silicon nitride, and silicon dioxide (Chapter 11); reflow of glass (Chapter 11); and the formation of silicide films (Chapter 12). We will focus on the equipment used for thermal oxidation processes.

There are three basic types of thermal processing equipment used:

- ◆ Horizontal furnace
- ◆ Vertical furnace
- ◆ Rapid thermal processor (RTP): single-wafer

The *horizontal furnace* was the workhorse in thermal wafer processing since the early days of the semiconductor industry. Its name derives from the horizontal position of the quartz tube where wafers are located and heated. This furnace was largely replaced in the early 1990s by the



Vertical Diffusion Furnace
(Photo courtesy of International SEMATECH)

vertical furnace, mainly because the vertical furnace is easier to automate, improves operator safety, and reduces particulate contamination.³⁴ The vertical furnace, also called a *vertical diffusion furnace*, or VDF, has better temperature control and uniformity than a horizontal furnace. Both the horizontal and vertical furnaces are considered conventional *hot wall* batch furnaces because both the wafer and the furnace walls are heated, and they generally process large quantities or batches of wafers (100 to 200 wafers). Conventional furnaces ramp up and down the temperature of the wafers at about 20°C/minute or less.

The *rapid thermal processor (RTP)* is a small, fast-heating system that typically processes a single wafer at a time with a radiant heat source and cooling source. The RTP is also referred to as a rapid thermal anneal (RTA) system when used to anneal the silicon substrate (see Chapter 17). Due to its extremely fast localized heating times, the RTP only heats the wafer (not the wall of the furnace). The typical RTP is able to achieve up and down ramp rates in the tens of degrees per second range, with models available that use dual-sided wafer heating to attain ramp-up rates up to 250°C/second.³⁵ The RTP has been in use since the late 1980s and is used in areas such as barrier-layer formation (see Chapter 12) and oxide reflow.

HORIZONTAL VERSUS VERTICAL FURNACES

A comparison of performance factors for the conventional horizontal and vertical furnaces is given in Table 10.3.³⁶

TABLE 10.3 Performance Comparison of Horizontal and Vertical Furnace Systems

Performance Factor	Performance Objective	Horizontal Furnace	Vertical Furnace
Typical wafer loading size	Small, for process flexibility	200 wafers/batch	100 wafers/batch
Cleanroom footprint	Small, to use less space	Larger, but has four process tubes	Smaller (single process tube)
Parallel processing	Ideal for process flexibility	Not capable	Capable of loading/unloading wafers during process, which increases throughput
Gas flow dynamics (GFD)	Optimize for uniformity	Worse due to paddle and boat hardware. Buoyancy and gravity effects cause nonuniform radial gas distribution.	Superior GFD and symmetric/uniform gas distribution
Boat rotation for improved film uniformity	Ideal condition	Impossible to design	Easy to include
Temperature gradient across wafer	Ideally small	Large, due to radiant shadow of paddle	Small
Particle control during loading/unloading	Minimum particles	Relatively poor	Improved particle control from top-down loading scheme
Quartz change	Easily done in short time	More involved and slow	Easier and quicker, leading to reduced downtime
Wafer-loading technique	Ideally automated	Difficult to automate in a successful fashion	Easily automated with robotics
Pre- and post-process control of furnace ambient	Control is desirable	Relatively difficult to control	Excellent control, with options of either vacuum or neutral ambient

Horizontal furnaces are still in use and have undergone advances in technology, leading to some renewed interest in their use in fabs. Their low cost relative to vertical furnaces makes them

appealing to wafers with geometries larger than $0.5\text{ }\mu\text{m}$. This capacity permits a mix-and-match approach, with horizontal furnaces used for certain less-demanding applications and vertical furnaces used for critical applications.

Both horizontal and vertical furnaces have been configured for atmospheric oxidation and diffusions, as well as low-pressure chemical vapor deposition (LPCVD) applications. Examples of materials LPCVD furnaces are used to deposit as thin layers are SiO_2 , Si_3N_4 , and polycrystalline silicon (poly) on silicon wafers. The topics of LPCVD and diffusion will be discussed in Chapter 11.

Vertical Furnace

The vertical furnace was first introduced in the early 1990s. The change from horizontal furnaces was driven primarily by the reduced cleanroom footprint achieved by vertical furnaces and the improvement in automated handling. To understand basic furnace technology, we will analyze a conventional vertical furnace. There are five major control systems to a vertical furnace system (these same five systems also apply to a horizontal furnace):

- ◆ Process chamber
- ◆ Wafer transfer system
- ◆ Gas distribution system
- ◆ Exhaust system
- ◆ Temperature control system

A block diagram of a vertical furnace system showing the five major systems is shown in Figure 10.16. Note that these are all under the control of a single microcontroller.

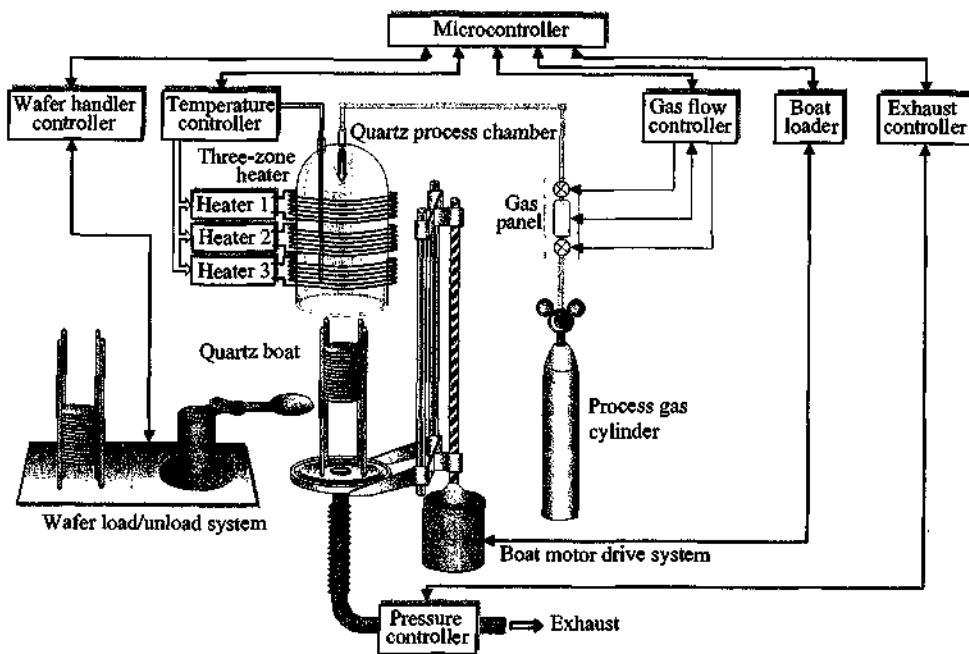


FIGURE 10.16 Block Diagram of Vertical Furnace System

Process Chamber ■ The *process chamber*, or *furnace tube*, is where wafers are heated in the furnace. A vertical furnace consists of a vertical quartz bell-jar style furnace tube that surrounds the wafers for thermal processing in the furnace, a heating element with multiple heat zones, and a heating jacket (see Figure 10.17 on page 242). Furnace tube is also a term used to describe the heated chamber on horizontal furnaces. The furnace tube must be easy to remove so that it can be cleaned when required.

Furnace Tube Materials. The wafers are placed in a vertical *wafer boat* that holds the wafers horizontal in the furnace tube. This boat and the other high-temperature components for the

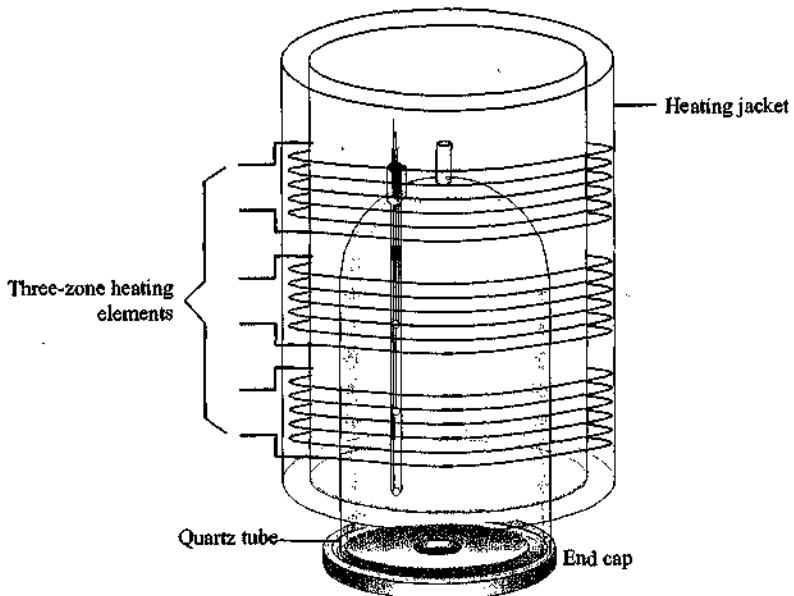


FIGURE 10.17 Vertical Furnace Process Tube

furnace tube are made of amorphous quartz that is resistant to high temperature. Quartz is single crystal SiO_2 , while amorphous quartz used for high-temperature applications is fused silica. Furnace containers are commonly called *quartzware*, and the supporting structure that holds wafers in the furnace tube is called a *quartz carrier*.

Films are deposited on the vertical wafer boat and also the inside wall of the furnace tube during processing, especially when the furnace is used for CVD deposition (see Chapter 11). The deposited films will crack and flake, sometimes after only a few cycles. The particles can become airborne and settle on the wafers causing defects and yield loss. This action requires furnace components such as the process tube to be removed and cleaned to reduce particles. Another option to reduce flaking of films is to manufacture the furnace components with silicon carbide (SiC), since the adhesion of the deposited thin film improves with SiC.³⁷ The drawback to SiC material for furnace components is that it is substantially more expensive than quartz.

Heat Zones. Each quartz process tube is surrounded by a heating element that can be controlled to produce multiple heating zones. It is common to have anywhere from three to seven zones, with 300-mm furnaces having up to nine zones.³⁸ The number of zones is important because it facilitates control of the furnace tube to attain a *flat zone* near the middle of the tube where thermal processing takes place. The heating zones on each side of the flat zone serve to optimize the wafer ramp up and ramp down to process temperature. This action allows the temperature in the flat zone to be controlled to less than 0.25°C , even at temperatures of over 1000°C .³⁹

Wafers are loaded horizontally into a quartz holder called a tower or boat, which is positioned on a quartz mount referred to as the pedestal. The pedestal and furnace tube rest on a water-cooled base plate. Some furnaces rotate the tower during processing for improved heating uniformity.

An important feature of an advanced vertical furnace is its control over the atmosphere around the wafers in the furnace tube. This atmosphere is referred to as the *ambient*. Some furnaces use load locks to keep the ambient in the furnace tube from being exposed to atmospheric gases, while others have nitrogen purge to remove any remaining gases after processing a batch of wafers. Some designs also include an extra quartz enclosure called a liner or inner tube around the wafers to improve the ambient temperature control.⁴⁰

Heating Element. The heating element in a vertical furnace is a metal resistance wire wrapped around the outside of the process tube to provide uniform heat throughout the zone it is heating. Heating elements for three zones are shown in Figure 10.18. The heater elements are switched on and off in response to signals from the temperature controller using a switching system consisting of silicon-controlled rectifiers (SCRs). This system permits control of the amount of power delivered to the heaters (e.g., 50% power or 100% power).

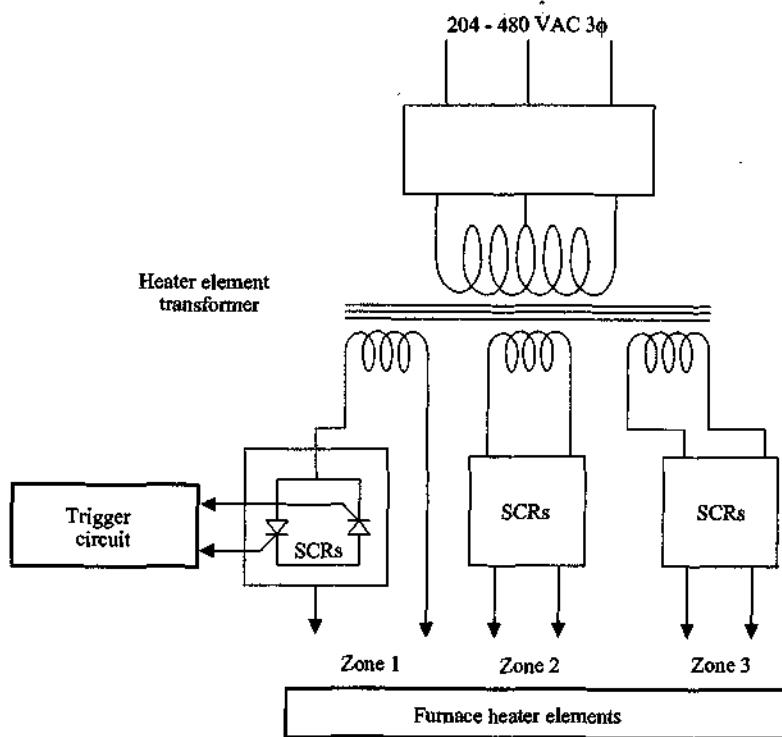


FIGURE 10.18 Heater Element Power Distribution
(Used with permission from International SEMATECH)

Temperature Control. The ability to precisely control temperature in the furnace tube is critical in a furnace. An important part of this temperature control is the sensor known as a *thermocouple (TC)* that detects the temperature and provides a corresponding millivolt signal to the furnace controller. Thermocouples are often used because they are rugged, accurate, inexpensive, and work over a wide range of temperatures.

There are multiple thermocouples for each zone of the process tube (see Figure 10.19). The *profile thermocouples* are positioned inside the process chamber in close proximity to the wafer stack, with one in each temperature zone. The profile thermocouples approximate the wafer surface temperature. There are *control thermocouples* (also referred to as spike thermocouples) positioned outside the process tube near the heater element windings in each temperature control zone. These

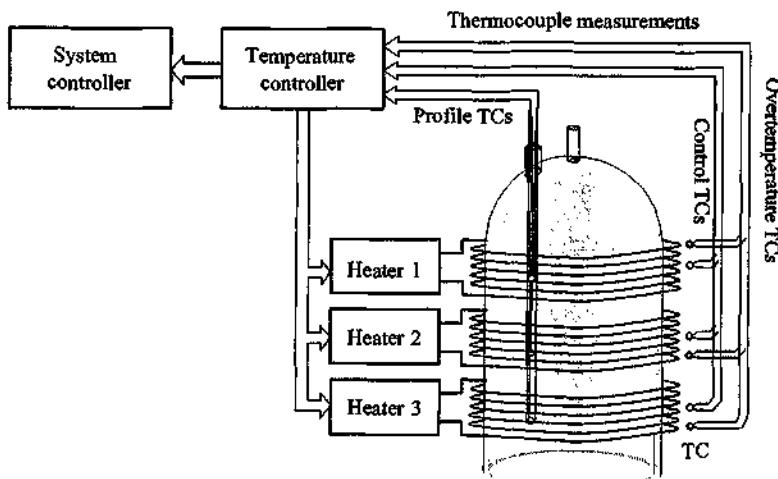


FIGURE 10.19 Locations of Thermocouples in the Furnace Chamber

thermocouples measure the heater element temperature.⁴¹ In addition, there are typically *over-temperature thermocouples* near the control thermocouples to monitor the maximum heater temperature to ensure the furnace never operates at an excessive temperature.

Wafer Transfer System ■ The primary function of the *wafer transfer system* in a vertical furnace is to load and unload wafers from the process chamber. All wafer loading and unloading during the wafer transfer is done using a wafer transfer robot. The robot moves wafers between four stations: cassette station, furnace station, load station, and cooling station. The load station during wafer transfer is maintained at a class 10 or better.

Gas Distribution System ■ The gas distribution system maintains the ambient inside the furnace by delivering the correct flow of gases to the process tube. Depending on the process, there are different bulk and specialty gases that are delivered to the furnace through the distribution system. Some of these typical furnace gases used for oxidation and other furnace processes are listed in Table 10.4.⁴²

TABLE 10.4 Common Gases Used in Furnace Processes

Gases	Classifications	Examples
Bulk	Inert gas	Argon (Ar), Nitrogen (N ₂)
	Reducing gas	Hydrogen (H ₂)
	Oxidizing gas	Oxygen (O ₂)
Specialty	Silicon-precursor gas	Silane (SiH ₄), dichlorosilane (DCS) or (H ₂ SiCl ₂)
	Dopant gas	Arsine (AsH ₃), phosphine (PH ₃) Diborane (B ₂ H ₆)
	Reactant gas	Ammonia (NH ₃), hydrogen chloride (HCl)
	Atmospheric/purge gas	Nitrogen (N ₂), helium (He)
	Other specialty gas	Tungsten hexafluoride (WF ₆)

It is important to properly remove gases and their by-products. In a vertical furnace, this removal is done through a port located at one end of the vertical process tube. The gases move into an exhaust manifold to control the direction of flow for each gas. Flammable gases, such as silane (SiH₄), phosphine, and hydrogen use a chamber known as a *burn box* to actually combust the gas in the presence of air far downstream of the process chamber, reducing the flammable gas to less harmful by-products (see Figure 10.20). Special particle filters are then used to remove the solids. The gases pass through a plant *scrubber* where the toxic gases are absorbed. Most furnace exhausts use wet scrubbers that use water solutions to absorb the gases.

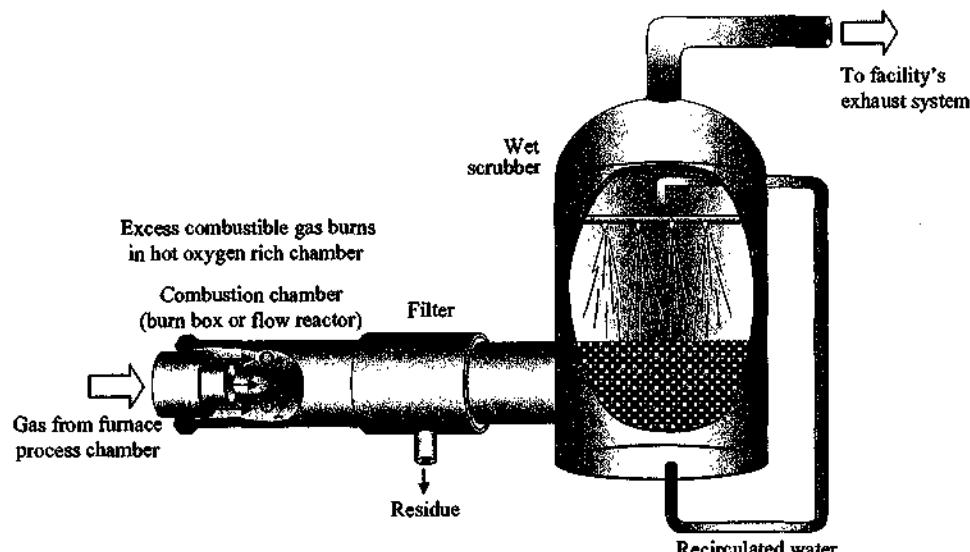


FIGURE 10.20 Burn Box to Combust Exhaust
(Used with permission from International SEMATECH)

Furnace Control System ■ A furnace microcontroller regulates all operations of the furnace, such as process time and temperature control, sequence of process steps, type of gas, the gas flow rates, temperature ramp rates, and wafer loading and unloading. The *temperature ramp rates* are the rate of temperature change that the wafers are exposed to during heating and cooling and are measured in °C/minute. A typical ramp rate for a vertical furnace is 10°C/minute while maintaining the flat zone temperature to within 0.6°C.⁴³ Other functions such as diagnostics and data collection are also performed by the microcontroller.

Each furnace microcontroller is typically interfaced into a host computer. The host computer is capable of downloading a specific wafer process recipe, which contains all the necessary data for the microcontroller. The host computer will also perform functions such as wafer lot tracking, recipe programming, and automatic scheduling of lots.

Fast Ramp Vertical Furnace

Key issues in furnace performance (and the resulting wafer throughput) are the heating and cooling times required by the furnace. New vertical diffusion furnaces, called *fast ramp furnaces*, are able to quickly raise the temperature of a batch of wafers to the processing temperature (ramp-up time), reducing the time needed to stabilize at the process temperature, followed by a rapid cooldown after processing (ramp-down time). The development of fast ramp furnaces makes it possible to process a wafer batch size of 100 wafers with a ramp-up rate of 100°C/minute and a cooling ramp-down of 60°C/minute.⁴⁴ This compares with a ramp-up rate of a few degrees per minute for a conventional vertical furnace.

This fast ramp performance becomes critical with large-diameter wafers due to the larger wafer mass required for heating and cooling. Figure 10.21 shows a typical thermal profile of a fast ramp vertical furnace versus a conventional vertical furnace. Another important aspect of fast ramp furnaces is control of the ambient atmosphere in the furnace. For instance, nitrogen ambient control during loading will inhibit uncontrolled oxide growth and produce more uniform thickness for oxide films.

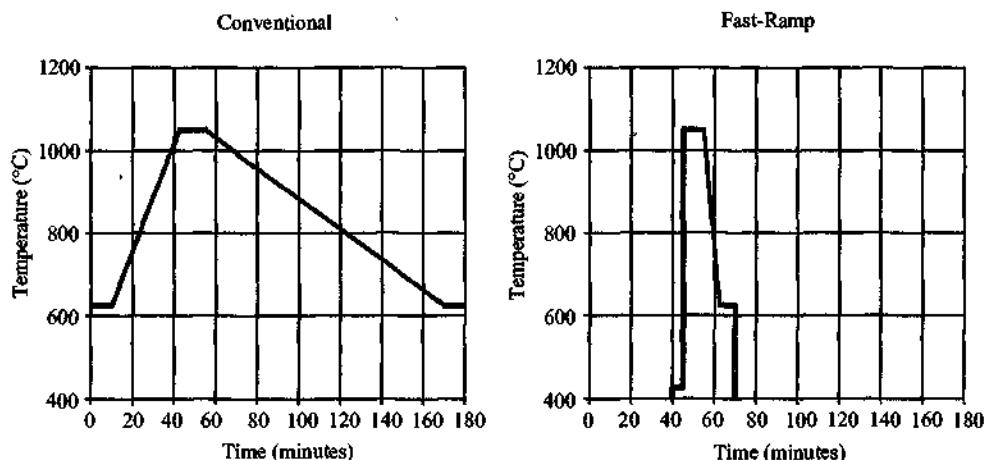


FIGURE 10.21 Thermal Profile of Conventional Versus Fast-Ramp Vertical Furnace
Printed from the June 1996 edition of *Solid State Technology*, copyright 1996 by PennWell Publishing Company

Advanced Temperature Control ■ Improved temperature control and thermal uniformity across the wafer during thermal processes are critical factors for reducing the thermal budget during wafer fabrication. A typical fast-ramp furnace can control the temperature uniformity to within $\pm 0.1^\circ\text{C}$ at process temperature, while maintaining $\pm 5^\circ\text{C}$ on all wafers during an 80°C/minute ramp cycle.⁴⁵ The key differences in a fast-ramp furnace are fast heating elements, special wafer carriers that increase the wafer-to-wafer gap for more uniform heating and cooling across the wafer, forced air cooling, and properly tuned temperature controllers.

Temperature in conventional furnaces is measured with thermocouples (see previous section), with a thermocouple controlling each furnace zone. A new technique of temperature control is

called *model-based temperature control*, which allows individual wafers to be controlled for heating and cooling rather than merely controlling the ambient of the furnace. This is done by optimizing the control software for each individual furnace with the size of the wafer load. The furnace controller is able to monitor each zone relative to one another to optimize the temperature at the wafer.

Another factor for fast ramp furnaces is the size of the load. There is a trade-off between a large batch size of 150 to 200 wafers and the ramp rate. Increasing the ramp rate for a large batch of wafers fast will create stress on the wafers because the edge circumference of the wafers will heat up as quickly as the furnace does, while the temperature at the center of the wafer will lag by several hundred degrees. Fast ramp furnaces will typically use a smaller batch size of 50–100 wafers to assist in increasing the ramp rate. The smaller batch size also improves the flow of parts through the furnace process because there are fewer wafers to batch together for a run.

Rapid Thermal Processor

The rapid thermal processor (RTP) is a method of heating a single wafer to a temperature range of 400 to 1300°C in a very short time (often fractions of a second). The main advantages for an RTP over a conventional vertical furnace are:

- ◆ Reduced thermal budget
- ◆ Minimized dopant movement in the silicon
- ◆ Ease of clustering multiple tools
- ◆ Reduced contamination due to cold wall heating
- ◆ Cleaner ambient because of the smaller chamber volume
- ◆ Shorter time to process a wafer (referred to as cycle time)

The comparison of a conventional vertical furnace to a rapid thermal processor is outlined in Table 10.5.⁴⁶

TABLE 10.5 Comparison of Conventional Vertical Furnace and RTP

Vertical Furnace	RTP
Batch	Single-wafer
Hot wall	Cold wall
Long time to heat and cool batch	Short time to heat and cool wafer
Small thermal gradient across wafer	Large thermal gradient across wafer
Long cycle time	Short cycle time
Ambient temperature measurement	Wafer temperature measurement
Issues:	Issues:
Large thermal budget	Temperature uniformity
Particles	Minimize dopant movement
Ambient control	Repeatability from wafer to wafer
	Throughput
	Wafer stress due to rapid heating
	Absolute temperature measurement

RTP Design ■ A schematic of a rapid thermal processor (RTP) is shown in Figure 10.22. A single wafer is rapidly heated in a chamber, commonly called a *reactor*, under atmospheric conditions or at low pressure. The RTP has a gas-handling system and a computer that controls system operation.

Most RTPs use multiple tungsten halogen lamps organized into zones as the heat source. The lamps are usually located on the top and bottom of the wafer and can range from 25 lamps to over 150 lamps. They are configured in multiple zones, such as from 4 to 14 zones to permit temperature contouring on the wafer. This contouring can compensate for nonuniform heating and cooling that may occur during ramp up and ramp down in cold wall systems. The silicon wafers are heated by selectively absorbing radiation from the lamps, which produce short-wavelength radiation. In this manner, RTP transfers energy between a radiant heat source and a wafer and does not heat the

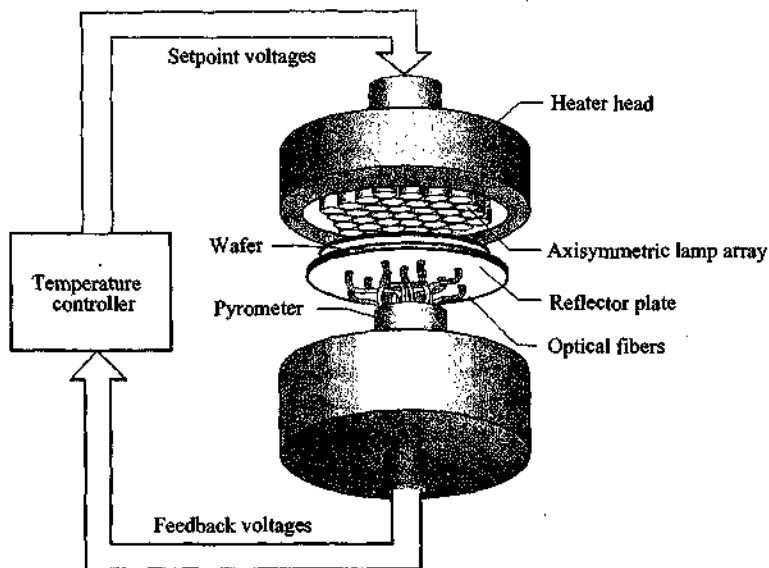


FIGURE 10.22 Rapid Thermal Processor (RTP)

surrounding reactor wall. This is why the term *cold wall* is used. A cold wall reactor is usually polished metal (such as stainless steel to increase reflectivity), with a quartz window to pass the radiation from the tungsten halogen lamp heat sources. Note that there are some systems that use lamps or resistance heaters to heat a susceptor, which then heats the wafer through thermal contact. In this case, the RTP would be classified as a warm or hot wall heater.

Temperature control in an RTP is done with thermocouple or optical pyrometers. Thermocouples are used for physical contact with the wafers to establish the actual wafer temperature. While relatively reliable, thermocouples have a slow response time and suffer from a shortened life span at higher temperatures.

An *optical pyrometer* can measure temperature from a distance with a fast response time. This measurement is done by sensing the emitted infrared radiation of the wafer while it is being heated. However, the emissivity (amount of emitted radiation) of the wafer surface can vary depending on the type of film on the top or bottom surface of the wafer. Recent advances in optical pyrometry



Rapid Thermal Processor
(Photo courtesy of Advanced Micro Devices)

have made emissivity less of a factor by utilizing solid-state pyrometers for better reliability and temperature measurement control.

The RTP has been hampered in the past with poor temperature uniformity across the wafer and nonequilibrium heating.⁴⁷ These problems lead to crystal slip and wafer warpage. There have been significant advances in RTP systems in the past few years, resulting in improved temperature control.

Equipment Integration ■ An RTP introduces flexibility into thermal processing because it can be integrated with other process steps into a multichamber cluster tool. This is beneficial because all processing steps occur in a vacuum environment to eliminate concerns with native oxide and contamination. The movement of single wafers through a cluster tool also reduces the time it takes to process wafers by reducing the waiting times associated with large wafer batches. Single wafers can also be processed in parallel with one another. These benefits occur in conjunction with the important action of reducing the thermal budget of the wafer.

RTP Applications ■ The RTP has become widely used in many processes throughout wafer fabrication. Advances in RTP chamber design and temperature uniformity are making this equipment acceptable with better temperature uniformity across wafers. Examples of operations that often use the RTP in wafer fabrication are:

- ◆ Anneal of implants to remove defects and activate and diffuse dopants
- ◆ Densification of deposited films, such as deposited oxide layers
- ◆ Borophosphosilicate glass (BPSG) reflow
- ◆ Anneal of barrier layers, such as titanium nitride (TiN)
- ◆ Silicide formation, such as titanium silicide ($TiSi_2$)
- ◆ Contact alloying

The first widely used application for RTP processing was annealing after ion implant. The advantage of using an RTP over a conventional diffusion furnace is that the wafer spends a much shorter time at temperature, thus reducing the thermal budget. However, the uniform temperature control of conventional furnaces makes this approach competitive.

OXIDATION PROCESS

The goal of thermal oxidation is the growth of a defect-free, uniform layer of SiO_2 at the thickness required. The type of oxidation process conditions used for a particular wafer fabrication step depends on the thickness of the oxidation layer and the properties required. Thin oxides, such as gate oxides, are usually grown in dry oxygen. Because sodium ion contamination is a concern, HCl is added to the O_2 supply during the oxidation of high-quality oxides. Where thick oxides are required, such as the field oxidation layer, then steam is used (HCl is not used in steam oxidation processes). Higher pressure during the growth process allows thick oxide growth to be achieved at a reduced temperature in reasonable time periods. The flow illustrated in Figure 10.23 shows the typical steps for thermal oxidation.

Pre Oxidation Cleaning

Cleanliness of the wafer is critical to achieve high quality oxidation (see Chapter 6). Contaminants such as particulates and mobile ionic contaminants (MICs) will have serious impact on device performance and yields. For example, the presence of MICs in the thermally grown gate oxide structure can result in long-term changes in the device threshold voltage as the MICs drift from the gate oxide to the Si/SiO_2 interface.⁴⁸ This condition is detrimental to device electrical performance. For a thermal oxidation process, the ability to keep MICs and particulate matter out of the system depends on maintaining the system to a high state of cleanliness control. The following areas are critical for minimizing contamination:

- ◆ Maintenance of the furnace and associated equipment (especially quartz components) for cleanliness
- ◆ Purity of processing chemicals
- ◆ Purity of oxidizing ambient (the source of oxygen in the furnace)
- ◆ Wafer cleaning and handling practices

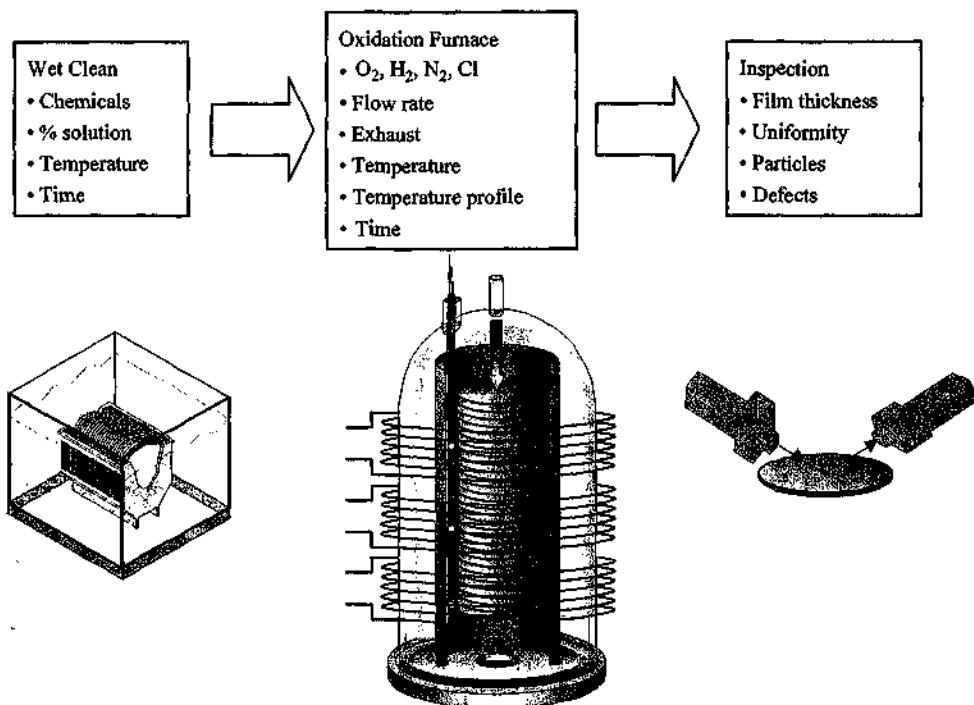


FIGURE 10.23 Thermal Oxidation Process Flow Chart

The basic wet-cleaning chemistries that have withstood the test of time are the RCA SC-1 and SC-2 clean systems and the piranha clean (sulfuric acid, hydrogen peroxide, and water mixture). These wafer cleaning methods and their modifications were discussed in Chapter 6. A wide selection of cleaning equipment is found in the oxidation area of the fab, including manual and automated wet sinks, ultrasonic systems (megasonics), acid spray processors, and rinser/dryer systems.

Oxidation Process Recipe

For thermal oxidation to occur, certain process conditions are followed in a specific format in the furnace equipment. This is known as a *process recipe*. For wafer fabrication, it is common that the parameters of a process recipe are stored in a software database and downloaded into the tool software for a particular wafer lot. We will review a sample process recipe for dry oxidation in a batch furnace.

Dry Thermal Oxidation ■ The most common reason for dry thermal oxidation is to grow the very thin layer of oxide used as the gate oxide.⁴⁹ As the devices become smaller in order to increase the component density on a chip (following Moore's law from Chapter 1), then the device dimensions and electrical parameters must also be scaled down proportionally. Gate oxide layers are considered the most critical oxidation application in wafer fabrication, with a gate oxide thickness on the order of 20–40 Å for 0.18-μm devices and 20–30 Å for 0.15-μm devices.⁵⁰ Dry thermal oxidation produces a high-quality oxide that is uniformly dense, free of pinholes, and reproducible. It is important for gate oxides to be produced in a clean, sodium-free system that is truly dry, since as little as 25 ppm of water will alter the oxide growth rate and properties of the oxide.⁵¹ There is ongoing research to evaluate other replacement materials that could function as a dielectric in the gate structure, but there is no suitable alternative in production.

Process Recipe ■ Table 10.6 shows a sample process recipe for dry oxidation of a gate oxide.⁵² During idle conditions, the process recipe shows there is a nitrogen (N_2) purge gas introduced into the chamber. It flows continuously until the process recipe is started and then the purge is turned off and N_2 process gas is turned on. Once the furnace tube is loaded, the temperature is raised from idle at 850°C to 1,000°C at a ramp rate of 20°C per minute. The wafers have a five-minute stabilization period. Since this is dry oxidation, oxygen at a flow rate of 2.5 slm is introduced into the process chamber. HCl flows in at 67 sccm and is used to reduce the interface charges and to getter MICs. During the anneal step, which is done to minimize charge buildup, the O_2 and HCl are turned off and N_2 is turned on. Oxygen still diffuses through the oxide even though the oxide flow has been cut off, which is taken into account to produce stoichiometric (uniform composition) silicon dioxide. A five-minute period is allowed to unload the furnace tube.

TABLE 10.6 Process Recipe for Dry Oxidation Process

Step	Time (min)	Temp (°C)	N_2 Purge Gas (slm)	N_2 (slm)	Process Gas O_2 (slm)	HCl (sccm)	Comments
0		850	8.0	0	0	0	Idle condition
1	5	850		8.0	0	0	Load furnace tube
2	7.5	Ramp 20°C/min		8.0	0	0	Ramp temperature up
3	5	1000		8.0	0	0	Temperature stabilization
4	30	1000		0	2.5	67	Dry oxidation
5	30	1000		8.0	0	0	Anneal
6	30	Ramp -5°C/min		8.0	0	0	Ramp temperature down
7	5	850		8.0	0	0	Unload furnace tube
8		850	8.0	0	0	0	Idle

Note: Gas flow units are slm (standard liters per minute) and sccm (standard cubic centimeters per minute).

QUALITY MEASUREMENTS

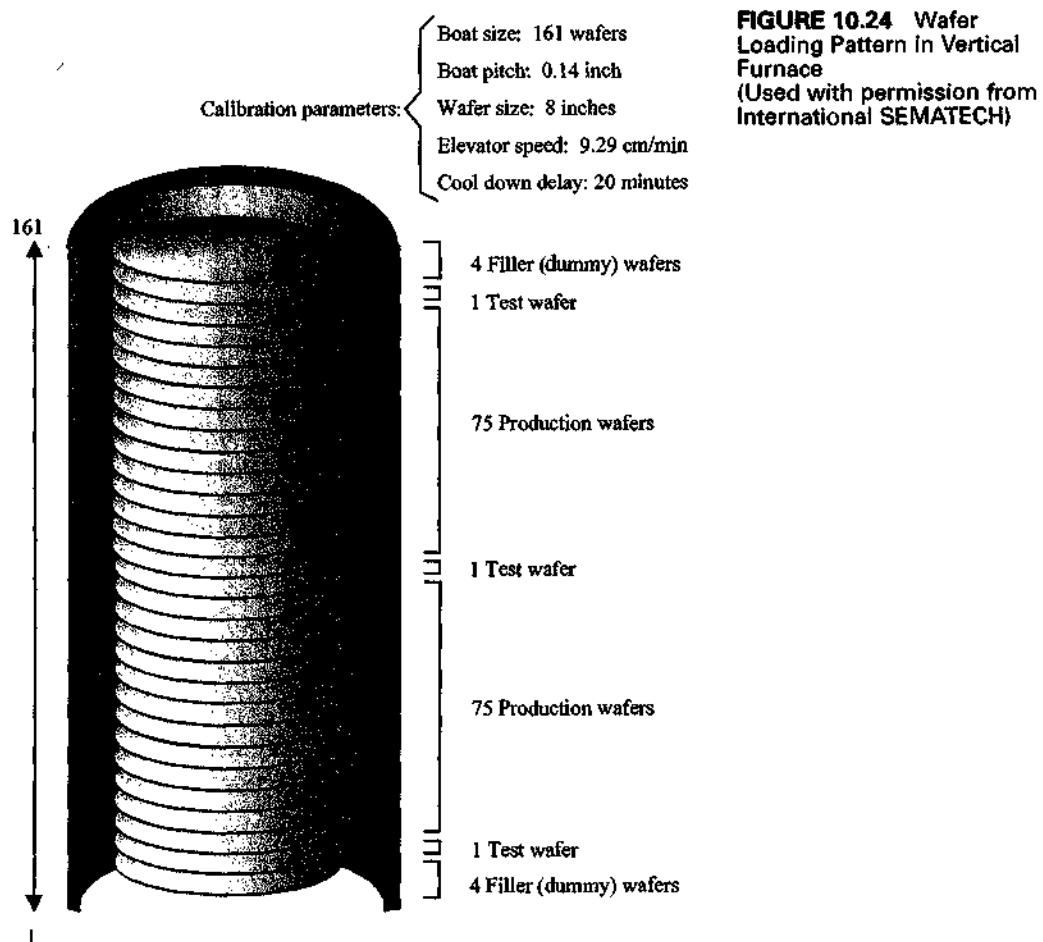
Representative quality measures for thermal oxide growth are listed in Table 10.7.⁵³

TABLE 10.7 Quality Measures for Thermal Oxide Layer

Quality Parameter	Types of Defects	Remarks
1. Oxide thickness.	A. Thickness of gate oxide outside of specification (representative gate oxide specification is $20 \pm 1.5 \text{ \AA}$).	Possible reasons for this problem are: <ul style="list-style-type: none"> • Incorrect process gas flow (e.g., MFC calibrated improperly). Since HCl enhances oxidation rate, verify HCl:O_2 ratio is correct. • Verify O_2 leak integrity of the furnace with a bare silicon test wafer. • Check the metrology equipment by measuring the oxide thickness against a standard thickness wafer. • Excess native oxide growth due to overexposure of wafer to air either before or after normal furnace oxidation.

Quality Parameter	Types of Defects	Remarks
2. Gate oxide integrity (GOI).	A. Gate oxide breakdown. B. Mobile ionic contamination (MIC) in film.	Gate oxide defects are often related to processing conditions: <ul style="list-style-type: none">• Perform a C-V test to demonstrate gate oxide integrity using an unpatterned test wafer.• Perform an oxide-charge analysis on a test wafer with a surface-charge analyzer.• Review the preoxidation clean steps to assess sources of contamination (e.g., particles or MICs).• Verify no contamination from an incoming gas line or defective filter.
3. Particles in the oxide film.	A. Contaminated quartzware. B. Wafer broken inside furnace. C. Contaminated carrier. D. Contaminated gas filter or line.	Actions to correct these sources of particles added during the oxidation process are: <ul style="list-style-type: none">• Check cleanliness of quartzware and carrier.• Verify the alignment of the robotic handling systems.• Check incoming gas filters.
4. Particles under the oxide film.	A. Contaminated preoxidation clean.	Source of particles prior to the oxidation step are: <ul style="list-style-type: none">• Verify the preoxidation clean steps are properly set up and performed.• Check cleanliness of quartzware and carrier.

When a batch of wafers is loaded into a furnace tube for oxidation, a number of test wafers with bare surfaces (also referred to as unpatterned wafers, as described in Chapter 19) are placed at strategic locations in the tube (see Figure 10.24). These test wafers are used to perform the various evaluations after the oxidation step to ensure the oxide has acceptable quality.



OXIDATION TROUBLESHOOTING

Common problems and corrective actions for thermal oxidation are described in Table 10.8.

TABLE 10.8 Common Oxidation Troubleshooting Problems

Problem	Probable Causes	Corrective Actions
1. Incorrect gas flow into furnace tube.	A. Incorrect process recipe. B. Malfunctioning MFC. C. Incorrect H ₂ :O ₂ ratio for steam process (O ₂ starved).	Possible corrective actions for this problem are: <ul style="list-style-type: none"> • Verify that correct process recipe is downloaded into furnace software for the wafer being processed. • Check process gas MFCs (O₂, N₂, H₂, Cl) to verify calibration was performed and operating properly. • Verify gas valves are functioning properly (e.g., no leaks, and so on). • Check that no room air is leaking into the furnace tube from outside the quartz or furnace door seal.
2. Incorrect temperature uniformity in vertical furnace chamber.	A. Wrong process recipe for product. B. Incorrect operation of thermocouples.	Possible corrective actions are: <ul style="list-style-type: none"> • Verify that the correct process recipe is loaded for the wafer being processed. • Assess whether the temperature nonuniformity is during the ramp up, flat zone, and ramp down for corrective action. • Check all thermocouples for proper operation. Verify no degradation due to excessive heat or corrosion. Replace defective thermocouples. Verify no drift in thermocouple reference temperature.
3. Inadequate temperature uniformity in rapid thermal processor (RTP).	A. Malfunctioning heating system (e.g., tungsten halogen lamps). B. Verify correct operation of temperature measurement sensor.	Possible corrective actions are: <ul style="list-style-type: none"> • Verify heating lamps are operating and have the correct lamp intensity. • Verify correct calibration and temperature measurement of the optical pyrometer temperature sensor. Check there is no variation in wafer emissivity by performing reflectivity measurement of the wafer surface.

SUMMARY

Silicon dioxide (or oxide) can be thermally grown or deposited on a wafer. SiO₂ has a tetrahedron cell atomic structure of a silicon atom surrounded by four oxygen atoms and is amorphous. Five uses of oxide film in wafer fabrication are (1) protection and isolation, (2) passivation, (3) dielectric, (4) dopant mask and (5) deposited insulator between metal layers. The thickness of an oxide layer depends on its applications. The chemical reaction for thermally grown dry oxide requires silicon and an oxidizer gas, consumes silicon during the reaction, and is speeded up with the use of wet oxidation. There is an undesirable oxide charge at the Si/SiO₂ interface that can be neutralized by the use of HCl. The growth of thermal oxide follows a linear growth stage up to about 150 Å, followed by a slower parabolic growth stage. Factors affecting the oxide growth rate are dopants, crystal orientation, pressure, and temperature. SiO₂ can be selectively grown or deposited.

The three basic types of thermal processing equipment are a horizontal furnace, vertical furnace, and rapid thermal processor (RTP). The vertical furnace has become the primary batch furnace in fabrication. It consists of a process chamber, wafer transfer system, gas distribution system, exhaust system, and temperature control system. Chamber materials are made of quartz or silicon carbide. The temperature is precisely controlled with the use of multiple thermocouples. The goal in thermal processing is to minimize the thermal budget. Fast-ramp vertical furnaces quickly ramp the temperature up or down with a smaller batch of wafers by using advanced temperature control. The rapid thermal processor (RTP) can reduce the thermal budget by heating wafers with ramp rates of up to hundreds of degrees per second and is often integrated with other process steps. Preoxidation cleaning is critical for attaining a high-quality oxide.

KEY TERMS

grown oxide layer
deposited oxide layer
topography (surface topology)

thermal budget
thermal oxide (thermal silicon dioxide [SiO_2])
tetrahedron cell

surface passivation
field oxide layer

gate oxide structure
gate oxide integrity

dry oxygen
wet oxidation

pyrogenic steam
diffusion

Fick's laws
fixed oxide charge

interface-trapped charge
mobile oxide charge

oxide-trapped charge
rate of oxide growth

linear stage
reaction-rate controlled

parabolic stage
diffusion controlled

local oxidation of silicon (LOCOS)
bird's beak effect

pad oxide

shallow trench isolation (STI)
oxidation-induced stacking faults (OISFs)

horizontal furnace
vertical furnace

hot wall batch furnaces
rapid thermal processor (RTP)

process chamber or furnace tube
wafer boat

quartzware
quartz carrier

flat zone
ambient

thermocouple (TC)
profile thermocouples

spike thermocouples
over-temperature thermocouples

wafer transfer system
burn box

scrubber
temperature ramp rates

fast ramp furnaces
model-based temperature control

reactor
cold wall

optical pyrometer
process recipe

REVIEW QUESTIONS

- What is the difference between a grown and a deposited oxide layer?
- Describe what topography is on the wafer surface.
- Define thermal budget and explain why it is undesirable.
- State what thermal silicon dioxide is and give another name for it.
- Describe the atomic structure for silicon dioxide.
- What is surface passivation and why is it beneficial?
- Describe the field oxide layer, and state its range of thickness.
- Why is the gate oxide thermally grown?
- Explain gate oxide integrity.
- Describe how SiO_2 can be used as a dopant barrier.
- List six applications for thermal oxides in wafer fabrication and give a purpose for each application.
- State the chemical reaction for dry oxidation. At what temperature range does this reaction usually take place?
- State the chemical reaction for wet oxidation. Is this faster or slower than dry oxidation? Why?
- If an oxide layer is thermally grown to be 2,000 Å thick, how much silicon is consumed?
- What is diffusion? How does this process occur in thermal oxidation?
- List the four types of oxide charges that can occur at the Si/SiO_2 interface. Are oxide charges desirable? Why or why not?
- Give two advantages to using chlorinated agents during oxidation.
- Describe rate of oxide growth. What parameters influence this rate?
- Describe the linear growth stage for thermal oxidation. What is the thickness range where it occurs? State the equation that describes this growth.
- What does it mean to be reaction-rate controlled for thermal oxidation growth?

21. Describe the parabolic growth stage for thermal oxidation. What is the thickness range where it occurs? State the equation that describes this growth.
22. What does it mean to be diffusion controlled for thermal oxidation growth?
23. What effect does doping have on oxide growth?
24. Explain the effect from crystal orientation on oxide growth.
25. Pressure has what effect on oxide growth?
26. What is the effect of plasma on oxide growth?
27. What is LOCOS and how does this process use thermal oxidation? What is a bird's beak in forming an oxidation layer, and why is this condition undesirable?
28. Explain shallow trench isolation (STI).
29. What causes stress in an oxide layer?
30. What are oxidation-induced stacking faults (OISFs)?
31. List the three basic types of thermal processing equipment.
32. What is a hot wall furnace?
33. List five performance factors for horizontal and vertical furnaces, and stipulate which type of furnace is the optimum.
34. What are the five components in a vertical furnace system?
35. Describe the process chamber (or furnace tube).
36. What two common materials are used for process tube materials?
37. How many zones are there typically in a furnace, and why is this number important?
38. What is a wafer boat?
39. What is meant by the term *ambient* when used in a high-temperature furnace?
40. What are the three locations for thermocouples in a vertical furnace? Describe the purpose of each thermocouple.
41. Explain the purpose of the wafer transfer system in a vertical furnace.
42. A gas delivery system in a vertical furnace serves what purpose?
43. Explain what a burn box is.
44. Describe the purpose of the furnace microcontroller.
45. Explain what a fast ramp vertical furnace is. What ramp-up and ramp-down rates can this furnace achieve?
46. Describe the advanced way to control temperature in a fast ramp vertical furnace.
47. What is a rapid thermal processor (RTP)? What are six advantages it has over the conventional furnace?
48. Describe how an RTP heats a wafer. Is an RTP typically a hot wall or cold wall heating system?
49. Describe the purpose of an optical pyrometer in an RTP and how it functions.
50. Why is preoxidation cleaning important?

FURNACE AND RTP EQUIPMENT SUPPLIERS' WEB SITES

Amtech Systems Inc.
Applied Materials
Asahi Glass Electronic Materials
ASM
Axcelis (formerly Eaton)
CVD Equipment Corporation
GaSonics International
Eaton Corporation
Heraeus Amersil Inc.
Kokusai Semiconductor Equipment
MRL Industries
Omega Engineering Inc.
Semitool Inc.
Silicon Valley Group
TEL, Tokyo Electron Ltd.
Tystar Corporation

<http://www.amtechsystems.com/>
<http://www.appliedmaterials.com/products/>
<http://www.agem-usa.com/>
<http://www.asm.com/>
<http://www.axcelis.com/>
<http://www.cvdequipment.com/>
<http://www.gasonics.com/>
<http://www.semiconductor.eaton.com/>
<http://www.heraeus-amersil.com/>
<http://www.ksec.com/>
<http://www.mrlind.com/>
<http://www.omega.com/>
<http://www.semitool.com/>
<http://www.svg.com/>
<http://www.teainet.com>
<http://www.tystar.com/>

REFERENCES

1. G. Moore, "The Role of Fairchild in Silicon Technology in the Early Days of 'Silicon Valley,'" *Proceedings of the IEEE* 86-1 (January 1998); p. 62.
2. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics for Future ICs," *Solid State Technology* (September 1998); p. 70.
3. G. Aner, *Planar Processing Primer*, (New York: Van Nostrand Reinhold, 1990), p. 146.
4. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: Wiley, 1994), p. 452.
5. Ibid.
6. SEMATECH, *Semiconductor Dictionary*, (Austin, TX: SEMATECH, 1995).
7. S. Ghandhi, *VLSI Fabrication*, p. 452.
8. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, (Boston: Kluwer Academic, 1995), p. 40.
9. SEMATECH, "Furnace Equipment Overview," Module 3 in *Furnace Processes and Related Topics*, (Austin, TX: SEMATECH), p. 3-2.
10. Ibid.
11. B. El-Kareh, *Fundamentals*, p. 47.
12. B. El-Kareh, *Fundamentals*, p. 45.
13. D. Schroder, *Semiconductor Material and Device Characterization*, 2nd. ed., (New York: Wiley, 1998), pp. 337-339.
14. J. Mayer and S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*, (New York: Macmillan, 1990), p. 265.
15. D. Schroder, *Semiconductor Material*, pp. 337-339.
16. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics*, p. 8.
17. G. Aner, *Planar Processing Primer*, p. 176.
18. S. Wolf and R. Tauber, *Process Technology*, vol. 1, *Silicon Processing for the VLSI Era*, 2nd ed. (Sunset Beach: Lattice Press, 2000), p. 269.
19. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics*, p. 6.
20. S. Ghandhi, *VLSI Fabrication*, p. 460.
21. Ibid.
22. Ibid.
23. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics*, p. 7.
24. B. Deal and A. Grove, "General Relationship for the Thermal Oxidation in Silicon," *Journal of Applied Physics* (1965); pp. 3770-78.
25. S. Ghandhi, *VLSI Fabrication*, p. 463.
26. S. Wolf and R. Tauber, *Process Technology*, p. 278.
27. S. Wolf and R. Tauber, *Process Technology*, p. 283.
28. B. El-Kareh, *Fundamentals*, p. 60.
29. S. Wolf and R. Tauber, *Process Technology*, p. 218.
30. S. Ghandhi, *VLSI Fabrication*, p. 462.
31. B. El-Kareh, *Fundamentals*, p. 69.
32. S. Wolf and R. Tauber, *Process Technology*, p. 296.
33. S. Ghandhi, *VLSI Fabrication*, p. 473.
34. P. Singer, "Furnaces Evolving to Meet Diverse Thermal Processing Needs," *Semiconductor International*, (March 1997), p. 85.
35. Industry Watch, "Thermal Processing: Meeting the Challenges of 300 mm," *Semiconductor International*, (April 1998), p. 19.
36. SEMATECH, "Furnace Equipment Overview," module 3 in *Furnace Processes and Related Topics*, pp. 3-6.
37. J. Tomanovich, "LPCVD Components Trend Toward SiC," *Solid State Technology*, (June 1997); p. 135.
38. P. Singer, "Furnaces Evolving," p. 86.
39. A. Helms et al., "Status and Future of Batch, Hot-Wall Furnaces," *Solid State Technology*, (November 1999); p. 83.
40. P. Singer, "Furnaces Evolving," p. 86.
41. C. Porter et al., "Improving Furnaces with Model-Based Temperature Control," *Solid State Technology*, (November 1996); p. 120.
42. SEMATECH, "Gas Delivery Systems," module 8 in *Furnace Processes and Related Topics*, (Austin, TX: SEMATECH, 1994), pp. 8-3.
43. A. Helms et al., "Status and Future," p. 83.
44. L. Peters, "Thermal Processing's Tool of Choice: Single-Wafer RTP or Fast Ramp Batch?" *Semiconductor International*, (April 1998); p. 84.
45. Ibid.
46. R. Fair, "Conventional and Rapid Thermal Processes," *ULSI Technology*, ed. C. Chang and S. Sze, (New York: McGraw-Hill, 1996), p. 151.
47. K. Reid and A. Sitaram, "Rapid Thermal Processing for ULSI Applications: An Overview," *Solid State Technology*, (February 1996); p. 64.
48. S. Wolf and R. Tauber, *Process Technology*, p. 225.
49. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics*, p. 11.
50. L. Peters, "Thermal Processing's Tool," p. 86.
51. S. Ghandhi, *VLSI Fabrication*, p. 466.
52. Adapted from SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics*, p. 13.
53. SEMATECH, "Troubleshooting Process and Equipment Problems," module in *Furnace Processes and Related Topics*, (Austin, TX: SEMATECH, 1994), pp. 1-16.

CHAPTER 11

DEPOSITION

Microchip fabrication is a planar process involving many different film layers on the wafer surface. Deposition is a process that places the film layers on the wafer. Conductor and insulator films are essential to the successful fabrication of semiconductor devices on silicon substrate wafers. Film layering technology is used to manufacture the circuits and to connect the different IC devices, primarily with metal conducting layers sandwiched between insulating dielectric layers.

Many different types of films are deposited on wafers during the fabrication process. In some cases, these deposited films become an integral part of the device

structure, while other sacrificial films are used for particular processing steps and then removed. Films deposited during microchip fabrication are often referred to as thin films because they are so thin that their electrical and mechanical properties differ from a thicker bulk film of the same material.

This chapter will discuss the processes and equipment used to deposit thin films, with a focus on dielectric films such as silicon dioxide and silicon nitride and the deposition of polysilicon. The deposition of metal and metal compound thin films is discussed in Chapter 12.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe multilayer metallization. Discuss the acceptable characteristics of a thin film. State and explain the three stages of film growth.
2. Provide an overview of the different film deposition techniques.
3. List and describe the eight basic steps in a chemical vapor deposition (CVD) reaction, including the different types of chemical reactions.
4. Describe how CVD reactions are limited, and explain reaction dynamics and the effect of dopant addition to CVD films.
5. Describe the different types of CVD deposition systems, explain how the equipment functions, and discuss the benefits/limitations of a particular tool for film applications.
6. Explain the importance of dielectric materials for chip technology, giving examples of applications.
7. Discuss epitaxy and three different epitaxial deposition methods.
8. Explain spin-on-dielectrics.

INTRODUCTION

The design and fabrication of early semiconductor wafers from the MSI and LSI eras was relatively straightforward. This process consisted of fabricating the semiconductor devices in silicon and interconnecting the devices to one metal conducting layer sandwiched between silicon dioxide as the dielectric material. Figure 11.1 on page 258 shows the deposited layers needed to make an early nMOS transistor. This technology was an extension of the first planar transistors made in the SSI era. The critical dimension was well over one micron. The wafer layers were not flat structures due

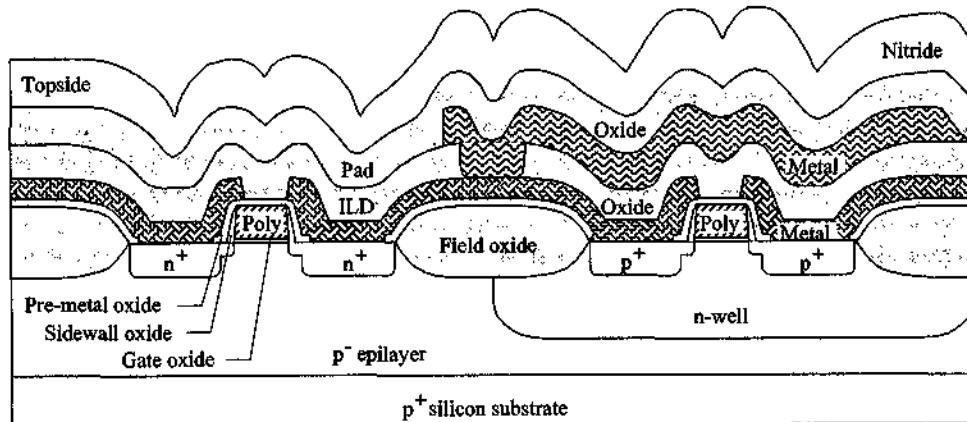


FIGURE 11.1 Film Layers for an MSI Era NMOS Transistor

to varying feature heights, which becomes a limiting factor for high-density chips with multiple metal layers needed in the ULSI era.

As wafers progress to higher density chips with shrinking geometries of $0.18 \mu\text{m}$ and below, the materials and processes used in wafer fabrication are undergoing dramatic changes. There is a concurrent scaling of all device features to maintain electrical performance. To make electrical connections in today's advanced microchips, six or more conducting metal layers are required. New metal conductor materials are needed to maintain electrical performance. Advanced dielectric materials are deposited between the metal films to provide adequate insulation protection. At the same time, each chip has literally tens of billions of electrical connections between the various metal layers and silicon devices. The ability to deposit reliable thin-film materials is a critical operation in a wafer fab. Thin films is a major process step in wafer fabrication, as shown in the process flow model in Figure 11.2.

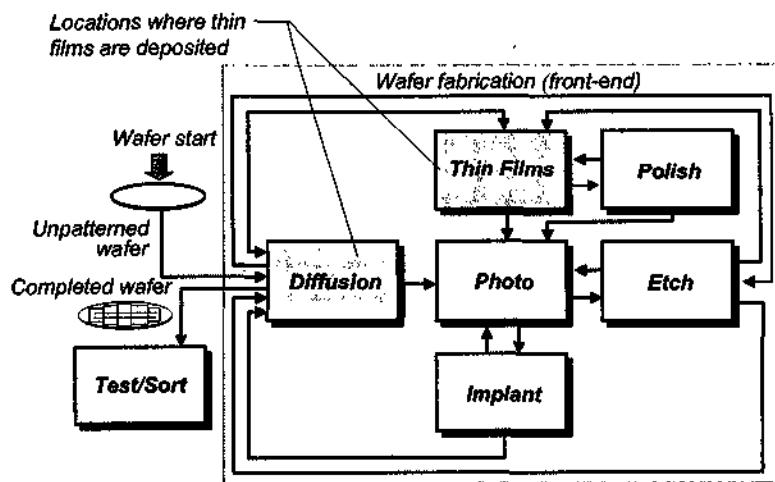


FIGURE 11.2 Process Flow in a Wafer Fab
(Used with permission of Advanced Micro Devices)

Film Layering Terminology

Multilevel metallization refers to the metal and dielectric layers needed to interconnect the densely packed devices on the wafer. An example of this process is shown in Figure 11.3. Without the dielectric insulating layers, electrical shorts would occur (similar to electrical wiring without its insulating cover). The metal layers are connected by openings in the dielectric film referred to as *vias*.

Adding metal levels is costly to manufacture. It is estimated that the addition of a single metal level in a CMOS process adds approximately 15% to the total wafer fabrication cost.¹ The

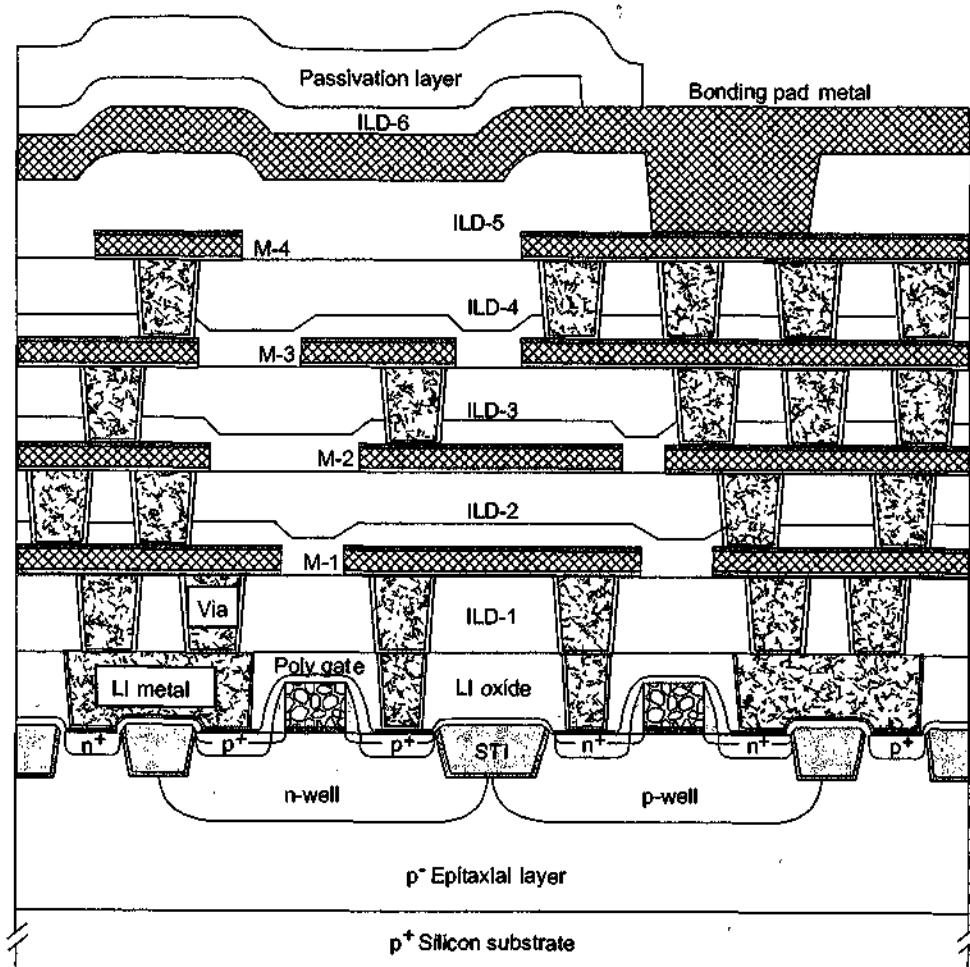


FIGURE 11.3 Multilevel Metallization on a ULSI Wafer

complexity added to wafer manufacturing by additional metal levels makes it important to reduce defects to ensure the chip yield is not impacted. The chip designer considers the tradeoffs among cost, complexity, and performance across the different wafer fab process areas for implementing additional metal levels.

Metal Layers ■ *Aluminum metallization* is the use of aluminum alloy for interconnect wiring. Aluminum alloy has been used since the beginning of semiconductor manufacturing. The Al metal is deposited on the entire surface of the wafer in a solid thin film and then etched to define the width and spacing of the interconnect lines. The industry is undergoing a transition to copper metallization to achieve increased chip speed with fewer process steps (copper metallization is reviewed in detail in Chapter 12). Each metal layer can be referred to as Metal-1, Metal-2, and so on. *Critical layers* are those metal layers with linewidths etched to the critical dimension of the device (e.g., CD of 0.15 μm). For ULSI, the CD typically occurs at the polysilicon features of the gate, the oxide structure, and the metal layers closest to the silicon surface. Critical layers are sensitive to particulate contamination (killer defects) and reliability issues such as electromigration are more pronounced in fine geometry linewidths. *Noncritical layers*, usually the upper metal levels, have much wider line widths, often 0.5 μm and larger, and are less sensitive to particulate contamination. However, factors such as long conductor lengths in the noncritical upper levels can affect chip speed and power consumption.²

Dielectric Layers ■ The dielectric layer between the active devices in silicon and the first metal layer is termed the *first interlayer dielectric (ILD-1)*. This layer is also called the *premetal dielectric (PMD)*. ILD-1 is typically a doped silicon dioxide, or glass (explained later in this chapter). The



Metal Layers in a Chip
(Micrograph courtesy of Integrated Circuit Engineering)

important function of the ILD-1 layer is to isolate transistor devices in two ways: electrically from the metal interconnect layers, and physically from contamination sources such as mobile ions. The ILD-1 has a restricted thermal budget in high-performance logic devices in order to avoid degrading transistor characteristics.³

The *interlayer dielectric (ILD)* is used between different metal layers in the device. The ILD serves as an insulating film between two conducting metals or adjacent metal lines. The ILD has traditionally been silicon dioxide (SiO_2) with a dielectric constant of around 3.9 to 4.0.⁴ The dielectric constant is an important property of deposited insulating films because it directly affects circuit speed performance (see the following section).

FILM DEPOSITION

A *thin film* is a thin, solid layer of a material created on a substrate. If a solid material has three dimensions (thickness, width, and length), then a thin solid film has one of its dimensions (usually thickness) much smaller than the other two (see Figure 11.4). The thin film is bonded to a wafer substrate, which has a much thicker dimension than the film. The thin-film surface is so close to the substrate that it has a profound influence on the physical, mechanical, chemical, and electrical properties of the thin film material.⁵ The most widely used unit to describe the thickness of a thin film in wafer fabrication is the angstrom (\AA).

Thin film deposition in semiconductor fabrication is any process that physically deposits a film on the wafer substrate. The film can be either a conducting, insulating, or semiconducting material. Examples of deposited films are polysilicon, silicon dioxide (SiO_2), silicon nitride (Si_3N_4), polysilicon (silicon in a polycrystalline structure), and metals such as copper and refractory metals (e.g., tungsten).

Thin-Film Characteristics

For a thin film to be acceptable in wafer fabrication, it must have desirable film characteristics. General characteristics of an acceptable thin film for device performance are:⁶

- ◆ Good step coverage
- ◆ Ability to fill high aspect ratio gaps (conformality)
- ◆ Good thickness uniformity
- ◆ High purity and density

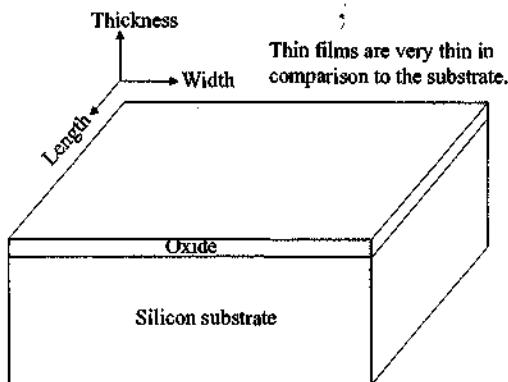


FIGURE 11.4 Solid Thin Film

- ◆ Controlled stoichiometries
- ◆ High degree of structural perfection with low film stress
- ◆ Good electrical properties
- ◆ Excellent adhesion to the substrate material and subsequent films

Film-Step Coverage ■ It is desirable for thin films to maintain a uniform thickness over surface features (see Figure 11.5). Steps on the wafer surface occur due to three-dimensional shapes from patterned features that create surface topography. If the film thins excessively at a step, this can cause high stress, electrical shorts, or undesirable induced charges in the device. Films with minimum stress are important, since stress can lead to substrate deformation in a convex or concave shape. All deposition techniques result in the formation of some stress in films.

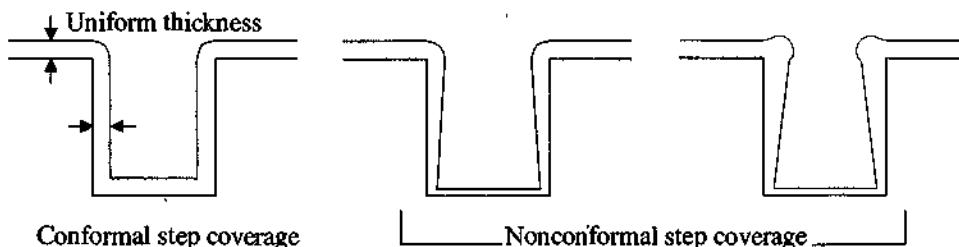


FIGURE 11.5 Film Coverage Over Steps

High Aspect Ratio Gaps ■ A small gap (trench or hole) is characterized by its *aspect ratio*, which is defined as the ratio of its depth to width (see Figure 11.6 on page 262). Aspect ratio is expressed as a ratio, such as 2:1, which in this case means the gap depth is two times the width. A gap can also be specified by its opening width, such as 0.25 μm , which may be different than the CD of other structures on the wafer. The ability to fill very small gaps and holes on the surface of the wafer has become one of the most important film characteristics for devices fabricated with sub-0.25 micron geometries. Examples of gaps that require effective gap-fill capability are vias passing through the interlayer dielectric (ILD) and trenches for shallow trench isolation (STI).⁷ A high aspect ratio is typically above 3:1, with some applications having aspect ratios of 5:1 or larger. High aspect ratio gaps make it difficult to uniformly deposit the film, leading to pinch-off and voiding. Deposition processes that can produce uniform, void-free films in high aspect ratio openings are critical in the sub-0.25 μm era of high-density IC circuits with shrinking geometries.

Thickness Uniformity ■ Acceptable thin films are conformal with good thickness uniformity, which means the film follows the outline of the base material in all areas. The resistance of a material will change with varying thickness, which would be undesirable. Thinner layers of film also

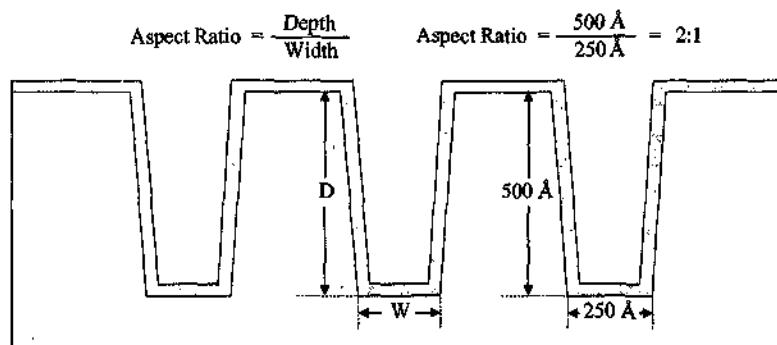
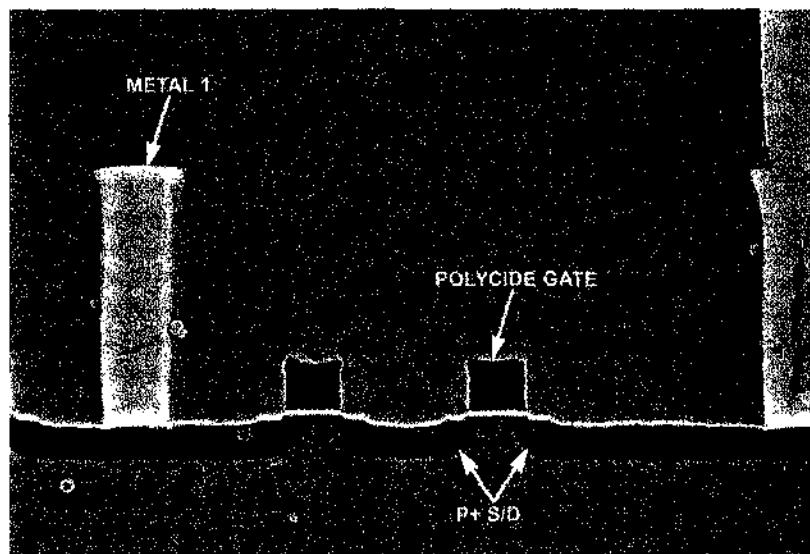


FIGURE 11.6 Aspect Ratio for Film Deposition



High Aspect Ratio Gap
(Micrograph Courtesy of Integrated Circuit Engineering)

tend to have more defects such as pinholes, which leads to less mechanical strength. It is desirable for thin films to have good surface flatness to minimize steps and cracking.

Film Purity and Density ■ High film purity means that the film does not have any unwanted chemical elements or molecules, which is important if the film is to function properly. This includes the avoidance of contaminants such as mobile ionic contaminants (MICs) and particulates. The occurrence of film contaminants such as hydrogen can degrade film properties. The density of the film is an important indicator of film quality because it means the film is free of pinholes and voids. A porous film has lower density and in some instances a lower refractive index than a nonporous film.

Stoichiometry ■ A desirable film will have uniform composition. As chemicals undergo change in a reaction, stoichiometry describes how the quantities of reactants and products are produced once the reaction ends or stabilizes. *Stoichiometry* is the ratio of one component amount to another component amount in a compound or molecule (e.g., the H₂O stoichiometric ratio is 2:1). If you know what is in a reaction, then stoichiometry tells you how much of the ingredients was needed. The chemical reactions that occur in deposition are complex and can lead to films that have a different composition than intended. A goal in deposition is to have the correct quantity of molecules in the reaction so that the deposited film approaches the ratio of the elements in the nominal chemical formula of the incoming gas.

Film Structure ■ The film structure is critical, especially related to the grain size. Materials tend to collect and grow as grains during the deposition process. If the grain size varies in a film, then it

will have varying electrical and mechanical properties that can affect the long-term reliability of the film, especially concerning electromigration (discussed in Chapter 12).⁸ The growth of a film can cause undesirable stress in the film layer that deforms the wafer substrate, leading to film cracking, delamination or void formation. An example of this condition is hydrogen contamination in nitride film deposition, which leads to a compressive stress.⁹ The objective during film deposition is always to minimize the amount of stress.

Film Adhesion ■ Film adhesion to the substrate material is important for thin films to avoid delamination and cracking. A cracked film can cause surface roughness and allows contamination to pass through the film. For insulating films, cracks can lead to electrical shorts or current-leakage paths. Adhesion of a film to a surface is determined by the cleanliness of the surface and also by the type of material to which the film can alloy. Metals such as chromium, titanium, and cobalt are often useful for their adhesion properties (see Chapter 12). Good adhesion of films in multilevel metallization is critical to maintaining the overall electrical and mechanical integrity of the structure. This is true in the as-deposited condition and after subsequent processing.

Film Growth

A deposited film grows in three distinct stages (see Figure 11.7).¹⁰ The first stage is *nucleation* where clusters of stable nuclei are formed. This stage occurs when the first few atoms or molecules of the reactants combine to form isolated patches of film that attach to the wafer surface. Nucleation occurs directly on the wafer and is essential to further thin film growth. The second stage is the *nuclei coalescence* into clusters, also referred to as *island growth*. These randomly oriented island clusters grow based on surface mobility and the density of the cluster. The island clusters continue to grow and eventually develop into the third stage, a *continuous film* where they meet and form a solid sheet that spreads across the substrate surface.

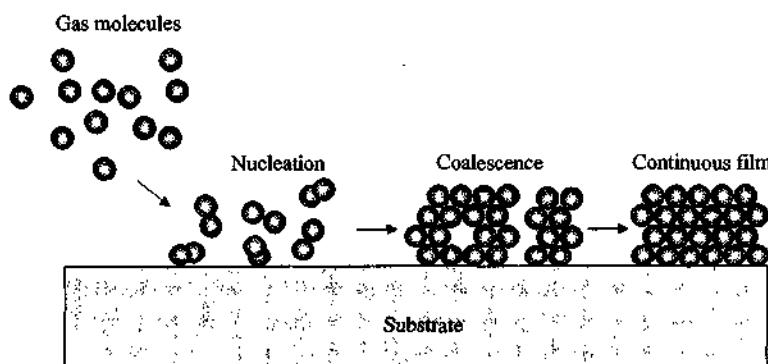


FIGURE 11.7 Stages of Film Growth

The size of the individual cluster, before meeting a neighboring cluster, is related to the surface mobility of the reacting species and the density of the nuclei. A high surface mobility and/or low nucleation rate promote the formation of relatively large clusters. This condition generally leads to polycrystalline films with short-range crystal order. On the other hand, low surface mobility combined with a high nucleation rate leads to amorphous film growth with no short-range order.¹¹ Low deposition temperatures usually favor amorphous films because the low thermal energy slows the surface mobility of the reacting materials.

Deposited films can be amorphous, polycrystalline or single crystalline.¹² Semiconductor films are used in all three forms. Insulator or metal films are usually amorphous or polycrystalline. Silicon deposited on an oxide layer is polycrystalline, as in the polysilicon used in the gate structure. Epitaxial single-crystal films are necessary for reliable semiconductor properties. To obtain a single-crystal deposited film, it is deposited on a single-crystal wafer substrate.

Film Deposition Techniques

Deposition of materials onto the wafer surface places a continuous thin film on the wafer. The film material derives from an external source, which may be a gas source that produces the material to be deposited through a chemical reaction or a solid target source that requires the physical removal of the material. With the increased complexity of semiconductor processing in the ULSI era, the number of deposited films on a wafer is high. As a result, the deposition of thin films is one of the most common processes in the semiconductor industry.

There are different techniques used in wafer fabrication to deposit films on a substrate. The major deposition methods shown in Table 11.1 are generally grouped into chemical or physical processes.

TABLE 11.1 Techniques of Film Deposition

Chemical Processes		Physical Processes		
Chemical Vapor Deposition (CVD)	Plating	Physical-Vapor Deposition (PVD or Sputtering)	Evaporation	Spin-On Methods
Atmospheric Pressure CVD (APCVD) or Sub-Atmospheric CVD (SACVD)	Electrochemical deposition (ECD), commonly referred to as electroplating	DC Diode	Filament and Electron Beam	Spin-on-glass (SOG)
Low Pressure CVD (LPCVD)	Electroless Plating	Radio Frequency (RF)	Molecular Beam Epitaxy (MBE)	Spin-on dielectric (SOD)
Plasma Assisted CVD: • Plasma Enhanced CVD (PECVD) • High Density Plasma CVD (HDPCVD)		DC Magnetron		
Vapor Phase Epitaxy (VPE) and Metal-organic CVD (MOCVD)		Ionized metal plasma (IMP)		

Adapted from F. Barlow III, A. Elshabini-Riad, and R. Brown, "Film Deposition Techniques and Processes," *Thin Film Technology Handbook*, eds. A. Elshabini-Riad and F. Barlow III (New York: McGraw-Hill), pp. 1-2.

This chapter reviews the dielectric thin films deposited with chemical vapor deposition (CVD), epitaxy, and spin-on-dielectric (SOD) methods. All metallization is discussed in Chapter 12, including chemical vapor deposition (CVD) for metals, sputtering, electroplating, and evaporation.

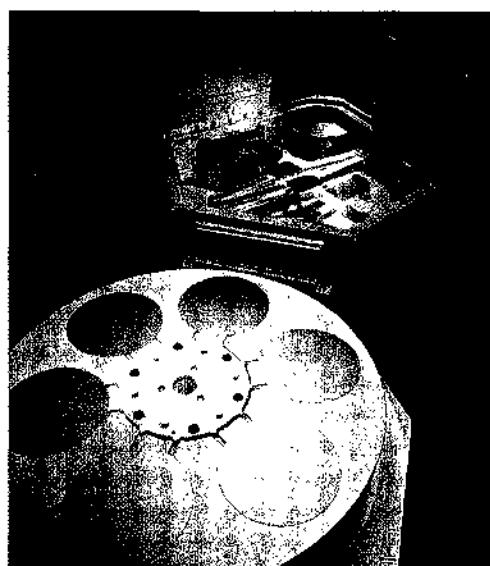
CVD is commonly used for depositing dielectric and metal films. Spin-on-dielectric (SOD) is used for applying liquid dielectric films followed by a high-temperature curing process. Sputtering, or PVD, is the most common method in wafer fabrication for applying metals such as aluminum. Electroplating, widely used in thin film deposition for magnetic recording heads in disk drives, has not been used in silicon wafer fabrication. It has recently become the most promising process for depositing copper metallization. Evaporation is the traditional method for applying a metal layer, but its poor gap-fill property led to its demise at the start of the VLSI era, when it was replaced by sputtering.

CHEMICAL VAPOR DEPOSITION

Chemical vapor deposition (CVD) is the process of depositing a solid film on the wafer surface through a chemical reaction of a gas mixture. The wafer surface or its vicinity is heated in order to provide additional energy to the system to drive the reactions. The essential aspects of CVD are:

1. Chemical action is involved, either through chemical reaction or by thermal decomposition (referred to as *pyrolysis*).
2. All material for the thin film is supplied by an external source.
3. The reactants in a CVD process must start out in the vapor phase (as a gas).

CVD deposition by use of a chemical reaction occurs when chemical compounds are mixed and then reacted in a deposition chamber, referred to as a *reactor*. The atoms or molecules deposit on the wafer surface and form the film.



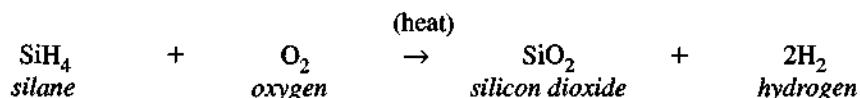
Chemical Vapor Deposition Tool
(Photo courtesy of Novellus Systems, Inc.)

CVD Chemical Processes

There are five basic chemical reactions that can be used in CVD:¹³

1. **Pyrolysis:** a compound dissociates (breaks bonds, or decomposes) with the application of heat, usually without oxygen.
2. **Photolysis:** a compound dissociates with the application of radiant energy that breaks bonds.
3. **Reduction:** a chemical reaction occurs by reacting a molecule with hydrogen.
4. **Oxidation:** a chemical reaction of an atom or molecule with oxygen.
5. **Reduction-oxidation (redox):** a combination of reactions 3 and 4 with the formation of two new compounds.

Among these five general types of reactions, there are many specific CVD reactions used to deposit films on a wafer substrate. The choice of a particular reaction is usually defined by parameters such as deposition temperature (which must be acceptable for the wafer materials), the film properties, and manufacturing issues. An example of a specific CVD reaction is the deposition of a silicon dioxide film by an oxidation reaction of silane with oxygen. This reaction produces silicon dioxide that is deposited on the heated wafer surface, with a by-product of hydrogen.



CVD Reaction

The CVD chemical reaction that deposits the solid film takes place on (or very close to) the wafer surface. This is a *heterogeneous reaction* (also referred to as *surface catalyzed*). Some reactions also occur in the gas phase above the wafer surface, which is termed a *homogeneous reaction*. Homogeneous reactions are undesirable because they form gas-phase clusters of the depositing material, which results in poorly adhering, low-density films with higher defects.¹⁴ In CVD chemical reactions, heterogeneous reactions at the wafer surface are favored for the production of high-quality films.

CVD Reaction Steps ■ The fundamental CVD reaction has eight major steps that explain the reaction mechanism. The steps are summarized in the following list and are shown in Figure 11.8:¹⁵

1. **Gas transport to deposition zone:** Mass transport of gas in the main gas flow region from the reactor inlet to the deposition zone of the wafer.
2. **Formation of film precursors:** Gas-phase reactions leading to the formation of the film precursors (initial atoms and molecules that will constitute the film) and by-products.
3. **Film precursors at wafer:** Mass transport of the film precursors to the wafer growth surface.
4. **Precursor adsorption:** Adsorption (binding) of film precursors to the surface.
5. **Precursor diffusion:** Surface diffusion of film precursors to the film growth sites.
6. **Surface reactions:** Surface chemical reactions leading to film deposition and by-products.
7. **By-product removal from surface:** Desorption (removal) of the by-products of the surface reactions.
8. **By-product removal from reactor:** Mass transport of the by-products in the bulk gas-flow region away from the deposition zone and towards the reactor exit.

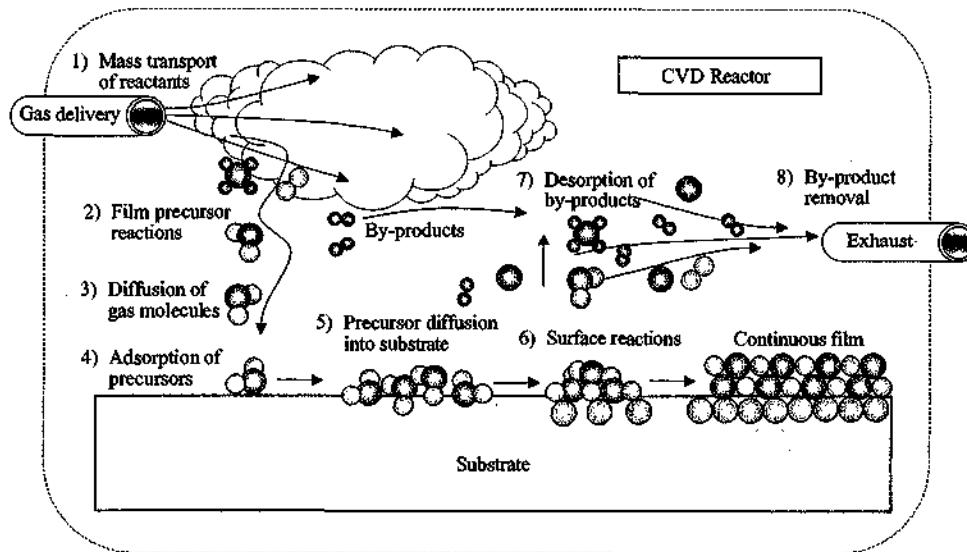
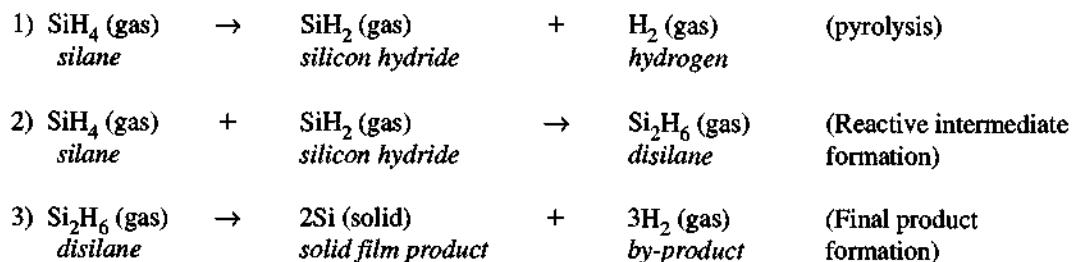


FIGURE 11.8 Schematic of CVD Transport and Reaction Steps
SEMATECH, "Deposition Processes," in *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 6.

Adsorption is the chemical binding that occurs during deposition, causing the gaseous atoms and molecules to chemically attach to the solid wafer surface. *Desorption* is the removal or release of by-products from the wafer surface reactions. In chemical reactions, *species* describes a particular chemical substance that is an atom, ion, or molecule. In gas-phase reactions, there are often intermediate reactions called *precursors* that form a gas species that does not contain the original gas components. In CVD, gas precursors are transported to the wafer surface for adsorption and reaction. For example, for the following three reactions, the first gas-phase reaction shows that when

silane decomposes, a silicon hydride precursor is produced. This precursor then reacts with silane to form disilane. During the intermediate CVD reaction, the silicon hydride is adsorbed onto the wafer surface along with silane. Disilane then decomposes to form the final silicon solid-film product.



Rate Limiting Step ■ The length of time it takes for the CVD reaction to occur is important for productivity in manufacturing. The surface reaction rate increases with increasing temperature. Since CVD reaction steps are sequential, the slowest step defines the bottleneck in the process. In other words, the step that occurs at the slowest rate will determine the rate of deposition.¹⁶

The rate of a CVD reaction cannot proceed more rapidly than the mass-transport rate at which reactant gases are supplied from the main gas stream to the wafer substrate. This is true no matter how high the temperature is increased. This situation is referred to as a *mass-transport limited* deposition process. A mass-transport limited process is only weakly dependent on temperature.¹⁷ This means that regardless of the temperature, there is insufficient reactant gas being supplied to the wafer surface to speed up the reaction. Because of this, a high-temperature, higher pressure CVD process that wants to proceed rapidly is usually mass-transport limited—it cannot supply enough reactant gas to the substrate because of diffusion or physical limitations.

At lower reaction temperatures and pressure, the surface reaction rate is reduced because there is less energy available to drive the surface reaction. Eventually the arrival rate of reactants will exceed the rate at which they are consumed by the chemical reaction process at the wafer surface. In this case, the deposition rate is *reaction-rate limited*.¹⁸ This condition is similar to pouring too much liquid through a funnel—the liquid will not flow faster than the funnel permits. Even though more reactant gas is supplied, the reaction will not speed up because the low temperature provides insufficient energy. A reaction-rate limited CVD process is also referred to as *kinetically controlled*, meaning the diffusivity of the reactants is not as critical as the kinetic energy of the reactants at the surface. For reaction-rate limited processes, it is important to maintain a uniform temperature in order to achieve a uniform deposition rate across the wafer surface.

CVD Gas-Flow Dynamics ■ The gas-flow dynamics for CVD are important for uniform film deposition. Gas flow refers to the physical transport of reactant gases to the CVD sites on the wafer (see Figure 11.9 on page 268). At the molecular level, this requires having a sufficient number of molecules at the right place and the right time. This condition is necessary for a stoichiometric reaction to occur between the reactants and products.

The principal factors for CVD gas flow are the mass transport of the reactant gases from the main gas stream to the wafer surface and the chemical reaction rate at the surface. It is assumed that the dominant transfer mechanism from the gas phase to the surface is diffusion.¹⁹ Considering the gas flow right at the wafer surface during a CVD reaction process, there is zero or near-zero gas flow. This condition creates a *boundary layer* of gas flow that increases from the zero gas-flow velocity at the wafer surface (due to frictional forces) to some typical gas flow farther from the surface.²⁰ These dynamics are shown in Figure 11.10 on page 268. The gas flow farther from the surface reaches some average gas-flow velocity that represents the average velocity of the main gas stream in the CVD reactor. If the boundary layer is narrow, then it can be treated as not moving at all near the wafer surface. In this case, it is referred to as a *stagnant layer*. The size of the stagnant layer will influence the design of the different types of CVD reactors considered later in this chapter.

Pressure in CVD ■ If a CVD reaction occurs at low pressure, then the diffusivity of the gas species increases significantly through the boundary layer to reach the surface. This diffusivity

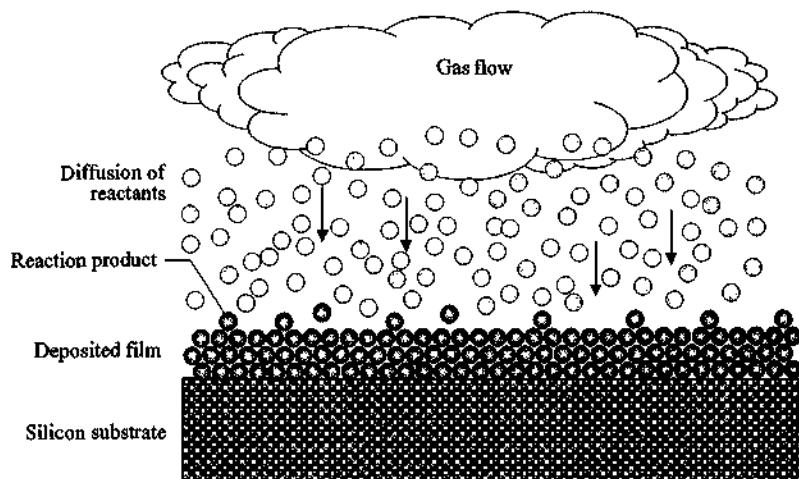
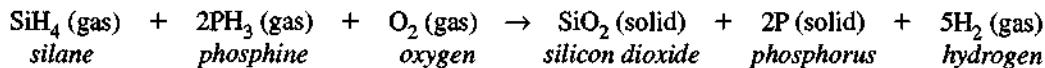


FIGURE 11.9 Gas Flow in CVD

increases the transport of the reactants to the substrate surface (and also increases the removal of by-products away from the substrate). The net effect of lower pressure in a CVD reactor is that the reactants reach the surface more quickly. In this case, the rate-limiting step becomes the surface reaction. At lower pressures, the CVD process is reaction-rate limited. This means that wafers can be stacked vertically and at very close spacing in the reactor since the transport of the reactants from the gas stream to the wafer is not limiting the process.

Doping During CVD ■ The introduction of dopants into SiO_2 during CVD deposition is desirable because it leads to important benefits for wafer fabrication. For instance, the introduction of phosphine (PH_3) to the gas stream during silicon dioxide deposition will lead to the formation of *phosphosilicate glass* (PSG). The chemical reaction is:



Within the glass, the phosphorous is in the form of a phosphorus pentoxide (P_2O_5), making the glass a dual compound as it codeposits with silicon dioxide. The P_2O_5 content in films is limited to about 4% by weight for films that are left permanently on the wafer because PSG becomes increasingly hygroscopic (water-absorbing).

PSG can be deposited at 600 to 650°C using high-density plasma CVD (HDPCVD) and has recently become popular for the first interlayer dielectric (ILD-1) layer because of its low deposition temperature, relatively planar surface, and excellent gap-fill characteristics. Incorporation of P_2O_5 into silicon dioxide films leads to a reduction in the film stress, which improves the film integrity. The dopant increases the moisture-barrier property of glass. Another benefit is that PSG layers are effective in serving as a getter to immobilize ionic contaminants. The ions become attached to the phosphorus and are thus restrained from diffusing through the film to the wafer surface.

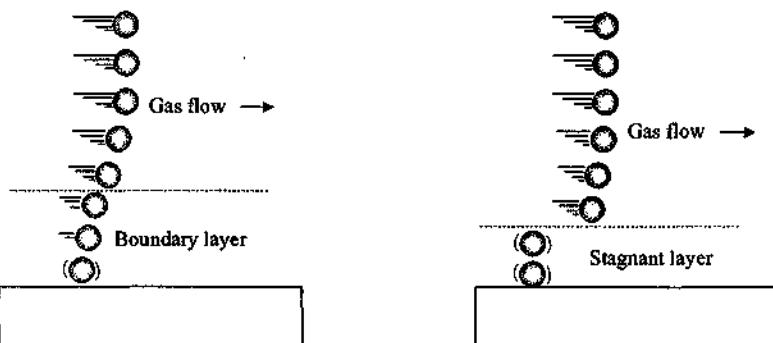


FIGURE 11.10 Gas-Flow Dynamics at the Wafer Surface

Borosilicate Glass. If diborane (B_2H_6) is used instead of phosphine (PH_3), then the result is *borosilicate glass (BSG)*. BSG requires reflow at a high temperature (e.g., 1000°C) to smooth out sharp steps in the wafer surface and densify the film. However, a high-temperature reflow is undesirable for the wafer thermal budget. BSG also has the disadvantage of not being a good barrier to mobile ionic contaminants (MICs).²¹

Borophosphosilicate Glass. Another method for doping oxide is to incorporate about 2 to 6 % weight each of boric oxide (B_2O_3) and phosphorus pentoxide (P_2O_5) into the silicon dioxide deposition to form *borophosphosilicate glass (BPSG)*. To achieve a dense oxide with good step coverage after deposition, the oxide is reflowed at a high temperature until it softens and flows. For BPSG, the reflow temperature is 800 to 1000°C for one hour. Reflow also improves its gettering characteristics to block mobile ionic contaminants (MICs). BPSG was commonly deposited on the wafer as the ILD-1 but has been recently replaced by PSG deposited by HDPCVD to reduce the thermal budget.²²

High temperatures adversely affect the wafer thermal budget in sub-0.25 μm devices. For example, refractory metals (e.g., titanium and cobalt) used to reduce contact resistance at junctions undergo material transformations at about 600°C that increase their contact resistance to unacceptable levels. In addition, the oxide planarity achieved by a high-temperature reflow is not sufficient for reduced depth of field requirements in ULSI photolithography. For this reason, reflow of doped oxide for planarity and gap fill has been largely replaced by high-density plasma CVD to achieve a stable film with good gap fill, followed by chemical mechanical planarization (CMP) to attain a flat surface (see Chapter 18).

Fluorosilicate Glass. *Fluorosilicate glass*, or FSG, is a fluorinated oxide that is being investigated as a first generation low- k dielectric for ILD deposition on 0.18 μm generation devices. By adding fluorine to silicon dioxide, there is a decrease in the dielectric constant of the material from about 3.9 for SiO_2 to about 3.5 for FSG.

To form an FSG film, silicon tetrafluoride (SiF_4) is added to the silane (SiH_4) and O_2 gases during the CVD reaction. A potential problem with using FSG is that the fluorine bonds are not always stable and can lead to corrosion, limiting the fluorine content to about 6%. If fluorine ions encounter moisture, hydrofluoric acid (HF), which etches oxide, is created. A method to avoid this problem is to remove weakly bonded fluorine atoms through the introduction of hydrogen during the deposition reaction.²³

Oxide Versus Silicon Doping. Note that the doping of silicon dioxide is not the same as doping of silicon. When silicon is doped, there is a donation or acceptance of electrons between the dopant and silicon, within a single crystal lattice structure. Deposited oxide is an amorphous crystal structure and the dopants do not act as donors or acceptors of electrons. Doping is done to modify a physical characteristic of SiO_2 , such as improving its ability to getter mobile ions.

CVD DEPOSITION SYSTEMS

There are different types of production systems for the deposition of dielectric and metal film layers. Some systems use large batch processing to perform CVD deposition on a large quantity of wafers at one time. However, the most advanced deposition systems currently in use for wafer manufacturing employ single-wafer processing in a cluster tool approach.

Table A.3 in Appendix A summarizes different gases used for CVD reactions. An example of a CVD gas is pure silane, which is used to deposit Si or SiO_2 in CVD systems. Silane reacts with air to form solid particles, causing particle contamination with the silane gas lines and leading to contaminated pipes or combustion. The gas line piping must be carefully installed to avoid any leaks. Many of the gases used in CVD film deposition are toxic. These gases are classified in Table A.3 based on four different safety hazards: pyrophoric (ignites spontaneously in air at or below 54.5°C), poisonous, corrosive, and dangerous combinations of gases, such as silane with hydrogen.

CVD Equipment Design

CVD has a range of equipment reactor designs, with each producing slightly different types of film quality. CVD reactors are broadly categorized based on the reaction chamber pressure regime used during operation: atmospheric-pressure CVD (APCVD) reactors and reduced-pressure reactors. The reduced-pressure CVD reactors have two general types. First, there are low-pressure CVD (LPCVD) reactors where the energy input is thermal. Second, there are plasma-assisted CVD reactors, either plasma-enhanced CVD (PECVD) or high-density plasma CVD (HDPCVD), where the energy is partially supplied by a plasma as well as thermally. Figure 11.11 depicts the different CVD reactor types.

CVD Reactor Types	Atmospheric	Low-pressure	Batch	Single-wafer
Hot-wall	✓	✓	✓	
Cold-wall	✓	✓	✓	✓
Continuous motion	✓		✓	
Epitaxial	✓		✓	
Plenum	✓		✓	
Nozzle	✓		✓	
Barrel	✓		✓	
Cold-wall planar		✓	✓	✓
Plasma-assisted		✓	✓	✓
Vertical-flow Isothermal		✓	✓	✓

FIGURE 11.11 CVD Reactor Types

CVD Reactor Heating ■ A major distinction in CVD reactors is whether they are hot-wall or cold-wall reactors. A *hot-wall reactor* employs a heating method that heats not only the wafer but also the wafer holder (referred to as a *susceptor*) and the walls of the reactor. Hot-wall reactors have film-forming reactions on both the reactor chamber walls and substrate, requiring frequent cleaning or in situ cleaning to minimize particle contamination. A resistance heater surrounding the reactor tube is an example of a hot-wall reactor. A *cold-wall reactor* heats only the wafers and susceptor, which is where film is deposited. The walls of the reactor are cold and therefore do not have sufficient energy to allow a deposition reaction to occur. RF induction heating or infrared lamps mounted in the reactor are examples of this type of reactor. Localized heating at the wafer minimizes particle formation in the reactor.

CVD Reactor Configuration ■ The geometry of the reactor design depends closely on the pressure attained during the deposition process. Atmospheric-pressure reactors operate in the mass-transport limited region, so they must be designed so that an equal amount of reactant gases are delivered to each wafer. To achieve this condition, wafers are usually laid flat on a horizontal surface. A drawback to this approach is that wafers are susceptible to contamination from falling particles.

Low-pressure reactors (LPCVD) are reaction-rate limited. This means that they can be vertically stacked at close spacing with a large number of wafers per run, since adequate reactant gas is always supplied to the wafer surface. Quartz wafer holders (boats) hold up to 200 wafers. At the same time, LPCVD reactors must have precise temperature control because of their limitation from the reaction rate.

CVD Reactor Summary ■ The different types of CVD processes and their principal characteristics are shown in Table 11.2.²⁴

TABLE 11.2 Types of CVD Reactors and Principal Characteristics

Process	Advantages	Disadvantages	Applications
APCVD (Atmospheric Pressure CVD)	Simple reactor, fast deposition, and low temperature.	Poor step coverage, particle contamination, and low throughput.	Low-temperature oxides (both doped and undoped).
LPCVD (Low Pressure CVD)	Excellent purity and uniformity, conformal step coverage, and large wafer capacity.	High temperature, low deposition rate, more maintenance intensive, and requires vacuum system.	High-temperature oxides (both doped and undoped), silicon nitride, polysilicon, W, and WSi_2 .
Plasma-Assisted CVD: • Plasma-Enhanced CVD (PECVD) • High-Density Plasma CVD (HDPCVD)	Low temperature, fast deposition, good step coverage, and good gap fill.	Requires RF system, higher cost, stress is much higher with a tensile component, and chemical (e.g., H_2) and particle contamination.	High aspect ratio gap fill, low-temperature oxides over metals, ILD-1, ILD, copper seed layer for dual damascene, and passivation (nitride).

APCVD (Atmospheric Pressure CVD)

The first type of chemical vapor deposition used in the semiconductor industry is *atmospheric pressure CVD (APCVD)*.²⁵ As previously discussed, APCVD generally operates in the mass-transport limited regime. At any given time, there may not be sufficient gas molecules present at the wafer surface for a reaction to occur. Therefore, the reactor must be designed to have optimum reactant gas flow to every wafer in the system. The basic system design should never permit the reaction to slow down because insufficient reactant gas is available. Since the reactor is at atmospheric pressure, the reactor design can be simple and allows for high deposition rates.

Two different types of continuous-processing APCVD systems are shown in Figure 11.12 on page 272. These equipment designs use a belt or conveyor to carry the wafer samples through the reactor gases, which flow through the center of the reactor.

Continuous-processing APCVD systems have high equipment throughput, good uniformity, and the capability to process large-diameter wafers. APCVD systems do have problems with high gas consumption and often need frequent reactor cleaning. Since film deposition also takes place on the conveyor as well as the wafer, the belt transport system must be cleaned (sometimes it is cleaned in situ, or during use). APCVD deposited films often exhibit poor step coverage.

Silicon Dioxide ■ The most common use of APCVD is in the deposition of silicon dioxide (SiO_2) and doped oxides (e.g., PSG, BPSG, and FSG, as discussed earlier in this chapter). These films have been used traditionally as an interlayer dielectric (ILD), as a protective overcoat, or to planarize (smoothen) a nonuniform surface.

SiO_2 Deposition With Silane. SiO_2 is deposited by oxidizing *silane* (SiH_4) with oxygen. Pure silane is a highly pyrophoric, unstable gas that burns on exposure to air. For this reason, it is commonly supplied in a low dilution (typically 2 to 10% by volume) in argon or nitrogen in order to make it safer to handle. This reaction can be done at a low temperature of 450 to 500°C, which is an advantage for depositing SiO_2 on the aluminum metal lines as an ILD.²⁶ However, due to the small mean free path and poor surface migration through the boundary layer, this method has poor step coverage and poor gap-fill characteristics. This makes it unacceptable for critical ULSI applications.

SiO_2 Deposition With TEOS-Ozone. A common APCVD application is to deposit oxide by reacting TEOS, which is tetraethylorthosilicate ($Si(C_2H_5O)_4$) or tetraethoxysilane, with ozone (O_3). TEOS is an organic liquid precursor. It usually uses a carrier gas, typically nitrogen, bubbled through it for delivery of the TEOS gas mixture to the reactor. There are also flow controllers for delivering the TEOS source in liquid state to the reactor. Ozone is a triatomic oxygen molecule that is much more reactive than oxygen; therefore, the process can be done at low temperatures (e.g., 400°C) without plasma. Because the decomposition of TEOS is induced by O_3 without a plasma, the reaction takes place at atmospheric pressure (APCVD) of 760 torr or subatmospheric pressure (SACVD) at about 600 torr. There appears to be no significant difference between APCVD and

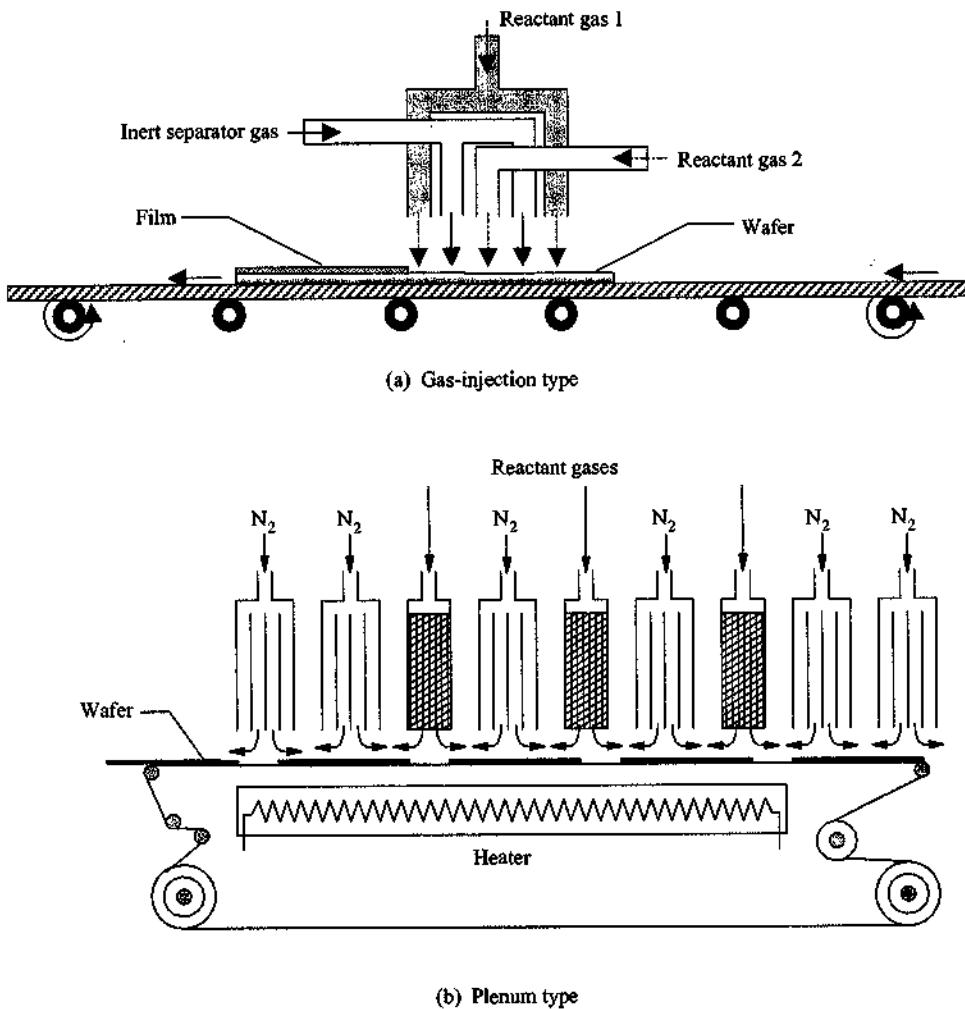
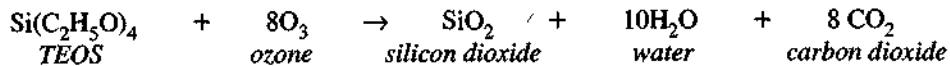


FIGURE 11.12 Continuous-Processing APCVD Reactors

SACVD.²⁷ The deposited oxide film has improved profiles over steps, is conformal, and has excellent electrical characteristics as a dielectric.²⁸ The reaction equation between TEOS and O₃ to obtain an oxide is:



An APCVD TEOS-O₃ as-deposited SiO₂ film is porous and often requires a reflow step to remove moisture as well as densify the film. This reflow adds a process step and reduces the thermal budget.²⁹ Doping uniformity is also a concern, as this can lead to problems for nonuniform removal rates during chemical mechanical planarization. The primary advantage of APCVD TEOS-O₃ is the improved step coverage for high-aspect ratio gaps for applications such as shallow trench isolation (see Figure 11.13). Another advantage to APCVD TEOS-O₃ is that it is just a basic thermal CVD process to deposit oxide, which could avoid wafer surface damage and wafer corner damage at steps. This type of damage is possible from high-density plasma CVD (discussed later in this chapter). An aspect of TEOS-O₃ films is that they are usually deposited in combination with other oxide film layers (e.g., PECVD oxide) for reasons such as reducing the TEOS-O₃ tensile stress on a thick film or to reduce sensitivity with the underlying material.³⁰

Doped Oxides. APCVD oxides are usually doped with chemicals such as phosphorus and boron (discussed earlier in this chapter). If SiO₂ is undoped, it is referred to as *undoped silicate glass (USG)* or *undoped oxide (UDOX)*. Phosphorus doped oxide is phosphosilicate glass (PSG). In the traditional deposition process, a high-temperature reflow (as high as 950°C for 15 to 30 minutes) heats and softens the doped oxide, causing it to planarize the surface (see Figure 11.14).

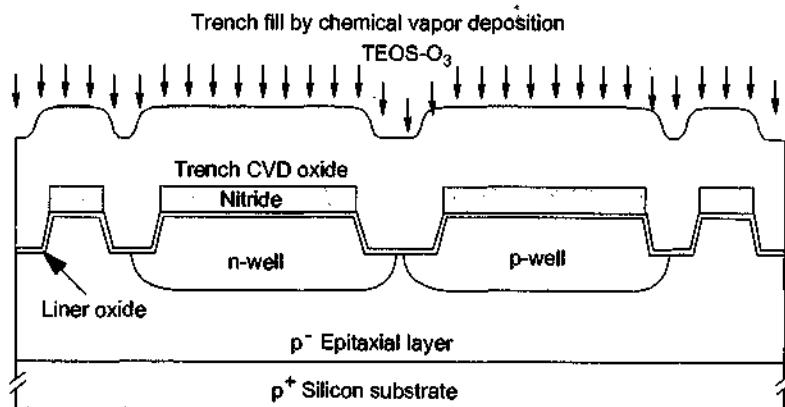


FIGURE 11.13 Improved Step Coverage of APCVD TEOS- O_3

Silane-based borophosphosilicate, or BPSG films, are deposited at low temperatures (400 to 450°C) and then immediately densified and stabilized at about 800°C for one hour. A smoother surface makes it easier to deposit and pattern the next film. Note that for the ULSI device technology, reflow has been replaced with chemical mechanical planarization (CMP) to planarize the surface. CMP is discussed in Chapter 18.

LPCVD (Low Pressure CVD)

Low pressure CVD (LPCVD) systems are more common now than APCVD because of their lower cost, higher production throughput, and superior film properties. LPCVD operates at a medium vacuum (about 0.1 to 5 torr), and employs temperatures between 300 and 900°C.³¹ Conventional oxidation type furnaces (horizontal or vertical) and multichamber cluster tools can be used for LPCVD processing.

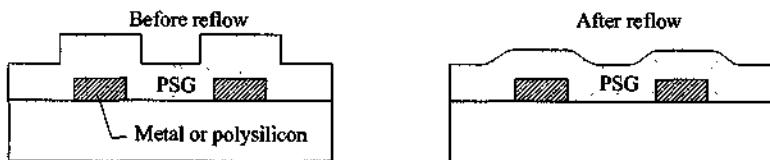


FIGURE 11.14 Planarized Surface after Reflow of PSG

LPCVD reactors typically operate in the reaction-rate limited regime. In this reduced-pressure regime, the diffusivity of the reactant gas molecules increases so that the mass-transfer of the gas to the wafer no longer limits the rate of the reaction. Because of this transfer state, the gas-flow conditions inside the reactor are not important, permitting the reactor design to be optimized for high wafer capacity (e.g., wafers can be closely spaced). Films are uniformly deposited on a large number of wafer surfaces as long as the temperature is tightly controlled.

The boundary layer in a LPCVD reactor is different from APCVD because it extends farther from the wafer surface due to the low-pressure condition (see Figure 11.15 on page 274). The boundary layer has a low molecular density, which permits incoming gas molecules to easily diffuse through this layer, exposing the wafer surface to more molecules than it can possibly consume. This condition explains why an LPCVD reactor operates in the reaction-rate limited regime. It is the reaction rate that is the limiting factor in the deposition rate, not the reactant supply. Furthermore, there are many collisions during LPCVD, so that material strikes the wafer in an undirected manner. This action is beneficial for conformal film coverage on high-aspect ratio steps and trenches. As a result, step coverage is usually good with LPCVD.³²

The LPCVD reactor design favors the hot-wall reactor type so that uniform temperature control is achievable over a large operating length. A representative reaction chamber is shown in Figure 11.16 on page 274.

Since LPCVD reactors are often hot wall, particles deposit on the reactor wall. These deposits are minimized by reducing the partial pressure of the gas-phase reactants, which leads to fewer

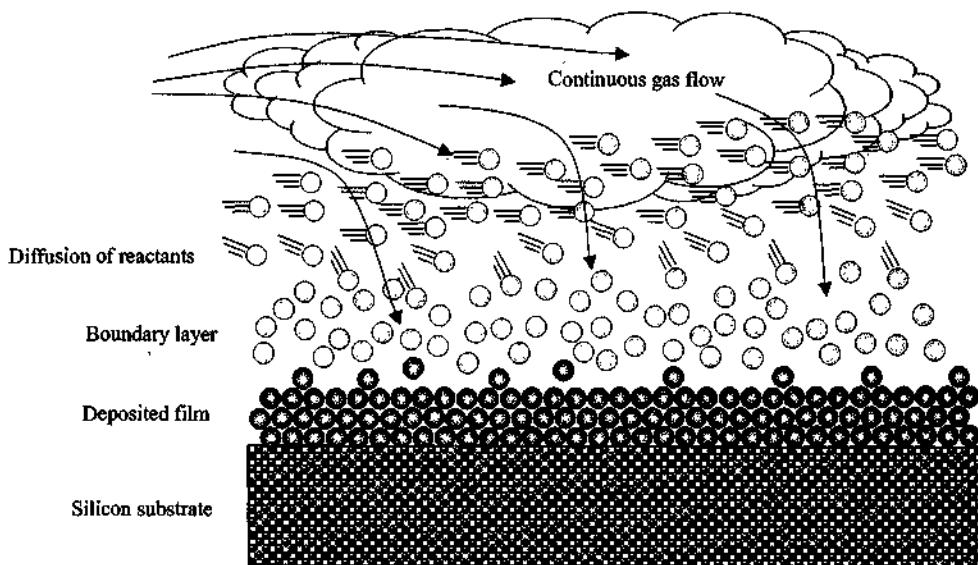


FIGURE 11.15 Boundary Layer at Wafer Surface

gas-phase reactions. For this reason, LPCVD reactors are usually operated at lower growth rates than APCVD systems.³³ Hot-wall reactors require periodic reaction chamber cleanup of particles by routine maintenance.

The traditional method of tube cleaning was to remove a dirty quartz tube, install a previously cleaned tube, and clean the dirty tube for future use. Cleaning involved manually rolling the tubes half-submerged in an acid bath such as aqueous HF. For production and safety reasons, an in situ clean is desirable. One in situ method is the use of plasma-generated fluorine gases that react with the solid residue in the reactor, forming volatile reaction products that are pumped from the system. Another in situ approach is to use chlorine trifluoride (ClF_3) with elevated temperature to activate a thermal clean. In situ tube cleaning for LPCVD reduces equipment downtime, lowers particle counts, and minimizes personnel exposure to chemicals.³⁴

LPCVD reactors with a large number of wafers (e.g., 150 to 200 wafers) experience reactant depletion as gases are transported along the reaction chamber. This depletion leads to a reduction in growth rate. Adjusting the reactor temperature so that it increases slightly from inlet to outlet (typically by 25 to 50°C) compensates for this temperature nonuniformity.

Silicon Dioxide ■ There are numerous applications for LPCVD oxides (doped and undoped) in ULSI multilevel metallization. LPCVD oxides are used for interlayer dielectric (ILD), shallow trench isolation oxide fill, and sidewall spacers.

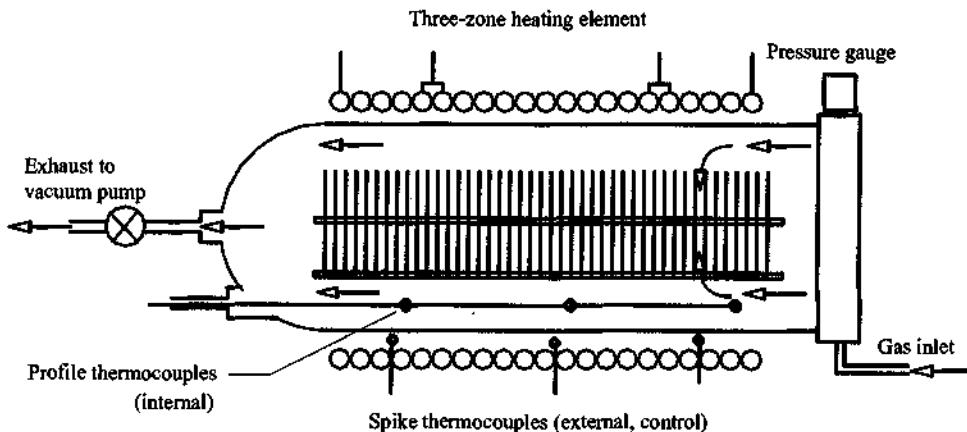


FIGURE 11.16 LPCVD Reaction Chamber

***SiO₂* With TEOS.** A common LPCVD method for depositing silicon dioxide is the pyrolysis (decomposition) of TEOS, with or without oxygen, at low pressure and at a temperature between 650 to 750°C. This method is sometimes referred to as LPTEOS, for low-pressure TEOS. LPTEOS yields very good oxide conformality due to rapid surface diffusion of gas molecules. A liquid TEOS source is used by bubbling a carrier gas through it (e.g., N₂, O₂, or He). The liquid source is heated by its own independent temperature source. The concentration of the liquid source vapor entering the reactor is controlled by the carrier gas-flow rate and liquid source temperature. A schematic of a typical TEOS deposition system is shown in Figure 11.17. The film growth rate for LPCVD oxides is significantly slower (100-150 Å/minute) than for APCVD films.³⁵

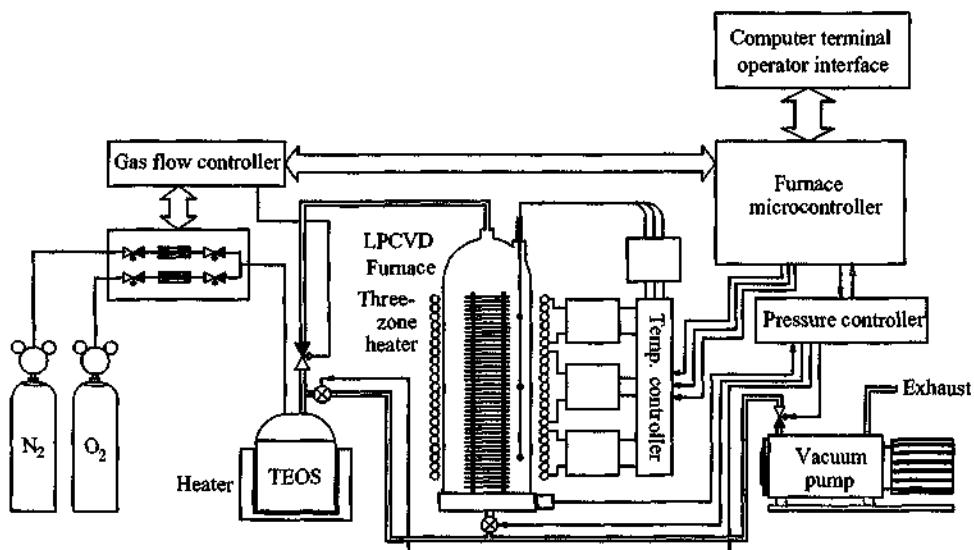
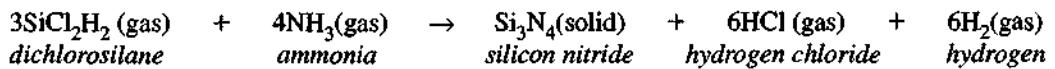


FIGURE 11.17 Oxide Deposition with TEOS LPCVD

***SiO₂* With SiH₄.** SiO₂ can also be deposited by LPCVD by oxidizing silane (SiH₄) at a relatively low temperature of about 450°C. Similar to APCVD oxide with SiH₄, this process has poor step coverage. A higher temperature deposition is done with dichlorosilane (SiH₂Cl₂ or DCS) and nitrous oxide (N₂O) at about 900°C to produce a good quality SiO₂ film. However, the high temperature is undesirable for the wafer thermal budget.

Silicon Nitride ■ Silicon nitride (Si₃N₄) is a film often used as the final wafer passivation layer because it provides good protection against the diffusion of impurities and moisture. It achieves excellent step coverage with high conformal coverage with LPCVD deposition. Silicon nitride is also used as a mask material (referred to as a hard mask), such as in the shallow trench isolation (STI) process. Silicon nitride has a high dielectric constant (i.e., a *k* value of 6.9, versus about 3.9 for CVD SiO₂), which makes it undesirable for the ILD dielectric due to the resultant high capacitance between conductor layers.

Reacting dichlorosilane (SiCl₂H₂) and ammonia (NH₃) in an LPCVD reactor produces silicon nitride at a reduced pressure and temperature between 700 and 800°C. The chemical reaction is:



The important variables that affect the properties of silicon nitride in an LPCVD process are total pressure, reactant concentrations, deposition temperature, and temperature gradients. For example, increasing the total pressure and partial pressure of dichlorosilane in the reaction chamber will increase the deposition rate.

Silicon nitride can also be produced in an APCVD process by reacting silane and ammonia, but better film uniformity and higher wafer throughput is achieved in the low-pressure process.³⁶ Deposited silicon nitride is an amorphous film that often contains large amounts of hydrogen. A high ammonia ratio and low deposition temperature will increase the hydrogen content.

Polysilicon ■ Polysilicon film (also referred to as *poly-si*, or *poly*) is commonly deposited using LPCVD. Recall that the term polysilicon means that the silicon is a polycrystal, with many small single-crystal regions separated by grain boundaries. The stress in thin polysilicon films is compressive.

Doped polysilicon serves as the gate electrode in MOS devices (see Figure 11.18). Key reasons for the use of doped polysilicon in the gate structure are:³⁷

1. Ability to be doped to a specific resistivity.
2. Excellent interface characteristics with silicon dioxide.
3. Compatibility with subsequent high-temperature processing.
4. Higher reliability than possible metal electrodes (e.g., aluminum).
5. Ability to be deposited conformally over steep topography.
6. Allows for self-aligned gate process (see Chapter 12).

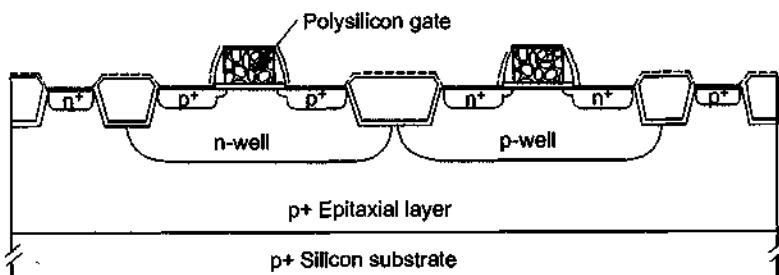
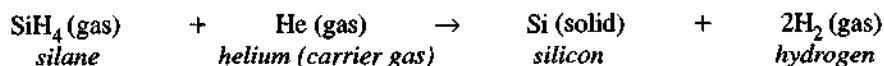


FIGURE 11.18 Doped Polysilicon as a Gate Electrode

Polysilicon is deposited in LPCVD by the pyrolysis (thermal decomposition) of silane at a temperature of 575 to 650°C. Low pressure reactors use either pure silane or 20 to 30% silane diluted with nitrogen, which is fed into the system at a pressure of 0.2 to 1.0 torr to fabricate polysilicon films. A practical deposition rate is about 100 to 200 Å/minute. The addition of diborane can enhance the polysilicon deposition rate because the diborane forms borane radicals, BH₃, which catalyze gas-phase reactions. The chemical reaction is:



The polysilicon film is usually doped *in situ* during the deposition process by adding arsine (AH₃), phosphine (PH₃), or diborane (B₂H₆) to the gas mixture. An inert gas is often accompanied to improve film uniformity. Doping can be done after deposition by ion implantation (see Chapter 17). The resistivity of the polysilicon after doping is highly dependent on parameters such as deposition temperature, dopant concentration, and annealing temperature effects on grain size.

A refractory metal is deposited over the polysilicon gate to form a *polycide* or *silicide* structure (see Chapter 12). This action is taken to reduce electrical resistivity at the polysilicon interface with the other material (e.g., tungsten) that forms the electrical connection to other device components. Lightly doped polysilicon also serves as a resistor in memory cells, capacitors, and thin-film transistors.

Silicon Oxynitride ■ Silicon nitride films that contain oxygen are referred to as *silicon oxynitride* (SiO_xN_y) and combine the advantages of oxides and nitrides. Oxynitride films also have improved thermal stability, cracking resistance, and reduced film stress when compared to silicon nitride. In general, increased oxygen in oxynitride film also decreases the film's refractive index, which makes it less reflective and a useful antireflective layer in photolithography (see Chapter 14).³⁸ Another benefit of a silicon oxynitride film is the nitrogen in the film accumulates at the silicon interface to reduce the concentration of strained Si-O bonds, potentially reducing the creation of hot electrons (undesirable charge carriers) by as much as three orders of magnitude. An oxynitride layer at the Si/SiO₂ interface of thin gate oxides has been found to be beneficial for improving the device electrical performance.

Oxynitride films are formed by different techniques, including the oxidation of silicon nitride, the nitridation of SiO_2 with ammonia (NH_3), and the direct growth of SiO_xN_y using nitrous oxide as the oxidant species. One way of forming oxynitride films is by reacting silane (SiH_4) with nitrogen dioxide (N_2O) and NH_3 . Nitride or oxynitride films are used frequently as an "etch stop" for selective etching of dielectric films.³⁹

Plasma-Assisted CVD

The third major type of CVD equipment relies on plasma energy, in addition to thermal energy, to initiate and sustain the chemical reactions necessary for CVD. The advantages of using plasma during CVD are:

1. Lower processing temperature (250 to 450°C).
2. Excellent gap-fill for high aspect ratio gaps (with high-density plasma).
3. Good film adhesion to the wafer.
4. High deposition rates.
5. High film density due to low pinholes and voids.
6. Wide range of applications due to lower processing temperature.

Film Formation ■ The plasma-assisted CVD reaction necessary to form a film occurs when RF power is used to break up gas molecules in a vacuum. The frequency of the RF power depends on the application, with typical frequencies found at 40 kHz, 400 kHz, 13.56 MHz, and 2.45 GHz (microwave power). The molecular fragments (radicals) are chemically reactive species and readily bond to other atoms to form a film at the wafer surface (see Figure 11.19). Gaseous by-products are removed by the vacuum pumping system. The wafer is usually heated in order to assist the surface reactions and reduce the level of undesirable contaminants, such as hydrogen.

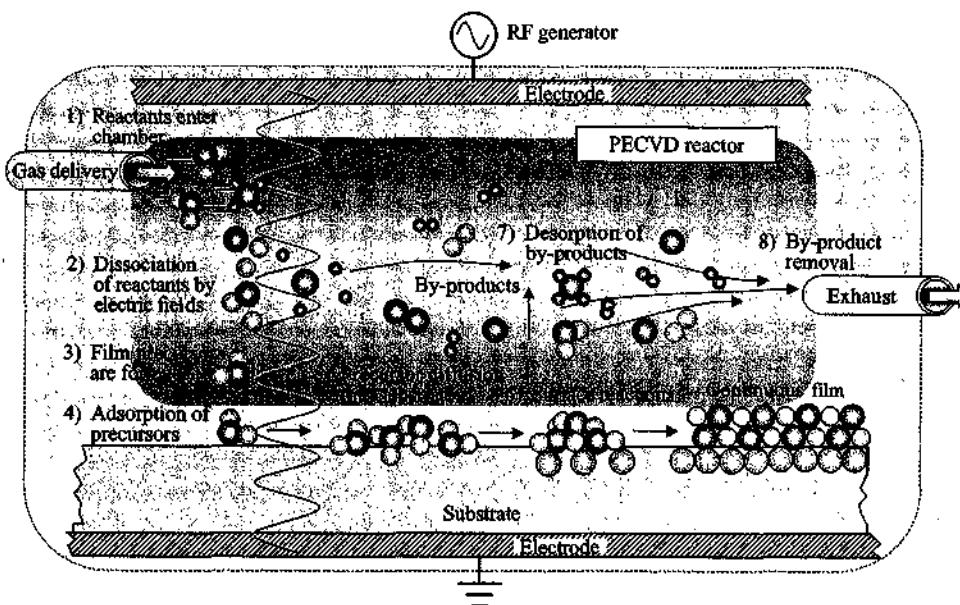


FIGURE 11.19 Film Formation During Plasma-Assisted CVD

The plasma-assisted CVD reactions at the wafer surface are very complex. The properties of plasma-deposited films depend on many variables, such as electrode configuration and separation, power level and frequency, gas composition, pressure and flow rate, and substrate temperature.⁴⁰ The essential aspect of these variables for production is embodied in the chamber design. The specific details of the intermediate reactions that occur in plasma glow discharge are not well understood. For these reasons, deposited films are often found to be not very stoichiometric. This means

that the deposited film will not necessarily have the same ratio of elements found in its nominal chemical formula.

There are two types of plasma processes used in CVD:

- ◆ Plasma-enhanced CVD (PECVD)
- ◆ High-density plasma CVD (HDPCVD)

Plasma-Enhanced CVD (PECVD) ■ The development of *plasma-enhanced CVD (PECVD)* uses plasma energy to create and sustain the CVD reaction. It is a natural follow-on to LPCVD since the system pressure for both types of CVD processes is comparable. The important difference is the much lower PECVD deposition temperature. For instance, silicon nitride (Si_3N_4) is deposited using LPCVD at 800 to 900°C, yet cannot be deposited over aluminum metallization because Al melts at 660°C. On the other hand, silicon nitride deposited with PECVD at a temperature of 350°C is suitable for this application.

PECVD is performed in a vacuum chamber between parallel conducting plates positioned several inches apart, typically with a variable gap for process optimization. A modern reactor is a multichamber cluster tool. The wafer (or wafers) is placed on the grounded bottom plate and RF power is applied to the top electrode. A plasma develops when the source gas is introduced through the gas manifold and the center of the electrodes. Exhaust gases are pumped out from the periphery of the bottom electrode. Sometimes the reactant gases are introduced at the periphery and pumped out at the center of the bottom electrode. A schematic of an overall PECVD system is shown in Figure 11.20.

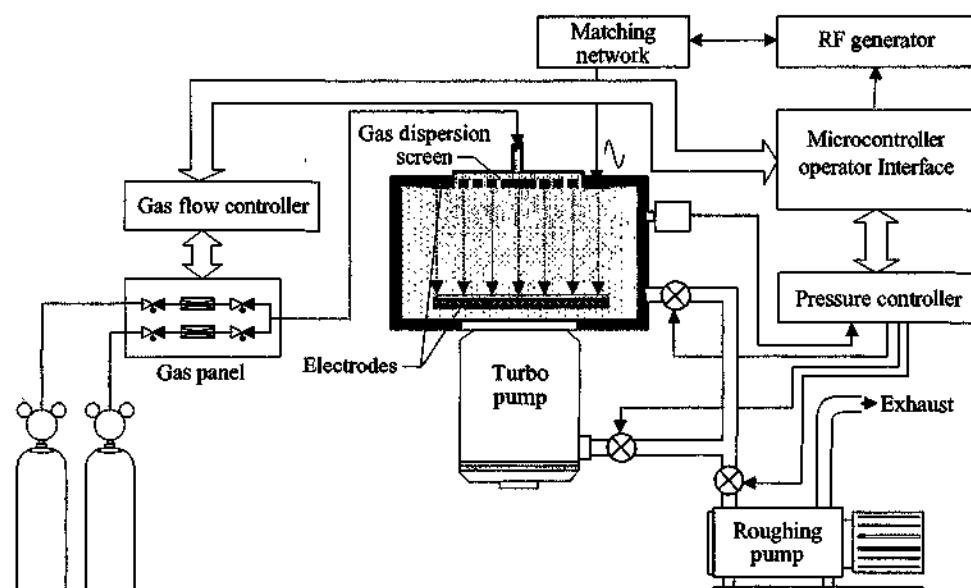
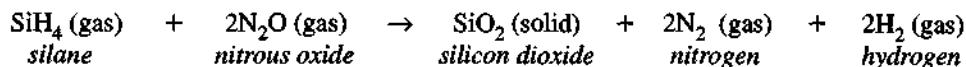


FIGURE 11.20 General Schematic of PECVD

A PECVD reactor is typically a cold-wall plasma reactor, with the wafer heated in its chuck while the remaining parts of the reactor are unheated. Deposition parameters must be controlled to ensure the temperature gradient does not affect the film thickness uniformity. Cold-wall reactors create fewer particles and require less downtime for cleaning. Deposition chambers often have in situ cleaning to reduce particles. With parallel-plate processing, there is the possibility that particles formed in the gas phase will fall on the wafer surface.

Silicon Dioxide PECVD. Oxide films with PECVD are usually formed by reacting silane (SiH_4) with either oxygen (O_2), nitrous oxide (N_2O), or carbon dioxide (CO_2) in a plasma. The processing temperature is about 350°C.⁴¹ The oxide can be doped with boron or phosphorus to form BSG or PSG, respectively, or with both to deposit BPSG. PECVD PSG tends to be more crack resistant, conformal, and free of pinholes than APCVD PSG.

A plasma gas mixture of silane and oxygen is not normally used because oxygen reacts readily in the gas phase, which promotes particle generation and poor film quality (e.g., pinholes). The silane and nitrous oxide gas mixture produces more uniform films than through the use of an oxygen reactant. The reaction of silane with nitrous oxide to form silicon dioxide is:



The silicon dioxide is nearly stoichiometric, although it does contain some hydrogen, with a small amount of nitrogen present in the film. Hydrogen can exist as a hydride (Si-H), silanol (Si-O-H), or as water (H-O-H). The presence of O-H groups is undesirable for MOS transistor electrical characteristics and must be minimized.⁴²

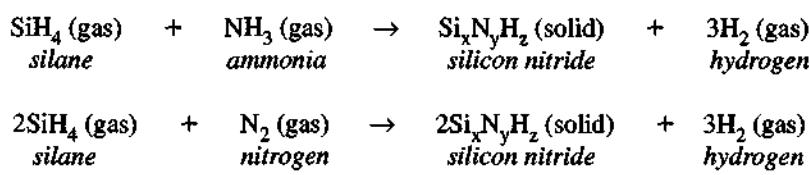
Parallel-plate PECVD can also deposit SiO₂ using TEOS, and is referred to as PETEOS. However, PETEOS is not acceptable for filling narrow-spaced metal lines due to voids. PETEOS is combined with APCVD TEOS or high-density plasma CVD (HDPCVD) to attain conformal gap fill, followed by chemical mechanical planarization. PETEOS has a relatively high SiO₂ deposition rate, which makes it desirable for wafer throughput in integrated tools.

Silicon Nitride PECVD. Silicon nitride PECVD is a film used as a final passivation layer on the chip as scratch protection, and a moisture barrier and as a barrier to prevent sodium diffusion. It is also widely used as a mask material in shallow trench isolation (STI) and in self-aligned contact technology (see Chapter 12). The Si₃N₄ nitride film deposited by PECVD is not very stoichiometric, and is sometimes written as Si_xN_yH_z. This equation highlights its nonstoichiometric composition and identifies that the film contains hydrogen (usually 9 to 30%). Table 11.3 compares the properties of silicon nitride for LPCVD vs. PECVD.

TABLE 11.3 Properties of Silicon Nitride for LPCVD vs. PECVD

Property	LPCVD	PECVD
Deposition temperature (°C)	700 to 800	300 to 400
Composition	Si ₃ N ₄	Si _x N _y H _z
Step coverage	Fair	Conformal
Stress at 23°C on silicon (dynes/cm ²)	1.2 – 1.8 × 10 ¹⁰ (tensile)	1 – 8 × 10 ⁹ (tensile or compressive)

PECVD nitride films are commonly produced from silane (SiH₄) and reacted with ammonia or nitrogen. The reaction equations are:



PECVD silicon nitride can have increased compressive film stress caused by ion bombardment during deposition that ruptures the Si-N or Si-H bonds. High compressive stress in a nitride layer can cause voids and cracks in the underlying aluminum metallization. There is usually substantial hydrogen content in the film. Hydrogen can reduce film stress, but it also degrades film properties.⁴³ When nitrogen is substituted for ammonia, the hydrogen content is reduced. However, ionization of nitrogen is difficult for plasma formation.

Silicon Oxynitride. Films of silicon oxynitride are deposited using PECVD by reacting Si₃N₄ with nitrous oxide (N₂O) at a temperature of 200 to 250°C.⁴⁴ Silicon oxynitride combines some properties from nitrides and oxides. It has good resistance to moisture and sodium penetration, with excellent mechanical, chemical, and electrical properties. These properties make silicon oxynitride suitable for topside passivation.

Gap Fill. A major limitation for PECVD in fine-geometry devices is gap fill. For gaps below 0.5 μm spacing (actual gap width), PECVD cannot fill the high aspect ratio gap without pinching off the top and leaving a void. A void in a dielectric material that is filling a gap is undesirable because of its effect on electrical performance and long-term reliability. For devices at the



High Density Plasma Deposition Chamber
(Photo courtesy of Applied Materials, Inc.)

0.25 μm technology node and below, HDPCVD has replaced PECVD because of its superior gap-fill properties.

It should be pointed out that other CVD methods can achieve fine-geometry gap-fill but have other limitations. For example, APCVD using TEOS oxide has good conformality and produces excellent gap fill. However, TEOS has different deposition characteristics depending on the underlying surface. One problem is that TEOS layers have dangling bonds that absorb moisture. To inhibit this condition, PECVD undercoat and overcoat layers are required. These layers add cost and also risk being removed during the subsequent chemical mechanical planarization (CMP).⁴⁵

High-Density Plasma CVD (HDPCVD) ■ A recent development in plasma assisted CVD is *high-density plasma CVD (HDPCVD)*. This deposition method became widely accepted in advanced wafer fabs in the mid-1990s. As its name implies, the plasma in HDPCVD is a high-density mixture of gases at low pressure that is directed toward the wafer surface in the reaction chamber. Its main benefit is that it can deposit films to fill high aspect ratio gaps with a deposition temperature range of 300 to 400°C. HDPCVD was initially developed for interlayer dielectric (ILD) applications, but is also being used for deposition in ILD-1, shallow trench isolation, etch-stop layers, and deposition of low- k dielectrics.⁴⁶

The HDPCVD reaction involves a chemical reaction between two or more gas precursors. For the deposition of oxide ILD, oxygen (or ozone) is often used with a silicon containing gas such as silane or TEOS, along with argon. To form the high-density plasma, a source excites the gas mixture with RF or microwave power (2.45 GHz) and directs the plasma ions into a dense region above the wafer surface. There are different high-density plasma sources, such as electron cyclotron resonance (ECR), inductively coupled plasma (ICP), and Helicon. These plasma sources are explained in detail in Chapter 16.

Wafer Bias and Heat Load. An RF bias (often between 1,500 to 3,000 kW of RF power) is applied to the wafer, pulling energetic ions out of the plasma and directing them toward the wafer surface. Biasing the wafer gives directionality to the energetic ions of the plasma. The high-density plasma ion density is about 10^{11} to 10^{12} ions/cm 3 at low pressures (between 2 to 10 mtorr). This high plasma density, in conjunction with directionality from wafer biasing, is an important reason for HDPCVD's ability to deposit films into narrow gaps with high-aspect ratio geometries of 3:1 to 4:1 and higher.

Much of the challenge for the use of high-density plasma is related not only to the performance of the plasma source but also to the details of the chamber design so that the technology works

in high-volume wafer fabrication.⁴⁷ A particular problem is that the high-density plasma will increase the thermal load to the wafer, since 2000W of RF bias can result in power density application of approximately 6 watts/cm² on the wafer surface. This condition leads to high wafer temperatures. However, the ILD must have process temperatures below 400°C to avoid harming the aluminum metallization; in addition, high heat loads cause thermal stress to the wafer.⁴⁸ This temperature limit requires cooling the wafer by applying a backside blanket of helium gas to the wafer through access ports in the electrostatic chuck (ESC). This action creates a thermal conductivity path between the wafer and the ESC, thus cooling the wafer and the chuck.

Simultaneous Deposition and Etching. An HDPCVD process uses a simultaneous deposition and etching action that is the basis for its ability to fill high-aspect ratio gaps with dielectric material without voids (see Figure 11.21). This is referred to as the *dep:etch (D:E) ratio*, which typically has a value of approximately 3:1 for HDPCVD.⁴⁹ Interpret this ratio as meaning the deposition rate (i.e., the rate material is deposited) is occurring three times faster than the etch rate. Increasing the ratio will increase deposition and therefore wafer throughput, but if the ratio is too high, then voids occur because the gap is not completely filled.

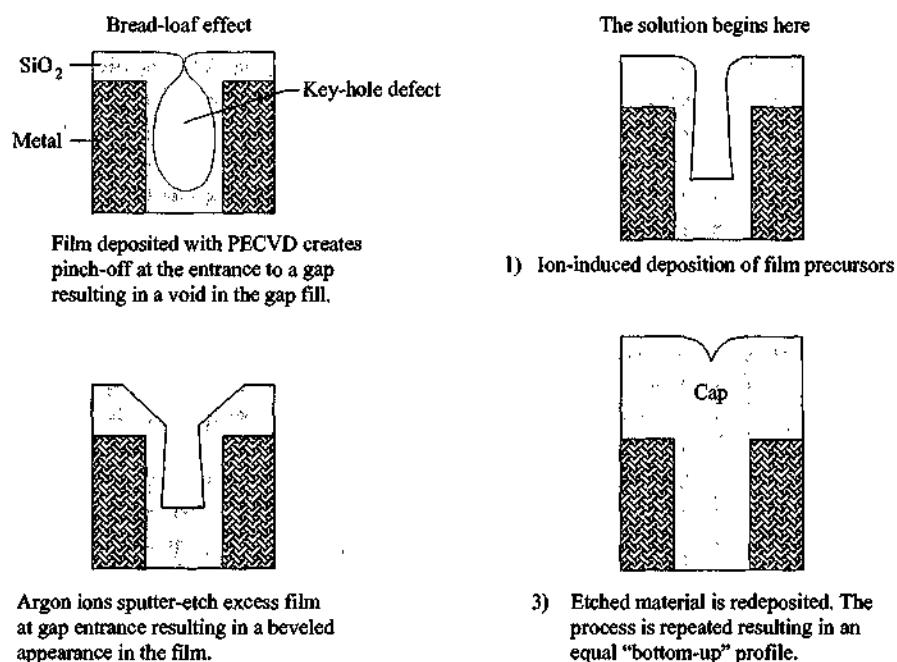


FIGURE 11.21 Dep-Etch-Dep Process

The surface reactions for the simultaneous deposition-etch include five steps, with the first three considered the dominant mechanisms:⁵⁰

1. **Ion-induced deposition:** The ion is pulled from the plasma and deposited to produce the main gap-filling phenomena. It is driven by the kinetic energy of reactant ions breaking surface bonds to form reaction sites.
2. **Sputter etch:** Energetic argon and reactant ions attracted to the surface due to wafer bias will bombard the surface and etch (dislodge) atoms.
3. **Redeposition:** Atoms are dislodged from the bottom of the gap and usually redeposit on sidewalls. This is important for uniform thickness on sidewalls and the bottom of the gap.
4. **Hot-neutral CVD:** This is a minor contribution to the reaction where thermal energy drives some deposition.
5. **Reflection:** Another minor contribution of ions reflecting off sidewalls and then deposited.

Simultaneous deposition and etching is a beneficial by-product of the plasma directionality created in the oxygen and argon gas mixture in HDPCVD. For a silicon dioxide (SiO_2) deposition process, the oxygen reacts with silane (SiH_4) to form SiO_2 , while argon acts to remove the deposited material away by sputtering (see Chapter 12 for an explanation of sputtering). Some factors that affect the dep:etch ratio are the ratio of the oxygen and argon gases, the chamber pressure, the ion energy, and the RF bias on the wafer.⁵¹ Low pressure is important to reduce the mean free path, which reduces collisions and maintains good directionality of the plasma. For acceptable deposition rates and wafer throughput in HDPCVD, high gas-flow rates are required. As shown in Figure 11.22, the wafer often will sit directly on the throat of a high-speed turbo pump (e.g., 4,000 to 5,000 liters/sec pump speed for 300-mm diameter wafers).

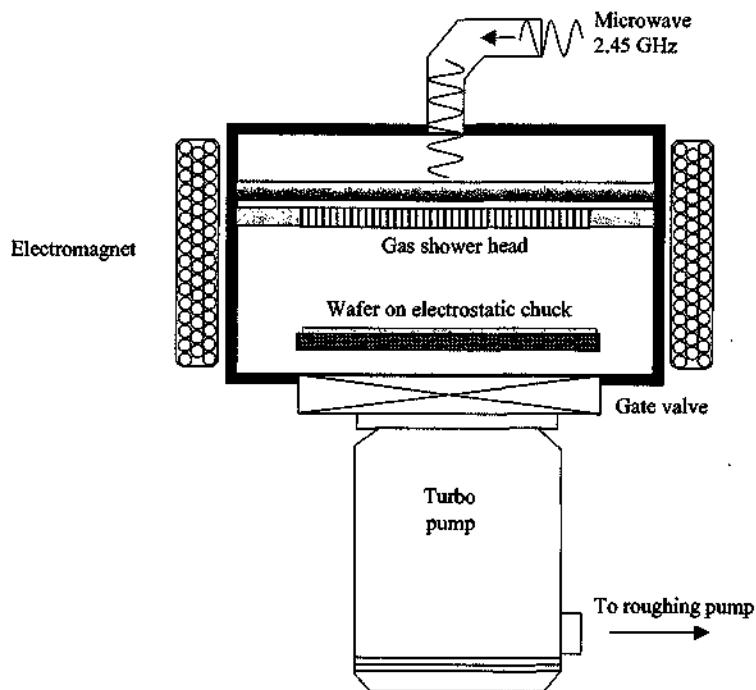


FIGURE 11.22 HDPCVD with Wafer at Throat of Turbo Pump

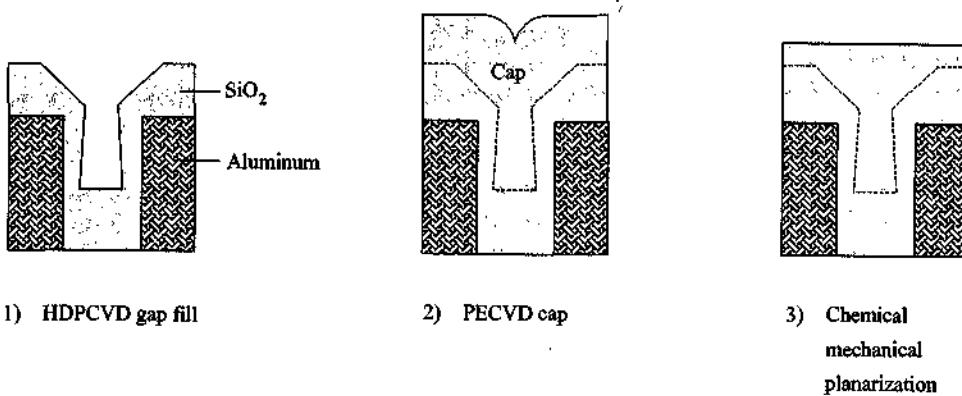
In practice, HDPCVD is sometimes used as the first step in a three-part process for dielectric gap fill.⁵² High aspect ratio gaps (> 3:1) are filled with HDPCVD, followed by a regular density PECVD to deposit a cap film (see Figure 11.23). This capping film is then planarized by chemical mechanical planarization (see Chapter 18). Planarization maintains a smooth and constant oxide thickness on top of the metal conductor lines.

DIELECTRICS AND PERFORMANCE

The dielectric materials used in different film layers directly affect the performance of a microchip. Two important aspects of dielectrics are the dielectric constant and device isolation.

Dielectric Constant

The *dielectric constant* (k) of a nonconductive material represents its effectiveness at storing potential electrical energy under the influence of an electric field. In other words, this is the insulating material's ability to act like a capacitor. The lowest attainable k is 1.0 and represents air. A high k dielectric stores more electrical energy. The dielectric constant of thermal silicon dioxide (SiO_2) has a k value of about 3.9. The k value of a plasma-enhanced CVD (PECVD) oxide is about 4.1 to 4.3.

**FIGURE 11.23** Three-Part Process for Dielectric Gap Fill

Doped SiO_2 has been traditionally the most common interlayer dielectric (ILD) material. Extensive research is underway to replace doped SiO_2 for the ILD with another dielectric material that has a lower dielectric constant (see Table 11.4).⁵³ Reducing the k value of the dielectric reduces capacitive losses between adjacent conductors because the dielectric stores less electric field and therefore takes less time to charge, allowing for an increase in speed performance of the metal conductors. A low- k dielectric material for the ILD becomes critical for smaller linewidths with less spacing between metal lines. As the metal linewidths decrease, the capacitive effects of the conductors and insulators increase, and using low- k materials compensates for this.

TABLE 11.4 Potential Low- k Materials for ILD of ULSI Interconnects

Potential Low- k Dielectric	Dielectric Constant (k)	Gap Fill (μm)	Cure Temp. ($^{\circ}\text{C}$)	Remarks
FSG (silicon oxyfluoride, Si_xOF_y)	3.4 to 4.1	<0.35	No issue	FSG has almost the same k -value as SiO_2 and a reliability concern that fluorine will attack and corrode tantalum barrier metal.
HSQ (hydrogen silsesquioxane)	2.9	<0.10	350 to 450	Silicon-based resin polymer available in solution as FOX (Flowable Oxide) for spin-on coating application. May require surface passivation to reduce moisture absorption. Cure is done in nitrogen.
Nanoporous silica	1.3 to 2.5	<0.25	400	Inorganic material with tunable dielectric constant that relies on pore density. Increased porosity reduces mechanical integrity—porous material must withstand polishing, etching, and heat treatments without degradation.
Poly(arylene) ether (PAE)	2.6 to 2.8	<0.15	375 to 425	Spin-on aromatic polymer with excellent adhesion and ability to be polished with CMP.
a-CF (fluorinated amorphous carbon or FLAC)*	2.8	<0.18	250 to 350	Leading candidate for CVD deposition with high-density plasma CVD (HDPCVD) to produce film with good thermal stability and adhesion.
Parylene AF4 (aliphatic tetrafluorinated poly-p-xyllylene)	2.5	<0.18	420 to 450	CVD film that meets adhesion and via resistance requirements with need to maintain gas delivery system at 200°C to control parylene precursor flow rate.

* P. Singer, "Technology News: Wafer Processing," *Semiconductor International* (October 1998): p. 44.

Chip Performance ■ An indicator of chip performance is the speed at which signals are transmitted. Continual die shrinks to achieve sub-0.25 μm technology nodes translate into reduced interconnect linewidths. This reduction leads to increased line resistance (R) for signals. Furthermore, reduced spacing between conductor lines creates more parasitic capacitance (C). The result is an increase in RC signal delay, which slows chip speed and lowers chip performance. This condition is a recent development due to sub-0.25 μm geometries and is often referred to as the *interconnect delay* (see Figure 11.24). In essence, reducing the interconnect dimensions results in more signal delay because of parasitic resistance and capacitance effects. This is the opposite of what occurs in transistors, where reducing gate length serves to reduce gate delay and increase transistor speed.⁵⁴

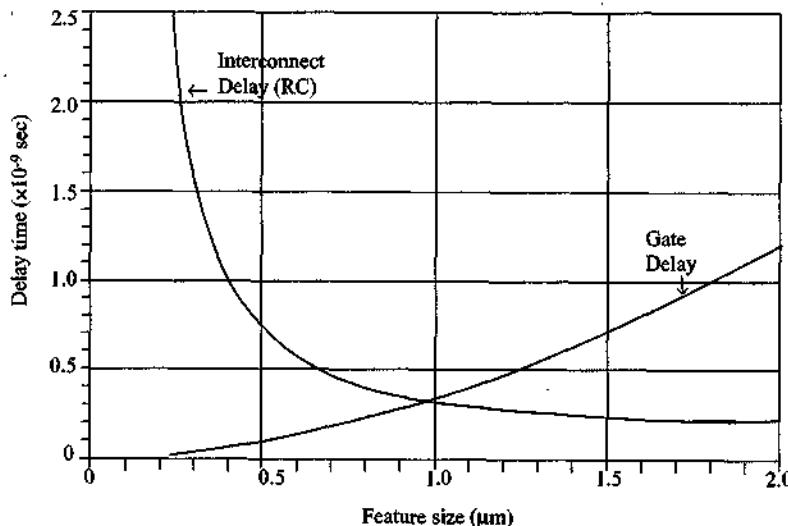


FIGURE 11.24 Interconnect Delay (RC) Versus Feature Size (μm)
(S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications, *Solid State Technology* (March 1996): p. 83.)

The line capacitance, C , is directly proportional to the k -value of the dielectric material. A low- k dielectric reduces the total interconnect capacitance of the chip (see Figure 11.25), reduces the RC signal delay, and improves chip performance. Lowering the total capacitance also decreases power consumption.⁵⁵ The use of a low- k dielectric material in conjunction with a low-resistance metal line provides the interconnect system with optimum performance for ULSI technology.

Low- k Dielectric Requirements ■ Table 11.5 outlines the typical requirements for a low- k dielectric constant film for introduction into wafer fabrication. It is anticipated in the semiconductor industry that the introduction of a low- k ILD material will follow the introduction of low-resistance metal lines made of copper (see Chapter 12 for a discussion of copper metallization). The investigation of alternative low- k dielectrics is being researched at this time.

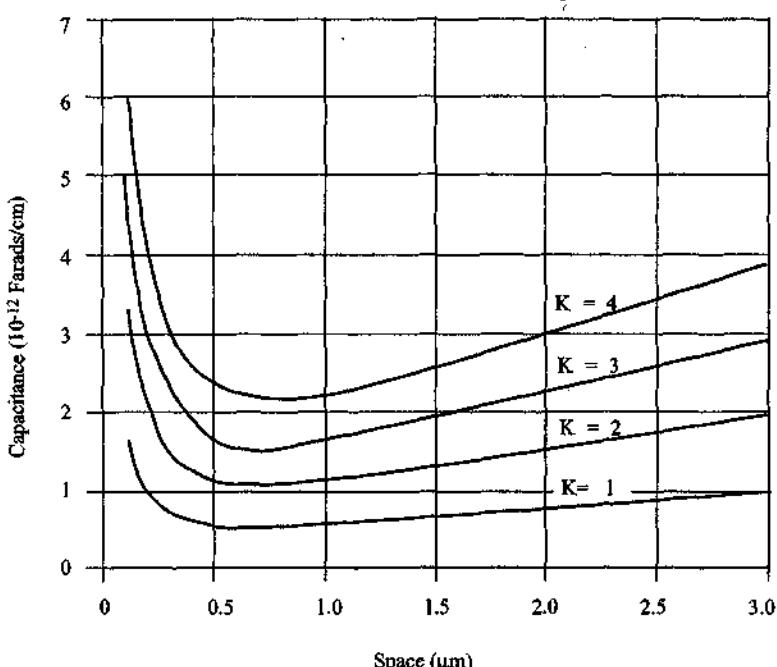


FIGURE 11.25. Total Interconnect Wiring Capacitance
(Used with permission from *Semiconductor International* (September 1998); p. 66.)

TABLE 11.5 Low-*k* Dielectric Film Requirements

Electrical	Mechanical	Thermal	Chemical	Processing	Metallization
Low dielectric constant	Good adhesion	Thermal stability	Resistant: acids and bases	Patternability	Low contact resistance
Low dielectric loss	Low shrinkage	Low coefficient of thermal expansion	Etch selectivity	Good gap fill	Low electromigration (corrosion)
Low leakage	Crack resistant	High conductivity	Low impurities	Planarization	Low stress voiding
High reliability	Low stress		No corrosion	Low pinhole	Hillock (smooth surface)
	Good hardness		Low moisture uptake	Low particulate	Compatible with barrier metals (Ta, TaN, TiN, and so on.)
			Acceptable storage life		

Adapted from P. Singer, "The Future of Dielectric CVD: High-Density Plasmas?" *Semiconductor International* (July 1997); p. 134.

High-*k* Dielectric Constant ■ There is ongoing investigation in the industry for new high-*k* dielectric materials, primarily for DRAM storage capacitors and as an eventual replacement for the extremely thin gate oxide (about 20 Å for 0.18 μm devices). DRAM technology has undergone a 4X increase in storage density every three years for the past 25 years due to processing and design improvements.⁵⁶ To attain the required charge storage with a SiO₂ and/or SiN_x dielectric, DRAM storage cell design has evolved into complex stacked capacitor structures (see Figure 11.26 on page 286). If the conventional SiO₂/SiN_x dielectric were replaced with a high-*k* dielectric, a simpler stack structure could be used for lower fabrication costs. A potential high-*k* material is tantalum pentoxide (Ta₂O₅), which has a value for *k* of 20 to 30 and can be easily integrated into the existing process. However, DRAM memory cells are sensitive to leakage and breakdown voltage of storage materials and interfaces. To compensate, the DRAM memory cell that requires a thicker Ta₂O₅

reduces its benefit. Another potential high- k material is barium strontium titanate ($(\text{BaSr})\text{TiO}_3$, or BST), which gives a significant improvement in capacitance per unit area over conventional dielectrics.

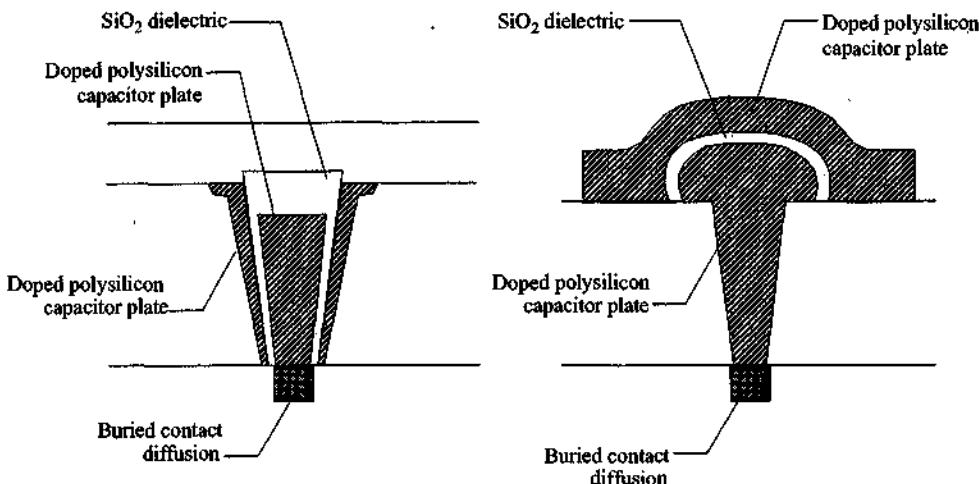


FIGURE 11.26 General Diagram of DRAM Stacked Capacitor

A new high- k gate dielectric is needed for the gate oxide because device scaling is forcing the existing SiO_2 to become extremely thin, with the future oxide thickness expected to be $<10 \text{ \AA}$ for the 50 nm device generation. In a MOS transistor, the gate dielectric material must withstand a significant voltage difference between the gate electrode and the underlying silicon substrate. Thin gate oxides are affected by *tunneling currents*, which becomes a problem when the SiO_2 dielectric layer is below about 15 to 20 \AA .⁵⁷ With the small gate dimensions in ULSI circuitry, electrons can tunnel through the gate material when the transistor is switched on and off. This leads to threshold voltage drift and eventual circuit failure because the device cannot switch states.

Device Isolation

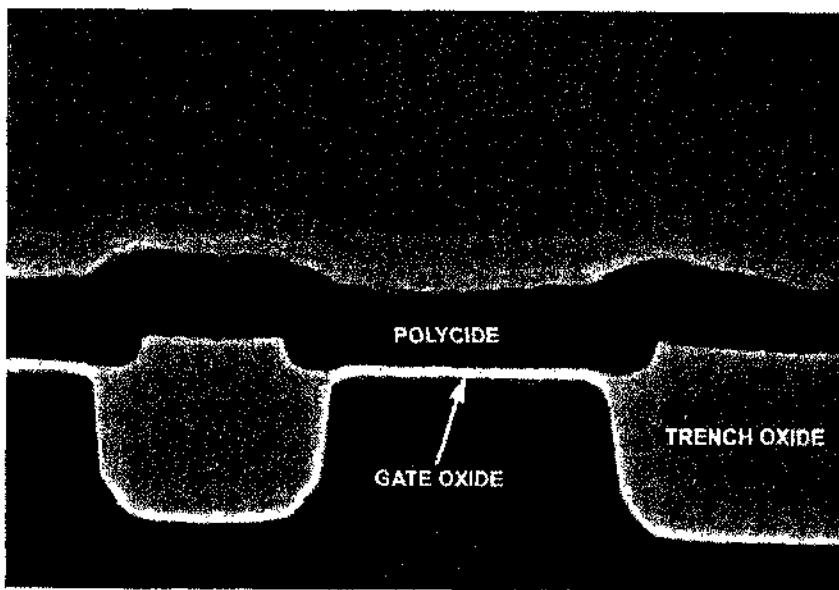
Device isolation in MOS wafer fabrication provides for electrical isolation between devices in the wafer. Isolation techniques serve to reduce or eliminate parasitic field transistors common in MOS planar fabrication (see Chapter 3). Isolation must accommodate scaling between the different device technologies (e.g., junction depth, gate oxide thickness, and so on.). This means that the space allocated for device isolation is being reduced for high-performance ICs. The two basic techniques for device isolation on MOS technologies are local field isolation through local oxidation of silicon (LOCOS), and shallow trench isolation (STI).⁵⁸

Local Oxidation ■ The *local oxidation of silicon (LOCOS)* has been the traditional isolation technique for wafer fabrication of devices with critical dimensions of 0.35 μm and larger (see Chapter 10). This technique uses patterned islands of silicon nitride (Si_3N_4) to define the regions for oxidation growth. LOCOS isolation structures are too large for isolation in devices with deep submicron scaling.⁵⁹ The major factor limiting LOCOS for 0.25- μm geometries and below is the lateral growth of oxide during silicon oxidation, which imposes natural limits on the minimum area and topography achievable. LOCOS is not acceptable for high-density ULSI technology.

Shallow Trench Isolation ■ *Shallow trench isolation (STI)* is the preferred isolation process for wafers fabricated at the 0.25 μm and below technology nodes. The reasons why STI replaced the LOCOS process are:⁶⁰

1. The need for more robust device isolation, especially in DRAM devices.
2. A significant reduction in the surface area for transistor isolation.
3. Superior latchup protection.
4. No channel encroachment.
5. Compatibility with chemical mechanical planarization (CMP).

In the basic STI process (see Chapter 10), trenches about 0.3 to 0.8 μm deep (up to several microns deep for DRAM trenches) are dry etched into the wafer substrate with nearly straight sidewalls and rounded corners (see Chapter 16 for an explanation of etching). The aspect ratio for STI trenches can vary from 2:1 to 5:1, with higher aspect ratios needed in DRAM devices due to their sensitivity to leakage. The trenches have a liner oxide grown on their surface, filled with CVD oxide and planarized using chemical mechanical planarization (CMP). STI technology is a more expensive process than LOCOS technology, primarily because it involves more complex process steps. Nevertheless, the benefits of STI outweigh its higher cost. The process steps for STI, including barrier layers and the liner oxide, are outlined in the CMOS process flow model in Chapter 9.



Shallow Trench Isolation
(Micrograph courtesy of Integrated Circuit Engineering)

SPIN-ON-DIELECTRICS

There are a wide range of low- k dielectrics that are designed to use wafer spin as an application method, referred to as spin-on-dielectric (SOD). The process for applying these materials is similar to a familiar process of spin-on-glass (SOG) used in low-cost IC fabrication, which we will review first.

Spin-On-Glass (SOG)

Spin-on-glass (SOG) was frequently used for gap fill and planarization of the ILD before the wide acceptance of chemical mechanical planarization (CMP) in the 1990s. SOG materials usually consist of two basic types: organic and inorganic.⁶¹ The organic is based on siloxane, while the inorganic is based on a silicate. The organic siloxane SOG has significant water absorption after curing, is thermally unstable, and does not tolerate plasma exposure. After curing, the silicate SOG behaves like SiO_2 , does not absorb excessive moisture, and is thermally stable. However, it shrinks significantly during curing, which leads to stress buildup and cracking if the layer is too thick. For these reasons, the siloxane SOG is usually used as a planarizing layer that undergoes an etchback process with a deposited oxide layer to render it smooth (see Chapter 18). The silicate SOG is used mainly for gap fill applications. There are many different commercial modifications of these two types.

Before applying the SOG, the wafer usually has a predeposited oxide layer, such as a PECVD oxide.⁶² The SOG is a liquid and is applied by spinning the wafer at a predetermined speed in either a closed or open bowl. The spin speed and parameters such as ramp-up rate define the SOG thickness. The film is cured, typically at 400°C, and then etched back to planarize the surface. A capping oxide film is applied to seal and protect the SOG, which minimizes water absorption

(see Figure 11.27). Without the cap oxide, there is substantial risk of reliability problems from absorbed moisture, such as through poisoning where the via between the metal layers is corroded and has high resistance.

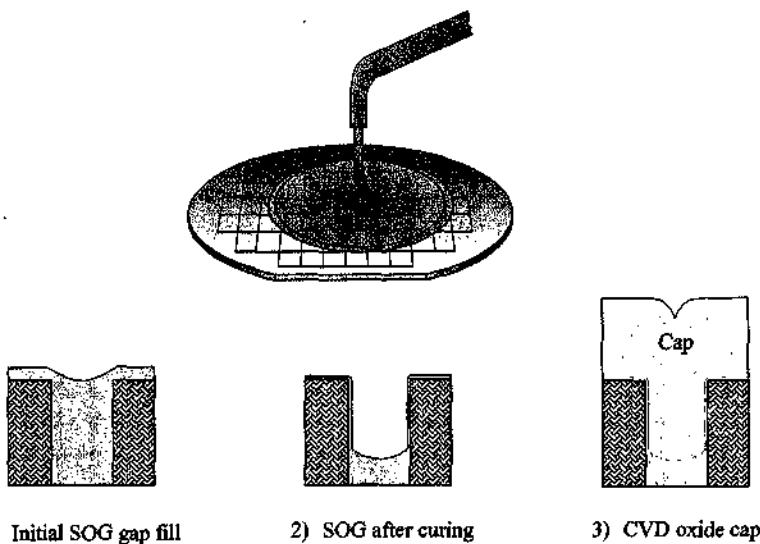


FIGURE 11.27 Gap-Fill with Spin-On-Glass (SOG)

Spin-On-Dielectric (SOD)

Low- k dielectric films are currently being investigated as *spin-on-dielectrics (SODs)*. An example of an SOD low- k film is hydrogen silsesquioxane (HSQ), presented earlier in this chapter in Table 11.4 on page 283. The SOD technology is being considered as a cost-effective alternative to CVD processes for depositing low- k films. Most SOD applications have used standard spin coaters (used in photolithography which is discussed in Chapter 13) with batch curing in generic furnace tubes. There has been investigation into a single-wafer cluster tool approach to spin coat and cure the film, using a high-temperature hotplate at 350 to 475°C.⁶³ A proposed spin coat and cure process for HSQ is outlined in Table 11.6.

TABLE 11.6 Proposed HSQ Low- k Dielectric Processing Parameters

Major Operation	Process Step	Parameter
Spin Coating	Apply bowl speed	50 rpm
	Maximum bowl speed	800 to 1500 rpm
	Backside rinse	800 rpm, 5 sec
	Topside edge bead removal	1000 rpm, 10 sec
	Spin dry	1000 rpm, 5 sec
Cure	Initial soft-bake cure	200°C, 60 sec, N ₂ purge
	In-line cure	475°C, 60 sec, N ₂ ambient

In some cases, the nature of the SOD material is an extension of an existing SOG material. Other SOD materials are spin-on polymers that are substantially different than SOG materials. New low- k materials have less tendency to absorb moisture and have superior crack resistance. Because of these features, single layer deposition is usually acceptable with no need for a cap oxide layer. In some cases, however, an adhesion promoter is required.⁶⁴

EPITAXY

Epitaxy is the deposition of a thin layer of single-crystal material upon the surface of a single-crystal substrate (see Figure 11.28). This epitaxial layer is often referred to as an epilayer. Epitaxy provides flexibility for the device designer to optimize performance by controlling epilayer doping thickness, concentration, and profile, independent of the wafer. This control is accomplished by the intentional addition of dopants during the epitaxial growth process. An epilayer also reduces the occurrence of latchup in CMOS devices (see Chapter 3). The most common epitaxial reactor in IC production is a high-temperature CVD system.⁶⁵

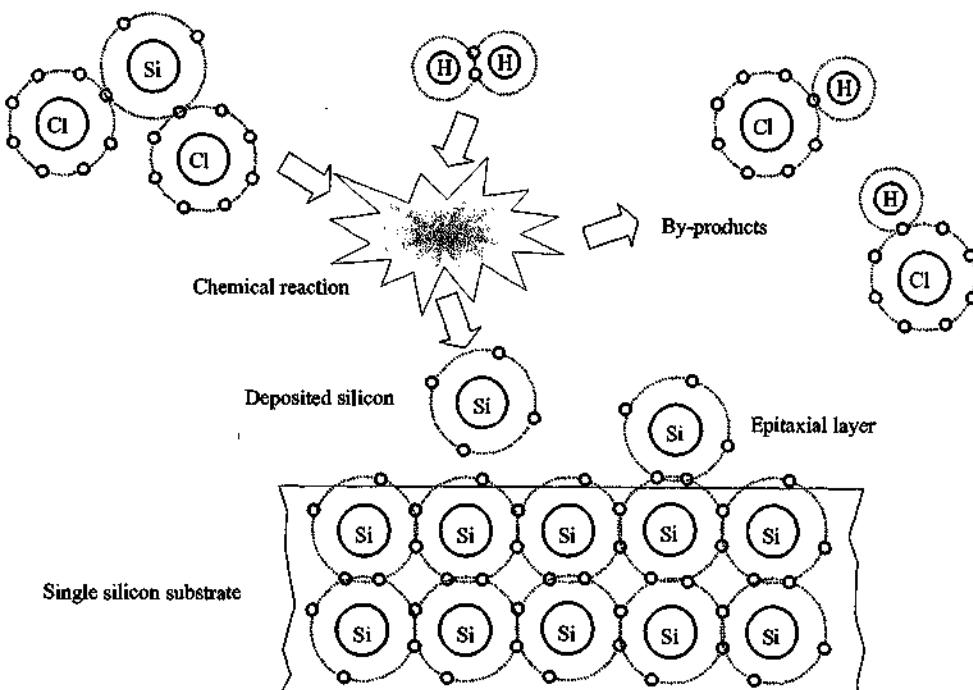


FIGURE 11.28 Silicon Epitaxial Growth on a Silicon Wafer

There are applications under investigation where epitaxy could contribute to advancing future IC performance. An example is *raised source/drain (S/D)* structures to achieve low contact resistance.⁶⁶ The raised S/Ds are formed by depositing epitaxial silicon on the source, drain, and gate areas of the device. This action effectively increases the surface areas of the S/Ds and therefore lowers the sheet resistance (similar to a larger diameter wire having lower resistance than the same wire with a smaller diameter). These deposits can decrease the contact resistance from 5 Ω/square on conventional devices to 1-2 Ω/square. The raised S/D structure is under investigation for possible implementation at the 0.15-μm technology node.

Undesirable conditions of nonuniform doping can occur during epitaxial deposition.⁶⁷ Since lightly doped epilayers are often grown on more heavily doped substrates, there is a condition of *autodoping* into the epitaxial layer. This condition occurs when dopant impurities evaporate from the wafer or are liberated due to chlorine etching of the wafer surface during deposition. These impurities move into the gas stream and cause unintentional doping of the epilayer. As the epilayer grows, less dopant is available from the wafer, and the impurities in the gas stream reach a constant level. Another form of irregular doping is the substrate acting as a source of dopant impurities that diffuse into the epilayer. This action is referred to as *out-diffusion*. Both autodoping and out-diffusion can affect the dopant transition between the substrate and epilayer, causing the transition to be less abrupt than desired.

If the film and substrate are of the same material (e.g., Si film on an Si substrate), then the film growth is *homoepitaxy*. A less common type of epitaxy is *heteroepitaxy*, which occurs when the film and substrate are of two different materials (e.g., Si on Al₂O₃).

Epitaxy Growth Methods

Epitaxial silicon is commonly deposited using CVD systems. Before epitaxy growth, the wafer must be cleaned of native oxide and any residual organic or metal impurities with a goal of achieving a perfect silicon surface (see Chapter 6). During the epitaxial deposition process, the atoms produced by the gas reaction strike the wafer surface and then move around until they find the correct location to bond to the surface atoms. This action forms the epitaxial layer with the same crystallographic arrangement as the substrate.

Possible gas sources for the epitaxy reaction are the hydrogen reduction of silicon tetrachloride (SiCl_4), silane (SiH_4), dichlorosilane (SiH_2Cl_2 , or DCS) or trichlorosilane (TCS). Deposition temperatures are from 1050 to 1250°C. Nearly all silicon epitaxy for wafer fabrication is done by the reduction of chlorosilanes ($\text{SiH}_x\text{Cl}_{4-x}$, where $x = 0, 1, 2$ or 3) that are diluted in hydrogen.⁶⁸ The temperature of the reaction can be lowered if there are fewer chlorine atoms in this precursor. Epitaxial silicon is not commonly grown with silane (SiH_4) because there is excessive particle formation when the silicon deposits on the warm surfaces of the reactor.

Different methods have been used to grow single-crystal layers on silicon wafers, including solid-phase, liquid-phase, vapor-phase, and molecular beam. Three methods used to grow epitaxial layers on silicon wafers for IC production are:

- ◆ Vapor-phase epitaxy (VPE)
- ◆ Metalorganic CVD (MOCVD)
- ◆ Molecular-beam epitaxy (MBE)

Vapor-Phase Epitaxy (VPE) ■ The most common method used for silicon epitaxial growth in wafer fabrication is *vapor phase epitaxy* (VPE), which is a subset of CVD.⁶⁹ Silicon VPE is achieved by passing gas compounds of the desired chemicals over single-crystal silicon wafers that are heated from 800 to 1150°C. The heat from the high temperature provides the energy necessary to drive the chemical reactions, which take place on the surface of the wafer. This process is shown in Figure 11.29.

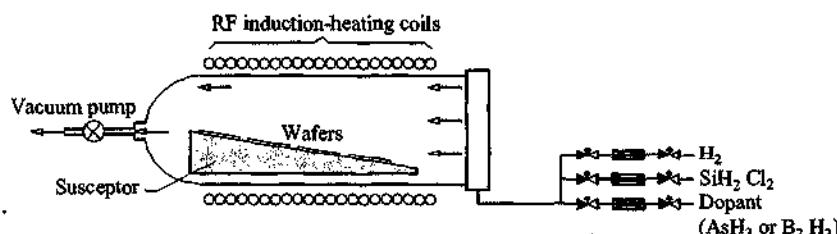


FIGURE 11.29 Illustration of Vapor-Phase Epitaxy

Before processing wafers in a VPE reactor, the system is purged with nitrogen or hydrogen, followed by a vapor HCl etching. The reactant gases, such as chlorosilane, along with dopant gases, are then introduced to the growth chamber where the wafer is heated to the desired temperature. Once the reactant and dopant gases are in the growth reactor, the chemical species undergo the necessary chemical and physical reactions to deposit the doped epilayer.

Typical equipment design for an epitaxial reactor consists of the gas distribution system, a reactor tube, a susceptor to hold and heat the wafer, a control system, and a gas exhaust system. Mass flow controllers and pneumatic valves are used in the gas distribution system to attain tight control over gas flows into the reactor chamber. The susceptor is typically made of graphite or a polysilicon that is coated with silicon carbide or silicon nitride. It must be strong and nonreactive to the reactants and their by-products.⁷⁰ Heating the susceptor is accomplished through inductive heating or by radiation from filament lamps. The horizontal and vertical reactors shown in Figure 11.30 are the most common. The susceptor in the horizontal reactor is tilted a few degrees to improve uniformity.

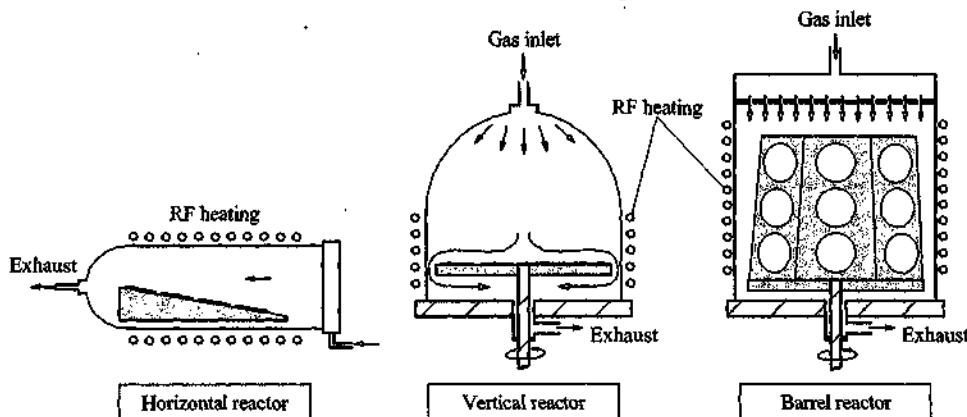


FIGURE 11.30 Silicon Vapor-Phase Epitaxy Reactors

Metalorganic CVD (MOCVD) ■ Another epilayer growth method is commonly referred to as *metalorganic CVD (MOCVD)*, but this is actually a broader term that is equally applicable to the deposition of polycrystalline and amorphous films of metals and oxides.⁷¹ MOCVD is a form of VPE and is not typically used for silicon epitaxy because no suitable sources are readily available. It has been used to deposit compound semiconductor epitaxial layers, such as the III-V compound of gallium arsenide (GaAs) from organometallic sources at low temperature. As with the VPE process, compounds of the desired materials are transported to a heated substrate where the complex chemical reaction takes place on the wafer surface. MOCVD is important for the controlled deposition of ultra-thin, doped, or undoped semiconductor heterolayers (such as GaAs/Si), primarily used for lasers, light-emitting diodes (LED), and optoelectronic integrated circuits.⁷² MOCVD is also used in development activity for depositing some organic low-*k* dielectrics for future IC fabrication.

Molecular-Beam Epitaxy (MBE) ■ *Molecular-beam epitaxy (MBE)* is one of the main methods used to deposit gallium arsenide (GaAs) heteroepitaxy with atomic resolution of thickness. It can also be used for depositing silicon onto a wafer substrate with very tight controls on epilayer thickness and doping uniformity. MBE takes place under conditions of high vacuum, usually at a base pressure of 10^{-10} to 10^{-11} torr or higher, typically with the use of a high vacuum cryopump. The reaction temperature range is between 500 to 900°C.

Most silicon MBE systems produce silicon atoms for the epitaxial reaction through the evaporation of silicon with an electromagnetically focused electron beam source. This process is similar to vacuum evaporation for deposition. The beam of silicon atoms leaving the silicon source travels through the evacuated chamber without collisions and deposits on the single-crystal wafer surface. Another more recent method is to deliver silicon atoms via a gas source at very low flow rates. The silicon growth rate of an MBE system is determined by measuring the number of atoms that leave the source and the fraction that actually strike the wafer surface and stick.

CVD QUALITY MEASURES

Important quality measures for CVD are listed in Table 11.7.

TABLE 11.7 Key Quality Measures for CVD

Quality Parameter	Types of Defects	Remarks
1. Voids during deposition of high-aspect ratio gap (>3:1) with PECVD SiO ₂ .	A. Keyhole voids are formed during deposition of high aspect ratio gaps. After chemical mechanical planarization (CMP) and removal of top surface, some voids become trenches (see Figure 11.31 on page 294). This trench can lead to an open circuit after metal deposition.	<ul style="list-style-type: none"> • Gap fill is critical as CDs are reduced and gaps have a high aspect ratio. • Voids are highly stressed regions and trap moisture or solvents that cause corrosion or outgas at high vacuum. • The root cause of this problem is trying to fill a high aspect ratio gap with a PECVD that is limited to deposition system. This application requires HDPCVD.
2. Film stress.	A. High film stress leads to cracking and delamination. B. Film stress can propagate silicon defects in the substrate. C. Stress may cause current leakage.	<p>The presence of dopant in a glass film can reduce stress. Deposition parameters that affect film stress are:</p> <ul style="list-style-type: none"> • RF power: adjust power to improve stress (e.g., reduce power by 10 watts to lower stress). • Pressure: adjust pressure to affect stress (e.g., increased pressure leads to higher film stress).
3. Film thickness.	A. Thickness of film exceeds requirements.	<p>Deposition parameters that affect thickness are:</p> <ul style="list-style-type: none"> • Time: reduce deposition time to reduce thickness (e.g., 1 sec reduction for each 100 Å of thickness). • Gas-flow rate: reduce gas flow to reduce thickness (e.g., reduce SiH₄ by 2 sccm for >5% thickness over nominal).
4. Refractive index (R/I).	A. R/I is a good monitor to assess quality of the film. B. R/I is highly dependent on the composition of the deposited film (stoichiometry).	<p>R/I is an optical property of a material. Compared to thermal oxides, CVD oxides are inferior in quality and integrity. CVD oxides exhibit more particles and pinholes. The R/I of CVD SiO₂ can be compared to the R/I of SiO₂ (1.46) to assess relative quality:</p> <ul style="list-style-type: none"> • High R/I means excessive Si in film. • Low R/I indicates a porous film, which tends to absorb moisture.

CVD TROUBLESHOOTING

Troubleshooting techniques for common CVD problems are described in Table 11.8. Plasma deposition systems have two areas of safety: chemical and electrical. Many plasmas use toxic gases (see Appendix A, Table A.3). There are high voltages and RF energy in the RF power supplies. Use caution when troubleshooting CVD systems.

TABLE 11.8 Common CVD Troubleshooting Problems

Problem	Probable Cause	Corrective Action
1. Particle contamination associated with film.	<p>Source of particles is isolated by determining whether particles are found on top of film, in film, or below film:</p> <ul style="list-style-type: none"> A. On film: particles formed after deposition. Check for particles on sidewalls of hot-wall reactors and belt-driven reactors. B. In film: Gas-phase nucleation particles from gas too rich in silane or silicon. Gas supply contamination causes particles. C. Under film: Particles from silicon carbide, quartz, or reactor walls fall on wafer before deposition. 	<ul style="list-style-type: none"> • Particles on film: Particles on sidewalls indicate need for more frequent wet cleaning of quartzware and chamber surfaces for preventive maintenance. Inspect <i>in situ</i> dry cleaning process. Also verify the correct procedures for cleaning (manually or <i>in situ</i>). • Particles in film: Improper gas flows from MFC calibration or problem in process recipe or software program. Check for leaks in gas supply system or O-rings. Verify point-of-use filters are acceptable. • Particles under film: Check wafer cleaning process before deposition.
2. Film thickness.	<p>Thickness is related to both equipment and process problems. Variables affecting thickness are:</p> <ul style="list-style-type: none"> A. Incorrect temperature control. B. System pressure is too high or low. C. System power needs adjustment. D. Improper gas flow. 	<ul style="list-style-type: none"> • Temperature controller may require calibration. A common problem is a defective thermocouple. • System pressure is controlled by process recipe. Check vacuum system for leaks. • Adjust RF power to optimize film thickness. • Check calibration of MFC to ensure proper gas flow. • If checking thickness with test wafer, verify the test wafer is clean and does not have a thickness variation before the test.
3. Cracking of dielectric material on top of electrostatic chuck (ESC).	<p>High thermal load at the ESC causes the dielectric material to erode or crack. This condition leads to plasma arcing or wafer chucking/dechucking problems.</p>	<ul style="list-style-type: none"> • Investigate the ESC backside cooling system to ensure it is functioning properly. • Inspect ESC materials to verify there is no breakdown due to high power exposure, temperature, or plasma clean conditions.

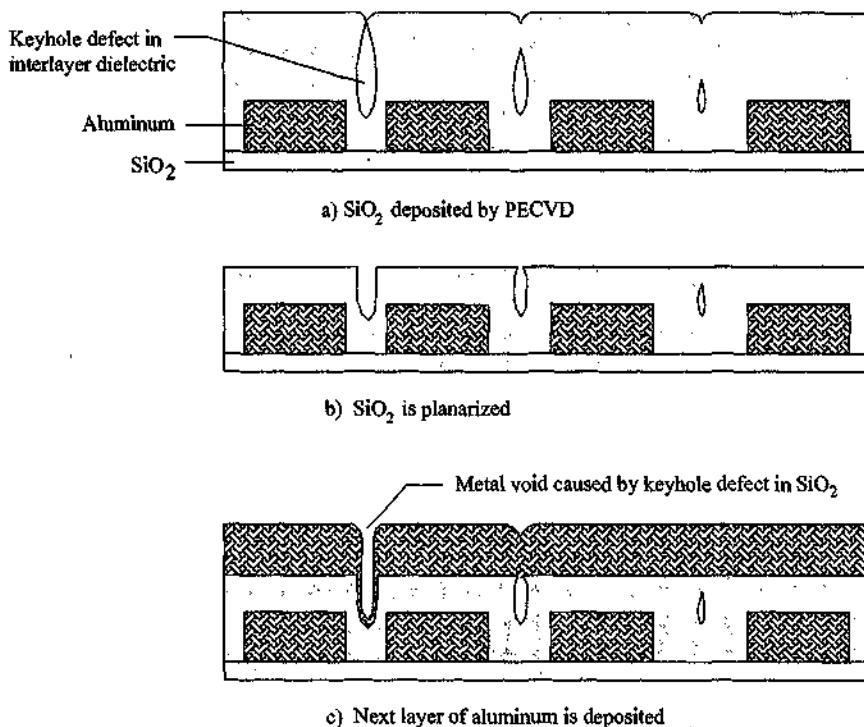


FIGURE 11.31 Effects of Keyholes in ILD on Metal Step Coverage

SUMMARY

Multilevel metallization with dielectric and metal layers is necessary for advanced ICs. Thin films have special characteristics to meet the requirements for wafer fabrication, including high aspect ratio gap fill. A film grows in stages from a cluster of nuclei to a continuous film. There is a wide range of tools for depositing film layers, with the focus on dielectric chemical vapor deposition (CVD). CVD deposits a thin film on the wafer surface through a chemical reaction that is assisted by thermal or plasma energy. CVD surface reactions include pyrolysis, oxidation, and reduction. The CVD chemical reaction follows defined steps and can be limited by the supply of chemicals or by the nature of the surface reaction. The thin film is sometimes doped to obtain improved performance. Reactors for CVD are categorized by atmospheric pressure CVD (APCVD), low pressure CVD (LPCVD), or plasma-assisted CVD. APCVD is simple in design and is primarily used to deposit conformal oxide with a TEOS-O_3 process.

LPCVD deposits various films, including oxide, silicon nitride, and polysilicon. The LPCVD process is reaction-rate limited (gas flow is not critical) and often occurs at lower temperatures. Plasma-assisted CVD is divided into plasma-enhanced CVD (PECVD) and high-density plasma CVD (HDPCVD). PECVD was the first plasma-assisted CVD but had limited use for filling gaps less than $0.5 \mu\text{m}$. HDPCVD is used for advanced ICs because of its excellent gap-fill capability for high-aspect ratios at relatively low temperatures. The deposited dielectric material increases chip performance by reducing the RC interconnect delay with a low- k dielectric. Dielectric material is also important for isolation technologies. Dielectric materials have been applied in liquid form by spinning, which may be an application method for future low- k dielectrics. Epitaxy is grown primarily by CVD methods, including vapor-phase epitaxy (VPE), metalorganic CVD (MOCVD), and molecular-beam epitaxy.

KEY TERMS

multilevel metallization

aluminum metallization

critical layers

noncritical layers

first interlayer dielectric (ILD-1)

premetal dielectric (PMD)

interlayer dielectric (ILD)

thin film

stoichiometry

aspect ratio

nucleation

nuclei coalescence

continuous film

chemical vapor deposition (CVD)

reactor

heterogeneous reaction (also surface catalyzed)

homogeneous reaction

adsorption

desorption

species

precursors

mass-transport limited

reaction-rate limited

kinetically controlled

boundary layer

stagnant layer

phosphosilicate glass (PSG)

borosilicate glass (BSG)

borophosphosilicate glass (BPSG)

fluorosilicate glass (FSG)

cluster tools

hot-wall reactor

cold-wall reactor

atmospheric pressure CVD (APCVD)

silane

TEOS oxide

low pressure CVD (LPCVD)

silicon nitride (Si_3N_4)

silicon oxynitride (SiO_xN_y)

plasma-enhanced CVD (PECVD)

PETEOS

silicon nitride PECVD

high-density plasma CVD (HDPCVD)

dep:etch ratio

dielectric constant (k)

device isolation

local oxidation of silicon (LOCOS)

shallow trench isolation (STI)

spin-on-glass (SOG)

spin-on-dielectric (SOD)

epitaxial growth

raised source/drain structures

autodoping

out-diffusion

homoepitaxy

heteroepitaxy

vapor-phase epitaxy (VPE)

metalorganic CVD (MOCVD)

molecular-beam epitaxy

REVIEW QUESTIONS

- What is multilevel metallization? Why is it necessary for chip fabrication?
- What is aluminum metallization? Describe a chip critical layer and a noncritical layer.
- Explain the purpose of an ILD layer. Where is the ILD-1 layer located on a chip?
- What is a thin film? List and describe eight characteristics of an acceptable thin film.
- Define the aspect ratio. Why is a high-aspect ratio important for ULSI devices?
- List and describe the three stages of thin film growth.
- List the five major techniques for deposition.
- Define chemical vapor deposition (CVD).
- What are the five basic CVD chemical reactions? Give a short description of each reaction.
- In a CVD reaction, is a heterogeneous or homogeneous reaction preferred? Why?
- Identify and describe the eight steps in a CVD reaction.
- Explain the difference between a mass-transport limited and a reaction-rate limited CVD process. Which process is temperature dependent?
- What is a boundary layer in CVD? Under what conditions does it become a stagnant layer?
- What is the beneficial effect of low pressure in a CVD reaction?
- What do the abbreviations PSG, BPSG, and FSG stand for? What are three benefits of adding dopants to oxide?

16. Describe the different types of CVD reactors and their principal advantages.
17. Explain the difference between a hot-wall reactor and cold-wall reactor for CVD.
18. Explain APCVD. What is the principal disadvantage with APCVD SiO₂, using silane as a source?
19. What is TEOS? Discuss the primary advantages to using APCVD TEOS-O₃.
20. Why is LPCVD more common than APCVD? Describe LPCVD processing.
21. Why does LPCVD operate in the reaction-rate regime?
22. State what film is deposited using LPCVD TEOS, and explain the advantage of this film.
23. Why is silicon nitride often used as a passivation layer?
24. What CVD tool is used to deposit the polysilicon gate material? List six reasons why polysilicon is used as a gate electrode.
25. What are the benefits to using a silicon oxynitride film?
26. State six advantages to using plasma during CVD.
27. What is PECVD? What is the main difference between PECVD and LPCVD?
28. Describe the main differences between silicon nitride PECVD and LPCVD.
29. Discuss a major limitation for using PECVD in ULSI microchips.
30. Explain what HDPCVD is. What are its main benefits in advanced ICs?
31. Describe the effect wafer biasing has on HDPCVD directionality.
32. Explain simultaneous deposition and etching for HDPCVD. What is the value for the typical dep-etch ratio?
33. List and describe the five steps used for simultaneous deposition-etching in CVD.
34. Discuss the importance of the dielectric constant of the ILD.
35. Explain the interconnect delay and why it is beneficial to lower the ILD constant dielectric.
36. What is an application for a high-k dielectric material and why is it needed?
37. What are LOCOS and STI? Why has STI replaced LOCOS for advanced ICs?
38. List the processing steps for STI.
39. What is spin-on-glass? Explain spin-on-dielectric and what it could be used for in future applications.
40. Describe what epitaxy is. Define the terms *autodoping* and *out-diffusion*.
41. List and discuss three methods for epitaxial growth.

DEPOSITION EQUIPMENT SUPPLIERS' WEBSITES

Amtech Systems Inc.
 Applied Materials
 ASM
 CVC Incorporated
 CVD Equipment Corporation
 Genus Incorporated
 Kokusai Semiconductor Equipment
 Novellus Systems Inc.
 TEL, Tokyo Electron Ltd.
 Tystar Corporation

<http://www.amtechsystems.com/>
<http://www.appliedmaterials.com/products/>
<http://www.asm.com/>
<http://www.cvc.com/>
<http://www.cvdequipment.com/>
<http://www.genus.com/>
<http://www.ksec.com/>
<http://www.novellus.com/index.htm>
<http://www.teainet.com>
<http://www.tystar.com/>

REFERENCES

1. R. Liu, "Metallization," *VLSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 379.
2. R. Jackson et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998): p. 49.
3. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics for Future ICs," *Solid State Technology* (September 1998): p. 69.
4. P. Singer, "The Future of Dielectric CVD: High-Density Plasmas?" *Semiconductor International* (July 1997): p. 134.
5. S. Sivaram, *Chemical Vapor Deposition: Thermal and Plasma Deposition of Electronic Materials* (New York: Van Nostrand Reinhold, 1995), p. 2.
6. Adapted from S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1 *Process Technology*, 2nd ed., (Sunset Beach, CA: Lattice Press, 2000) pp. 108–114.
7. J. Baliga, "Options for CVD of Dielectrics Include Low-k Materials," *Semiconductor International* (June 1998): p. 139.
8. C. Rye et al., "Barriers for Copper Interconnections," *Solid State Technology* (April 1999): p. 53.
9. P. Singer, "Film Stress and How to Measure It," *Semiconductor International* (October 1992): p. 54.
10. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: John Wiley & Sons, 1994), p. 522.
11. Ibid.

12. G. Anner, *Planar Processing Primer* (New York: Van Nostrand Reinhold, 1990), p. 361.
13. G. Anner, *Planar Processing Primer*, p. 361.
14. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology* (Sunset Beach, CA: Lattice Press, 1986), p. 151.
15. A. Jones and P. O'Brien, *CVD of Compound Semiconductors: Precursor Synthesis, Development and Applications* (Weinheim, Germany: VCH, 1997), p. 31.
16. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, p. 151.
17. Ibid., p. 154.
18. Ibid., p. 154.
19. G. Anner, *Planar Processing Primer*, p. 365.
20. SEMATECH, "Desposition Processes," *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 6.
21. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 532.
22. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics," p. 69.
23. J. Baliga, "Options for CVD of Dielectrics," p. 140.
24. H. Cheng, "Dielectric and Polysilicon Film Desposition," *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 211.
25. M. Hammond, "Introduction to Chemical Vapor Deposition," *Solid State Technology* (December 1979): p. 61.
26. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 529.
27. G. Schwartz and K. Srikrishnan, "Interlevel Dielectrics," *Handbook of Semiconductor Interconnection Technology*, ed. G. Schwartz, K. Srikrishnan, and A. Bross (New York: Marcel Dekker, 1998), p. 231.
28. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies* (Boston: Kluwer Academic Publishers, 1995), p. 118.
29. J. Baliga, "Options for CVD of Dielectrics," p. 140.
30. G. Schwartz and K. Srikrishnan, "Interlevel Dielectrics," p. 233.
31. H. Cheng, "Dielectric and Polysilicon Film Deposition," *ULSI Technology*, p. 211.
32. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 524.
33. Ibid., p. 520.
34. C. Gugliemini and A. Johnson, "Properties and Reactivity of Chlorine Trifluoride," *Semiconductor International* (June 1999): p. 162.
35. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 529.
36. H. Cheng, "Dielectric and Polysilicon," p. 225.
37. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, Vol. 1, *Process Technology*, 2nd ed., p. 175.
38. Y. Trouiller et al., "Inorganic Bottom ARC SiO_xN_y for Interconnection Levels on 0.18 μm Technology," *SPIE* 3508, (1998), p. 122.
39. L. Peters, "Pursuing the Perfect Low- k Dielectric," *Semiconductor International* (September 1998), p. 122.
40. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 125.
41. Ibid.
42. SEMATECH, "Desposition Processes," *Furnace Processes and Related Topics*, p. 35.
43. C. Applett et al., "Silicon Nitride Growth in a High-Density Plasma System," *Solid State Technology* (November 1995): p. 73.
44. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 535.
45. E. Korczynski, "HDP-CVD: Trying to Lasso Lightning," *Solid State Technology* (April 1996): p. 63.
46. P. Singer, "Future of Dielectric CVD," p. 128.
47. P. Burggraaf, "Advanced Plasma Sources: What's Working?" *Semiconductor International* (May 1994): p. 57.
48. J. Bondur et al., "Impact of Electrostatic Chuck Performance on HDP CVD SiO_2 Films," 52nd Symposium on Semiconductors and Integrated Circuits Technology, Osaka, Japan.
49. P. Singer, "Future of Dielectric CVD," p. 127.
50. E. Korczynski, "HDP-CVD," p. 64.
51. P. Singer, "Future of Dielectric CVD," p. 127.
52. Ibid.
53. L. Peters, "Pursuing the Perfect Low- k Dielectric," p. 68.
54. M. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *Solid State Technology* (September 1996): p. 105.
55. S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications," *Solid State Technology* (March 1996): p. 83.
56. D. Kotecki, "High- k Dielectric Materials for DRAM Capacitors," *Semiconductor International* (January 1998): p. 38.
57. P. Singer, "Wafer Processing: New Gate Dielectric Material Needed," *Semiconductor International* (January 1998): p. 38.
58. P. Van Cleemput et al., "HDPCVD Films Enabling Shallow Trench Isolation," *Semiconductor International* (July 1997): p. 180.
59. Ibid., p. 182.
60. L. Peters, "Choices and Challenges for Shallow Trench Isolation," *Semiconductor International* (April 1999): p. 69.
61. R. Liu, "Metallization," p. 419.
62. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 569.
63. T. Batchelder et al., "In-Line Cure of SOD Low- k Films," *Solid State Technology* (March 1999): p. 29.
64. L. Peters, "Pursuing the Perfect Low- k Dielectric," p. 66.
65. S. Campbell, *The Science and Engineering of Microelectronic Fabrication* (New York: Oxford University Press, 1996), p. 340.
66. L. Peters, "Is the 0.18 μm Node Just a Roadside Attraction?" *Semiconductor International* (January 1999): p. 52.
67. R. Jaeger, *Introduction to Microelectronic Fabrication* vol. V, (Reading, MA: Addison-Wesley, 1988), p. 125.

68. S. Campbell, *Science and Engineering*, p. 347.
69. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., p. 283.
70. P. J. Wang, "Epitaxy," *VLSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 124.
71. A. Thompson, R. Stall, and B. Kroll, "Advances in Epitaxial Deposition Technology," *Semiconductor International* (July 1994): p. 173.
72. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 102.

CHAPTER 12

METALLIZATION

Metalization in wafer fabrication is the process of depositing metal film over a dielectric film and later patterning it to form the interconnecting metal lines and plugs of integrated circuits. This is similar to using insulated copper wire in an automobile to interconnect all the electrical components to make a fully-functional electrical system. The metal lines are sandwiched between dielectric layers for electrical integrity. High-performance microprocessors use metal lines to interconnect tens of millions of devices on one chip. Transistor densities are projected to reach 1 billion transistors per chip by the year 2010, with a corresponding increase in interconnect complexity.

Microchip interconnect technology has become a critical challenge for future IC performance due to the need to decrease signal propagation delay. Because the density of ULSI circuit elements is increasing, interconnect resistance and parasitic capacitance increase, thereby slowing the signal propagation.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the terminology for metallization.
2. List and describe the six categories of metals used in wafer fabrication. Discuss the performance requirements and give applications for each metal category.
3. Explain the benefits for using copper metallization in wafer fabrication. Describe the challenges for implementing copper.
4. State the advantages and disadvantages to sputtering.
5. Describe the physics of sputtering and discuss different sputtering tools and applications.
6. Describe the benefits and applications for metal CVD.
7. Explain the fundamentals of copper electroplating.
8. Describe a process flow for dual-damascene processing.

A significant change currently underway in wafer fabrication technology is the reduction in interconnect metal resistivity, ρ . This reduction is achieved by the replacement of aluminum alloy with copper as the primary conducting metal. With deep submicron linewidths, there is also a need for a low- k interlayer dielectric (ILD). Lowering the dielectric constant will reduce parasitic capacitance, which contributes to signal delay.

Traditionally, the methods used to deposit metals on a wafer have been physical processes. On the other hand, processes used to deposit insulating and semiconducting layers usually involved the CVD chemical reactions studied in Chapter 11. This separation into physical and chemical processes is less defined with the introduction of new IC metallization technology.

INTRODUCTION

Wafer metallization is the deposition of a thin film of conductive metal onto a wafer by use of a chemical or physical process. This process is closely linked to dielectric deposition. Metal lines conduct the signal through the IC circuit while the dielectric layers ensure signals are not influenced by adjacent lines. Both metal and dielectrics are thin film processes. In some cases, metal and dielectrics are deposited by the same type of equipment.

Metallization has unique terminology for the different metal connections. The term *interconnect* describes the conductor materials, such as aluminum, polysilicon, or copper, that create the metal wiring that carries electrical signals to different parts of the chip. Interconnect is also used as a general term for the wiring between devices on a die and the overall package. A *contact* is the electrical connection at the silicon surface between the devices in the silicon wafer and the first metal layer. *Vias* are openings that pass through the various dielectric layers to form an electrical pathway from one metal layer to the adjacent metal layer. A metal *plug* fills the vias to form the electrical connection (interconnect) between the two metal layers. These connections are illustrated in Figure 12.1.

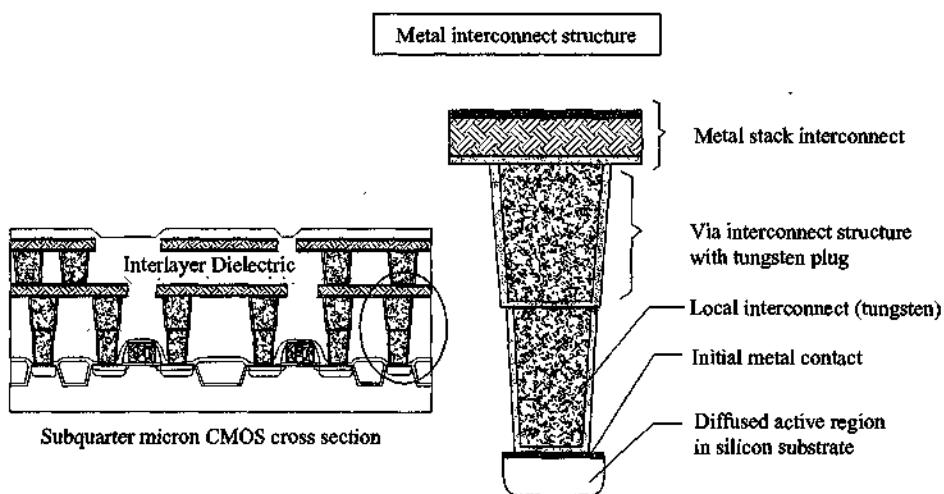


FIGURE 12.1 Multilevel Metallization

The interlayer dielectric (ILD) is an insulating material that electrically separates the metal levels. Once deposited, the ILD is patterned and etched to form via pathways for the various metal layers and the silicon. The vias are filled with a metal, traditionally tungsten (W), to form the via plug. Appreciate that there are many vias on a wafer, up to 10^{11} vias on each individual layer of a 300-mm product wafer.¹ This process of creating vias in the ILD is repeated for every ILD layer on the die. In traditional metallization, a blanket layer of aluminum alloy metal is deposited on the dielectric layer and then patterned and etched to form metal lines. Metal etch is an important technology for traditional metallization.

Metallization is in a transition period and is undergoing rapid change with the introduction of copper metallurgy to replace aluminum alloy. This change is due to the difficulty of etching copper. To overcome this problem, copper metallurgy uses the dual damascene process to form vias and copper interconnects (explained later in this chapter). This metallization process is literally the opposite approach of traditional metallization (see Figure 12.2). Damascene starts with deposition of a blanket dielectric, planarizes the dielectric, patterns and etches holes or trenches in the dielectric for vias and connecting metal lines, deposits blanket metal into the trenches, and then planarizes the metal down to the dielectric to define the metal interconnects.

Metallization technology is critical for increasing performance in advanced ICs, as explained in Chapter 11. Reducing chip performance signal delay caused by interconnect lines was not a significant concern for older IC technologies. The dominant signal delay has traditionally been caused by the device. However, for newer ULSI products manufactured with denser wiring, signal delay

due to interconnect has become a larger portion of the clock cycle time and has more effect on limiting the IC performance.²

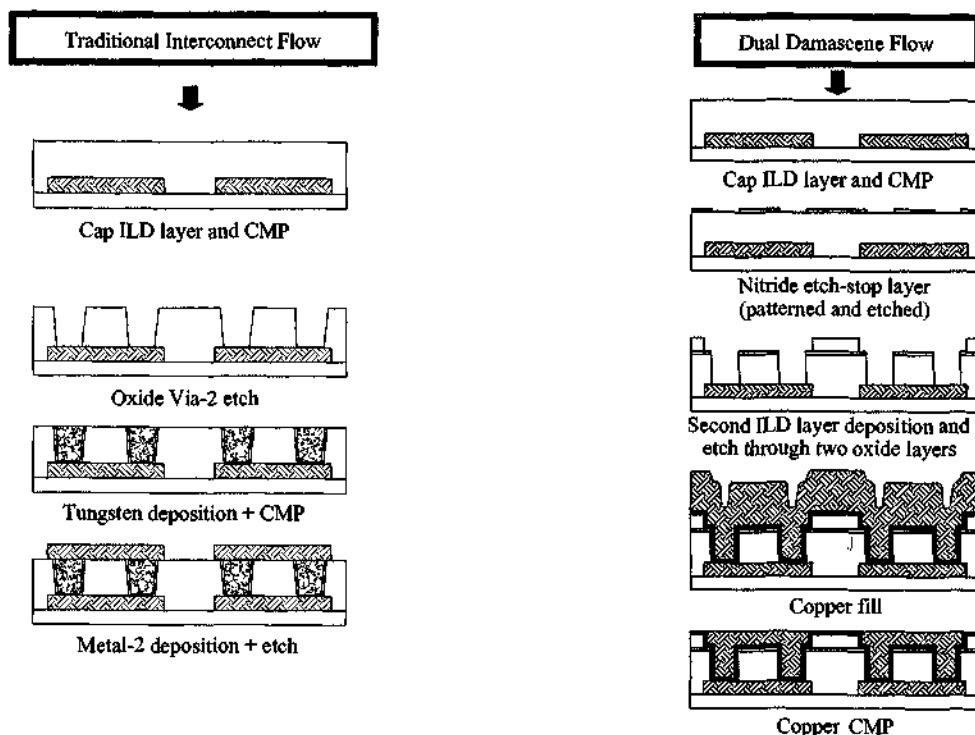
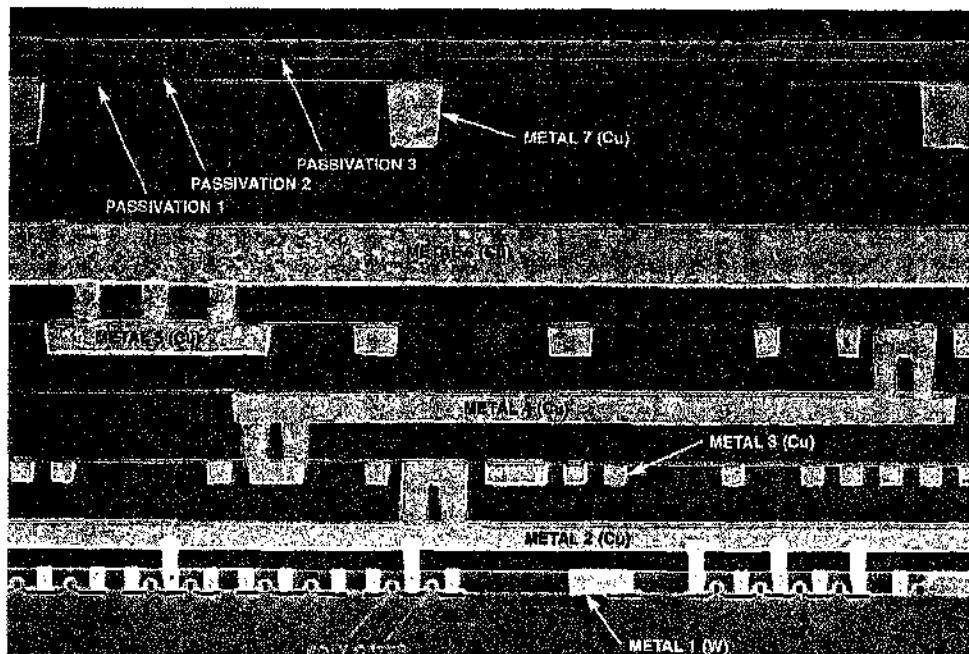


FIGURE 12.2 Traditional Versus Damascene Metallization



Copper Metallization
(Micrograph courtesy of Integrated Circuit Engineering)

TYPES OF METALS

The types of metals and metal alloys used for chip interconnects are evolving based on the performance requirements. Requirements for a successful metal material are:^{3, 4}

1. **Conductivity:** Must be highly conductive and capable of handling high current densities while maintaining electrical integrity.
2. **Adhesion:** Ability to adhere to the underlying substrate and easily connect to external connections. Low contact resistance to semiconductor and metal interfaces.
3. **Deposition:** Readily deposited with a uniform structure and composition (for alloys) by a relatively low-temperature process. Deposition into high-aspect ratio gaps for damascene metallization technique.
4. **Patterning/Planarization:** High-resolution patterning for traditional Al-based metallization without etching the underlying dielectric. Ease of planarization for damascene metallization.
5. **Reliability:** The metal is relatively soft and ductile in order to withstand cyclic temperature variations during processing and service.
6. **Corrosion:** High resistance to corrosion with minimal chemical interactions with the adjacent layers and underlying device regions.
7. **Stress:** Resistance to mechanical stress to reduce wafer distortion and material failures for reasons such as cracking, void formation, and stressed-induced corrosion.

The melting temperature and resistivities of different metals found in wafer fabrication and silicon are shown in Table 12.1.⁵

TABLE 12.1 Silicon and Select Wafer Fab Metals (at 20°C)

Material	Melting Temperature (°C)	Resistivity ($\mu\Omega\cdot\text{cm}$)
Silicon (Si)	1412	$\approx 10^9$
Doped Polysilicon (Doped Poly)	1412	≈ 500 to 525
Aluminum (Al)	660	2.65
Copper (Cu)	1083	1.678
Tungsten (W)	3417	8
Titanium (Ti)	1670	60
Tantalum (Ta)	2996	13 to 16
Molybdenum (Mo)	2620	5
Platinum (Pt)	1772	10

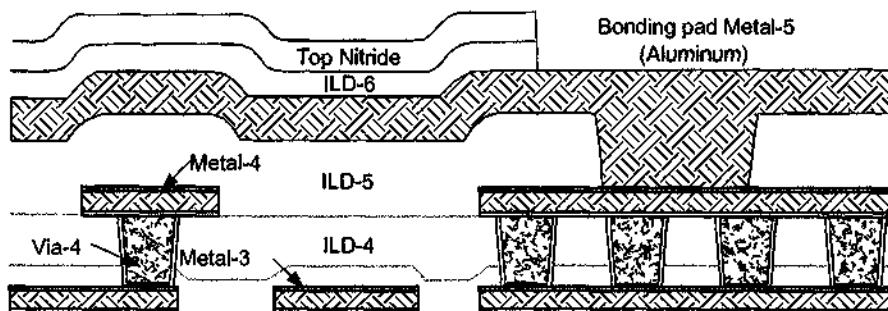
The various metals and metal alloys used in wafer fabrication can be grouped into the following categories:

- ◆ Aluminum
- ◆ Aluminum-copper alloys
- ◆ Copper
- ◆ Barrier metals
- ◆ Silicides
- ◆ Metal plugs

Aluminum

The earliest interconnect metal in semiconductor fabrication was aluminum, and it is still the most common interconnect metal in wafer fabrication. Aluminum is projected to continue as the dominant fab interconnect metal for several more years.⁶ Copper interconnect metal is expected to replace aluminum for high-performance IC fabrication during the early 2000s. Nevertheless, it is beneficial to review the background into the selection of aluminum for metallization because many of the same basic technical challenges exist today.

Selection of Aluminum ■ Aluminum has been one of the major materials used in wafer fabrication, along with silicon and silicon dioxide. It is used during wafer fabrication in thin film form to interconnect the different devices in the silicon wafer (see Figure 12.3). At the same time, aluminum is one of the thickest films deposited on the wafer, with the first metal layer deposition about 5,000 Å thick. The upper noncritical layers on a wafer (e.g., metal layer with bonding pads) can range up to 20,000 Å thick.



Metal-4 is preceded by other vias, interlayer dielectric, and metal layers.

FIGURE 12.3 Aluminum interconnect

Aluminum's low resistivity of $2.65 \mu\Omega\text{-cm}$ at 20°C has similar but slightly higher resistivity than copper, gold, and silver. However, silver and copper are both prone to corrosion and have high diffusivity in Si and SiO_2 , a fact that keeps them from being used for semiconductor fabrication. Gold and silver are much more expensive than aluminum and do not adhere well to oxide films. Gold was sparingly used in the beginning of wafer fabrication, but its high contact resistance with silicon required a platinum barrier metal. On the other hand, aluminum reacts readily with silicon dioxide (SiO_2) when heated to form a thin layer of aluminum oxide (Al_2O_3), which promotes adhesion between the silicon dioxide and aluminum. Aluminum is easily deposited on the wafer and etches in solutions that do not attack underlying films. For these reasons, aluminum was selected as the preferred metal for metallization.

Overall, aluminum is compatible with the major processes in silicon IC fabrication and is relatively inexpensive, making it the choice for metallization since the early days of IC fabrication. However, because of increased circuit density, increased number of metal layers on a wafer, and the reduction in linewidth due to scaling, the metallization technology has evolved from the simple one-level metal layer to a scheme of multiple metal layers. With its lower resistivity, copper is desirable to replace aluminum as the primary interconnect material.

Ohmic Contact ■ Before VLSI circuitry, pure aluminum was used for metallization. Silicon melts at a temperature of 1412°C and pure aluminum melts at 660°C . However, aluminum and silicon alloyed together actually have a lower melting temperature depending on their composition. For an alloy of 88.7% aluminum and 11.3% silicon, the alloy melts at 577°C , which is referred to as the *eutectic temperature*.⁷ The eutectic temperature is the lowest temperature that an alloy melts at a particular eutectic composition.

To form a contact between aluminum and silicon, it is necessary to heat the interface, usually in an inert or reducing H_2 atmosphere at a temperature of 450 to 500°C . This thermal bake is also referred to as a low temperature *anneal* or *sinter*. Baking the aluminum on silicon forms a desirable electrical interface referred to as an *ohmic contact*. An ohmic contact has very low resistance (the voltage-current characteristics of the contact interface behave according to Ohm's law). However, there is a small resistance associated with the ohmic contact and it is inversely proportional to the area of the contact. That is, smaller contacts have higher resistance. In modern chip design, ohmic contacts use a special refractory metal (e.g., titanium in the form of a silicide) as a contact at the silicon interface to reduce electrical resistance and improve adhesion (see Figure 12.4 on page 304). Ohmic contacts are fabricated with a salicide (self-aligned silicide) process to achieve optimal positioning above the source/drain and in close proximity to the gate structure. The salicide process is discussed later in this chapter.

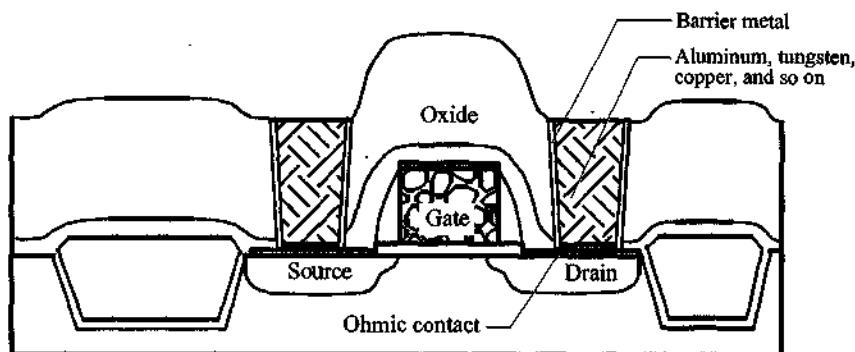


FIGURE 12.4 Ohmic Contact Structure

Given the hundreds of millions of contacts that can exist on a particular die, a reliable contact interface with low resistance and excellent adhesion is extremely important for optimum electrical performance. A failure in any individual contact can cause the entire chip to fail during test or use.

A problem encountered during the earliest work to develop ohmic contacts was undesirable interaction between the aluminum and silicon during the thermal step. This condition led to the microalloying of the contact metal to the silicon, and was referred to as *junction spiking*. Spiking occurs when the interface of pure aluminum and silicon is heated (see Figure 12.5), causing the diffusion of silicon into the aluminum. The amount of silicon dissolved into the aluminum is not uniform and will depend on the time and temperature involved in the heating process. If pure aluminum is heated to 450°C and a source of silicon is present, silicon will begin to dissolve in the aluminum until it reaches a concentration of about 0.5%.⁸ The problem is that the source of this silicon is the wafer. As the silicon is dissolved in the aluminum, it leaves behind voids in the wafer, permitting spikes of aluminum to form and penetrate into the silicon contact region. If the aluminum forms an ohmic contact to a shallow junction, the spike may cause a junction short.

The problem of junction spiking was addressed by two methods: the addition of silicon to the aluminum and barrier metallization. The first approach was to use an alloy of aluminum and silicon instead of pure aluminum. If there is already silicon in the aluminum, this will slow down the dissolution of additional silicon from the substrate. However, silicon alloying in aluminum is limited and can lead to silicon nodule formation (small regions of high silicon concentration) in the aluminum due to silicon condensation. Nodule formation can significantly increase the contact resistance and pose a serious reliability concern from local heating at the nodules. The main method to resolve junction spiking was the introduction of barrier metallization to inhibit diffusion (see the following section).

It is important that the oxide at the contact interface be as thin as possible. Wafers with exposed contact regions are often dipped in a dilute hydrofluoric (HF) acid solution immediately before placing the wafers in the deposition chamber. This serves to remove the native oxide layer prior to contact formation.

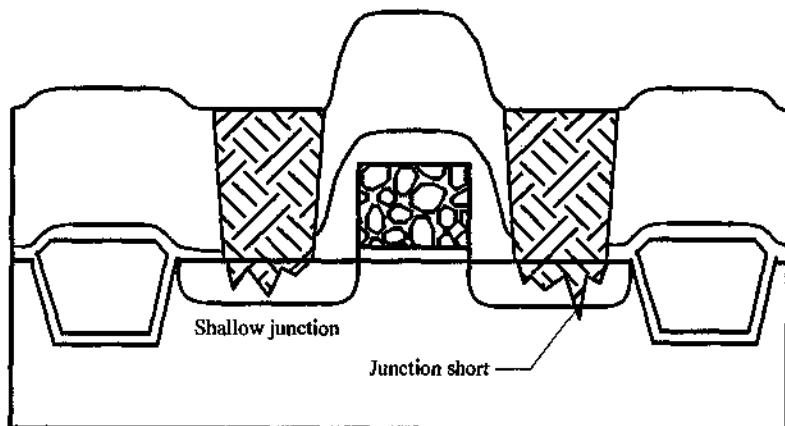


FIGURE 12.5 Junction Spiking

Aluminum-Copper Alloys

Aluminum metal was selected as the primary interconnect material for ICs because of its low resistivity and its compatibility with silicon and the wafer fabrication process. However, aluminum suffers from a reliability problem known as *electromigration*. This is the movement of aluminum atoms in the conductor due to momentum transfer from the electrons carrying the current.⁹ Under high current-density conditions, electrons collide with aluminum atoms, causing the atoms to gradually move. This movement of atoms leads to a depletion of atoms at the negative end of the conductor. In conductor regions where depletion occurs, this action leads to voids, thinning of lines, and a potential open circuit. In other regions of the conductor that have an accumulation of metal atoms, metal atoms pile up and form hillocks (see Figure 12.6).¹⁰ Hillocks are protrusions on the surface of metal films due to electromigration. If excessive or large hillocks form, adjacent lines or lines on two levels can short together. With advanced circuitry designs in ULSI technology, there is increased current density and chip temperature, both of which make the aluminum chip metallization more prone to electromigration.

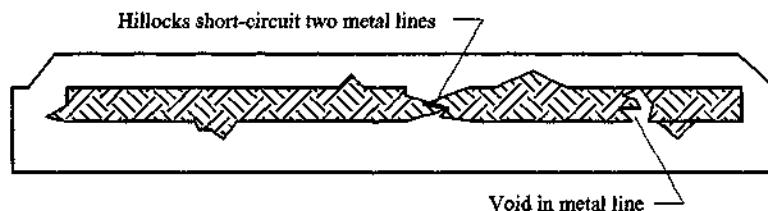


FIGURE 12.6 Hillock on a Metal Line

Electromigration in interconnects is controlled by alloying the aluminum with between 0.5 to 4% copper. This alloying essentially increases the conductor's current-carrying capacity by reducing the grain boundary diffusion effects in the aluminum. At the same time, it has been shown that values of copper in aluminum greater than 8% will actually increase electromigration.¹¹ The exact reason why the addition of copper reduces the chance of electromigration failure in Al-Cu alloys is not well understood.¹² If copper is alloyed with aluminum, more care must be given to plasma etching this alloyed aluminum (etching is discussed in Chapter 16). Copper is difficult to etch, and any residual copper remaining after etching the aluminum alloy interconnect can promote corrosion.¹³ It is also possible to have electromigration in contacts and vias. Electromigration failure in contacts is resolved by the use of barrier metallization.

Electromigration is probably one of the most widely studied failure mechanisms of integrated circuits. Studies of this failure began in the 1950s with the development of the semiconductor and intensified in the late 1960s with the discovery of crack formation in the aluminum conductor leading to failures.¹⁴ The significance of electromigration is that it usually occurs after the chip has been in service for an amount of time, which means that failure occurs catastrophically during customer usage.

Copper

A major transformation is underway in IC design and fabrication by means of the introduction of *copper interconnect* technology. Copper interconnects will replace aluminum metallization with copper to achieve significant benefits in chip performance. A legitimate question that someone may pose is why introduce copper if aluminum has performed so admirably over the years.

Need for Copper ■ The benefits for introducing copper for IC interconnect metallization are:¹⁵

1. **Reduction in resistivity.** The interconnect wiring resistivity reduces from $2.65 \mu\Omega\text{-cm}$ for aluminum to $1.678 \mu\Omega\text{-cm}$ at 20°C for copper, reducing RC signal delay and increasing chip speed.
2. **Reduction in power consumption.** Narrower lines consume less power.

4. **Tighter packing density.** Narrower lines permit tighter circuitry packing, which means that fewer levels of metal are needed.
5. **Superior resistance to electromigration.** Copper does not have a concern for electromigration.
6. **Fewer process steps.** Potential for 20 to 30% fewer processing steps with damascene processing of copper.

As wafer fabrication design rules reduce to 0.15-μm linewidths and below, the increased packing density of devices on the chip permits more electrical signal speed from device to device (the transistors are closer; therefore, the signal has less distance to travel). This density leads to improved chip performance. However, this improved chip performance is only possible if the interconnect system between the devices is optimized. Narrower linewidths lead to increased line resistance. Tightly spaced conductor lines with a dielectric material between them act as capacitors, leading to a degradation in performance from an increase in resistance (R) and capacitance (C). This condition is the signal delay or interconnect delay discussed in Chapter 11. If either or both of these two parameters are reduced, then the signal delay reduces, leading to increased chip performance.

One possible method reducing interconnect resistance is to increase the conductor cross section. However, this contradicts the goal of increased packing density since wider conductors will take substantially more space. Smaller IC feature sizes are not achievable with larger linewidths. This is why the semiconductor industry is placing emphasis on copper—to lower the interconnect resistance, which lowers R and therefore the overall signal delay. The optimum improvement to RC signal delay is gained when resistance (R) is reduced and the capacitance (C) is lowered by using a low- k dielectric along with thinner barrier metals (see Table 12.2).¹⁶

TABLE 12.2 Change in Interconnect Delay Compared to 0.25-μm Device Generation

Technology	0.25 μm	0.18 μm	0.13 μm
Conventional Interconnect Technology: • Al/Cu interconnect alloy and TiN barrier metal	0	+21%	+93%
New Technology Introduced by Generation: • Reduced barrier thickness • Low- k (3.0) dielectric • Dual damascene Cu interconnect and plugs	-10%	-27%	-16%

Another benefit from the implementation of copper is that smaller linewidths can carry the same amount of current, permitting a tighter packing density on each metal level. This condition produces fewer overall levels of metal on a chip, leading to significantly reduced manufacturing costs.¹⁷

Copper is a relatively soft metal. It has superior resistance to electromigration, a common reliability problem with aluminum.¹⁸ This means that chips fabricated with copper can handle higher electrical power densities, which permits the development of new product applications. Table 12.3 compares Cu and Al for different properties and processes common in wafer fabrication.¹⁹

TABLE 12.3 Comparison of Properties/Processes Between Al and Cu

Property/Process	Al	Cu
Resistivity ($\mu\Omega\text{-cm}$)	2.65 (3.2 for Al—0.5% Cu)	1.678
Electromigration resistance	Low	High
Corrosion resistance (in air)	High	Low
Etch processing	Yes	No
CMP (chemical mechanical planarization) processing	Yes	Yes

Copper Challenges ■ There are three major challenges involved with using copper as semiconductor interconnects that distinguish it from traditional aluminum interconnects. These three challenges differ significantly from aluminum technology and must be resolved before the implementation of copper for IC interconnects:²⁰

1. Copper diffuses quickly into oxides and silicon. This is a concern if copper diffuses into the active region of the silicon (i.e., source/drain/gate region of the transistor) because it will damage the device by creating junction or oxide leakage.
2. Copper cannot be easily patterned using regular plasma etching techniques (see Chapter 16 for a discussion of etching). Copper dry etching does not produce a volatile by-product during the chemical reaction that is necessary for economical dry etching.
3. Copper oxidizes quickly in air at low temperatures ($<200^{\circ}\text{C}$) and does not form a protective layer to stop further oxidation.

These challenges are addressed by converting to dual damascene processing with special barrier metals optimized for copper. Damascene processing eliminates the need to etch copper. Furthermore, tungsten plugs are expected to be used as the first level of metal to contact the source, drain, and gate regions.²¹ This use of tungsten overcomes the concern of copper contaminating the silicon (referred to as copper poisoning). The tungsten may even be patterned for wiring of the local interconnect (LI). All other metal lines and vias for the multiple metal layers are expected to be copper.

Another major hindrance to converting to copper in semiconductor manufacturing is the natural reluctance to introducing new materials into production. New materials means there are new sources of contamination, new equipment and procedures, unpredictable results, and so on. Change always brings unforeseen problems that pose increased risks to the semiconductor manufacturer. However, these risks are balanced by the benefits from changing the interconnect material in order to improve IC performance.

Barrier Metals

A more effective way to make reliable ohmic contacts to shallow junctions without material diffusion or problems such as junction spiking is to use barrier metallization. A *barrier metal* is a thin layer of deposited metal or metals that is designed to prevent intermixing of the materials above and below the barrier (see Figure 12.7). The thickness of barrier metals has typically been around 100 nm for features sizes in the $0.25\text{-}\mu\text{m}$ generation, which is substantially thinner than the 400 to 600 nm thickness required for $0.35\text{-}\mu\text{m}$ generation of devices. Thickness of barrier metals is projected to decrease to 23 nm or less for the $0.18\text{-}\mu\text{m}$ technology node and beyond.²²

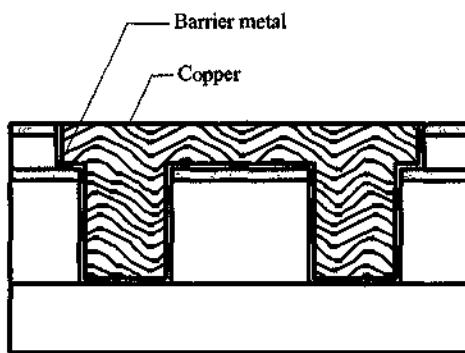


FIGURE 12.7 Barrier Layer for Copper Interconnect Structure

Barrier metals are widely used in semiconductor manufacturing. For contacts between the tungsten (W) plug that connects the aluminum interconnect wiring and the silicon source/drains, the barrier metal prevents the silicon and W from coming into contact with each other. This barrier prevents diffusion between the W and silicon and any junction spiking.

The essential properties of an acceptable barrier layer metal are:

1. Good diffusion barrier properties so that the diffusivity of the two interface materials (e.g., tungsten and silicon) is low at the sintering temperature (*sintering* refers to joining the materials by thermal means).
2. High electrical conductivity with low ohmic contact resistance.
3. Good adhesion between the semiconductor and the metal.

4. Resistance to electromigration.
5. Stability when thin and at high temperature.
6. Resistance to corrosion and oxidation.

Metals commonly used for barrier metals are a class of high melting point metals known as *refractory metals*. Common refractory metals used in multilevel metallization for wafer fabrication are titanium (Ti), tungsten (W), tantalum (Ta), molybdenum (Mo), cobalt (Co), and platinum (Pt). Refractory metals have been used in wafer fabrication since the 1960s, such as in the formation of Schottky barrier diodes in bipolar processes. The benefits of using titanium as a barrier metal for aluminum alloy wiring are improved adhesion, reduced contact resistance, reduced stress, and controlled electromigration. To achieve good barrier properties, wafers undergo a clean step (referred to as *sputter etch* and explained in a later section) before deposition in the processing chamber to remove native oxides and oxide residues on the wafer.

Titanium tungsten (TiW) and titanium nitride (TiN) are two common barrier metal layers that inhibit diffusion between the silicon substrate and aluminum. TiN is widely used in ULSI manufacturing for its superior barrier performance in aluminum alloy interconnect processing. TiN functions as a barrier metal for both tungsten and aluminum. TiN is also widely used as an antireflective coating on aluminum to improve the photolithography patterning process (see Chapter 14). However, TiN does not produce good contact resistance to silicon. To resolve this problem, a thin layer of titanium (e.g., several hundred angstroms or less) is typically deposited before the TiN so that it can react with the underlying material as a silicide (see the following section for an explanation of silicide) and lower the contact resistance. The Ti and TiN are typically deposited in a cluster tool to keep oxides from forming between the two layers.

Copper Barrier Metals ■ Barrier metals are critical for copper metallurgy. Cu has high diffusivity in silicon and silicon dioxide, which will destroy device performance. Traditional barrier metals are not sufficient barriers for copper. Cu requires complete encapsulation by a thin-film barrier layer that functions as an adhesion promoter and effective diffusion barrier.²³ There is a trade-off between these two requirements since good adhesion requires some reaction with Cu, whereas a good barrier metal should not react with Cu. The special barrier metal requirements for copper are:²⁴

1. Prevent copper diffusion.
2. Low film resistivity.
3. Good adhesion to both dielectric material and copper.
4. Compatible with chemical mechanical planarization (CMP).
5. Metal layer is continuous and conformal with good step coverage and deposition in high aspect ratio gaps.
6. Minimal thickness to allow the copper to occupy the maximum cross-sectional area.

For Cu interconnect metallurgy, tantalum (Ta), tantalum nitride (TaN), and tantalum silicon nitride (TaSiN) are candidate materials for barrier metals (see Figure 12.8). The diffusion barrier must remain thin (about 75 Å) so that it does not affect the resistivity of the high-aspect ratio plug while still acting as a barrier metal. This condition is difficult to maintain as geometries continue to shrink and metals are deposited into high-aspect ratio vias. Research has shown Ta has good barrier and adhesion properties for copper, whereas TiN, a traditional barrier metal for Al/SiO₂ interconnects, has good barrier properties but poorer adhesion.²⁵ If TaN is used, it is obtained by small amounts of nitrogen doping or by deposition with a tantalum nitride compound. Investigations have also shown tungsten nitride (WN) functions as an effective barrier for Cu metal.²⁶ There is ongoing research in the development of barrier metals for Cu metallurgy.

Copper barrier layers can be deposited by using high-density plasma CVD (HDPCVD) or ionized metal plasma physical vapor deposition. Ionized metal plasma PVD for tantalum achieves good step coverage. For deposition with high-aspect ratio gaps, then HDPCVD barrier deposition is often the choice.²⁷

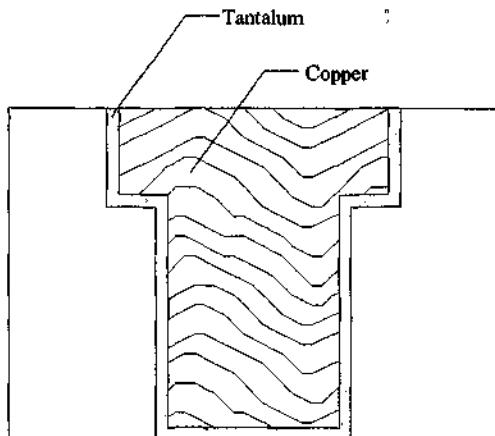


FIGURE 12.8 Ta for Copper Barrier Metal

Silicides

Refractory metals react with silicon when alloyed together to form a silicide. A *silicide* is a metal compound that is thermally stable and provides for low electrical resistivity at the silicon/refractory metal interface (see Figure 12.9). Refractory metal silicides are important in wafer fabrication because of the need to reduce the electrical resistance of the many silicon contacts in the source/drain and gate regions for chip performance. Titanium and cobalt are common refractory metal used for contacts in aluminum interconnect technology.

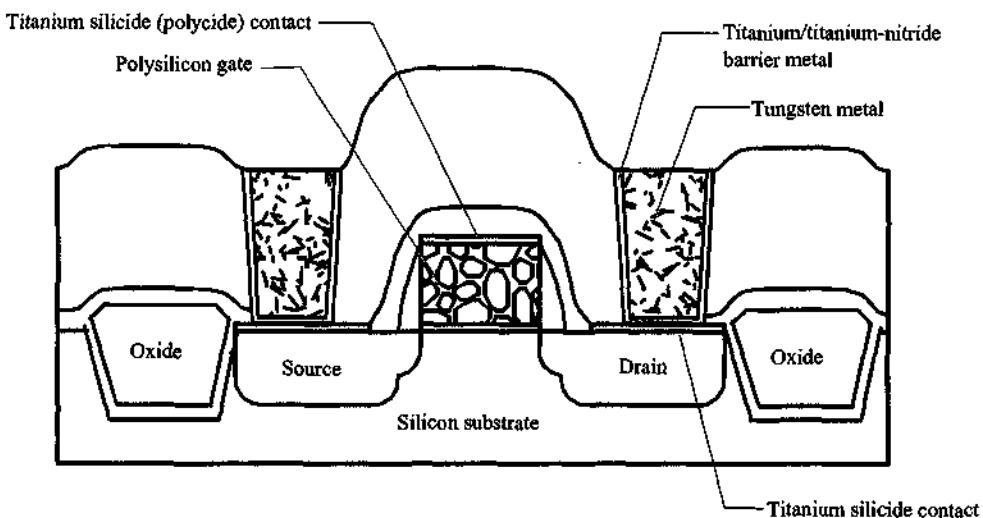


FIGURE 12.9 Refractory Metal Silicide at a Silicon Contact

If the refractory metal is reacted with polysilicon, it is called a *polycide* (see Figure 12.10 on page 310). Doped polysilicon is used as the gate electrode and has a relatively high resistivity (about $500 \mu\Omega\text{-cm}$), which leads to an undesirable RC signal delay. A polycide is beneficial for reducing the series resistance of an interconnection to polysilicon. At the same time, it maintains polysilicon's good interfacial characteristics to oxides.

Silicides improve contact resistance by reducing residual oxides on the silicon surface during silicide formation. Silicides form extremely good metallurgical contact to the silicon, serving as a critical adhesion layer between the contact metal and the silicon junction regions. Many silicides are stable at temperatures exceeding 1000°C with a relatively high eutectic temperature. Table 12.4 on page 310 lists some properties of common silicides used in wafer fabrication.²⁸ Recall that the eutectic temperature is the lowest temperature where the alloy melts. It is undesirable for the silicide to melt because the liquid alloy can extend into the substrate material and cause junction spiking.

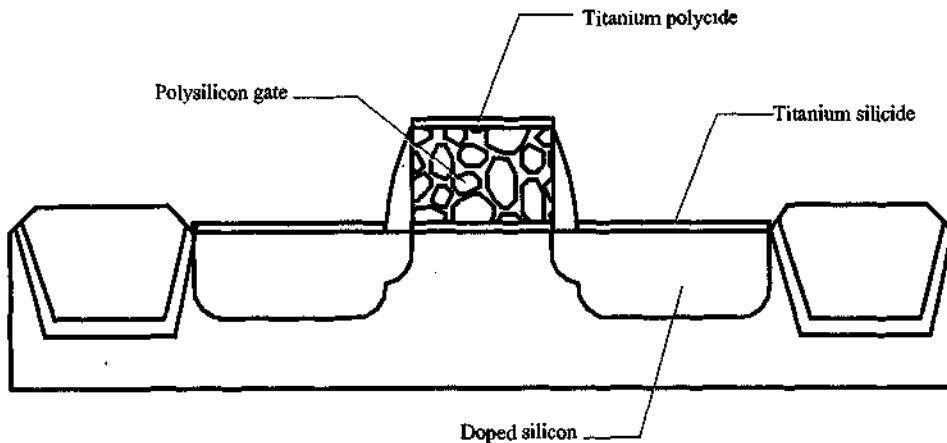


FIGURE 12.10 Polycide on Polysilicon

TABLE 12.4 Some Properties of Select Silicides

Silicide	Lowest Eutectic Temperature (°C)	Typical Forming Temperature* (°C)	Resistivity ($\mu\Omega\text{-cm}$)
Cobalt/silicon (CoSi_2)	900	550–700	13–19
Molybdenum/silicon (MoSi_2)	1410	900–1100	40–70
Platinum/silicon (PtSi)	830	700–800	28–35
Tantalum/silicon (TaSi_2)	1385	900–1100	35–55
Titanium/silicon (TiSi_2)	1330	600–800	13–17
Tungsten/silicon (WSi_2)	1440	900–1100	31

* B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, (Boston: Kluwer Academic Publishing 1995), p. 537.

Silicide formation typically requires the refractory metal be deposited on the silicon wafer, followed by a high temperature thermal anneal process to produce the silicide material. In areas where there is silicon, the metal reacts to form a silicide. In other areas on the wafer surface, such as where there is SiO_2 , there is little or no silicide formation. Often this thermal anneal step uses a rapid thermal anneal (RTA) process in a multichamber cluster tool.

TiSi_2 has traditionally been the most common contact silicide for wafer fabrication, serving as the contact between the transistor active regions of silicon and the tungsten plug. It is often referred to as the glue that holds the tungsten to the silicon. Its beneficial properties are high temperature stability, compatibility with self-aligned contact processing (discussed in the following section), low resistivity when compared to other silicides, and its compatibility with TiN barrier metals.²⁹ TiSi_2 forms two different grain phases (a phase is a physically homogeneous material state) as it is annealed, a low-temperature C49 phase and a high-temperature C54 phase (see Figure 12.11). Both of these phases are TiSi_2 .

The C49 phase of TiSi_2 forms at an anneal temperature of 625 to 675°C with a resistivity of 60 to 65 $\mu\Omega\text{-cm}$. The C54 phase is formed in a second anneal following the C49 phase formation and requires a temperature of 800°C or more. It has a much lower resistivity of only 10 to 15 $\mu\Omega\text{-cm}$, which is desirable to lower the overall contact resistance.³⁰

However, at the time of this writing, TiSi_2 appears limited in its future use as a contact silicide in sub-0.25 μm technologies. Contacts for ultrashallow source/drain junctions are becoming thin. TiSi_2 is undesirable for thin contacts because the silicide resistivity increases.³¹ It is also difficult to form this contact during the second anneal because of critical time at temperature requirements.

The silicide that appears promising for the 0.18 μm and below technology nodes is cobalt silicide (CoSi_2). This silicide maintains a reduced contact resistance of 13 to 19 $\mu\Omega\text{-cm}$ after its

	Sintering Temperature	Resistivity
TiSi ₂ -C49	625 – 675°C	60 – 65 $\mu\Omega\text{-cm}$
TiSi ₂ -CS4	800°C	10 – 15 $\mu\Omega\text{-cm}$

FIGURE 12.11 Anneal Phases of TiSi₂

anneal operation, even with deep submicron geometries of 0.18 μm and less. This reduced resistivity occurs because the grain size of CoSi₂ is about ten times smaller than the grain size of TiSi₂. Therefore, the low resistance phase is completely nucleated and grown during the thermal anneal operation. A CoSi₂ contact is also easier to form because of the smaller grain size.

Note that a silicide is not a barrier metal. In some silicides it is found that silicon diffuses rapidly through them. This diffusion occurs during heat treatment of the metal-silicide-silicon system where silicon diffuses through the silicide to the metal, which then reduces the integrity of the system. The solution to this problem is to place a metal barrier layer between the silicide and the metallization layer (discussed in the preceding section). A common silicide barrier film is titanium nitride (TiN), which is effective for both tungsten and aluminum. A tantalum-based barrier layer is anticipated for copper metallurgy.

Salicide ■ Because of the need to scale device sizes in VLSI and ULSI to optimize performance, the cross-sectional area of the source/drain (S/D) contact between the silicon and the first metal layer is small. This small size leads to increased resistance. A technique to decrease contact resistance at the source and drain areas by providing stable contact structures is referred to as a salicide. A *salicide* approach (taken from the expression *self-aligned silicide*) is used to create silicides that are properly aligned with the exposed silicon of the source, drain, and polysilicon gate. There are numerous chip performance issues that depend on salicide formation (see Figure 12.12).

The basic salicide steps are shown in Figure 12.13 on page 312. This process flow corresponds to Step 6 of the CMOS process flow described in Chapter 9. To form a salicide, oxide has been previously deposited and etched back with a dry plasma etch to leave oxide sidewall spacers

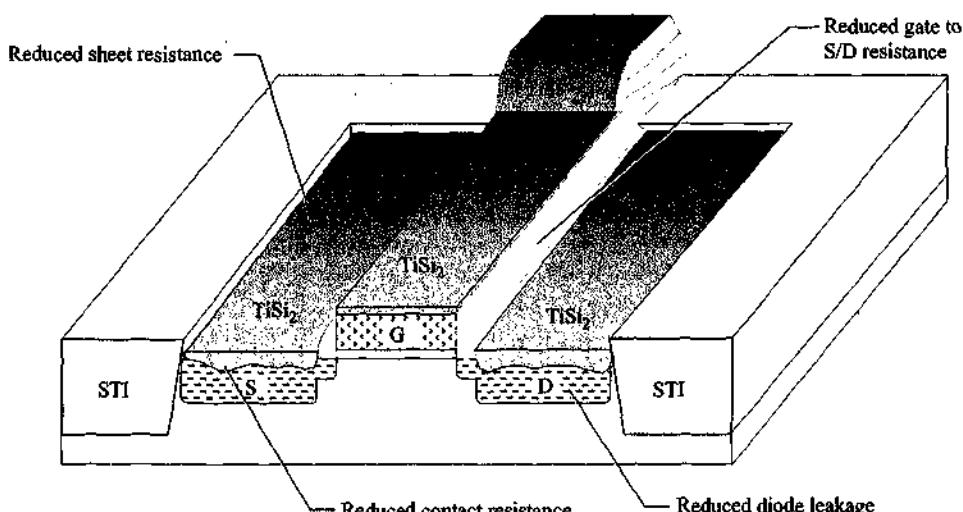


FIGURE 12.12 Chip Performance Issues Related to a Salicide Structure

on the side of the polysilicon gate (Step 5 of the CMOS process flow described in Chapter 9). With the sidewall spacers, only the top poly surface of the gate is exposed. After an HF-dip clean step to remove native oxides, a 250 to 350 Å thin layer of titanium metal is deposited on the wafer. The refractory metal undergoes an RTP rapid anneal at 600 to 800°C to create the high-resistivity C49 titanium silicide phase wherever the refractory metal contacts silicon. The sidewall spacer keeps the S/D TiSi_2 from shorting out to the sides of the gate poly, while the oxide of the shallow trench isolation separates the devices. After the first RTP anneal, all unreacted titanium is removed by a wet chemical etch in ammonium hydroxide (NH_4OH) and hydrogen peroxide (H_2O_2). TiSi_2 remains and covers the S/D areas and top surface of the poly gate. A second RTP silicide anneal at 800 to 900°C creates the low-resistivity C54 metallic silicide. A major benefit of the salicide process is the avoidance of alignment tolerances that would occur if patterning were required to align the refractory metal on the silicon contact. Care must be taken during the anneal steps to avoid oxygen contamination in the furnace tube. Titanium reacts easily with oxygen to form undesirable titanium oxides. This oxide contamination can promote silicide formation on top of the oxide spacer regions, leading to a short between the polysilicon gate and source or drain.

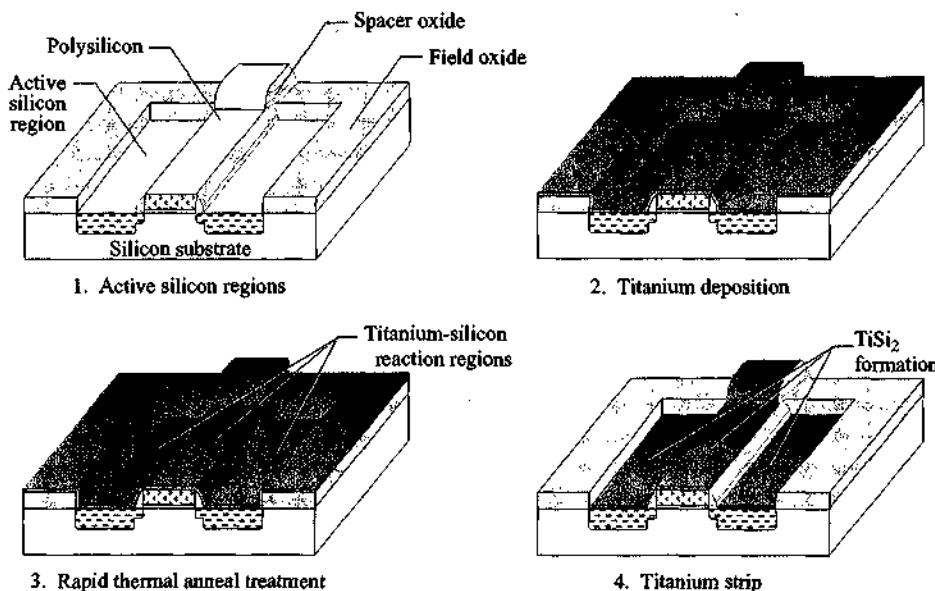


FIGURE 12.13 Formation of Self-Aligned Metal Silicide (Salicide)

Metal Plugs

Multilevel metallization creates the need for billions of vias filled with metal plugs to form electrical pathways between two metal layers. A contact plug is also used to connect the silicon devices in the wafer to the first level of metallization. The most common metal currently used for plugs is tungsten (W), which is why plugs are often referred to as *tungsten plugs* (see Figure 12.14). Tungsten has been traditionally used as a plug material because of its ability to uniformly fill the high-aspect ratio vias when deposited by chemical vapor deposition (CVD) methods. Tungsten is resistant to electromigration failure. It also serves as a barrier to inhibit diffusion and reaction between Si and the first metal layer. Tungsten is a refractory material with a melting point of 3,417°C and a bulk resistivity of 52.8 $\mu\Omega\text{-cm}$ at 20°C.

Aluminum would be desirable as a plug material because of its lower resistivity (2.65 $\mu\Omega\text{-cm}$ at 20°C), but sputtered aluminum cannot fill high-aspect ratio vias (see the next section for a description of sputtering). For this reason aluminum is used as the interconnect material and tungsten is limited to the plug. There has been recent interest in reflow aluminum plugs, where aluminum is sputter deposited into the via and then reflowed by a high-temperature anneal with a rapid thermal processor (RTP).³²

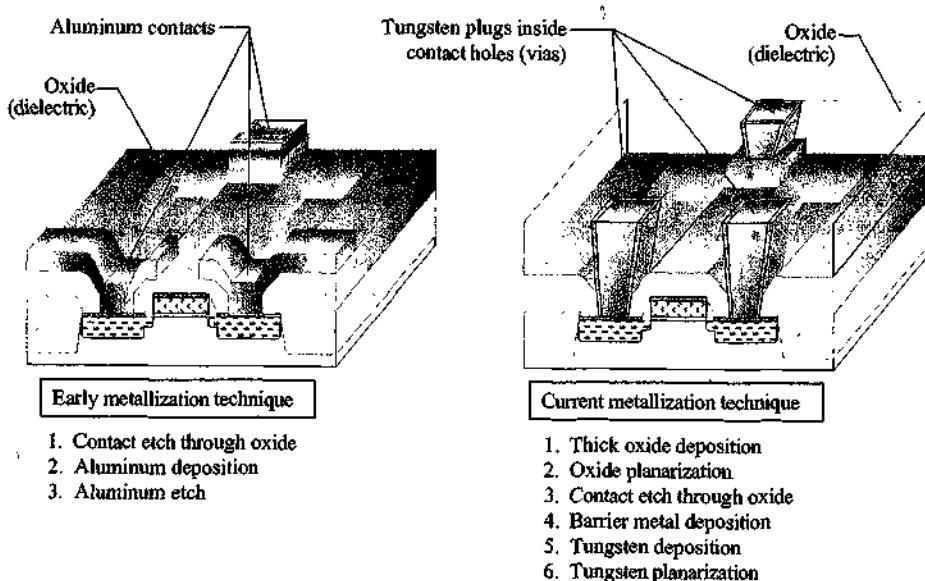


FIGURE 12.14 Tungsten Plug for Multilevel Metal Layers

METAL DEPOSITION SYSTEMS

The traditional metallization techniques used in semiconductor manufacturing fall under a category called *physical vapor deposition (PVD)*. PVD has been done with filament evaporation, followed by electron beam evaporation and, most recently, sputtering. Chemical vapor deposition (CVD) has recently become a frequently used technique to deposit metal films. Each change in the deposition system has brought improved control of thin film properties and quality.

During the SSI and MSI eras of semiconductor manufacturing, evaporation was the method for metallization. It was replaced by sputtering primarily because evaporation has poor step coverage. Evaporation is still used in research applications and for semiconductors using III-V technologies. It is also used in some specialized areas, such as for C4 bump deposition during packaging (see Chapter 20).

Electroplating technology has been employed in various applications. In recent years it has been used for thin film head metallization in the disk drive industry. However, electroplating is just now being introduced for semiconductor manufacturing as a copper deposition method. It is the early deposition method of choice for creating copper interconnects.

The different metal deposition systems used in traditional and dual damascene metallization are:

- ◆ Evaporation
- ◆ Sputtering
- ◆ Metal CVD
- ◆ Copper electroplate

Evaporation

In the early days of semiconductor manufacturing, all metal layers were deposited by the PVD method of *evaporation*. Since the late 1970s, sputtering has replaced evaporation in most silicon wafer technologies in order to gain improved step coverage, gap fill, and yield. It is beneficial to review evaporation to understand how it functions and why the silicon industry changed to sputtering.

The process of evaporation consists of placing the material to be deposited in a crucible and heating it inside a vacuum chamber until it vaporizes (see Figure 12.15 on page 314). The most typical method of heating was the use of an electron beam to heat the metal placed in a crucible. By maintaining a high vacuum in the evaporator, the mean free path of the vapor molecules is increased, and the vapor travels in a straight line in the chamber until it strikes a surface and condenses to form a film.

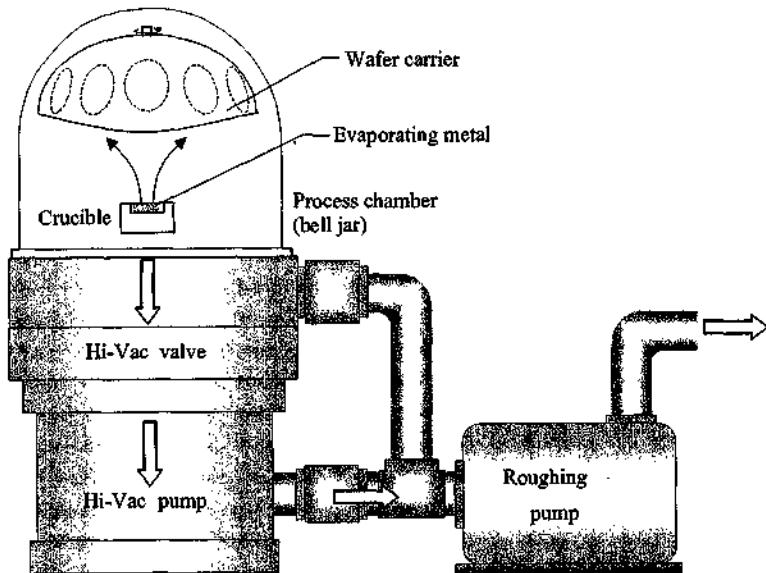


FIGURE 12.15 Simple Evaporator

The biggest drawback to evaporation was the inability to produce uniform step coverage. As the industry progressed to VLSI and ULSI technology, metallization needed to be capable of filling higher aspect ratio holes and producing conformal step coverage. Evaporators made some progress for step coverage by heating the wafers and rotating the wafers in the chamber with a hemispherical cage. However, evaporator technology is not able to form a continuous film over features with an aspect ratio greater than 1.0:1, and is marginal for an aspect ratio of 0.5:1 to 1.0:1. This drawback to evaporation led to its demise in IC production.

Another serious drawback of evaporation is its limitation for depositing alloys. The evaporator requires multiple crucibles for the different materials, which is a problem because of the different vapor pressures of the materials. This limitation makes it difficult to control the ultimate composition of the deposited alloy with any degree of accuracy.

Ultimately, the wafer fabrication industry quickly replaced evaporators with sputtering systems due to their improved step coverage capability. The evaporator is still used in research applications and for some compound semiconductor technologies that can actually use the poor step coverage of evaporators to their advantage during specialized processing. In addition, evaporators are still sometimes used in the chip packaging process to deposit solder bumps on the chip surface (see Chapter 20).

Sputtering

Sputtering, a form of physical vapor deposition (PVD), was discovered in 1852 by Sir William Robert Grove and developed as a thin film deposition technique in the 1920s by Langmuir.³³ As its name implies, sputtering is mainly a physical, rather than chemical, process. In the sputtering process, high-energy particles strike a solid slab of high-purity target material and physically dislodge atoms. These sputtered atoms migrate through a vacuum and eventually deposit on a wafer.

The advantages of sputtering are:³⁴

1. Ability to deposit and maintain complex alloys.
2. Capability to deposit high-temperature and refractory metals.
3. Ability to deposit controlled, uniform films on large wafers (200 mm and larger).
4. Capability of multichamber cluster tools to clean the wafer surface for contamination and native oxides before depositing metal (referred to as *in situ sputter etch*).

While sputtering was a big improvement in gap fill over the earlier metallization method of evaporation, its limited capability for step coverage and filling high-aspect ratio gaps makes it insufficient for ULSI applications.³⁵ Over the years, improvements have been made to the sputtering

process to obtain better step coverage, including recent advances with ionized metal plasma PVD. For critical applications, such as the step coverage in tungsten plugs, deposition will usually be done with a CVD metal process. Sputtering continues to be used for depositing critical barrier and seed layers, such as tantalum/tantalum nitride used in copper metallization (when the aspect ratio is reasonable).

Basic Sputtering Steps ■ There are six basic steps to sputtering:³⁶

1. Positive argon ions are generated in a plasma in a high vacuum chamber and accelerated toward a target material at a negative potential.
2. During acceleration, the ions gain momentum and strike the target.
3. The ions physically dislodge (sputter) atoms from the target, which has the desired material composition.
4. The dislodged (sputtered) atoms migrate to the wafer surface.
5. The sputtered atoms condense and form a thin film on the wafer surface with essentially the same material composition as the target, following the stages for thin film growth outlined in Chapter 11.
6. Excess material is removed from the chamber by vacuum pump.

To illustrate sputtering, Figure 12.16 shows the basic configuration of a simple, DC diode sputter deposition chamber with parallel plates. It consists of the solid *target* material, a substrate (wafer), and a vacuum enclosure. The target is electrically grounded and referred to as the *cathode*, while the substrate has a positive potential and is termed the *anode*. The anode and cathode are both called *electrodes*. The sputtering target is composed of the necessary material for deposition. Targets are manufactured to achieve a homogeneous composition, fine grain size, and specific crystallographic orientation, all of which contribute to achieving a uniform film deposition rate across the wafer.³⁷ Target purity levels of 99.999% (5 nines) or better are required for achieving acceptable film purity for sub-0.25 μm geometries.

A high density of positive ions from an argon gas glow discharge is strongly attracted to the negative target material, striking it at high velocity and dislodging the atoms to be deposited. The atoms are sputtered (knocked off) from the target material and scatter in the chamber; eventually some come to rest on the wafer or the chamber walls. This action necessitates cleaning of the chamber in some systems. The atoms on the wafer nucleate and grow a thin film. An important aspect of sputtering is that this process can be used to sputter-deposit alloys, in particular aluminum-silicon

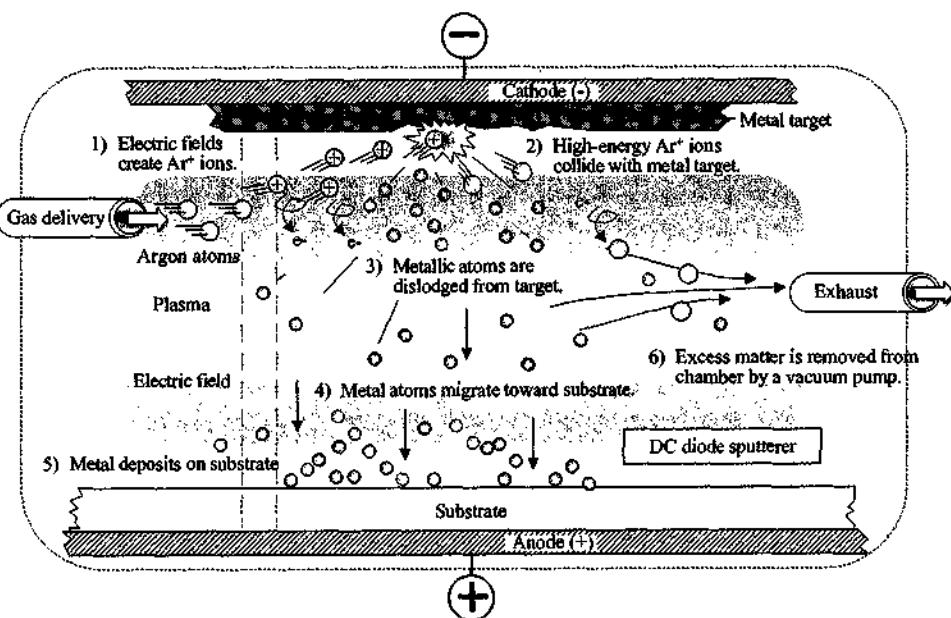


FIGURE 12.16 Simple Parallel Plate DC Diode Sputtering System

and aluminum-copper alloys. For instance, if the target material is aluminum with 1% copper, then the deposited film will be aluminum with 1% copper (at steady-state conditions).

The latest technology of sputtering equipment systems are multichamber cluster tools. The system has a loadlock that moves single wafers from the wafer cassette to its ultimate base vacuum pressure. The loadlock may consist of dual loadlocks to move the wafer through a staged vacuum sequence from ambient atmosphere to base vacuum pressure. A robot transfer system designed specifically for the cluster tool environment moves wafers from the various process chambers and a buffer chamber. It is important that the robot operate without generating particles and be extremely reliable. Cluster tool robots are often magnetically coupled, which permits the robot drive motor to remain on the outside of the process chambers to reduce particles.

Vacuum conditions are important in the sputtering chamber to create the plasma and maintain purity of the deposited films, with base vacuum typically at 10^{-7} torr during initial pump down. The rate that argon gas enters the process chamber is critical because it causes the pressure to rise in the chamber. With the argon and sputtered material in the chamber, the pressure rises to about 10^{-3} torr. The hot, high-vacuum environment of the process chamber promotes the formation of the sputtered atoms into thin films on the wafer surface.

Physics of Sputtering ■ An essential aspect of sputtering is the formation of the ionized argon gas into a plasma (see Chapter 8). Argon is used as the sputtering ion species because it is relatively heavy and is a chemically inert gas, which keeps it from reacting with the growing film or the target. If a high-energy electron strikes the neutral argon, the collision knocks off outer electrons and creates a positively charged argon ion. This energetic particle is used to strike the negatively charged target material to be sputtered.

Sputtering Mechanism. Positively charged argon ions in the plasma are strongly attracted to the negative potential of the cathode target. The charged ions accelerate and acquire kinetic energy (or energy of motion) as they pass through the voltage drop of the glow-discharge dark space (see Chapter 8 for a description of glow discharge). When the Ar ions strike the target surface, the momentum of the Ar ion transfers to the target material to dislodge one or more atoms. This action is referred to as *momentum transfer*. The ejected neutral atom or atoms move through the plasma (with a small chance of being ionized) and land on the wafer. The incident ion energy must be large enough to dislodge target atoms but not so large that the ion penetrates into the target material. Typical sputtering ion energies range from 500 to 5,000 eV.

An analogy for dislodging metal atoms from the surface of the target material during sputtering is hitting billiard balls in a pool game. Even though the cue ball in pool is travelling in one direction, the billiard balls can be knocked in some other direction. This same effect occurs in sputtering, where the argon ion impacts the target and dislodges one or more atoms from the surface of the target (see Figure 12.17).

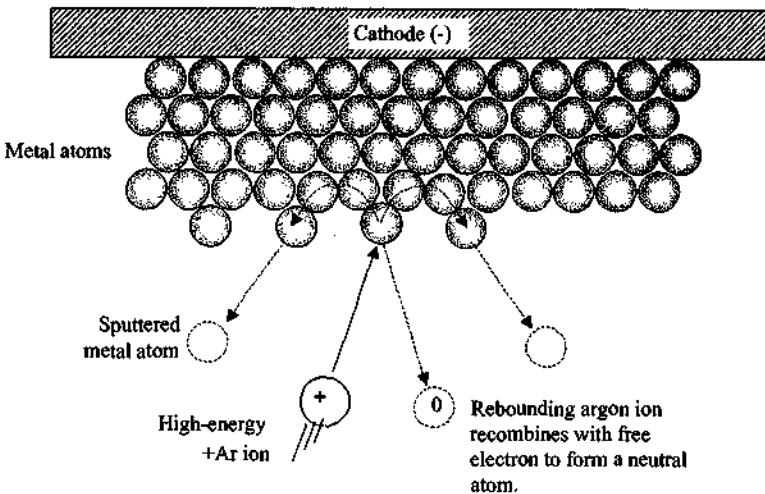


FIGURE 12.17 Dislodging Metal Atoms from Surface of Sputtering Target

The *sputtering yield* is defined as the number of atoms ejected by the target (cathode) per incident ion that strikes it. The yield largely determines the rate of sputter deposition. It varies from about 0.5 to 1.5. A sputter yield of 0.5 means that, on the average, two ions must strike the target for one atom to be ejected. The sputter yield depends on the following conditions:³⁸

1. Incident angle of the bombarding ions.
2. Composition and geometry of the target material.
3. Mass of bombarding ions.
4. Energy of the bombarding ions.

A method of increasing the sputtering deposition rate is to confine the plasma to the region between the target and wafers. Due to the ionization process, there are dark spaces just to the front and the sides of the target. Dark space shields are placed on the sides of the target to prevent target material from being sputtered on the sides, since this material will never deposit on the wafers. Shields require periodic changing because sputtered target material builds up on the surface and causes particulate contamination.

Targets undergo erosion from the ion bombardment and are replaced when about 50% or more of the target material is removed.³⁹ Much of the energy consumed during the sputtering process is dissipated as heat in the target, or by emission of secondary electrons and photons by the target. For this reason, cooling of the target material is done to maintain a low target temperature.

In addition to being struck by the sputtered atoms, the substrate has other species landing on it (see Figure 12.18). These species cause heating of the substrate (up to 350°C) which leads to uneven film deposition. The high temperature can also create an undesirable aluminum oxide during aluminum deposition, which interferes with the sputtering process. The many species impinging on the wafer surface during diode sputtering also increase the possibility of damage due to radiation to sensitive devices.

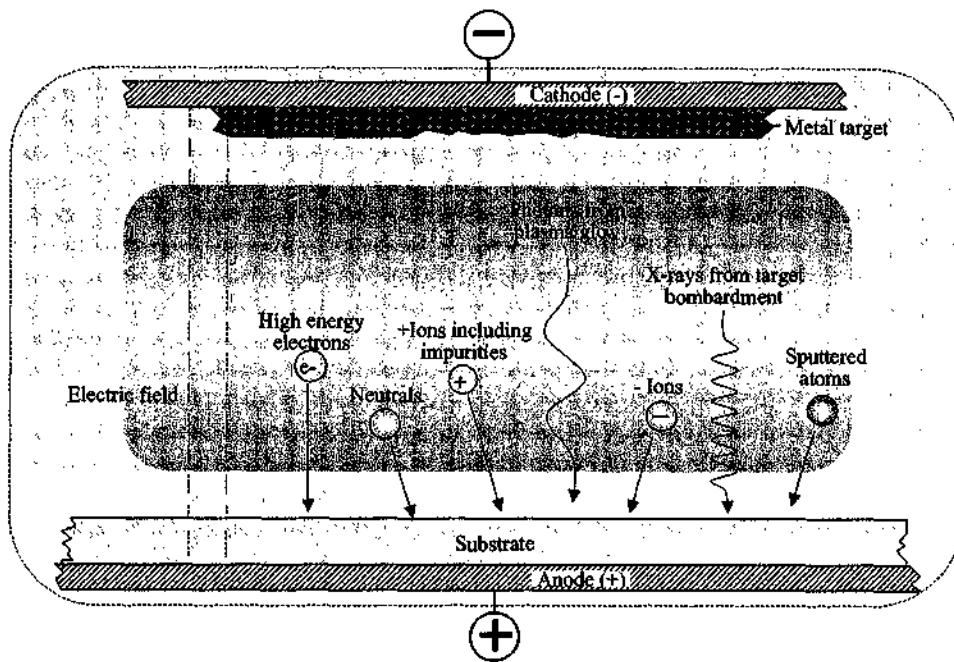


FIGURE 12.18 Different Species Landing on Substrate

Note the presence of the impurity gas atoms during diode sputtering. This is a problem if the impurity atoms are incorporated into the growing film on the substrate. The sources of contaminants are: (1) impure sputtering gas, (2) outgassing from the wafer holder, (3) outgassing from chamber walls, and (4) vacuum leaks.⁴⁰ It is necessary to use ultrahigh purity (UHP) sputter gases and to verify there are no chamber leaks.

The spacing between the cathode and anode must be optimized for each sputtering system. The goal is to have as many sputtered atoms as possible be deposited on the wafers and form the desired film.

The sputtering system described thus far is a simple DC diode system that has a DC voltage applied between two electrodes (cathode and anode). The DC diode sputtering system has serious drawbacks that limit its use in manufacturing. A DC diode type sputtering system cannot be used to sputter dielectrics because the electrodes become coated with the dielectric and the glow discharge cannot be sustained. The target rapidly builds up a positive charge, which repels incoming positive ions. DC diode type sputtering is also not capable of performing a sputter etch. *Sputter etch* (or *reverse sputter*) is a preclean step where the sputter process is reversed and argon atoms are used to remove thin native-oxide layers and remaining etch residues that contaminate contacts and vias. In other words, the wafer is sputtered instead of the target. This sputter etch preclean is important in multichamber cluster tools because of the benefit of cleaning followed by deposition without removing the wafer from the vacuum environment.

Three types of sputtering systems are covered in the following pages:

- ◆ RF (radio frequency)
- ◆ Magnetron
- ◆ IMP (ionized metal plasma)

The simple RF sputter system is not used in wafer fabrication because of its inherent inefficiency. The magnetron has traditionally been the most widely used sputter system, and ionized metal plasma (IMP) is becoming more common for sub-0.25 micron technology.

RF Sputtering ■ In an *RF sputtering* system, an RF field is used to create the plasma instead of the DC field described above. The RF frequency has commonly been 13.56 MHz. It is applied to the back surface of the target electrode and capacitively coupled to its front surface (see Figure 12.19). Both the electrons and ions in the plasma are exposed to the RF field, but due to the high frequency, the electrons respond most strongly. The chamber and electrode behave like a diode, creating a high amount of electron flow and resulting in a negative charge on the target electrode. This negative charge (produced by a self-bias) attracts positive argon ions, which sputter material from the insulator or non-insulator target.⁴¹

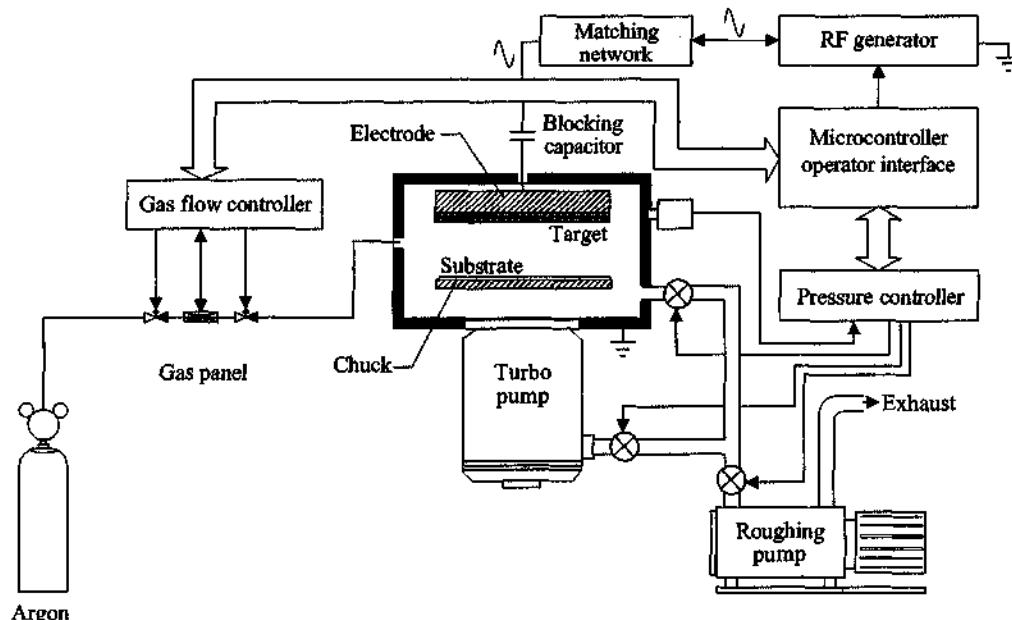


FIGURE 12.19 RF Sputtering System

The wafer can be electrically biased to put it at a different field potential than the argon. The bias applied to the wafer causes the argon atoms to strike the wafer directly. This RF biasing permits the exposed wafer surface to be sputter etched and cleaned.

In practice, an RF sputtering system is limited because it does not have a high sputtering yield, leading to a low deposition rate. Many of the secondary electrons emitted by the target pass through the discharge region without contributing to the creation of plasma. The sputtering rate of the target would be much higher if these secondary electrons were involved in ion collisions, leading to the creation of more ions to bombard the target. To overcome this inefficiency and to achieve high metal deposition rates necessary for wafer fabrication, the concept of magnetron sputtering was developed.

Magnetron Sputtering ■ *Magnetron sputtering* employs magnets configured around and behind the target to capture and restrict the electrons in front of the target (see Figure 12.20). This setup increases the ion bombardment rate on the target in order to produce more secondary electrons, which then increases the ionization rate in the plasma. The end result is that more ions cause more sputtering of the target, which increases the deposition rate of the system. The invention of magnetron sputtering to improve the deposition rate is one of the main reasons why sputtering has become the leading process for single-wafer deposition systems for aluminum and contact alloys used in metallization.

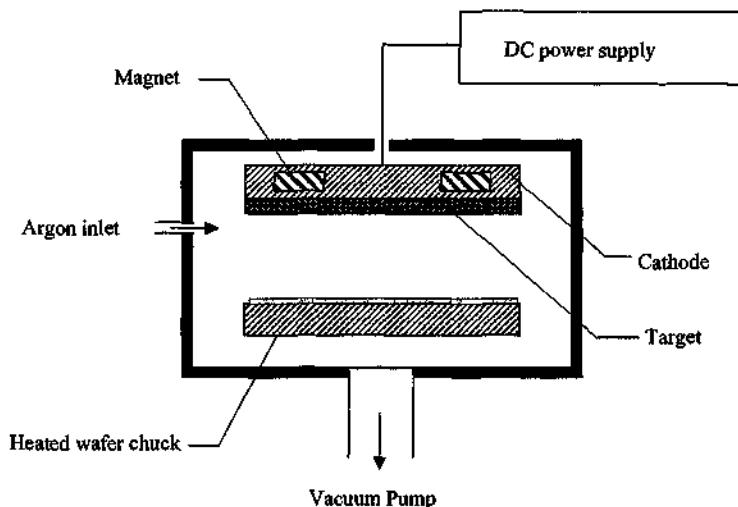


FIGURE 12.20 Magnetron Sputtering

The magnetron sputtering design requires a substantial amount of power supplied to the argon plasma to maximize the sputtering rate (about 3kW to 20 kW of power). Because the sputtering target absorbs most of this power and is in contact with the cathode, cooling of the cathode is required. An important challenge is uniformity of the sputtered deposit on large wafers. To achieve high deposition rates and film uniformity in single-wafer cluster tools, new cathodes have been developed with rotating, rare earth, high-strength permanent magnets.⁴²

Step Coverage. Sputtering requires a vacuum environment with high-purity argon to avoid contamination from residual gas. The vacuum during the sputtering process is around 1 mtorr, with a mean free path of several centimeters. This is approximately the distance between the target and wafer. Because of this distance, atoms dislodged from the target essentially pass on to the wafer by following a line-of-sight path.

This line-of-sight path for the ejected atoms exiting at many different angles from the target causes poor coverage on steps and sidewalls in contacts and vias. Due to their geometry, the sidewalls and bottoms of high-aspect ratio contact windows and vias will receive only 10% or less of the metal deposited on the top surface.⁴³ For this reason PVD is not usually chosen when depositing a material on high aspect ratio steps and trenches.

Collimated Sputtering. To achieve increased coverage on the bottom and side of a contact or via, directional enhancement can be achieved by the use of *collimated sputtering* (see Figure 12.21). The collimator is configured so that it appears as the electrical ground for the plasma. In this manner, any neutral species sputtered from the target at a high angle is intercepted and deposited on the collimator. Other atoms that are ejected straight back from the target will pass by the collimator and deposit on the bottom of the contact hole. At the same time, a collimator reduces sidewall coverage in the contact or via.

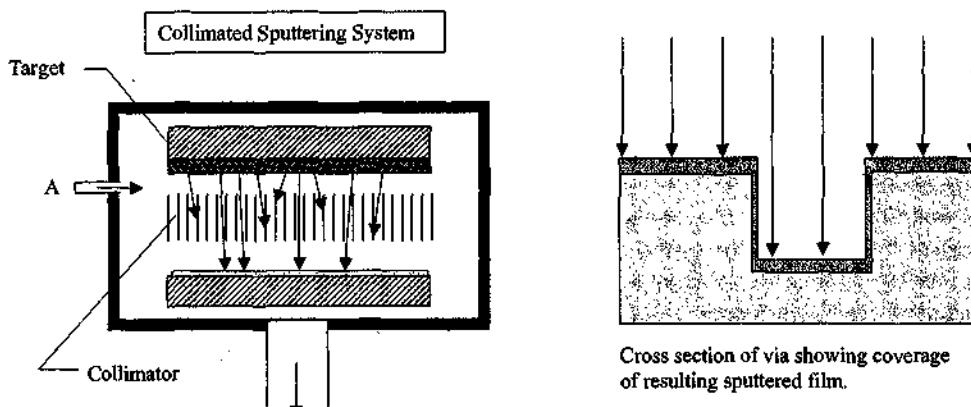


FIGURE 12.21 Collimated Sputtering

The use of a collimator means that a large portion of the sputtered material will not reach the wafer, since most of the sputtered material ends up on the collimator. This result lowers the sputter yield and increases the cost of deposition, since fewer sputtered atoms reach the wafer. At the same time, maintenance costs increase because the target is less efficient. Given these problems, it is more efficient to use ionized metal plasma (IMP) or CVD deposition over magnetron sputtering if step coverage is a critical factor.

Ionized Metal Plasma (IMP) ■ A problem for sputtering technology in wafer fabrication for high performance ICs is that as feature sizes shrink, the capability to sputter into high aspect ratio vias and narrow trenches becomes limited. A recent development in PVD to overcome this problem is *ionized metal plasma PVD (IMP or ionized PVD)*, which was introduced in the mid-1990s. In this approach, the sputtered metal is ionized in an RF plasma at a pressure of 20 to 40 mtorr (see Figure 12.22). The positive metal ions then travel in a highly directional, vertical path toward a wafer configured with a negative voltage bias. This voltage bias can also be used to control the energy of the incoming metal ions, which helps reduce damage to the surface of the wafer.⁴⁴ Biasing the wafer enables a higher degree of film conformality in the bottom and corners of high aspect ratio gaps.

Ionized PVD achieves good hole-fill for Ti and TiN on structures with 0.25-micron contacts and vias and 6:1 aspect ratios, with 70% bottom coverage, 10% sidewall coverage, and excellent sidewall integrity. This coverage is critical in the bottom, bottom corner, and lower sidewall areas of holes with a typical 85° sidewall angle.

Ionized PVD is used in production to deposit titanium and titanium nitride barrier films for devices at the 0.25 μm and below generations with high-aspect ratio contact and via structures.⁴⁵ It is also capable of depositing tantalum, tantalum nitride, and copper, which are important metals for copper metallization. The development of ionized PVD has further improved sputtering to extend its usefulness in high-aspect ratio gap fill.

Metal CVD

The use of chemical vapor deposition (CVD) for metal deposition is increasing because of its superior conformal step coverage and void-free filling of high-aspect ratio contacts and vias. These factors are critical for wafer fabrication as feature sizes decrease to the 0.15- μm generation and

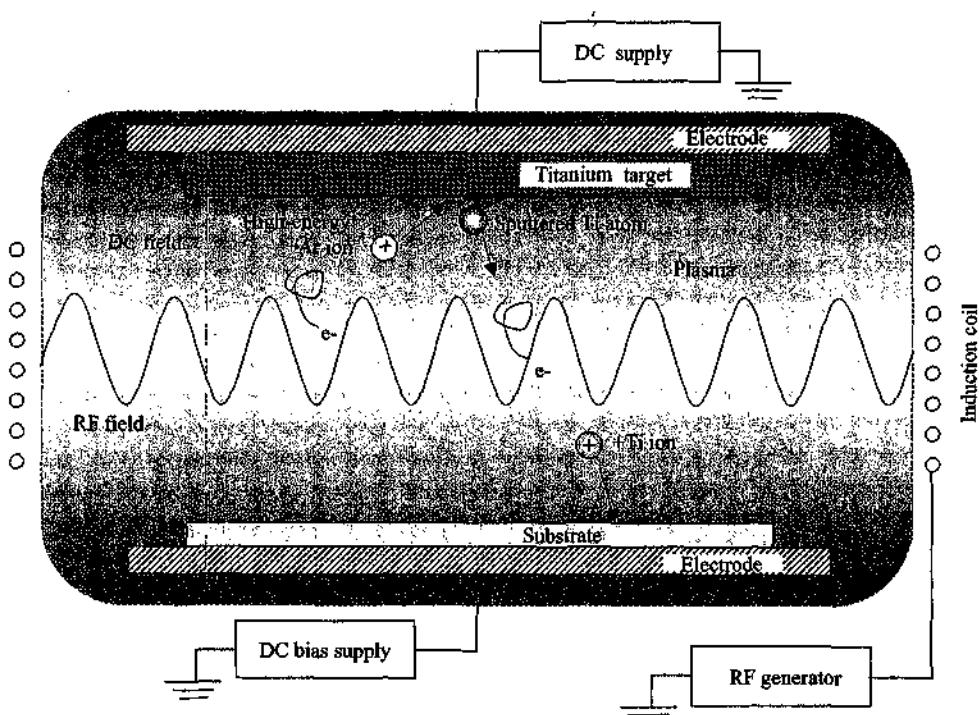


FIGURE 12.22 Concept of Ionized Metal Plasma PVD

beyond. The projected aspect ratio for vias in DRAM memory is projected to be 7:1 for the 0.15- μm generation, and 2.4:1 for logic.⁴⁶

Tungsten CVD ■ *Tungsten (W) CVD* is used in multilevel interconnection technology as the process to deposit the via plug and to act as the contact plug between the first metal layer and the silicide contact. Tungsten CVD is also used as the local interconnect (LI) because of its low resistivity. Tungsten has performed reliably, with billions of W CVD plugs used on one microchip. However, its usefulness as a vertical interconnect may be limited with the introduction of the new damascene copper metallization scheme.

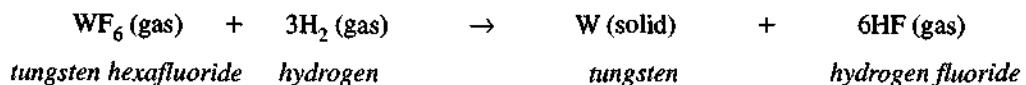
Tungsten CVD is common as a plug material due to:

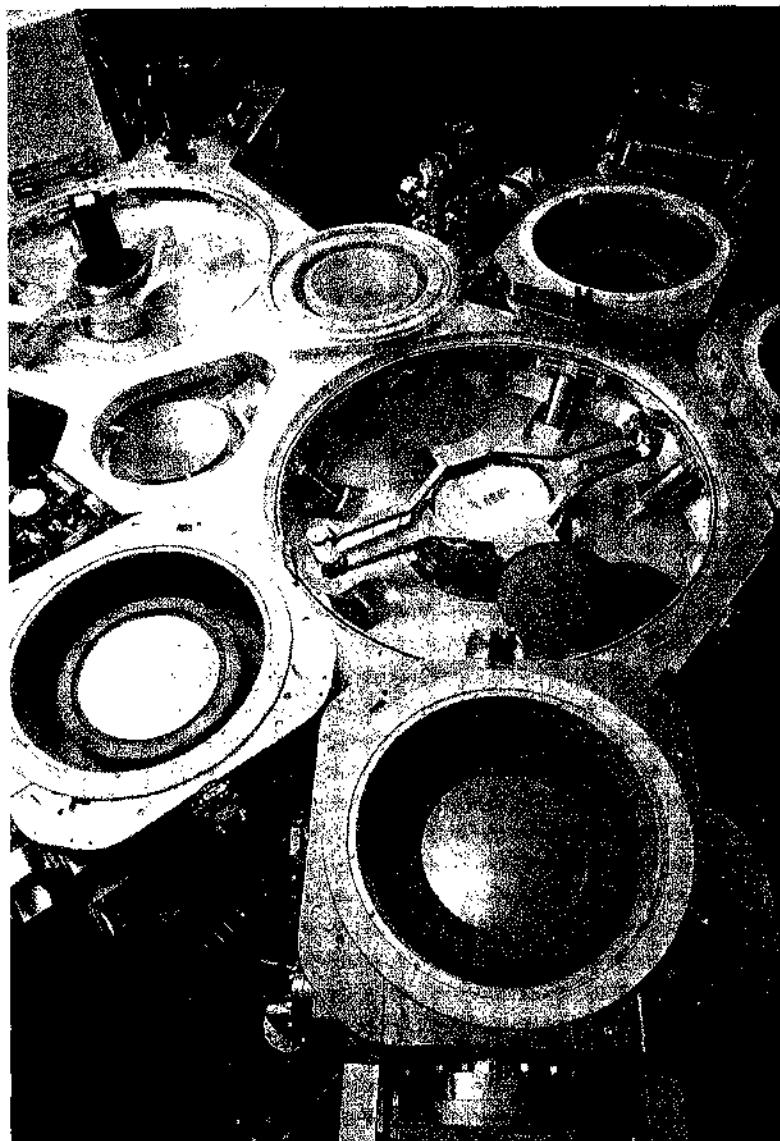
1. Excellent step coverage and gap-fill, particularly in high-aspect ratio vias.
2. High electromigration resistance.

Low-pressure CVD (LPCVD) is the usual method for depositing W plugs, along with plasma-enhanced CVD (PECVD). Reactors are both hot- and cold-wall and configured as batch or multi-chamber cluster tools. The most common source gas for W is WF₆.⁴⁷ Sputtering costs less to operate than CVD for W deposition but has traditionally poorer directional control. This coverage produces nonuniform tungsten deposition quality in vias, which is why CVD has been the preferred method. This may change with the recent development of ionized PVD.

Tungsten CVD is typically deposited in blanket films. Blanket deposition deposits tungsten nonselectively on the entire wafer surface, including the via pathways, and is the most common approach.

Blanket Tungsten CVD Deposition. Blanket tungsten deposition followed by chemical mechanical planarization (CMP) is the common method used to fill vias to connect between metal layers. As always in CVD, the properties of the film deposited with CVD depend on the chemical reactions that occur at the wafer's surface. For blanket tungsten CVD, the most frequently used reaction is the reduction of tungsten hexafluoride (WF₆) by hydrogen to produce tungsten and hydrogen fluoride:





PVD Cluster Tool
(Photo courtesy of Applied Materials, Inc.)

Typically the first step in blanket tungsten CVD is to deposit a titanium/titanium nitride (Ti/TiN) barrier layer (see Figure 12.23). The Ti is deposited before TiN to react with the underlying material and lower the contact resistance. If this is a contact in the first interlayer dielectric (ILD-1), then Ti reacts with silicon to form TiSi_2 silicide. The TiN functions as a barrier metal and adhesion promoter (glue layer) for tungsten. It needs a minimum bottom thickness of about 50 Å with continuous sidewall coverage to be an effective barrier metal and keep the tungsten from attacking underlying layers.⁴⁸ The Ti/TiN barrier layer may be deposited by IMP PVD. The TiN barrier metal may also be deposited either by CVD using ammonia and one of a number of precursor molecules, including the tetrakis dimethylamino titanium (TDMAT) or tetrakis diethylamino titanium (TDEAT). CVD has usually been preferred for TiN because of improved step coverage.

It is necessary to remove the excess blanket tungsten above the dielectric layer. This was previously done by a tungsten etchback process to remove the excess tungsten and leave a planarized plug. In 0.25- μm device generations and below, planarization of tungsten by chemical mechanical planarization (CMP) is the preferred process (CMP and etchback are addressed in Chapter 18).⁴⁹ Barrier metals such as TiN and Ti are used to prevent diffusion between the tungsten and the silicon.

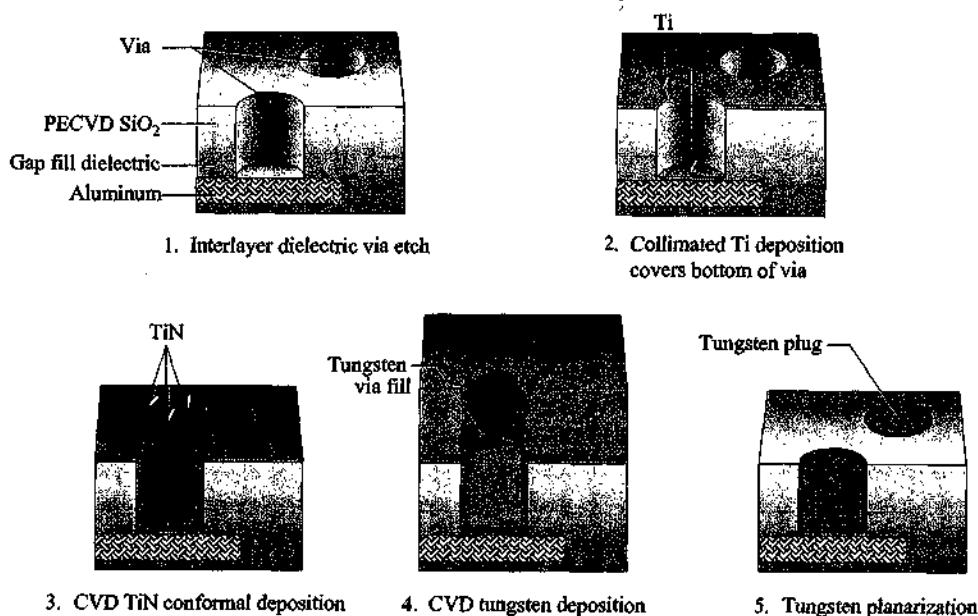


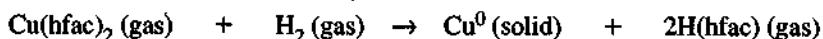
FIGURE 12.23 Blanket Tungsten CVD with Ti/TiN Barrier Metal

Copper CVD ■ *Copper (Cu) CVD* is a potential process for depositing a thin copper seed layer necessary for copper electroplating. The seed layer, or strike layer, is thin (about 500 to 1000 Å) and deposited on top of a diffusion barrier (most likely a Ta-based barrier metal). For successful electroplating, it is critical that this seed layer be continuous and free of pinholes and voids along the feature sidewalls and bottom. If a seed layer is not continuous, this can create a void in the electroplated copper. The excellent conformality of CVD is why this process has been investigated for the seed layer.

Cu Precursors. There are two precursors for CVD Cu, Cu(I), and Cu(II). Note that Cu(I) indicates a copper ion with a +1 positive charge and Cu(II) indicates a +2 positive charge. The most widely used Cu(I) precursor is Cu(hfac)(TMVS), chemical name trimethylvinylsilylhexafluoroacetylacetone copper (I), chemical formula C₁₀H₁₃CuF₆O₂Si, and referred to by its trade name of CurpraSelect.⁵⁰ This molecule combines copper in a +1 oxidation state with TMVS and hfac ligands to make a clear yellow, liquid precursor (a ligand is a molecule or ion surrounding a central metal cation). TMVS is flammable, but the compound Cu(hfac)(TMVS) does not ignite easily. Cu(hfac)(TMVS) is compatible with industrial stainless steel delivery vessels. The CVD deposition proceeds by reducing a copper atom to metal while oxidizing another atom to Cu(hfac)₂ and releasing free TMVS as by-products. The chemical reaction is:



The most common Cu(II) precursor is Cu(hfac)₂. For example, the Cu(II) reaction can proceed by reduction in H₂:



Copper Electroplate

The transformation of IC fabrication to copper metallization is underway for all chip manufacturers. Initially, high-performance microprocessors and fast, static RAMs are being converted to copper technology. *Copper electroplate*, also known as *electroplating*, *electrochemical deposition (ECD)*, or *electrofill*, is the first-generation deposition method to be used for copper metallization. Cost and performance will be important issues that affect when copper will replace aluminum as the mainstream interconnect metallization. Electroplating has not traditionally been used in semiconductor manufacturing. However, it has been an important metallization process in other areas of electronic manufacturing, such as thin film heads used in disk storage devices and copper wiring for printed circuit boards. Therefore, its process and equipment requirements are well understood.



Copper Electroplating Tool
(Photo courtesy of Novellus)

Electroplating Fundamentals ■ The basic principle of electroplating copper metal is to immerse a wafer with a conductive surface into a solution of copper sulfate ($\text{Cu}(\text{SO}_4)_2$) which contains copper ions to be deposited (see Figure 12.24). The wafer and its seed layer are connected electrically to an external power supply as a negatively charged plate, or cathode. A solid piece of copper is immersed in the solution and configured as the positively charged anode. Electrical current passes through the wafer, into the solution, and through the copper anode.⁵¹ When current flows, the following reduction reaction occurs at the wafer surface to deposit copper metal:

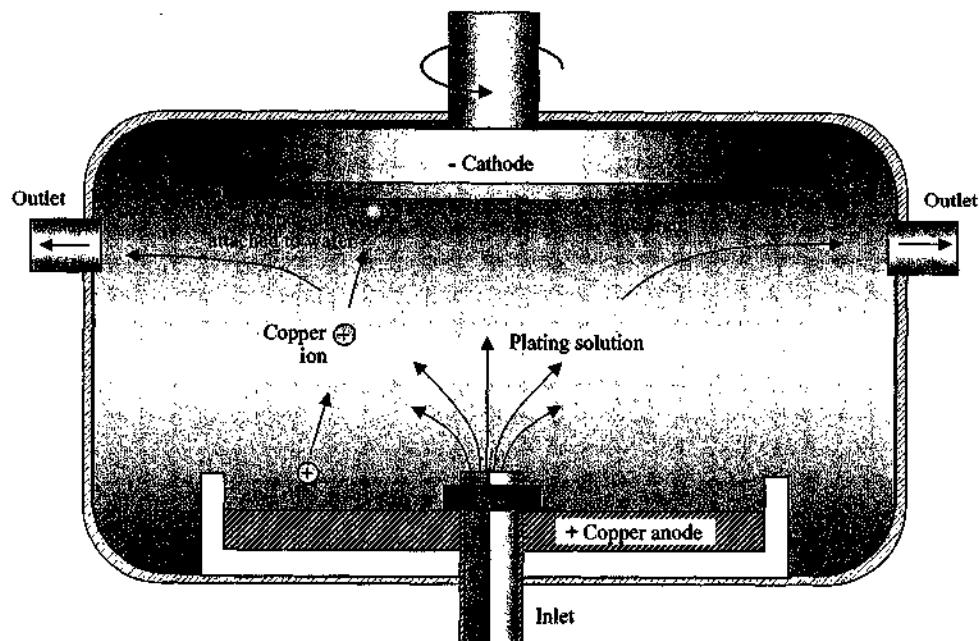


FIGURE 12.24 Copper Electroplating

Metal copper ions on the cathode wafer surface are reduced to metal copper atoms during electroplating. At the same time, an oxidation reaction occurs at the copper anode that balances the current flow in the cathode. This reaction maintains electrical neutrality in the solution.

The quantity of copper deposited is directly proportional to the current delivered to the conductive wafer surface (Faraday's law of electrolysis).⁵² With this relationship, the basic parameters to control in electroplating are current and time. With no current flow, there is an equilibrium potential among the anode, cathode, and solution. When an external power source applies a voltage, a current flows between the anode and cathode and metal is deposited on the cathode proportional to the amount of current.

In practice, electroplating is complicated to control, especially the electroplating solution and how the current is applied. Given the high-aspect ratio holes and trenches that must be filled on high-performance ICs, electroplating is challenged to maintain a uniform current density in the holes. This condition is important because the plating rate is a direct function of current density. If there is a high current density at the top of the hole and a lower current density at the bottom of the hole, then copper will plate faster at the top.

Different types of voltage waveforms applied to the cathode/anode system can assist in plating high-aspect ratio holes. For instance, by applying an oscillating electric field and controlling the amplitude of the waveform, a deposit/etch (referred to as dep/etch) sequence is obtained. In this manner, the copper can be slightly removed (etched) in the high-density current regions to balance the copper gap-fill capability.

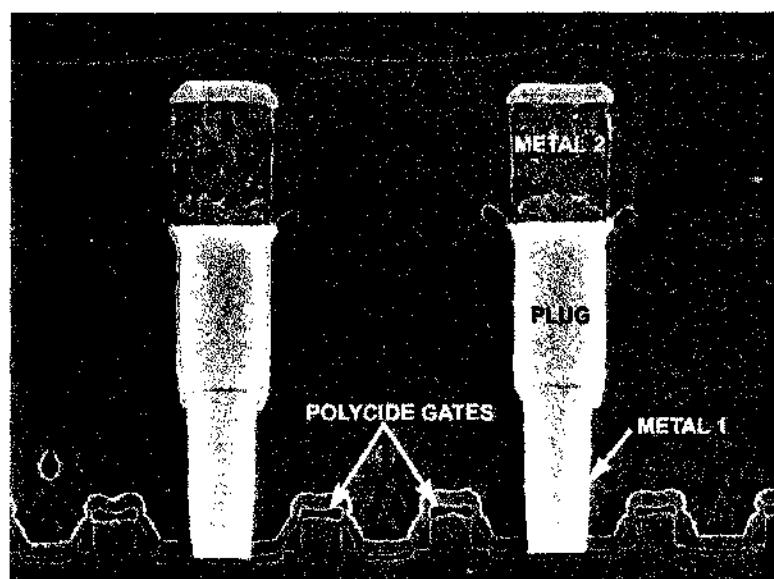
Because the copper electroplating metallization is a new process for semiconductor manufacturing, there are a host of other issues to address. For instance, it is important that no copper plate on the backside of the wafer. Electroplating also poses concerns for chemical control and disposal. It does simplify the metallization process somewhat by avoiding a high vacuum or complex heating of the wafer.

METALLIZATION SCHEMES

The strategy for metal interconnects for IC fabrication is undergoing dramatic change. After many years of development activities, copper is being implemented as the main chip interconnect conductor for a wide range of ICs.⁵³ Either aluminum or copper metallization is expected to be used for chip applications during a transition period until copper is fully implemented. The type of metallization scheme used depends on how the chip design is optimized for performance and price.

Traditional Aluminum Structure

The traditional interconnect metal has been an aluminum-copper alloy with an interlayer dielectric of SiO_2 to insulate between the metal layers. The process steps for the traditional aluminum



Blanket Aluminum Etch
(Micrograph courtesy
of Integrated Circuit
Engineering)

interconnect technology for an ILD-2 and metal-2 (one complete level) are described in Steps 11 and 12 of the process flow described in Chapter 9. This traditional process can be viewed as an *aluminum subtractive process* because the aluminum is deposited as a blanket film and then etched away (subtracted) to form the circuitry. The critical step that defines the width and spacing of the aluminum interconnect lines is metal etch, while SiO_2 is deposited into the narrow gaps between the lines.

New materials and processes are being rapidly introduced into advanced wafer fabs to support the 0.18- μm technology node with traditional aluminum alloy metallization. This is a transitional period until copper metallization is eventually implemented. As a rule, most fabs will change only one major microchip material at a time (e.g., metal or insulator) to minimize the inherent risk to production. With these new fabrication materials and their stringent processing requirements, the general trend is for increased use of high-density plasma CVD (HDPCVD) for dielectrics, a shift from sputtered metals to metal CVD, and increased use of chemical mechanical planarization (CMP). The goals are to increase product performance and manufacturing productivity.

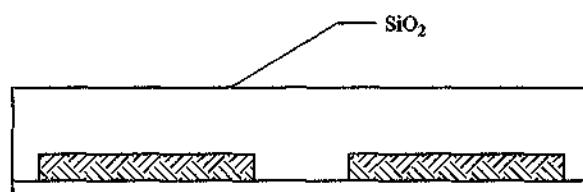
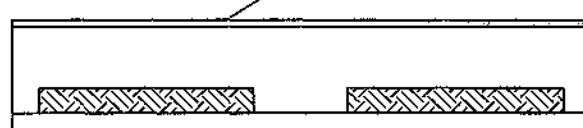
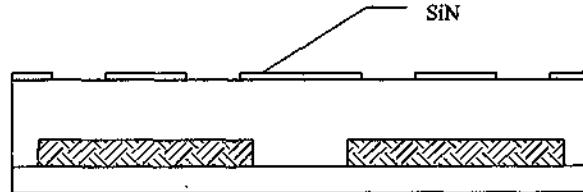
Copper Damascene Structure

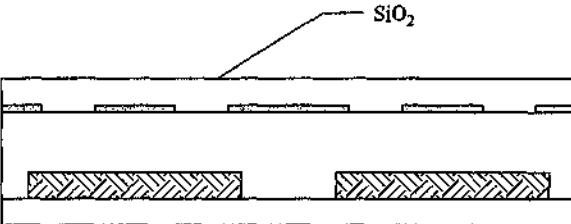
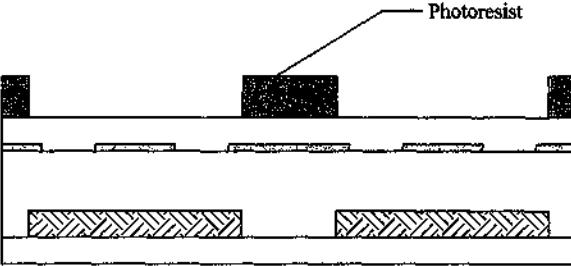
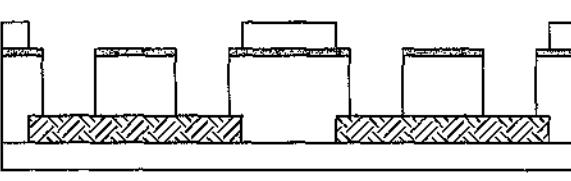
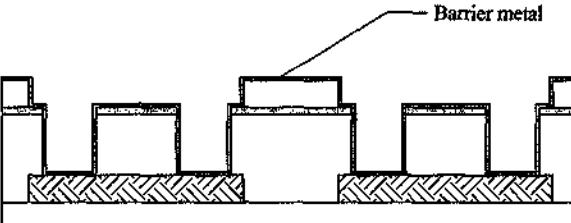
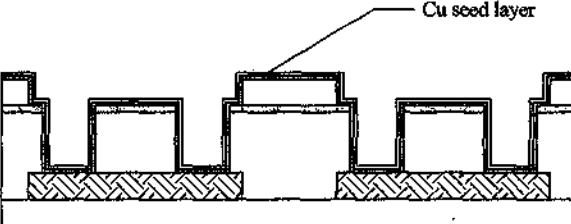
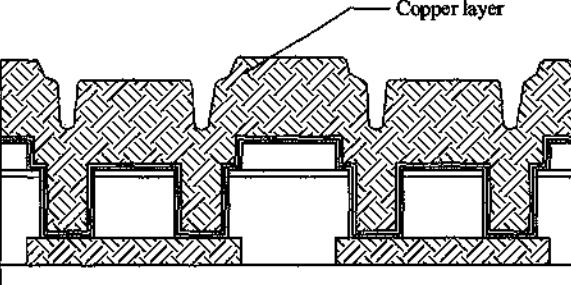
The semiconductor industry is implementing copper as the interconnect material for microchip fabrication. The traditional process flow will not be used for copper metallurgy because copper is not suitable for dry-etching. Copper also has special requirements to minimize its diffusion into silicon. To form copper interconnect wiring, a dual-damascene process is used to avoid copper etching. In the damascene process, it is no longer metal etch that defines the critical line width and spacing, but rather dielectric etch.

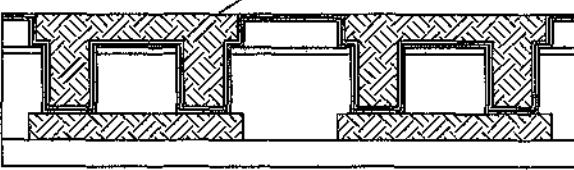
Dual Damascene Process ■ Since copper is difficult to etch, early developers of copper technology were forced to consider alternative approaches to creating metal lines. The *dual-damascene process* became the consensus process for copper metallurgy. It creates both the vias and lines for each metal layer by etching holes and trenches in the ILD, depositing copper in the etched features and using chemical mechanical planarization to remove the excess copper (referred to as *copper overburden*).

There are many possible scenarios for a dual-damascene process flow. Table 12.5 explains a basic process flow with the essential technologies that may or may not include all the steps performed by a particular manufacturer.

TABLE 12.5* Copper Metallization using Dual Damascene

Process Step	Description	Structure
1. SiO_2 deposition	ILD oxide deposition with PECVD to desired thickness for via. There is no critical gap fill; therefore, PECVD is acceptable.	
2. SiN etch stop deposition	Thin (250 Å) SiN etch stop is deposited on ILD oxide. The SiN needs to be dense and pinhole-free; therefore, HDPCVD is used.	
3. Via patterning and etch	Photolithography to pattern and dry etch via openings into SiN. Strip photoresist after completion of etch.	

Process Step	Description	Structure
4. Deposit remaining SiO_2	PECVD oxide deposition for remaining ILD oxide.	
5. Interconnect patterning	Photolithography to pattern SiO_2 trench with resist. Previously patterned via openings are located in trench.	
6. Etch trench for interconnect and hole for via	Dry etch trench in ILD oxide, stopping on the SiN layer. Etch continues to form via opening by passing through opening in SiN .	
7. Deposit barrier metal	Deposit Ta or TaN diffusion layer with ionized PVD on bottom and sidewalls of trench and via.	
8. Deposit Cu seed layer	Deposit continuous Cu seed layer with CVD. The layer must be uniform and free of pinholes.	
9. Deposit Cu fill	Deposit Cu fill with electrochemical deposition (ECD). Fill both via opening and trench.	

Process Step	Description	Structure
10. Remove excess Cu with CMP	Remove excess Cu using chemical mechanical planarization (CMP). This planarizes the surface and prepares for the next level. The resulting surface is a planar structure with metal inlays in the dielectric to form the circuitry.	

* P. Singer, "Making the Move to Dual Damascene Processing," *Semiconductor International* (August 1997): p. 79.

For dual-damascene processing, a two-tiered metal inlay is created that includes a series of holes (for contacts or vias) in addition to the interconnect trench for metal lines. The vias and metal line levels undergo copper fill concurrently, which saves a process step and eliminates the interface between the via and metal line.

There are different approaches to performing dual-damascene processing, each with their own set of challenges. Most of the manufacturing methods integrate the use of etch-stop materials to control how deep etching occurs for the via and trench. Etch stop is typically done by using a barrier metal such as silicon nitride (Si_3N_4 or Si_xN_y) as a hard mask at the bottom of the via or trench. Si_3N_4 etches at a much slower rate than the dielectric material, thus effectively stopping the etch process. Some approaches may even have two separate Si_3N_4 layers, which adds process complexity. In addition, Si_3N_4 has a high- k and increases the interlayer capacitance of the interconnect stack, which is why the Si_3N_4 layer needs to be thin.

The dual-damascene flow shown above has the via etch done in the same etch step as the trench, with the SiN etch stop to define the bottom of the trench. In other variations of dual-damascene, the via is etched first, then the trench is patterned and then etched. Trench structures created over topography from a previous via etch pose a significant challenge for photolithography.

The most important reason for using a damascene approach to copper metallization is to avoid metal etch. A second advantage to damascene processing is that there is no further need for dielectric gap fill between etched metal lines, since the dielectric is applied as a blanket and then etched.

The dual-damascene process in wafer fabrication potentially uses 20 to 30% fewer steps than traditional aluminum metal interconnect. Not only are there fewer manufacturing steps, but dual damascene also eliminates or reduces some of the most difficult steps of traditional aluminum metallization, including aluminum etch and many of the tungsten and dielectric chemical-mechanical polish steps.⁵⁴ Reducing process steps in wafer fabrication is important for improving the yield of the process, since fewer process steps translate to fewer sources of error that reduce assembly yield.

METALLIZATION QUALITY MEASURES

Wafer fabrication quality measures for metallization are provided in Table 12.6.

TABLE 12.6 Quality Measures for Metallization

Quality Parameter	Types of Defects	Remarks
1. Sputtered metal adhesion.	A. Metal layer does not adhere to substrate.	<p>Parameters affecting film adhesion are:</p> <ul style="list-style-type: none"> • Wafer contamination • Stress • Types of materials • Temperature of substrate • Argon pressure
2. Sputtered film stress.	<p>A. Excessive stress leads to:</p> <ul style="list-style-type: none"> • Cracks in the film surface. • Loss of film adhesion. • Increased resistivity in some materials. 	<p>Film stress can be caused by excessive temperature of the wafer substrate. Ways to reduce wafer temperature are:</p> <ul style="list-style-type: none"> • Lower deposition rate • Increased backside cooling <p>Cracks can cause:</p> <ul style="list-style-type: none"> • Peeling of the film layer • Contaminant migration • Electrical opens or short. <p>Measure film stress (bow of wafer surface) before and after deposition. Units for stress are dynes.</p>
3. Sputtered film thickness.	A. Metal layer does not meet film thickness specifications (e.g., sheet resistance out of specification).	<p>Parameters that affect film thickness are:</p> <ul style="list-style-type: none"> • Incorrect recipe • Improper flow rates • Improper substrate temperature • Improper chamber pressure • Improper power supply energy • Wrong time setting
4. Uniformity of electroplated (electrochemical deposition, or ECD) metal film.	<p>A. Nonuniform film thickness indicated by:</p> <ul style="list-style-type: none"> • Inadequate gap fill and step coverage for high aspect ratio openings (nonuniform metal thickness at top and bottom of high aspect ratio holes). • Nonuniform deposition thickness across the wafer and from wafer to wafer. • Voids in film. 	<p>Critical parameters for electroplating film uniformity are:</p> <ul style="list-style-type: none"> • Depositing a uniform CVD seed layer without voids. • Maintaining proper electroplating bath chemistry (composition and concentration) for organic additives, primarily brighteners and suppressors, to attain void-free deposition that fills bottom and sidewalls of gaps.

METALLIZATION TROUBLESHOOTING

Common troubleshooting problems for metallization are provided in Table 12.7.

TABLE 12.7 Common Metallization Troubleshooting Problems

Problem	Probable Cause	Corrective Actions
1. Degradation in step coverage of metal film.	A. Decrease in substrate temperature. For Al alloy sputter deposition, step coverage is dependent on the wafer temperature during deposition.	Heating the substrate to improve step coverage because surface mobility of deposited metal atoms improves.
	B. Increase in deposition rate.	Decrease deposition rate. Increase in sputter deposition rate may degrade step coverage due to reduced surface mobility because of arrival of more atoms at surface.
2. Vacuum chamber integrity.	A. Chamber cleanliness or moisture in chamber.	<ul style="list-style-type: none"> Check for vacuum leaks or chamber outgassing. Residual gas in chamber that can change film reflectivity. Check for O₂ or N₂ in metal films that can change resistivity and film stress. Clean chamber and do H₂O bake-out.
	B. Outgassing or system leaks.	<ul style="list-style-type: none"> Use helium leak detection to inspect for system leaks. Assess the chamber conditions before deposition with residual gas analyzer.
3. Contamination of metal film.	A. Particles on surface of film.	<p>Check the following common particle sources:</p> <ul style="list-style-type: none"> Dirty input wafers. Problems in hand-off between robot and chamber mechanism. Incomplete chamber cleaning between runs. Dirty cassettes. Contaminated nitrogen backfill.
4. Voids in copper trench fill after dual-damascene electroplating.	A. Excessive seed layer thickness on the wafer surface (field) that pinches off the via or trench, creating a center void in the film.*	Optimize the CVD Cu seed layer deposition by evaluating the electroplating current from the field to the bottom of vias and trenches. The goal is to deposit adequate seed layer at the bottom of a high-aspect ratio feature without increasing field thickness.
5. Excessive copper dishing after dual-damascene CMP.	A. Cu dishing for Cu CMP is often caused by tantalum diffusion barrier, which must be polished back in the presence of Cu.**	<p>After Cu is polished, the Ta layer must be removed from the level. The Cu pad/slurry field does not effectively remove Ta.</p> <p>Options are:</p> <ul style="list-style-type: none"> Optimize the pad/slurry for Ta. Minimize the Ta field level thickness.

* R. Jackson et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998); p. 56.

** Ibid.

SUMMARY

Metallization deposits a thin metal film to form the interconnect wiring and contact or via connections on the chip. There are six metal categories used in wafer fabrication, with specific requirements for optimum performance. Aluminum has been the traditional metal for interconnect lines. An ohmic contact is the low-resistance contact between silicon and the interconnect metal. Aluminum is sometimes alloyed with silicon to reduce junction spiking at the ohmic contact. Aluminum is often alloyed at 0.5 to 4% copper to minimize electromigration reliability problems. New interconnect metallization is based on copper metallurgy to decrease the metal resistivity. Copper combined with low- k dielectric will decrease the chip interconnect delay, but copper does require a new processing approach since it does not etch well. Barrier metals are used when joining metals. Different barrier metals have optimum properties depending on the application. A silicide is a refractory metal alloyed with silicon to improve contact resistance and adhesion. A salicide is a particular silicide that is aligned to the source, drain, and

gate structure. Metal layers are connected with vias filled with tungsten plugs. The most widely used metal deposition system is sputtering. Sputtering physically bombards a target to dislodge atoms and deposit the atoms as a thin film on the wafer surface. The three most common types of sputtering are RF, magnetron, and ionized metal plasma. RF sputtering is inefficient, while magnetron has a higher deposition rate but limited gap-fill capability. Ionized metal plasma has improved directionality for high-aspect ratio deposition. CVD metal has the best conformality and is used for tungsten plug fill and copper seed layer. Copper electroplate is the process chosen by major semiconductor manufacturers for copper wiring deposition. Traditional aluminum metallization deposits a blanket aluminum film and then etches line patterns to form interconnect wiring. Copper metallurgy will use a dual-damascene process that etches vias and trenches in dielectric, deposits copper fill, and then removes excess copper through chemical mechanical planarization.

KEY TERMS

- interconnect
- wiring
- contact
- vias
- plug
- eutectic temperature
- anneal or sinter
- ohmic contact
- junction spiking
- electromigration
- hillocks
- copper interconnect
- barrier metal
- refractory metals
- silicide
- polycide
- salicide
- tungsten plugs

- evaporation
- physical vapor deposition (PVD)
- target
- momentum transfer
- sputtering yield
- sputter etch (reverse sputter)
- RF sputtering
- magnetron sputtering
- collimated sputtering
- sputtered aluminum
- ionized metal plasma PVD (IMP or ionized PVD)
- tungsten (W) CVD
- copper CVD
- electroplate (electroplating, electrochemical deposition (ECD), or electrofill)
- evaporation
- aluminum subtractive process
- dual-damascene process

REVIEW QUESTIONS

1. Define the following terms: interconnect, contact, via, and plug.
2. What metal has traditionally been used for interconnect wiring, and what is its replacement?
3. List and describe the seven performance requirements for metals used in wafer fabrication.
4. List the categories of metals and metal alloys used in semiconductor manufacturing.
5. Explain why aluminum has been the interconnect metal of choice for microchips.
6. What is an ohmic contact and what is its advantage?
7. Describe junction spiking and list the two primary ways it was resolved.
8. Discuss electromigration and how it affects reliability. How is electromigration controlled in aluminum wiring?
9. List and discuss the five benefits for introducing copper metallization.
10. What are the three major challenges for converting to copper for interconnects?
11. Explain what a barrier metal is and the essential properties of a barrier material. What metals are commonly used as a barrier metal?
12. List the special requirements for a copper barrier metal.
13. Define silicide and explain why refractory metal silicides are important in wafer fabrication.
14. What are the benefits of titanium silicide? Explain the C49 and C54 phase structures.
15. What silicide is promising for future wafer technology? Why?
16. What is the salicide process?
17. Describe a tungsten plug, and discuss how it is used in multilayer metallization.
18. Why was evaporation replaced as a metal deposition system?
19. Provide a short explanation about sputtering and how it works.
20. Is sputtering suitable for alloy deposition?
21. What are the advantages to sputter deposition?
22. List and explain the six steps to sputtering.
23. Explain the physics of sputtering.
24. Describe sputter yield.
25. What is the purpose of sputter etch?
26. Describe an RF sputter system. What is its major limitation?
27. Discuss how a magnetron sputtering system improves the deposition rate.
28. Explain why sputtering has poor step coverage. How does collimated sputtering improve step coverage?
29. Describe ionized metal plasma. How does this process improve high-aspect ratio gap fill?
30. What is the typical method for creating a tungsten plug in advanced ICs?
31. Why could CVD be used for the copper seed layer?
32. Explain the fundamental process of copper electroplate.
33. Why is traditional aluminum metallization a subtractive process?
34. What is the primary reason for implementing the dual-damascene process for copper?
35. List ten steps for a dual-damascene metallization process.

**METALLIZATION EQUIPMENT
AND MATERIALS SUPPLIERS'
WEB SITES**

Angstrom Sciences Inc.
Applied Materials
Genus Incorporated
Materials Research Corporation
Nordiko USA Inc.
Novellus Systems Inc.
Process Materials Inc
TEL, Tokyo Electron Ltd.
Veeco-CVC

<http://www.angstromsciences.com/>
<http://www.appliedmaterials.com/products/>
<http://www.genus.com/>
<http://www.materialsresearch.com/>
<http://www.nordiko.com/>
<http://www.novellus.com/index.htm>
<http://www.processmaterials.com/>
<http://www.teainet.com/>
<http://www.veeco.com>

REFERENCES

1. C. Weber, D. Jensen, and E. Hirleman, "What Drives Defect Detection Technology?" *Micro* (June 1998): p. 60.
2. M. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *Solid State Technology* (September 1996): p. 105.
3. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed. (New York: Wiley, 1994): p. 548.
4. K. Bachmann, *The Materials Science of Microelectronics*, (New York: VCH Publishers, 1995), p. 484.
5. Compiled from multiple sources, including: S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, 2nd ed., (Sunset Beach: Lattice Press, 1986), p. 399.
6. L. Peters, "Advanced Aluminum Interconnect Technology," *Semiconductor International* (November 1998): p. 83.
7. R. Jaeger, *Introduction to Microelectronic Fabrication*, (Reading, MA: Addison-Wesley, 1988), p. 136.
8. S. Campbell, *The Science and Engineering of Microelectronic Fabrication*, (New York: Oxford University Press, 1996), p. 406.
9. Ibid., p. 412.
10. E. Amerasekera and F. Najm, *Failure Mechanisms in Semiconductor Devices*, 2nd ed., (New York: Wiley, 1997), p. 103.
11. J. Yue, "Reliability," Sze, *VLSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 672.
12. K. Rodbell, "Reliability," *Handbook of Semiconductor Interconnection Technology*, ed. G. Schwartz, K. Srikrishnan, and A. Bross (New York: Marcel Dekker, 1998), p. 480.
13. S. Campbell, *The Science and Engineering of Microelectronic Fabrication*, p. 413.
14. J. Yue, "Reliability," p. 663.
15. P. Singer, "Tantalum, Copper and Damascene: The Future of Interconnects," *Semiconductor International* (June 1998): p. 91.
16. "Technology News," *Solid State Technology*, (February 1998): p. 26.
17. P. Singer, "Tantalum, Copper and Damascene," p. 91.
18. Ibid.
19. A. Sethuraman, J. F. Wang, and L. Cook, "Copper vs. Aluminum: A Planarization Perspective," *Semiconductor International* (June 1996): p. 178.
20. X. Lin and D. Pramanik, "Future Interconnect Technologies and Copper Metallization," *Solid State Technology* (October 1998): p. 63.
21. P. Singer, "Tantalum, Copper and Damascene," p. 94.
22. Semiconductor Industry Association, *National Technology Roadmap for Semiconductors: Technology Needs*, (San Jose: SIA, 1997), p. 101.
23. C. Ryu et al., "Barriers for Copper Interconnections," *Solid State Technology* (April 1999), p. 53.
24. B. Chin et al., "Barrier and Seed Layers for Damascene Copper Metallization," *Solid State Technology* (July 1998), p. 141.
25. C. Ryu, "Barriers for Copper Interconnections," p. 56.
26. A. Braun, "Copper Electroplating Enters Mainstream Processing," *Semiconductor International* (April 1999), p. 64.
27. R. Jackson, et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998): p. 56.
28. S. Ghandhi, *VLSI Fabrication Principles*, p. 576.
29. D. Campbell, "Semiconductor Contact Technology," *Handbook of Semiconductor Interconnection Technology*, ed. G. Schwartz, K. Srikrishnan and A. Bross (New York: Marcel Dekker, 1998), p. 176.
30. K. Rodbell, *Reliability*, p. 480.
31. L. Peters, "Is the 0.18 μm Node Just a Roadside Attraction?" *Semiconductor International* (January 1999), p. 50.
32. R. Liu, "Metallization," *VLSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 410.
33. S. Campbell, *Science and Engineering*, p. 292.
34. R. Liu, "Metallization," p. 379.
35. X. Lin and D. Pramanik, "Future Interconnect Technologies," p. 74.
36. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, 2nd ed., pp. 443–50.
37. M. Rittner, "Growth Predicted for Sputtering Target and Sputtered Film Markets," *Solid State Technology* (January 2000): p. 26.
38. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, p. 445.
39. A. Braun, "Sputtering Targets Adapt to New Materials and Shrinking Architectures," *Semiconductor International* (June 1998): p. 130.
40. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, p. 449.
41. Ibid., pp. 348–51.
42. R. Liu, "Metallization," p. 382.
43. Ibid., p. 384.
44. B. Chin et al., "Barrier and Seed Layers," p. 141.
45. A. Braun, "Sputtering Targets," p. 128.
46. Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors: Technology Needs*, (San Jose: SIA, 1997), p. 101.
47. G. Schwartz, "Metallization," *Handbook of Semiconductor Interconnection Technology*, ed. G. Schwartz, K. Srikrishnan and A. Bross, (New York: Marcel Dekker, 1998), p. 320.
48. J. Baliga, "Depositing Diffusion Barriers," *Semiconductor International* (March 1997): p. 77.
49. K. Wijekoon, et al., "Tungsten CMP Process Developed," *Solid State Technology* (April 1998): p. 53.
50. B. Zorich and M. Majors, "Safety and Environmental Concerns of CVD Copper Precursors," *Solid State Technology* (September 1998): p. 101.
51. P. Singer, "Tantalum, Copper and Damascene," p. 94.
52. Ibid.
53. Ibid., p. 91.
54. Ibid.

CHAPTER 13

PHOTOLITHOGRAPHY: VAPOR PRIME TO SOFT BAKE

The essence of photolithography is to print temporary circuit structures on a wafer that will later be used to assist etch and ion implant processes. The structures are first created as patterns on a quartz template that is referred to as a mask. Ultraviolet light passes through the mask to transfer the pattern to a light-sensitive film on the wafer surface. As with common photography work, the photographic film is developed and the image appears on the wafer. Later a chemical etching process is used to image the film pattern on the underlying wafer surface, or the wafer is sent to the ion implant area where selective doping can be performed through the patterns on the wafer. The various patterns transferred to the wafer define the numerous device features, vias, and wiring interconnects necessary for the device layers as well as determine areas of the silicon to be doped.

Photolithography is closely related to the cost and performance of a microchip. The processing cost of a

wafer is largely independent of the number of chips on a wafer. That is, a wafer with few chips costs almost as much as the same wafer with many chips since the number of processing steps, amount of materials, wafer handling, and so on, are nearly identical for the two wafers. More chips at the same cost means that the overall cost of each chip is lower, simply by putting more chips on the wafer.

Since the beginning of semiconductor manufacturing, photolithography has been recognized as the driving force behind the integrated circuit fabrication process. This trend continues today as the industry strives to pack more devices and the associated circuitry on a chip. More than any other single fabrication technology, photolithography has contributed to the revolutionary advancement of chip performance, affirming Moore's law over the fifty-plus years of semiconductor manufacturing.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the basic concepts for photolithography, including process overview, critical dimension generations, light spectrum, resolution, and process latitude.
2. Discuss the difference between negative and positive lithography.
3. State and describe the eight basic steps to photolithography.
4. Explain how the wafer surface is prepared for photolithography.
5. Describe photoresist and discuss photoresist physical properties.
6. Discuss the chemistry and applications of conventional i-line photoresist.
7. Describe the chemistry and benefits of deep UV (DUV) resists, including chemically amplified resists.
8. Explain how photoresist is applied in wafer manufacturing.
9. Discuss the purpose of soft bake, and explain how it is accomplished in production.

INTRODUCTION

Photolithography produces a three-dimensional pattern on the surface of the wafer using a light-sensitive photoresist material and controlled exposure to light. Other terms for the photolithography process are *photo*, *lithography*, *masking*, and *patterning*. In general, lithography refers to any printing process that transfers a pattern to a planar surface. For this reason, photolithography is sometimes referred to as “printing.” For semiconductor manufacturing, *microlithography* describes the process used to pattern the ultraminiature features necessary for submicron wafer fabrication. The terms microlithography and photolithography are used interchangeably in this text.

Photolithography Concepts

Photolithography is at the center of the wafer fabrication process, as evidenced by how wafers repeatedly flow into and out of the photolithography operation in the fabrication process (see Figure 13.1). Photolithography is often considered the most critical step in the IC process, requiring superior performance to achieve high yield in other process areas. It is estimated that lithography accounts for nearly one third of the total wafer fabrication cost.¹

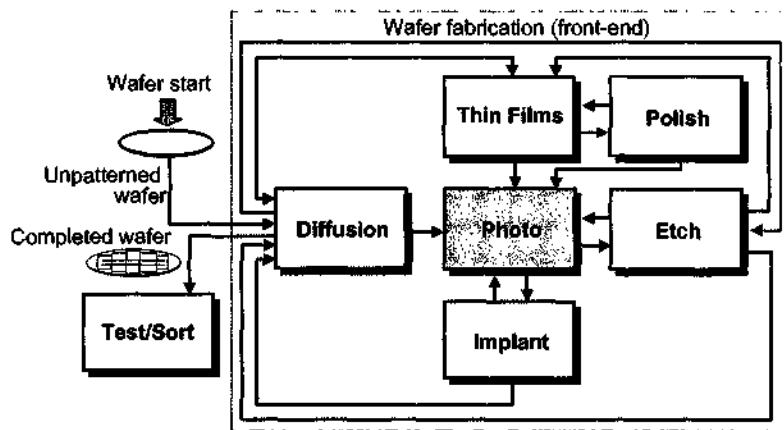


FIGURE 13.1 Wafer Fabrication Process Flow
(Used with permission from Advanced Micro Devices)

The shape of the photolithography pattern transferred to the wafer surface depends entirely on the wafer layer being constructed. Possible patterns are semiconductor devices in the silicon wafer, isolation trenches, contacts, the metal interconnects, and vias to interconnect metal layers. The patterns are transferred into a light-sensitive photoresist material to prepare the substrate for etch (see Chapter 16) or ion implantation (see Chapter 17). The patterned photoresist images are three-dimensional because there are width, length, and height to the pattern in the photoresist (see Figure 13.2). There may be hundreds of identical chips on a wafer, each needing the appropriate pattern to be transferred to each die.

Photolithography technology employs a light-sensitive photoresist, or resist, that is applied on the substrate surface as a polymer solution. The resist is then baked, which drives out the solvent. Next, the resist is exposed to a controlled light. The light passes through a reticle that defines the desired pattern. The photoresist is a temporary material placed on the wafer surface solely to transfer the necessary pattern and is removed once the pattern is etched or implanted onto the wafer surface material.

The *reticle* is a quartz plate that contains the pattern to be reproduced on the wafer, just as the negative of a camera film has the pattern for a photograph. This pattern may contain a single die or several die. A *photomask*, often called a *mask* and often used interchangeably with the term reticle, is a quartz plate that contains the full die array needed to define one process layer for the entire wafer. For this text, reticle refers to the pattern for a die or set of die, whereas mask refers to the entire array or matrix of die necessary for an entire wafer layer. Since light is critical for transferring the pattern to the resist and is controlled through optics, photolithography is sometimes referred to as *optical lithography*.

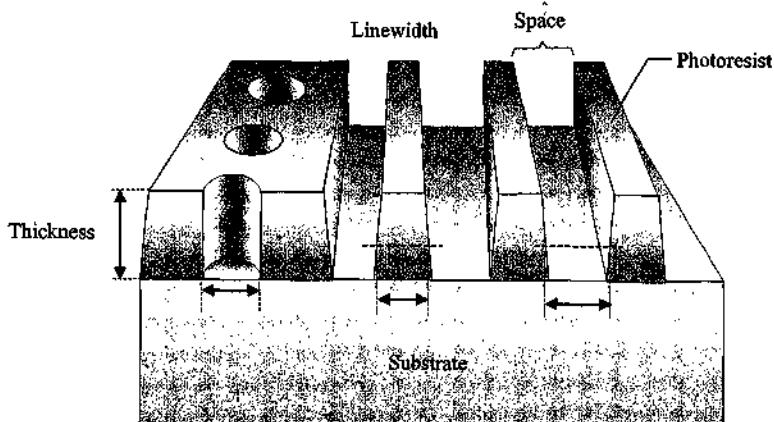
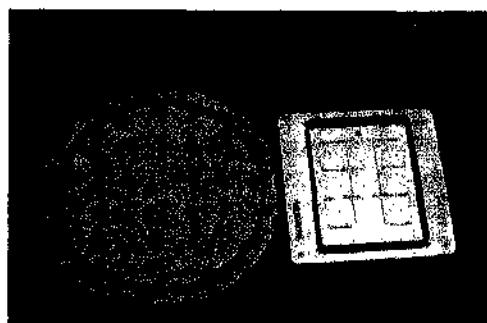


FIGURE 13.2 Three Dimensional Pattern in Photoresist

Advanced CMOS ICs can have up to 30 mask layers needed to pattern the multiple layers on a chip.² Each reticle has the unique pattern needed for the particular image, or feature, to be placed on the wafer surface and is stepped across the entire wafer to complete one layer. To fabricate an entire layer, the wafer must be processed into the photolithography operation and then out to undergo subsequent operations (e.g., etch or implantation).



Reticle versus Photomask for Microlithography
(Photo courtesy of Advanced Micro Devices)

Critical Dimension Generations ■ Microlithography has traditionally been the technology in semiconductor manufacturing that defines the critical dimension (CD) on a wafer (e.g., the polysilicon gate length on a microprocessor). Since the CD is the most difficult dimension to control during fabrication, it becomes the required dimension for other process areas to achieve. The critical dimension in photolithography is often used to describe device technology nodes or generations. Some sub-0.25 μm technology nodes are 0.18 μm , 0.15 μm , and 0.1 μm . The ability to reduce the critical dimension permits more die to be placed on a single wafer, thus lowering fabrication costs significantly and improving profitability.

Light Spectrum ■ Energy is required to activate the photoresist and transfer the pattern from the reticle. The source of energy is in the form of radiation, typically an ultraviolet (UV) light source. Light-sensitive photoresists are designed to chemically respond to a certain UV wavelength. UV light has been the most common energy source for photolithography patterning and this will probably continue for the near future (including devices fabricated at the 0.10- μm technology node and possibly lower).

The electromagnetic spectrum is used to introduce the UV light spectrum of most interest for photolithography, as shown in Figure 13.3 on page 338. The different UV wavelengths important for photolithography are shown in Table 13.1 on page 338. In general, deep UV (DUV) describes wavelengths less than 300 nm. Additional details about light and the UV spectrum for photolithography will be provided in Chapter 14.

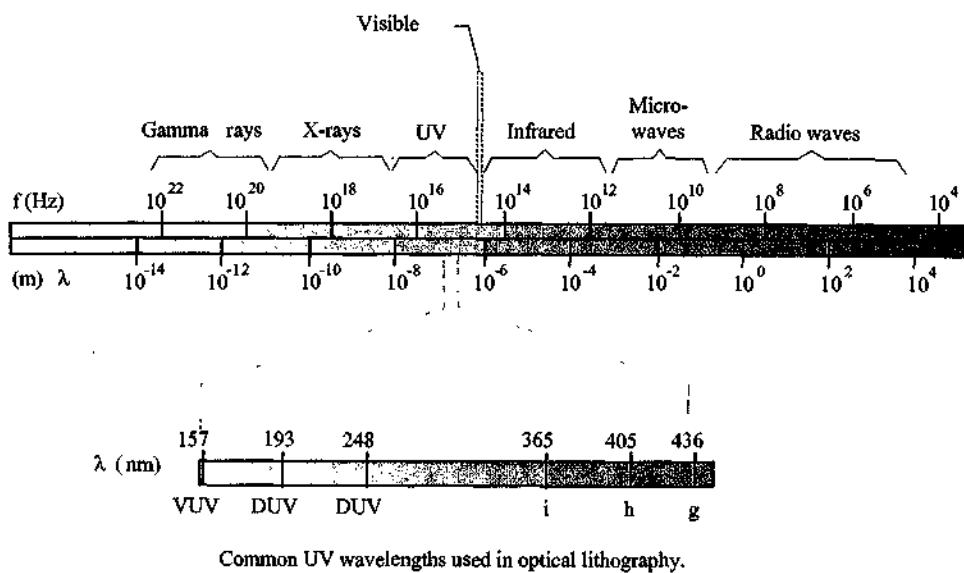


FIGURE 13.3 Section of the Electromagnetic Spectrum

TABLE 13.1 Important UV Wavelengths for Photolithography Exposure

UV Wavelength (nm)	Wavelength Name	UV Emission Source
436	g-line	Mercury arc lamp
405	h-line	Mercury arc lamp
365	i-line	Mercury arc lamp
248	Deep UV (DUV)	Mercury arc lamp or krypton fluoride (KrF) excimer laser
193	Deep UV (DUV)	Argon fluoride (ArF) excimer laser
157	Vacuum UV (VUV)	Fluorine (F_2) excimer laser

Resolution ■ An important performance measurement in photolithography is the resolution of each image. *Resolution* is the ability to differentiate between two closely spaced features on the wafer surface. The actual dimensions of the patterned images on the wafer are the feature sizes. The minimum feature size dimension is the critical dimension (CD), such as $0.18 \mu\text{m}$. However, this does not mean that every feature on the wafer has exactly this CD. On a critical layer, there are critical dimensions, such as a contact opening in the first interlayer dielectric or the polysilicon gate length, and other features that are noncritical and larger than the CD. Resolution is important for critical dimensions.

Over the past 30 years, the CD in wafer fabrication has decreased at about 11% per year. The reduction of device feature sizes has been done to achieve improved performance from device scaling, increased circuit density and decreased die size. However, the CD can only be reduced as much as the photolithography process will permit. An important challenge of lithography for achieving the smaller resolution has been the need to reduce the wavelength of the exposing light to be about the same length as the CD (to be discussed in Chapter 14).

Overlay Accuracy ■ As we shall see in Chapter 14, photolithography requires precise alignment between the pattern on the mask and the existing features on the wafer surface. This quality measure is known as *overlay accuracy*. Alignment is critical because the mask pattern must be precisely transferred to the wafer from layer to layer (see Figure 13.4). Since multiple masks are used during patterning, any overlay misalignment contributes to the total placement tolerances between the different features on the wafer surface. This condition is known as the *overlay budget*. A large overlay budget essentially reduces the circuit density, which limits device feature sizes and, therefore, IC performance.

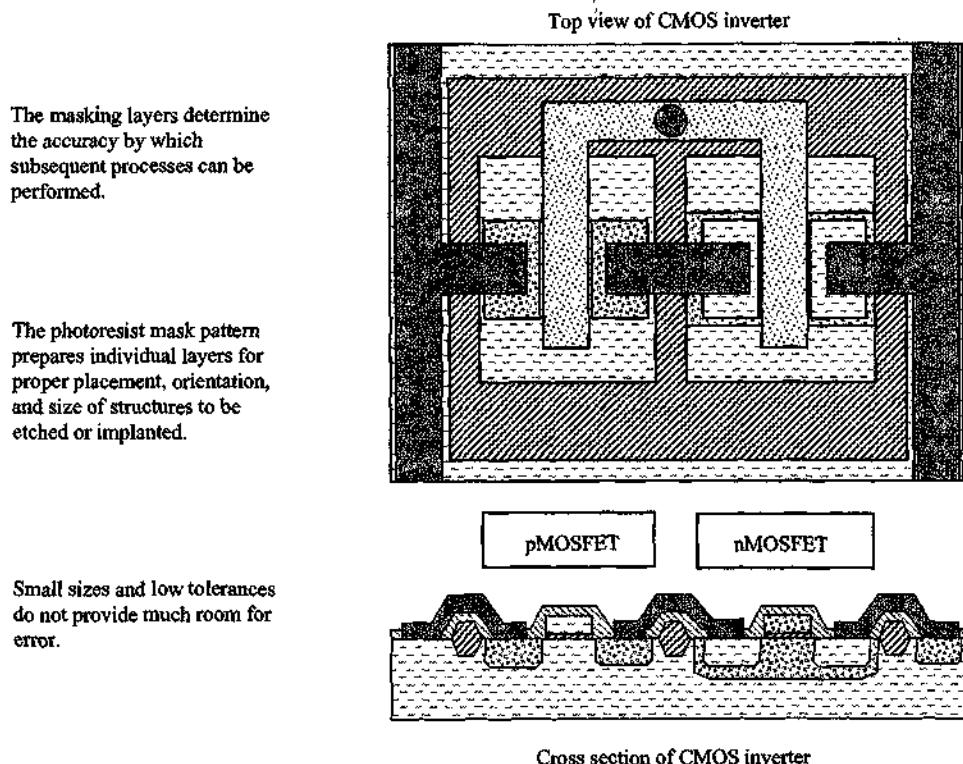


FIGURE 13.4 Importance of Mask Overlay Accuracy

Different types of misalignments affect the overlay budget. Misregistration is caused by poor alignment between the mask and the wafer. *Runout* is a net difference in the distance from die to die.³ It can be caused by temperature variations. These types of misalignment are controlled in modern alignment systems by global wafer alignment and site-to-site alignment schemes (discussed in Chapter 14).

Process Latitude ■ There are many process variables involved in photolithography. Examples of process variables are equipment settings, types of materials, human performance, machine alignments, and the stability of materials over time. *Process latitude* represents the capability of the photolithography process to consistently produce products that meet the specified requirements. The objective is to achieve the largest process latitude possible to make the process more capable of producing good parts. Process engineers define the settings for the different process variables to obtain the largest process latitude. For photolithography, high process latitude means critical dimensions can be met within the specified limits even with all the process variations encountered during production.

PHOTOLITHOGRAPHY PROCESSES

Photolithography consists of two basic types of processes: negative and positive lithography. *Negative lithography* prints a pattern on the wafer surface that is the opposite of the pattern in the mask. *Positive lithography* prints a pattern on the wafer that is the same pattern as that on the mask. The major difference between these two basic processes is the type of photoresist used. How the photoresist material reacts when exposed to light depends on whether it is a negative or a positive resist material.

Negative Lithography

The primary feature of negative lithography is that the photoresist will become insoluble and harden by crosslinking when exposed to light. Once hardened, the crosslinked resist cannot be washed away in solvents. Resists of this type are known as *negative resists* because the image formed in the

resist is the negative of the pattern found on the reticle (see Figure 13.5). Negative resists were the earliest types of photoresist used in semiconductor photolithography.

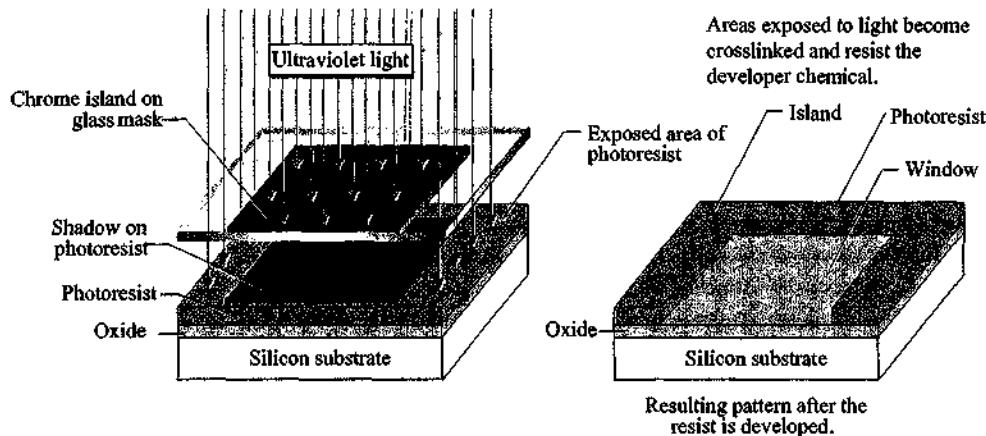


FIGURE 13.5 Negative Lithography

It can be seen in the diagram in Figure 13.5 that the mask for the negative resist is a transparent plate of quartz. The dark portion of the mask is a thin film of chromium deposited and patterned to form the desired mask artwork. The chrome is opaque (not transparent) and does not permit the passage of the ultraviolet light. For a negative resist, the regions underneath the opaque chrome of the mask are not exposed to light and therefore remain unchanged. The resist remains soft and dissolves when exposed to the developer chemical. UV light passes through the clear areas of the mask and hardens the resist so that it does not dissolve in the subsequent developer chemical. In this manner, the negative resist has a pattern that is the opposite of the pattern on the mask.

Positive Lithography

In positive lithography, the pattern printed on the wafer surface has the same image as the pattern on the mask. The areas of the resist exposed to the UV light undergo a photochemical reaction and become soluble and soften in the developer. In this manner, regions of the positive resist exposed to light are removed in a developer solution, whereas resist not exposed to light underneath the opaque mask pattern remains on the wafer (see Figure 13.6). This type of resist is referred to as a *positive resist* because the image formed in the resist is the same pattern formed on the reticle. The resist that remains was already hardened before light exposure and remains on the wafer surface to protect the top layer during subsequent processing, such as etch. The resist is removed when the follow-on processing is done. During the 1970s, positive resist became the dominant resist type for submicron microlithography and remains so today (see the following section for a discussion of the benefits from positive resist).

The pattern transferred into the photoresist can be considered a window or an island. A summary of the results from the two different types of masks and photoresists is shown in Figure 13.7.

Another way to describe masks is by their outward appearance. A mask is referred to as a *dark-field mask* if much of the quartz plate is coated with chrome. A *clear-field mask* is one that has very thin patterns of chrome with large areas of clear quartz. When comparing masks for positive and negative lithography, if a clear-field mask is required for a specific masking layer using a positive resist, then a dark-field mask of the same pattern would be used for negative resist.

Using positive lithography as an example, some common dark-field masks are the masks that are used prior to source/drain implant, LDD implant, and contact etch. A clear-field mask is used for applications such as preceding gate etch and metal interconnect etch. Figure 13.8 illustrates examples of the two types of masks based on positive photoresist lithography.

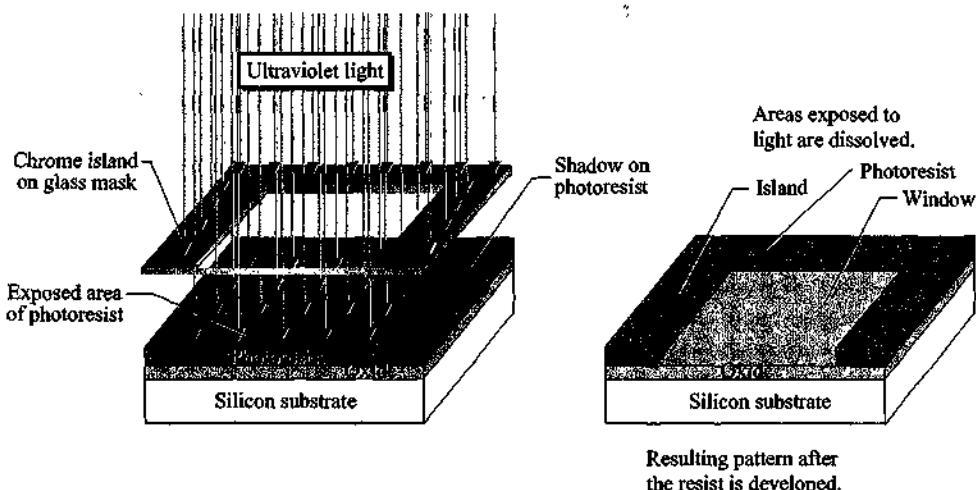


FIGURE 13.6 Positive Lithography

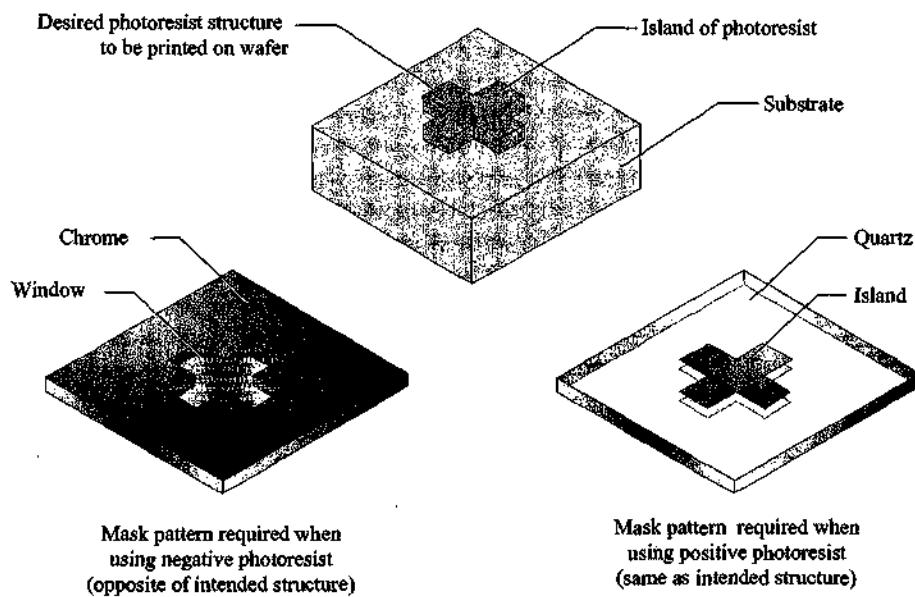


FIGURE 13.7 Relationship Between Mask and Resist

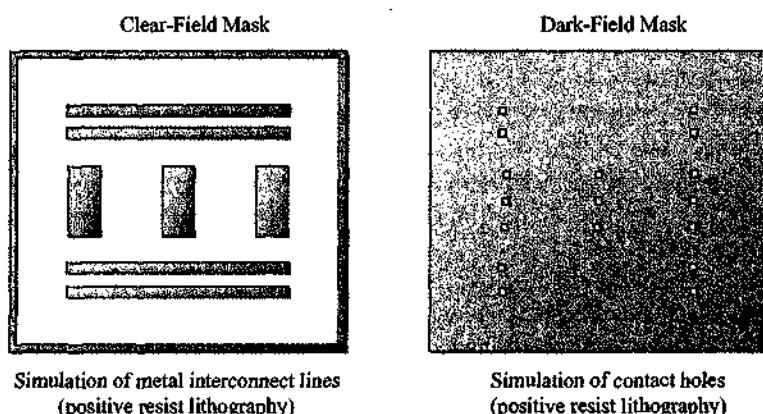


FIGURE 13.8 Clear-Field and Dark-Field Masks

EIGHT BASIC STEPS OF PHOTOLITHOGRAPHY

Photolithography is a complex process with many variables contributing to its process latitude, such as reduced feature size, alignment tolerance, number of masking layers, and cleanliness of the wafer surface. For convenience, we can divide the patterning process of photolithography into an eight-step procedure (see Table 13.2). Figure 13.9 provides a pictorial overview of the eight process steps. In the wafer fab these steps are often referred to as operations. Breaking the large photo process into these eight steps simplifies the various aspects of microlithography. This chapter will first provide a short overview of the eight basic steps, followed by an in-depth analysis of the materials, equipment, and processes used in each photolithography step. The in-depth analyses of the steps are covered at the end of this chapter and in the following three chapters. Table 13.2 indicates the specific steps that are covered in a chapter.

TABLE 13.2 Eight Steps of Photolithography

Step	Chapter
1. Vapor prime	13
2. Spin coat	13
3. Soft bake	13
4. Alignment and exposure	14
5. Post-exposure bake (PEB)	15
6. Develop	15
7. Hard bake	15
8. Develop inspect	15

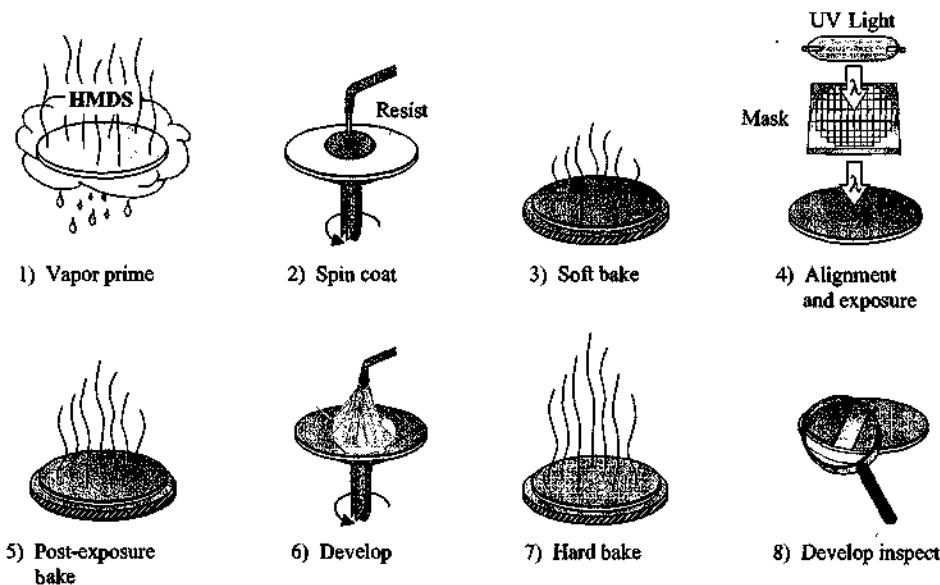


FIGURE 13.9 Eight Steps of Photolithography

The photolithography process has undergone significant equipment integration. Automated equipment, referred to as a *coater/developer track system*, or *tracks*, employs robots, automated material handling, and computers to perform all eight steps without human intervention. Integrated tracks offer many benefits over the previous stand-alone manual tooling for photolithography. This integration improved process control by controlling delays between process steps, processing wafers efficiently, increasing flexibility, reducing contamination due to environmental control and minimal operator handling, and increasing safety due to reduced operator exposure to chemicals.



Photolithography Track System
(Photo courtesy of Advanced Micro Devices)

Step 1: Vapor Prime

The first step in photolithography is to clean, dehydrate, and prime the surface of the wafer. The purpose of these steps is to promote good adhesion between the photoresist and the wafer surface.

Wafer cleaning may involve a wet clean and DI water rinse to remove contaminants. Most wafer cleaning is performed before entering the photolithography workbay. A dehydration dry bake is done in a closed chamber to drive off most of the adsorbed water on the surface of the wafer. The wafer surface must be clean and dry. Immediately after the dehydration bake, the wafer is primed with hexamethyldisilazane (HMDS), which acts as an adhesion promoter.

Step 2: Spin Coat

After priming, the wafer is coated with the liquid photoresist material by a spin coating method. The wafer is mounted on a vacuum chuck, which is a flat metal or teflon disc that has small vacuum holes on its surface to hold the wafer. A precise amount of liquid photoresist is applied to the wafer and then the wafer is spun to obtain a uniform coating of resist on the wafer (see Figure 13.10).

Different resists require different spin coating conditions, such as an initial slow spin (e.g., 500 rpm), followed by a ramp up to a maximum rotational speed of 3,000 rpm or higher. Some important quality measures for photoresist application are time, speed, thickness, uniformity, particulate contamination, and resist defects such as pinholes.

Process Summary:

- Wafer is held onto vacuum chuck
- Dispense ~ 5ml of photoresist
- Slow spin ~ 500 rpm
- Ramp up to ~ 3000 to 5000 rpm
- Quality measures:
 - time
 - speed
 - thickness
 - uniformity
 - particles and defects

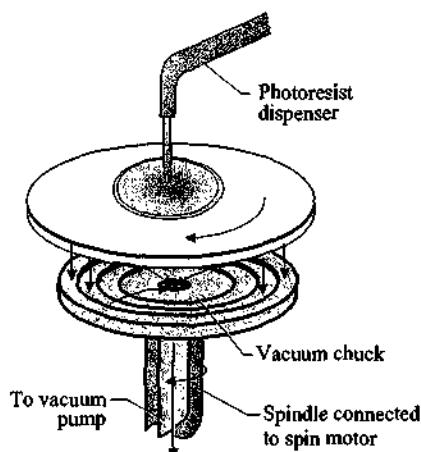


FIGURE 13.10 Spin Coat

Step 3: Soft Bake

After the resist has been applied to the wafer surface, it must undergo a soft bake. The purpose of this bake is to drive off most of the solvent in the resist. The soft bake process improves adhesion, promotes resist uniformity on the wafer, and yields better linewidth control during etching. Typical soft bake temperatures are 90 to 100°C for 30 seconds on a hot plate, followed by a cooling step on a cool plate to achieve wafer temperature control for uniform resist characteristics.

Step 4: Alignment and Exposure

The next step is referred to as alignment and exposure. The mask is aligned to the correct location of the resist-coated silicon wafer. The wafer surface could be bare silicon but usually has an existing pattern previously defined on its surface. Once aligned, the mask and wafer are exposed to controlled UV light to transfer the mask image to the resist-coated wafer (see Figure 13.11). The light energy activates the photosensitive components of the photoresist. Important quality measures for alignment and exposure are linewidth resolution, overlay accuracy, and particles and defects.

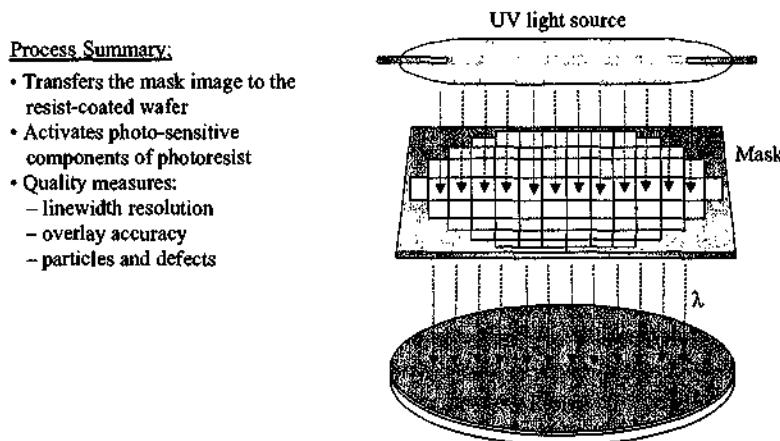


FIGURE 13.11 Alignment and Exposure

Step 5: Post-Exposure Bake (PEB)

There is a post-exposure bake on a hot plate at 100 to 110°C that is necessary for the deep UV (DUV) resists. This bake follows immediately after the photoresist exposure. A few years ago, it was an optional step for non-DUV conventional resists, but it has become a virtual standard even for conventional resists.

Step 6: Develop

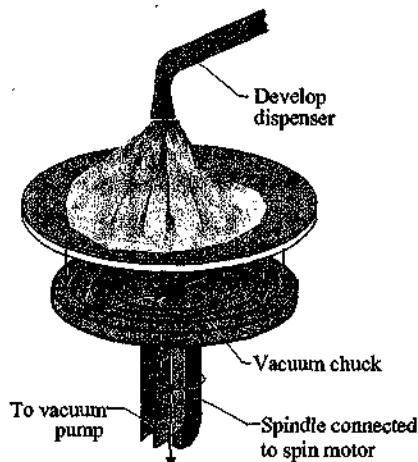
Develop is the critical step for creating the pattern in the photoresist on the wafer surface. The soluble areas of the photoresist are dissolved by liquid developer chemicals, leaving visible patterns of islands and windows on the wafer surface. The most common methods for development are spin, spray, and puddle (see Figure 13.12). Following development, the wafers are rinsed in deionized (DI) water and then spin-dried.

Step 7: Hard Bake

A post-development thermal bake, referred to as hard bake, is required to evaporate the remaining photoresist solvent and improve the adhesion of the resist to the wafer surface. This is the critical step for stabilizing the resist for the following etch or implant processing. The hard bake temperature for positive resists is about 120 to 140°C. This is a higher temperature than soft bake, but it cannot be too high or else the resist will flow and deform the pattern.

Process Summary:

- Soluble areas of photoresist are dissolved by developer chemical
- Visible patterns appear on wafer
 - windows
 - islands
- Quality measures:
 - line resolution
 - uniformity
 - particles and defects

**FIGURE 13.12** Photoresist Development**Step 8: Develop Inspect**

Once the resist is patterned on the wafer, an inspection is done to verify the quality of the resist pattern. The inspection system is nearly always automated for patterning on highly integrated critical layers. The purpose of this inspection is twofold: to identify wafers that have quality problems with the resist and to characterize the performance of the photoresist process to make improvements. If the resist is determined to be defective, it can be removed through resist stripping and the wafer can be reprocessed.

The goal of the photolithography process, as in any manufacturing process, is a defect-free product. However, it would be catastrophic to not inspect and leave defects in the resist. Develop inspect is one of the few areas in the wafer fabrication process where errors can be detected and corrected. Once defective wafers are moved to the next patterning step, usually etch, then there is no further chance to correct mistakes. If a wafer is etched incorrectly, it has a fatal defect, is considered scrap, and is of no further value to the company. This is why inspection data to characterize and improve the photoresist process is so important.

VAPOR PRIME

To successfully manufacture integrated circuits, wafers must be meticulously clean at all steps of the wafer fabrication process. Cleaning steps are necessary due to the inevitable contamination that occurs during storage and handling between process steps. Wafer surface preparation is critical to achieving a high-yield photolithography process, since many types of defects can be traced back to contaminated wafers.

Wafer Cleaning

The first step of photolithography is to clean and prepare the surface of the wafer, usually before the wafers arrive at the photolithography workbay. Wafer contaminants and the appropriate cleaning processes were discussed in Chapter 6. As a review, undesirable surface contaminants are particles, metallic impurities, organic contamination, and native oxide. There are many sources of these contaminants, including people, process chemicals, process equipment, packaging and storage, wafer handling methods, and environmental conditions. Thin layers of contaminants on the wafer can consist of ionic (metallic) impurities and atomic and polymeric (organic) layers, which are all difficult to detect. Refer to Chapter 6 for a discussion of types of contamination and cleaning processes.

One of the major effects of contaminants on the wafer surface during photolithography is poor adhesion of the photoresist to the wafer. This condition creates a yield problem of resist lifting during development or the subsequent etch operation. Resist lifting leads to undercutting of the underlying film layer during the etch process (see Figure 13.13 on page 346). Particulate contamination in the resist can lead to uneven resist coating and pinholes in the resist.

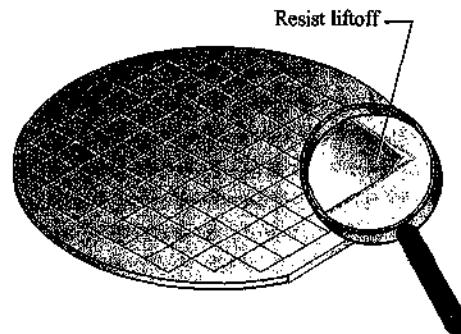


FIGURE 13.13 Effect of Poor Resist Adhesion Due to Surface Contamination

Often wafers entering the photolithography process will have just completed an oxidation or deposition operation and will be in a clean condition. The best situation is to coat these clean wafers with resist as soon as possible. To achieve this, some fabs will place time limits on how long wafers can wait before undergoing photolithography processing to minimize contamination adsorption on the wafer surface.

Dehydration Bake

Silicon wafers readily adsorb moisture on their surface. A wafer surface exposed to moisture is called *hydrophilic* (also called *hydrated*). For resist adhesion, it is important to have a dry, or *hydrophobic*, wafer surface. A hydrophobic wafer is also called *dehydrated* and promotes good resist adhesion. One way to maintain a hydrophobic wafer surface is to coat the wafer with photoresist as quickly as possible after leaving the previous process. Another method is to maintain a controlled room humidity below 50% relative humidity.

Due to the criticality of having a dry wafer surface for resist adhesion, a *dehydration bake* is usually done before priming and resist spin coating. The actual bake temperature varies, with 200 to 250°C commonly used. Dehydration bake temperatures usually do not exceed 400°C, and often cannot attain this temperature due to the temperature sensitivity of the underlying layers (e.g., aluminum). Total wafer surface dehydration occurs at temperatures above 750°C. However, moisture from the atmosphere is quickly reabsorbed back on the wafer surface during cooling; therefore, there is little benefit to heating to this temperature.⁴

Dehydration bake is typically done in a convection oven with a dry inert gas (such as nitrogen) or in a vacuum oven. It can also be done on a hot plate, which is followed by a cold plate to quickly reduce the wafer temperature. Since nearly all wafer fabs use automated wafer track systems, the dehydration bake process is integrated into the wafer handling system.

Wafer Priming

Right after dehydration bake the wafer is primed with *hexamethyldisilazane (HMDS)*, which serves as an adhesion promoter. This process is similar to a paint primer being used to prepare wood for a coat of paint. The HMDS reacts with the silicon surface to tie up molecular water, while also forming a bond with the resist material. It essentially serves as a link between the silicon and the resist so that these materials become chemically compatible.

An important aspect of wafer priming is that the wafer should be coated with resist quickly after the prime operation to minimize moisture problems. It is recommended that coating be performed no later than 60 minutes after completing the priming step.⁵ This priming is typically controlled by software on the automated track system.

Priming Techniques ■ HMDS may be applied by puddle, spray, and vapor methods. The method used to prime the wafer is integrated onto the wafer track system.

Puddle Dispense and Spin. The puddle dispense and spin method is used for single wafer processing (see Figure 13.14). The temperature and volume are easily controlled and the system requires a drain and exhaust. This method consumes a large amount of HMDS, which can be a disadvantage.

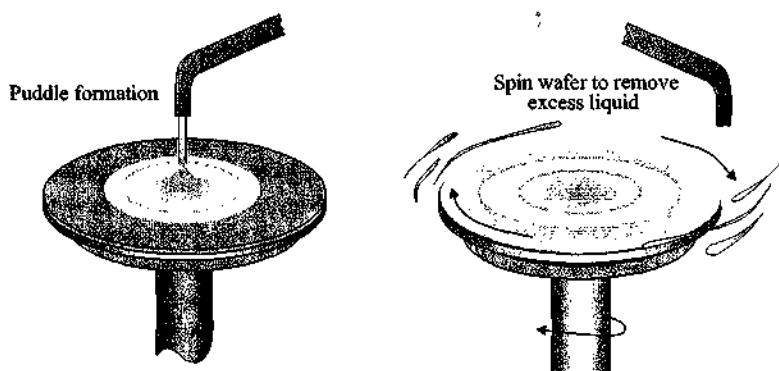


FIGURE 13.14 HMDS Puddle Dispense and Spin

Spray Dispense and Spin. Spray dispense uses a nozzle spray to deposit a fine mist of HMDS on the wafer surface. An advantage of this approach is the spray will assist in particle removal from the wafer surface. A disadvantage is a longer application time and a high consumption of HMDS.

Vapor Prime and Dehydration Bake. The most common method for applying HMDS to the wafer surface is with a *vapor prime coating*. The vapor priming is done at a typical temperature and time of 200 to 250°C for 30 seconds. An advantage of vapor priming is there is no contact with the wafer, which reduces the possibility of particulate contamination from the liquid HMDS. Vapor priming also minimizes consumption of HMDS. Adequate priming of the wafer surface is confirmed with a contact angle meter (see Chapter 7).

One approach is to first perform a dehydration bake followed by a vapor prime of single wafers by thermal conduction heating on a hot plate (see Figure 13.15). The wafer holder is typically made of quartz. Advantages of this approach are inside-out baking of the wafer, low defect density, uniform heating, and repeatability.

Process Summary:

- Dehydration bake in enclosed chamber with exhaust
- Hexamethyldisilazane (HMDS)
- Clean and dry wafer surface (hydrophobic)
- Temp ~ 200 to 250°C
- Time ~ 60 sec.

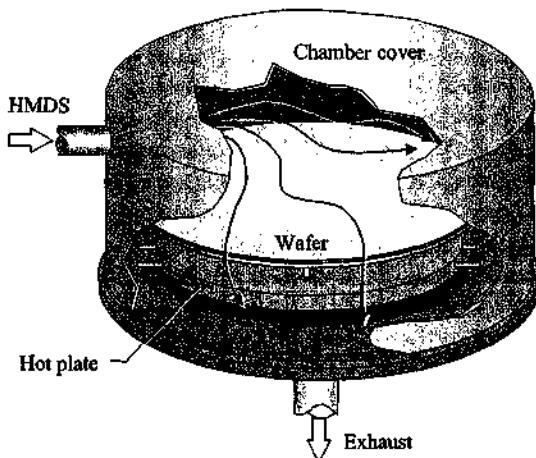


FIGURE 13.15 HMDS Hot Plate Dehydration Bake and Vapor Prime

Another method for dehydration bake in conjunction with vapor priming is to use a vacuum chamber with a nitrogen carrier gas. This is a batch process where the wafers are placed in a quartz holder in the oven chamber. The heated chamber is evacuated and back-filled to a preset pressure with HMDS vapor in the nitrogen carrier gas. At the completion of the pretreatment, the oven is evacuated and back-filled with nitrogen at atmospheric pressure.

SPIN COAT

Once the wafer surface has been cleaned, dehydrated, baked, and primed, it is ready for the application of photoresist. Photoresist is applied by spin coating, where the wafer is spun to form a thin film of resist on the wafer surface. Before considering the photoresist spin process, we will review the different chemistries of photoresists.

Photoresist

Photoresist is an organic compound that experiences a change in solubility in a developer solution when exposed to ultraviolet (UV) light. Photoresists used in wafer fabrication are applied to the wafer surface as a liquid but dried into a film. The purpose of photoresist in wafer fabrication is:

1. To transfer the mask pattern to the resist on the top layer of the wafer surface.
2. To protect the underlying material during subsequent processing (e.g., etch or ion implantation barrier).

As successive generations of device circuit density have reduced critical dimensions, resist technology has undergone improvements in order to transfer the submicron linewidth patterns to the wafer surface. These resist improvements include:

1. Better image definition (resolution).
2. Better adhesion to semiconductor wafer surfaces.
3. Better uniformity characteristics.
4. Increased process latitude (less sensitivity to process variations).

Types of Photoresists ■ The two major types of optical photoresists are negative resist and positive resist. This classification is based on how the resist material responds to UV light. For a negative resist, the UV-exposed regions become crosslinked and hardened. This makes the exposed photoresist less soluble in the developer solution; the photoresist is not removed in the developer liquid. A negative mask image is patterned in the resist. For a positive resist, UV-exposed regions of the resist become more soluble and a positive mask image appears in the resist. A positive resist breaks down during exposure to light and the exposed areas are easily washed away in the developer solution.

We can also group photoresists based on the smallest CD the resist can pattern. A major group is conventional resists capable of patterning linewidth dimensions down to and including 0.35 μm . A new resist technology introduced in the late 1990s was chemically amplified (CA) resist for deep UV (DUV) wavelengths (see the following section). Chemically amplified resist technology can pattern fine-geometry CDs of 0.25 μm and below in high-volume production. This patterning has also been demonstrated in a lab environment for a CD of 0.05 μm . For fabrication of high-performance ICs, a wafer may be patterned with conventional resist for noncritical layers and CA resist for critical layers. We will refer to all optical resists that are not chemically amplified as conventional resists. There are also special resists used for the nonoptical photolithography techniques of e-beam pattern reproduction and X-ray exposure systems (discussed in Chapter 15). Our focus will be on the conventional and chemically amplified optical resists used in wafer fabrication.

Negative Versus Positive Resists ■ As previously outlined, negative photoresist hardens when exposed to light and becomes insoluble. When developed, the image is the opposite of the original mask pattern. Positive photoresist softens when exposed to light and becomes soluble. When developed, the image is the same as the original mask pattern.

The earliest resists in semiconductor microlithography were primarily negative resists until the mid-1970s. Negative resists of this era exhibited good adhesion to the wafer substrate and good resistance to etching. However, due to the swelling and distortion during develop, negative resists typically have a resolution limit of about 2 μm .⁶ As long as the circuit linewidth CD remained above this dimension, then negative resists were acceptable. With the introduction of VLSI and ULSI technology and their associated micron and submicron circuit image sizes, the negative resist

type was replaced by positive resist. Positive resist had been around for many years, but it had poorer adhesion and its capability for improved resolution was not needed.

This change in the 1970s from negative to positive resist presented a fundamental change in the photolithography process. It required changing the polarity of the photomasks, meaning clear-field regions became dark (opaque) and dark-field regions became transparent. However, this change did not simply involve changing the fields in the mask-making process. Mask dimensions print differently with the two resists. As an example, for a negative resist with a clear-field mask, the dimensions in the resist are smaller than the corresponding photomask dimensions because of light scattering (diffracting) around the image. It is the opposite for a positive resist and a dark-field mask, as diffraction will tend to enlarge the image in the resist. Thus, changing from negative to positive resist required new masks and different adjustments to the mask during construction. Positive resists remain the dominant type of resist today.

Photoresist Physical Properties

In its most elemental form, photoresist is a polymer solution in an organic solvent. The photoresist must perform well under all the different process conditions, including coating, spinning, baking, develop, ion implantation, and etch. There are many different types of photoresists used in wafer applications, each with its own physical properties that are directly related to the photolithography process requirements. A particular resist is selected based on the following physical properties:

- ◆ Resolution
- ◆ Contrast
- ◆ Sensitivity
- ◆ Viscosity
- ◆ Adhesion
- ◆ Etch resistance
- ◆ Surface tension
- ◆ Storage and handling
- ◆ Contaminants and particles

Resolution. Resolution is the ability to differentiate between two or more closely spaced patterned features on the wafer surface. A practical way to interpret resolution is by the smallest feature that can be printed on a wafer and meets specified quality requirements. The smaller the critical dimension produced, the better the resolution capability of the resist and the photolithography system.

Contrast. The *contrast* represents the sharpness of the transition from exposure to non-exposure in photoresists (see Figure 13.16). Contrast represents the ability of the resist to become exposed only in those areas defined by the clear region of the photomask. It is desirable to have high-contrast to produce vertical resist sidewalls.

Poor Resist Contrast

- Sloped walls
- Swelling
- Poor contrast



Good Resist Contrast

- Sharp walls
- No swelling
- Good contrast



FIGURE 13.16 Resist Contrast

Sensitivity. *Sensitivity* is the minimum amount of light energy (measured in millijoules/cm² or mJ/cm²) at a certain wavelength that is needed to produce a good pattern in the photoresist on the wafer surface. The amount of light energy supplied to a resist is commonly called the *exposure dose*. Sensitivity is important for new resists because of the low energy levels emitted at the shorter UV wavelengths (e.g., deep UV). Resist manufacturers sensitize their resist formulations for maximum absorption in the areas of optimum energy output of the radiation sources available.

Viscosity. The *resist viscosity* represents a quantitative measure of flow characteristics for the liquid resist. Viscosity is time-dependent in that it increases as the solvent in the resist evaporates during usage. Viscosity is important because it affects resist thickness and the uniformity of the resist over the wafer surface topography such as steps and small gaps. As viscosity increases, the resist has less tendency to flow and its thickness on the wafer increases. A thicker resist provides better step coverage and better dry etch resistance but makes the resolution of small openings more difficult. Thick resist also has less probability of having pinholes, which are microscopically small voids that pass completely through the resist to the substrate material. Pinholes are undesirable for subsequent etch processing because the etch chemicals can pass through the small voids and damage the underlying substrate material. Low viscosity resist has more tendency to flow and will lead to thinner coverage of the wafer surface.

Solid content represents the amount of the liquid resist that would remain as a solid if the solvent were evaporated. Specific gravity (SG) is a measure of the resist density and is related to the solid content in the resist. A higher SG means there are more solids in the resist, which indicates a higher viscosity and lower tendency to flow.

Viscosity is measured in the unit of poise. Photoresist viscosity is measured in centipoise (1/100th of a poise), or cps. Viscosity may be reported in centistokes (cs), which is the kinematic viscosity, whereas the centipoise is the absolute viscosity. The kinematic viscosity is found by dividing the absolute viscosity (centipoise) by the density (specific gravity). For instance, a 70-cps resist with an SG of 0.8 would produce a kinematic viscosity of:

$$\begin{aligned}\text{Kinematic viscosity} &= [\text{absolute viscosity (cps)}] / (\text{specific gravity}) \\ &= (70 \text{ cps}) / 0.8 \text{ SG} \\ &= 87.5 \text{ centistokes (cs)}\end{aligned}$$

Adhesion. *Resist adhesion* describes how strongly the resist sticks to the substrate. The resist must adhere to many different types of surfaces, including silicon, polysilicon, silicon dioxide (doped and undoped), silicon nitride, and different metals. Lack of resist adhesion leads to distorted patterns on the wafer surface. The resist adhesion must withstand exposure, development, and subsequent processing (e.g., etch and ion implantation).

Etch Resistance. The resist film must maintain its adhesion and protect the substrate surface from the subsequent wet and dry etch processes (see Chapter 16). This property is known as *etch resistance*. Some dry-etch processes are done at higher temperatures (e.g., 150°C), which require a resist that has thermal stability to maintain its shape.

Surface Tension. *Surface tension* refers to the molecular attraction forces in a liquid that tend to pull surface molecules toward the body of the liquid. An example of high surface tension is water beading up on a waxed automobile surface (see Figure 13.17). Resist has molecular forces that create a relatively high surface tension so that the resist molecules hold together during the various process steps. At the same time, the surface tension of resist must be low enough to provide for good flow and wafer coverage during application.

Storage and Handling. Photoresist chemicals are activated in the presence of energy, either as light or heat, which requires carefully controlled storage and usage conditions. Resists are specified with a shelf life and storage temperature environment. Crosslinking of negative resists and sensitizer-delay of positive resists can occur with extended storage time or elevated temperature. If the solvent on the resist was allowed to evaporate due to an unsealed container, the viscosity would change rapidly and solids could precipitate from the liquid.

Resist manufacturers have developed automated dispense mechanisms to control contamination and evaporation. The resist containers are often a closed system, such as collapsible pouches that are opaque to light and loaded directly into the wafer tracks to avoid opening in the atmosphere.

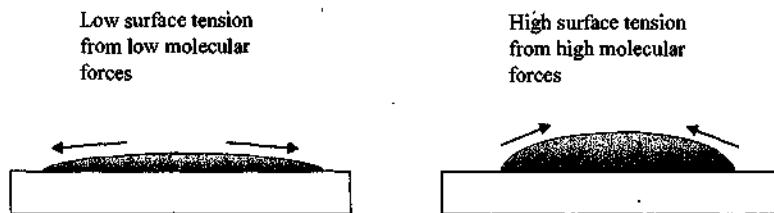


FIGURE 13.17 Surface Tension

Contaminants and Particles. As with any chemical used in wafer fabrication, the purity of the resist material is important. The most critical purity concerns for resist are mobile ionic contaminants (MICs) and particles. The resist is applied to the wafer surface material and can easily introduce detrimental contamination to the wafer. To control contamination and particles, resist suppliers meet extremely tight filtration and packaging procedures. Localized point-of-use resist filtration with membrane filters just before resist application can control contamination in the resist to < 1 ppb.

Conventional I-Line Photoresists

Conventional i-line photoresists for optical microlithography are those suitable for i-line UV wavelengths (365 nm), which corresponds to noncritical layers with a CD down to 0.35- μm . This condition includes both positive and negative resists, although positive resists are the most common. Note that the characteristics of i-line photoresist also represent the same basic properties for resists used at g-line (436 nm) and h-line (405 nm) wavelengths.

I-line photoresists are composed of three basic components,⁷ and frequently have a fourth component of special additives (see Figure 13.18):

1. Resin (polymer material)
2. Sensitizer
3. Solvent
4. Additives (optional)

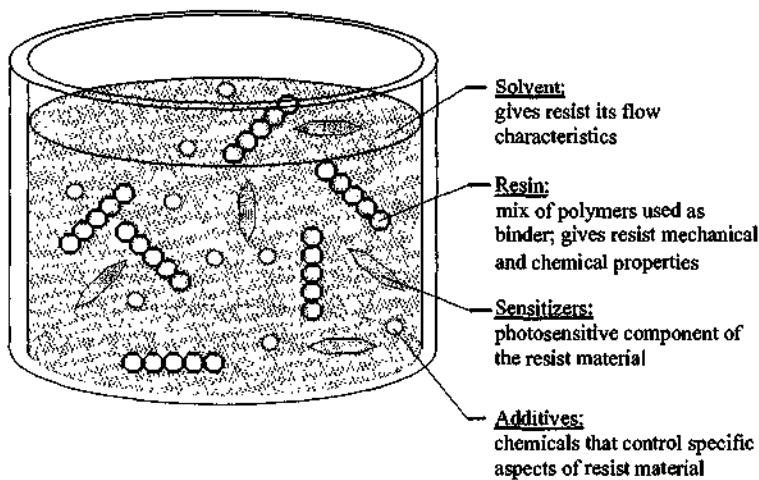


FIGURE 13.18 Components of Photoresist

Resin. The resist resin is an inert matrix of polymers (large organic molecules containing carbon, hydrogen, and oxygen) which is used as a binder to hold together the different materials in the resist. Resin gives the resist its mechanical and chemical properties, such as adhesion, etch resistance film thickness, flexibility, and thermal flow stability. The resin material is typically insensitive to light, meaning it does not undergo chemical change upon exposure to UV light.

Sensitizer. The resist sensitizer is the photosensitive component of the resist material that reacts in response to radiant energy in the form of light, especially in the UV region. The sensitizer reacts photochemically in response to light energy.

Solvent. The resist solvent keeps the resist in its liquid state until it is applied to the wafer substrate. Most solvents are evaporated from the resist before the exposure is done and have little effect on the photochemistry of the resist.

Additives. Resist additives are usually proprietary chemicals, which means the contents are developed by the manufacturer and not disclosed to the public for competitive reasons. Additives are used to control or modify specific chemical or light response aspects of the resist material. It also includes dyes to control the reflectivity of the resist.

Negative I-Line Photoresists ■ The resin for i-line negative resist is typically a chemically inert polyisoprene polymer that is a natural rubber. The polyisoprene is suspended in a solvent such as xylene. The resist sensitizer is a photoactive agent that releases nitrogen gas on exposure to UV light with the proper wavelength, generating free radicals that form crosslinks between the rubber molecules. This crosslinked rubber formed from UV light exposure becomes insoluble in the organic developer solution (see Figure 13.19). At the same time, the unexposed (and therefore un-polymerized) areas of the resist are rinsed away in the developer. These basic steps for crosslinking of negative resist are shown in Table 13.3.

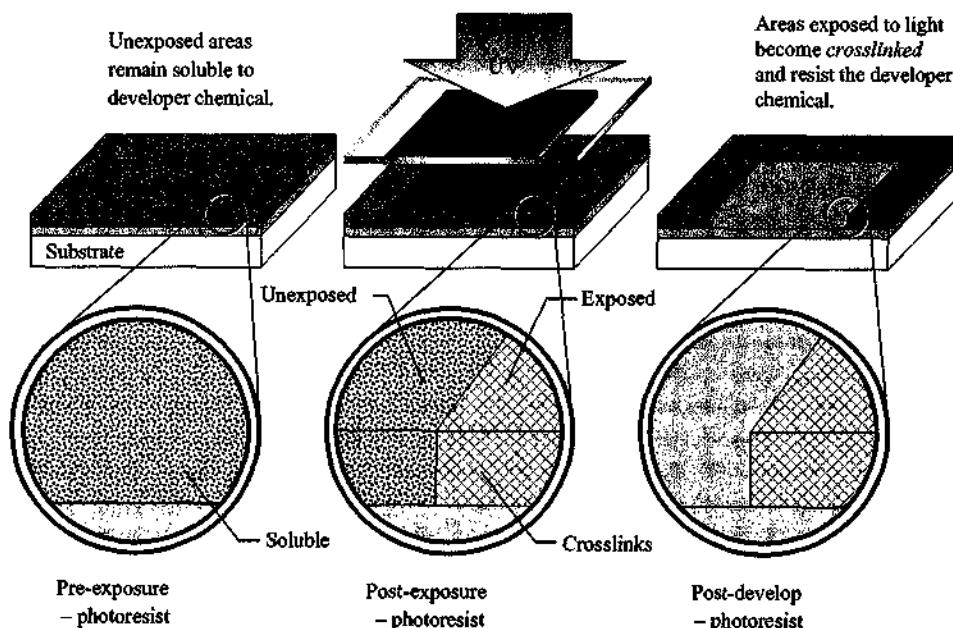


FIGURE 13.19 Negative Resist Crosslinking

TABLE 13.3 Steps to Negative Resist Crosslinking

1. Resist resin is polyisoprene rubber polymer suspended in solvent (soluble in developer solution).
2. Photoactive sensitizer agent releases N_2 during exposure.
3. Release of N_2 generates radicals.
4. Radicals polymerize resist by crosslinking rubber polymer (insoluble in developer solution).

A major disadvantage of conventional negative resists has been their solvent-induced swelling of the exposed resist regions during development. This swelling distorts the resist pattern on the wafer surface and is unacceptable for fine-geometry features with micron and submicron critical dimensions. Another disadvantage is the resist's reactivity with oxygen during UV exposure that can inhibit crosslinking. There is, however, research into new negative resist formulations that

do not swell or distort patterns.⁸ A negative resist has high exposure speed (which translates into higher throughput of wafers) and good wafer adhesion.

Positive I-Line Photoresists ■ Positive i-line photoresist is the most common resist in IC photolithography. The resin material in positive i-line photoresists is a phenol-formaldehyde polymer called *novolak*. An example of a common use of novolak is the glue that binds the different layers of plywood lumber. Novolak is a long chain polymer that forms the key resist film properties, such as good adhesion and chemical resistance, making it an acceptable barrier to subsequent processes such as etch. Novolak resin dissolves in developer solution when there is no dissolution inhibitor present.

The sensitizer in positive i-line photoresist is referred to as a *photoactive compound (PAC)*, the most common being a diazonaphthoquinone (DNQ). Before exposure to light, the DNQ serves as a strong dissolution inhibitor, reducing the dissolution rate. After exposure to UV light, DNQ promotes dissolution in the developer by a factor of about 100 or more (see Figure 13.20).⁹ This i-line resist is often called *DNQ-novolak*, representing the mixture of diazonaphthoquinone (DNQ) and novolak resin. It originally evolved from materials used to make blueprints. Note the mask in Figure 13.20 is the direct opposite of the mask used in Figure 13.19.

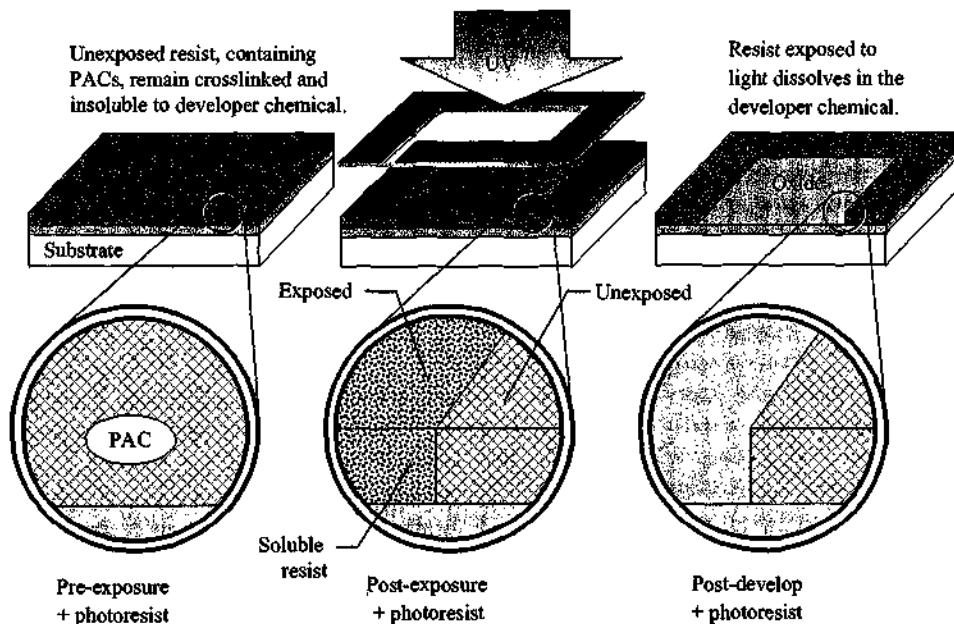


FIGURE 13.20 PAC as Dissolution Inhibitor in Positive I-Line Resist

When exposed to light energy of the proper wavelength, the diazonaphthoquinones (DNQs) photochemically decompose in the resist. The exposed DNQ now becomes a dissolution promoter in exposed regions of the resist. This change is accomplished when decomposition of the DNQ creates carboxylic acid, which is highly soluble in the developer solution. In this manner, exposed positive resist changes from an insoluble to a soluble form and dissolves in the developer solution. This is the dissolution of the positive resist (see Table 13.4 on page 354). With this process, areas of the resist that correspond to transparent mask areas will dissolve, thus creating a positive image. In the DNQ resist, regions exposed to light are about 100 times more soluble in the developer than unexposed resist regions.¹⁰

TABLE 13.4 Steps to Positive I-line Resist Dissolution in Developer

1. Resin is phenol-formaldehyde polymer (novolak) suspended in solvent.
2. Sensitizer photoactive compound DNQ added to novolak as strong dissolution inhibitor (insoluble in developer).
3. During exposure, DNQ photochemically decomposes and creates carboxylic acid.
4. Carboxylic acid promotes dissolution of novolak in exposed areas of resist (resist becomes soluble in developer).

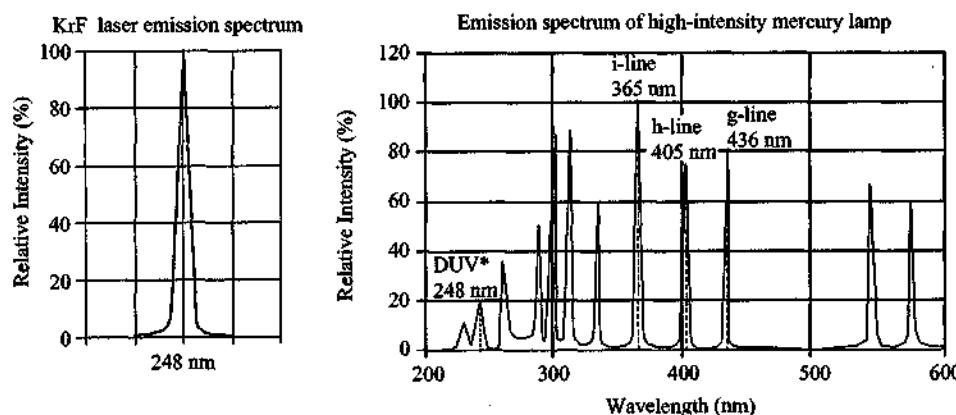
A major advantage of a positive resist is that the unexposed areas of the resist are not affected by the developer, since the resist is initially insoluble and remains as such. Thus the pattern of fine lines that is transferred to the resist will maintain the linewidths and shape during the photolithography process, producing good linewidth resolution. Older positive resists had marginal adhesion strength, but modern positive resists have been designed with more acceptable adhesion.

A reason why positive resist has good resolution is contrast. The positive resist can distinguish better between the light and dark areas of the mask, making a sharp transition from exposure to nonexposure in the resist. Positive photoresist has good contrast characteristics, which produces improved line resolution due to sharp resist sidewalls (see Figure 13.21).

**FIGURE 13.21** Good Contrast Characteristics of Positive I-Line Photoresist

Deep UV (DUV) Photoresists

For optimum pattern resolution, a goal in microlithography has been for the wavelength of the exposure light to be proportional to the critical linewidth dimension (this is explained in Chapter 14). For exposure with an i-line wavelength of 365 nm in the mid-1990s, a CD of 0.35 μm was achievable. In order to achieve a CD of 0.25 μm on critical layers in the latter 1990s, it was necessary to reduce the wavelength of the exposure light source to a value of around 250 nm (0.25 μm). This value corresponds to the 248 nm UV wavelength referred to as *deep UV (DUV)*. This wavelength is shown in Figure 13.22.



* Intensity of mercury lamp is too low at 248 nm to be usable in DUV photolithography applications. Excimer lasers, such as shown on the left provide more energy for a given DUV wavelength.

FIGURE 13.22 DUV Emission Spectrum Mercury Lamp Spectrum
(Used with permission from USHIO Specialty Lighting Products)

Standard i-line photoresists are generally not acceptable at the DUV wavelengths necessary for CDs of 0.25 μm and below due to lack of sensitivity to smaller wavelengths. Resists have a sensitivity to the minimum amount of light energy required to transfer an acceptable pattern. The DNQ i-line resist system has excessive light absorption and lacks sensitivity at the low light intensity of the shorter DUV wavelength.¹¹ If the i-line resist has excessive light absorption, then the light never penetrates the full thickness of the resist, leading to incomplete exposure and poor image formation. The lower light sensitivity of the i-line resist would require longer exposure time, causing a decrease in wafer throughput. This low sensitivity is a serious problem when the light source is a mercury lamp with its limited output at the DUV wavelength of 248 nm. For this reason, i-line photoresists are used on noncritical wafer layers of 0.35 μm linewidths and above.

Chemical Amplification for DUV Resist ■ A major change in photoresist technology occurred with the introduction of a new type of photoresist for DUV wavelengths based on chemical amplification. *Chemical amplification (CA)* is a means to increase the sensitivity of the resist substantially over that of the DNQ-novolak i-line resists. All positive and negative tone 248-nm resists are chemically amplified.¹² The researchers who developed chemically amplified resist in the 1980s at IBM (International Business Machines) realized that the light absorption and sensitivity of the i-line resist system was unacceptable for shorter DUV wavelengths, creating the need for a new photoresist that offered the benefits of chemical amplification.¹³

Chemically amplified DUV resists are formulated to react at an accelerated rate to the exposing DUV light source based on an acid-catalyzed reaction. This action is accomplished by boosting the resist sensitivity with the use of a sensitizer known as a *photoacid generator (PAG)*. The PAG produces an acid upon exposure to radiation from deep UV light. This acid is only produced in the resist regions exposed to light. In the areas of resist where there is no light exposure, the acid is not created. Most well-known acid generators used in DUV resists are onium salts, such as iodonium and sulfonium salts.¹⁴

CA DUV resist technology has undergone significant development over the years. The basic principle for all CA DUV resist resins is that the resin requires a chemical protecting group that makes it insoluble in aqueous (water-based) developer. The acid generated by the PAG in the DUV-exposed areas of the resist then removes the protecting group (referred to as *deprotection*) through a catalytic reaction while heated during the post-exposure bake step. After heating the DUV resist, the exposed resin becomes soluble in the developer.

An early CA DUV resin developed and used at IBM in the 1980s to manufacture DRAMs using DUV lithography was t-BOC (tert-butoxycarbonyl).¹⁵ The t-BOC is the protecting group for the resin known as PHS or PHOST (poly[hydroxystyrene]). Once the photo-generated acid removes the t-BOC-protecting group (referred to as an acid-catalyzed t-BOC deprotection reaction), then the PHS resin becomes soluble in developer.

PHS resin is well-established for use in CA DUV positive resists (note that all major resist manufacturers are actively researching “new generation” CA DUV resists).¹⁶ PHS resin is a phenolic copolymer that has protecting groups that make it dissolve poorly in an aqueous-based solution. During deep UV exposure, the PAG acid is photochemically generated in all exposed resist regions. This acid is the chemical catalyst that removes (cleaves) the protecting group during post-exposure thermal bake. This operation renders the exposed PHS highly soluble in the aqueous developer (see Figure 13.23 on page 356). Essentially, these chemical reactions transform the exposed regions of the DUV resist from an insoluble resist into a highly soluble resist in an aqueous-based developer. Table 13.5 on page 356 outlines the four steps to developing a CA DUV resist.

The first commercial CA DUV resist was a negative system based on crosslinking a phenolic resin with an acid catalyst. All negative CA resists are built on this phenolic resin crosslinking. Negative CA resists have poor process latitude in imaging small contact holes because of the clear-field image during exposure. At the same time, negative DUV resists are resistant to thermal flow for temperatures below 150°C, due to the crosslinking of the phenolic resin polymer.¹⁷ This condition can be beneficial in some etch applications.

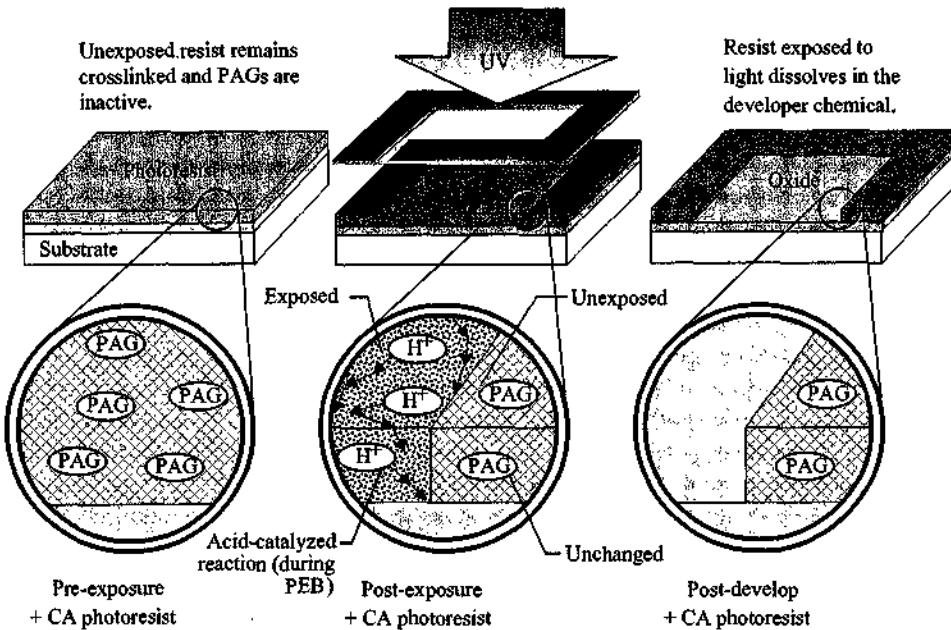


FIGURE 13.23 Chemically Amplified (CA) DUV Resist

TABLE 13.5 Exposure Steps for Chemically Amplified DUV Resist

1. Resin is phenolic copolymer with protecting group that makes it insoluble in developer.
2. Photoacid generator (PAG) generates acid during exposure.
3. Acid generated in exposed resist areas serves as catalyst to remove resin-protecting group during post-exposure thermal bake.
4. Exposed areas of resist without protecting group are soluble in aqueous developer.

Positive CA resists are the dominant type for patterning wafers with 0.18- μm generation devices using DUV wavelengths. One of the main benefits of CA DUV resists is the speed of resist exposure. CA chemistry is capable of providing a 10-fold improvement in the exposure speed over DNQ-novolak-based resist systems without a major compromise in any of the resist lithographic performance areas.¹⁸ Chemical amplification in DUV resist improves the sensitivity of the resist for low-intensity light sources. In addition, DUV offers high-contrast imaging with vertical sidewall profiles and high resolution suitable for resists used on wafers with critical dimensions of 0.25 μm and below.¹⁹

193 nm DUV Resists. Suitable chemically amplified DUV resists are being introduced into production at this time for 193 nm wavelength exposure light capable of imaging a CD resolution of 0.18 μm . The biggest change is the unacceptability of resists based on phenolic polymers (DNQ i-line resist) or hydroxystyrene (248 nm DUV resist). Some early 193 nm resists are based on polyvinyl phenol and polymethylmethacrylate (PMMA).²⁰ Another chemical platform under investigation is cyclic olefin polymer materials, which offer superior etch properties.²¹ These resists may be used with multilayer resist technology using top surface imaging (see the advanced lithography section in Chapter 15).

DUV Process Requirements ■ DUV resists are very sensitive to contamination, specifically amines (an organic compound such as ammonia) that are found in the ambient atmosphere. Less than several parts per billion (ppb) exposure to amines commonly found in air will lead to undesirable CD variation in the top of the resist profile. This resist variation leads to unacceptable linewidths after etch. Amine sensitivity in CA DUV resists requires that the air in all photolithography equipment be chemically filtered.²² Monitoring equipment used in photolithography can detect down to 500 parts per trillion amines.

CA resists for DUV are also dependent on the bake temperature at the post-exposure bake (discussed in Chapter 15) for reducing CD variation. Early DUV resists had significant temperature sensitivities, whereas newer resists have less temperature sensitivity. In addition, early DUV resists required the time between resist exposure and post-exposure bake processing to be limited to a few minutes. Recent DUV resists can be delayed slightly longer, up to about 30 minutes.

Photoresist Dispensing Methods

For semiconductor microlithography, the most widely used method to apply the liquid photoresist to achieve a uniform coating on the wafer surface is *spin coating*. There are four basic steps to applying resist by spin coating (see Figure 13.24):²³

1. **Dispense.** The resist is dispensed onto the wafer while it is stationary or spinning very slowly.
2. **Spin-up.** Quickly accelerate the wafer rotation to a high rpm (revolutions per minute) spin speed to spread the resist over the entire wafer surface.
3. **Spin-off.** Throw off excess resist to obtain a uniform resist film coating on the wafer.
4. **Solvent evaporation.** Continue to spin the coated wafer at a constant rpm until the solvent evaporates and the resist film is nearly dry.

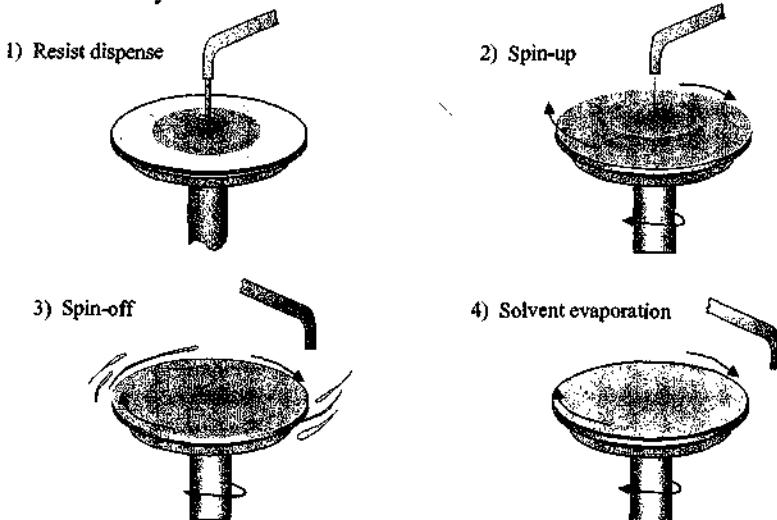


FIGURE 13.24 Four Steps of Photoresist Spin Coating

The two goals in resist spin coating are to have a uniform film coating on a wafer and to achieve a repeatable resist thickness that is maintained from wafer to wafer over extended periods. The target resist thickness is specified for a particular application and is typically on the order of 1 μm . Thickness variation of the resist film on a wafer should be measured less than 20 to 50 \AA across the entire wafer surface. Wafer-to-wafer resist thickness control is typically less than 30 \AA measured over many wafers.²⁴

Spin Coating Equipment ■ Wafer spin coating takes place in an automated wafer track system with wafer handling equipment to move the wafers from operation to operation (see Figure 13.25 on page 358). Robotic handling is preferred over conveyors to minimize particle generation and wafer damage.

The complexity of the automated wafer track is that there are many different operations being performed simultaneously. At any one time, the wafer track could be processing 15 to 20 wafers through different photolithography operations, such as vapor prime resist spin coat, develop, baking, and chilling. Track systems are connected to the wafer stepper and must interface with the stepper so that wafers are delivered and picked up at the correct time with limited waiting.

To spin coat the wafer, robotic handlers position the wafer on a vacuum chuck at the spin coat station in the wafer track. The chuck is a hollow, flat metal disc with holes in its surface that is

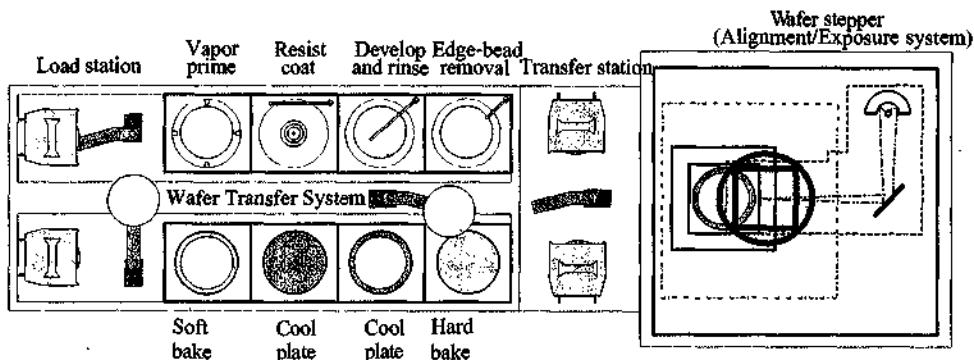


FIGURE 13.25 Automated Wafer Track System for Photolithography

connected to a vacuum. When the wafer is placed on the chuck surface, the vacuum pulls the wafer into intimate contact with the chuck to hold it while spinning.

The resist is dispensed on the wafer through a dispenser nozzle that should deliver a steady stream of resist (see Figure 13.26). The nozzle has a suckback feature to prevent after-dispense drips on the wafer surface. The nozzle is controllable in the X, Y, Z, and θ directions, giving the process engineer flexibility to position the nozzle in a location to optimize thickness, uniformity, and the amount of resist dispensed. A particular dispense method is called radial dispense, where the arm is actually moved across the wafer while the resist is dispensed. This action minimizes the amount of resist dispensed. The movement is computer controlled and can be from edge-to-center or center-to-edge.

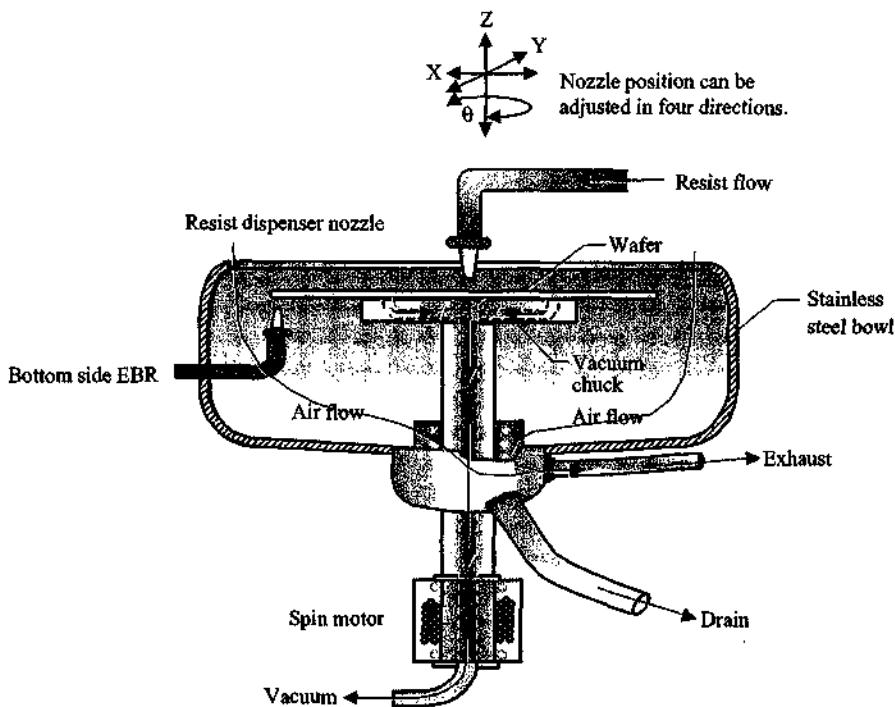


FIGURE 13.26 Photoresist Dispense Nozzle

Spin Coating Parameters ■ The manner of dispensing the liquid resist varies depending on the parameters defined for the application by the process engineer. The resist can be dispensed on a wafer while it is not rotating, known as *static dispense*. After the static dispense, the wafer is first spun at a low rpm to uniformly spread the resist. Once the resist approaches the wafer edge, the rpm is accelerated to the final spin speed (e.g. a typical final speed may be 4,000 rpm). Another approach

is to dispense the resist on a wafer that is spinning slowly (e.g., 100 to 200 rpm) in order to uniformly coat the wafer, followed by acceleration to the final spin speed. This is referred to as a *dynamic dispense*.

The amount of resist dispensed depends to a large degree on the viscosity of the resist. For example, a high-viscosity resist with a viscosity of 55 cps may require 2.4 cc of resist for a good coat. On the other hand, a low-viscosity resist with an 8 cps viscosity may need only 1.3 cc of resist. Lower viscosity resists typically have a thinner target thickness.²⁵

Resist thickness and uniformity on a wafer are critical quality measures. Thickness is not controlled by the amount of resist deposited since most of the resist flies off the wafer (less than 1% remains on the wafer). The most critical parameters for resist thickness are spin speed and resist viscosity. Many resist manufacturers publish thickness versus spin speed data for their resist formulation (see Figure 13.27). This data assists in determining the optimum final spin speed. Higher viscosities and slower spin speeds will produce thicker layers of photoresist. In general, the resist thickness has been found to vary with spin speed as:²⁶

$$\text{Resist thickness} \propto \frac{1}{(\text{RPM})^{1/2}}$$

Where RPM is the spin speed in revolutions per minute.

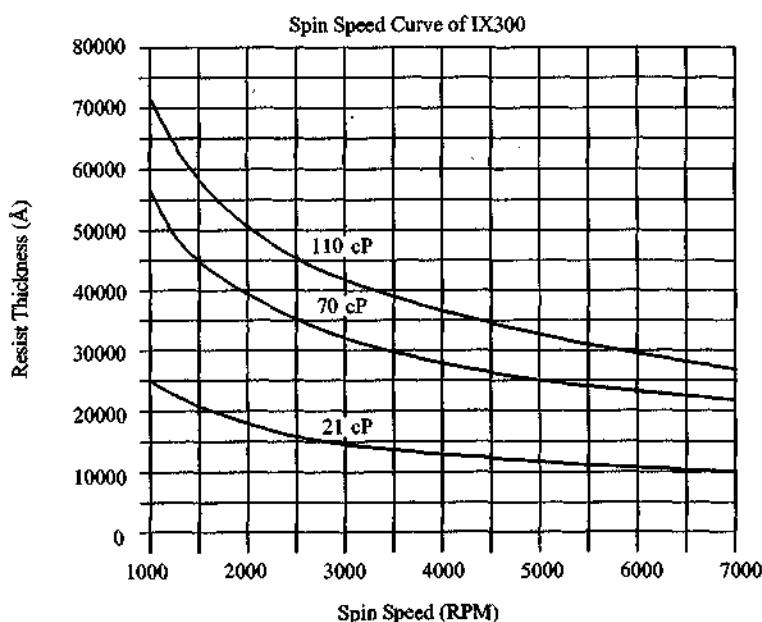


FIGURE 13.27 Resist Spin Speed Curve
(Used with permission from JSR Microelectronics, Inc.)

Different parameters affect resist thickness and uniformity. The acceleration ramp rate to attain the final spin speed is controllable on most spin coaters. High ramp rates usually produce better film uniformity than slow ramp rates. Another parameter is the height of the nozzle and its location during the dispense.

Another important parameter for resist dispense is environmental control. The wafer track system is a closed environment in order to control temperature, humidity, exhaust, and particulate contamination during application of the photoresist coating. Temperature and humidity strongly affect resist uniformity across the wafer. The minienvironment of a wafer track system for photolithography is often controlled for particulate contamination to a class 0.1 level. In addition, the exhaust air flow rate around the spin coater must be properly controlled to avoid negatively impacting resist thickness and uniformity. Too much air flow due to strong exhaust can cause excessive drying of the resist, while poor exhaust flow can cause particles to deposit on the wafer instead of being carried away by the air flow.

Edge-Bead Removal ■ During the wafer spin process, resist flows outward due to centrifugal force toward the wafer edge and onto the backside. This ridge of resist on the wafer edge and backside is called edge bead. When dry, the resist flakes off and causes particles. These particles can land on active circuit areas, on wafer carriers, and inside process equipment, leading to increased defect densities on the wafer. Furthermore, the resist on the back of the wafer can create problems by causing it to stick to wafer chucks.

Resist spin coat stations are equipped with an *edge-bead removal (EBR)* device (see Figure 13.26 on page 358). A common approach is to use a spray nozzle assembly to spray a small amount of solvent on the underside of the spinning wafer. The solvent laps up over the beveled edge to the topside, with careful control to ensure the solvent does not reach the topside of the resist (this control is important for defect density reduction). Typical solvents to remove the edge bead are propylene glycol monomethyl ether acetate (PGMEA) or ethylene glycol monomethyl ether acetate (EGMEA).

Some photolithography tools use a laser to expose the edge of wafers right after the normal wafer exposure has taken place. The exposure softens the resist and allows the edge-bead resist to be removed during the normal develop step or by a designated solvent nozzle. This step can be performed in the same track system that was used to spin coat the wafers.

SOFT BAKE

Following the application of resist on the wafer by spin coating, wafers are subjected to an elevated temperature step referred to as *soft bake* (also called *pre-bake*). The reasons for the soft bake of the resist are:²⁷

1. Drive off the solvent from the coated resist on the wafer.
2. Improve the adhesion of the resist so that it adheres adequately during the development step.
3. Relieve stresses in the resist film that occurred during the spin process.
4. Prevent resist from getting all over the equipment (keep the tool clean).

The soft bake temperature and time depend on the particular resist and process conditions. The starting point for setting the soft bake parameters is the process recommended by the resist manufacturer. After that, the process is optimized to achieve the adhesion and dimensional control required for the product. The soft bake temperature is usually in the range of 85 to 120°C. The duration of the soft bake varies for different resists, but 30 to 60 seconds is a typical time.

If the resist film were not soft baked and continued directly to alignment and exposure after resist apply, the following problems would occur:²⁸

1. The resist film would be tacky and susceptible to particulate contamination.
2. Inherent stresses in the resist film from spin coating would lead to adhesion problems.
3. The high solvent levels would cause inadequate discrimination between exposed and unexposed resist for proper dissolution during development.
4. Outgassing from the resist (from heat during exposure) could contaminate the lens of the optical system.

Before spin coating, photoresist typically contains between 65 to 85% solvents. After spinning, the solvent has been reduced to 10 to 20%, yet the film should still be considered in a liquid state.²⁹ The ideal amount of solvent after soft bake is about 4 to 7%. Because the solvent is reduced in the soft bake process, the thickness of the resist film also decreases.

Soft Bake Equipment

The preferred method for resist soft bake is heat conduction from a wafer on a vacuum hot plate³⁰ (see Figure 13.28). In this method, heat is quickly conducted from the hot plate through contact with the backside of the wafer to the resist. The resist is heated from the wafer-resist interface outward, which minimizes the potential for solvent entrapment. Because of the short cycle time (e.g., 30 to 60 seconds), this single-wafer hot plate method is suitable for the flow of multiple wafers through the process steps of an automated wafer track system. In the wafer track process flow, the heating

step is usually followed by a cooldown step on a cooling plate. This step rapidly cools the wafer for the next operation. The vacuum hot plate design is the same type as that used for dehydration bake. There is also the option of infrared (IR), microwave, and convection heating for soft bake, but these methods are not commonly used.

- Purpose of Soft Bake:
- Partial evaporation of photoresist solvents
 - Improves adhesion
 - Improves uniformity
 - Improves etch resistance
 - Improves linewidth control
 - Optimizes light absorbance characteristics of photoresist

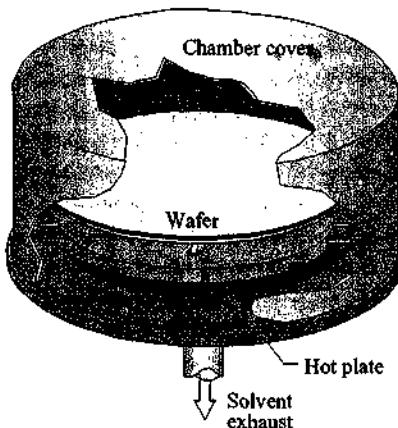


FIGURE 13.28 Soft Bake on Vacuum Hot Plate

Process Characterization

The soft bake process requires characterization to determine optimum settings. Reasons to optimize soft bake settings are to have better control over critical dimensions, improve the resist sidewall angle, increase resolution, improve contrast, and achieve wider process latitude.

The solvent content of the resist can be measured as a function of temperature with a *thermogravimetric analysis* (TGA). TGA uses the principle of volatilization, or polymer weight loss during heating, to represent the amount of solvent loss and thermal decomposition as a function of temperature.³¹ To perform this test, a sample is coated with resist and placed immediately into an oven without soft baking. While heating, the sample's weight is measured and plotted (see Figure 13.29).

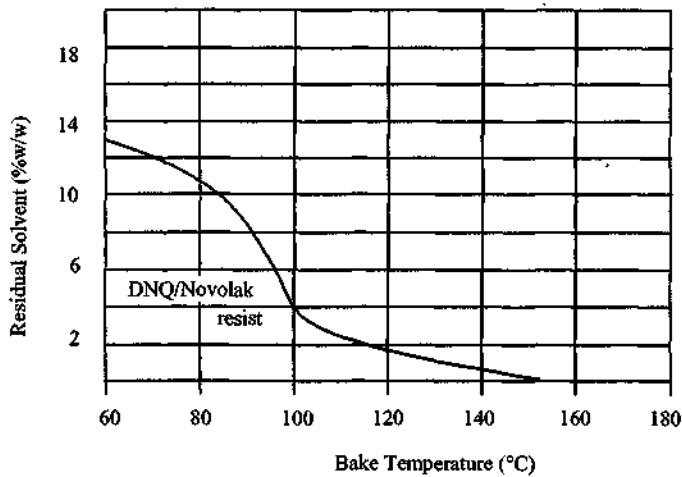


FIGURE 13.29 Solvent Content of Resist versus Temperature During Soft Bake

Another method to characterize the amount of solvent removal is to measure the change in resist thickness as a function of temperature. As solvent evaporates, the resist will become thinner. The temperature at which the resist thickness stabilizes is where most solvents are removed.

PHOTORESIST QUALITY MEASURES

Quality measures for photoresist processing are provided in Table 13.6.

TABLE 13.6 Key Quality Measures for Photoresist

Quality Parameter	Types of Defects	Remarks
1. Resist adhesion.	A. Resist dewetting, which is also called lift-off. The resist does not adhere to the substrate, causing problems during subsequent etch or ion implantation processing.	Possible causes of resist lift-off are: <ul style="list-style-type: none"> • Contamination on wafer surface. • Inadequate HMDS priming or dehydration bake resulting in moisture on wafer surface (measured by contact angle meter as described in Chapter 7). • Excessive HMDS priming can cause resist “popping,” which is failure of poorly adhered resist. • SiO_2 is a difficult surface to achieve good adhesion because it is a hydrophilic surface (water attracting).*
2. General quality of resist coating on wafer.	A. Pinholes (very small holes) in the resist. B. Splashback (drops of resist fall on resist coating).	<ul style="list-style-type: none"> • Particulate contamination on mask/reticle (evident after exposure) or on wafer. Check surface preparation cleaning. • Improper exhaust level in spin coater. • Adjust vacuum suckback on dispense nozzle to stop drops from forming and dripping on resist.
3. Thickness of resist coating.	C. Resist skinning (thin layer of insoluble resist dried on top of resist coating). A. Coated resist thickness is out of control. Resist thickness must be uniform: wafer-to-wafer mean thickness requirement is often $<30 \text{ \AA}$ total indicated runout (TIR, which is maximum minus minimum thickness).	<ul style="list-style-type: none"> • Spin rate is too high. High spin speeds can cause striations (lines) in the resist. • Exhaust rate too high in spin coater. • For DUV resists, ensure coated resist is not exposed to amines inside track system (check carbon filtration). <p>Parameters that affect thickness of resist are:</p> <ul style="list-style-type: none"> • Check spin acceleration and speed. Higher spin speeds produce thinner resist (recall that thickness is inversely proportional to the square root of spin speed). Verify spin speed versus thickness characteristic curve of resist manufacturer. At low spin speeds, irregular solvent loss leads to thickness nonuniformity. • Verify correct spin time. Resist reaches stable thickness in several seconds and requires additional time for thickness uniformity. • Check for correct resist type and viscosity. • Verify no mechanical vibration or air turbulence during high speed spin dry.

* B. Smith, “Resist Processing,” *Microlithography, Science and Technology*, ed. J. Sheats and B. Smith, (New York: Marcel Dekker, 1998), p. 529.

PHOTORESIST TROUBLESHOOTING

Common troubleshooting problems for photoresist are provided in Table 13.7.

TABLE 13.7 Common Photoresist Troubleshooting Problems

Problem	Probable Cause	Corrective Actions
1. Excessive resist usage during spin coating (DUV resist costs \$2,000 to \$5,000 per gallon).	A. Sub-optimum setting of process variables during spin coating process.*	Optimize process variables for spin coating: <ul style="list-style-type: none"> First, determine minimum volume of resist needed to coat the wafer (referred to as cut-off volume). Second, evaluate the following critical parameters using design of experiments: dispense spin speed, exhaust flow rate during dispense and drying steps, dispense rate, resist temperature, cool plate temperature, and ambient air temperature.
	B. Improper spin coater tool setup.	Check the spin coater setup for the following: <ul style="list-style-type: none"> Incorrect nozzle size, height, or position. Calibration of coater equipment. Wrong software process recipe for wafer.
2. Wafers damaged (broken or chipped) during normal processing in the wafer track.	A. Improper setup of different tools used in track for processing wafers.	Check the following equipment for proper setup: <ul style="list-style-type: none"> Improper calibration of robot arms during wafer handoff from one track tool to another. Loss of chuck vacuum during spin. Spin coat machine is not leveled properly and has mechanical vibration. Excessive spin speed. Verify calibration of cassette indexer to index wafers in cassette for robot pickup. Warped cassettes cause equipment malfunction.
3. DNQ-novolak resist becomes unstable and contaminated with articles.	A. DNQ-novolak resist has aged due to shelf life.	Excessive shelf life of DNQ-novolak resist (> several months, with maximum of six months to one year) can lead to the following:** <ul style="list-style-type: none"> Increase in absorption at longer wavelengths. Susceptible to thermal degradation of the DNQ, leading to crosslinking and increase of high molecular-weight resist components. Formation of undesirable acids in the resist. Precipitation (falling out of solution) of sensitizer to form crystallized particles that contaminate resists. This especially occurs at high-temperature storage. Point-of-use filtration is common practice for production applications to control resist consistency. Typical filter size is 0.05 µm for 0.25 µm feature size.

* B. Lorefice et al., "How to Minimize Resist Usage During Spin Coating," *Semiconductor International* (June 1998); p. 182.

** B. Smith, "Resist Processing," *Microlithography, Science and Technology*, ed. J. Sheats and B. Smith,

SUMMARY

Photolithography transfers a pattern from a mask to a UV light-sensitive photoresist on the wafer surface. Resolution is the smallest feature that can be patterned. Photolithography requires process latitude to consistently pattern wafers that meet the requirements. Negative lithography prints a pattern in the wafer that is opposite the mask pattern, whereas positive lithography prints the same mask pattern in the wafer. Photolithography can be divided into eight basic steps: (1) vapor prime, (2) spin coat, (3) soft bake, (4) alignment and exposure, (5) post-exposure bake, (6) develop, (7) hard bake, and (8) develop inspection. The wafer is prepared for resist application by cleaning, dehydration bake, and HMDS vapor prime. Negative photoresist hardens (crosslinks) with UV exposure, while positive resist softens (dissolves). Soft areas are washed away in developer solution. Positive resist is used for submicron

lithography. Conventional i-line resist responds to 365 nm UV light for a CD of 0.35 μm and is common on noncritical layers. Positive i-line resist has DNQ-novolak chemistry and is activated by a photo-active compound (PAC). Wafer critical layers (CD of 0.25 μm and below) are patterned with a deep UV (DUV) resist that is chemically amplified. The DUV wavelength is 248 nm and is being reduced to 193 nm. A chemically amplified resist has a chemical protecting group that makes it insoluble in aqueous developer. A photoacid generator (PAG) generated in the DUV-exposed areas deprotects the resist when heated and makes the exposed resist soluble in developer. Liquid photoresist is applied to the wafer by spin coating, followed by edge-bead removal. A soft bake is used to drive off solvents and improve resist adhesion.

KEY TERMS

photolithography (also photo, lithography, masking, or patterning)
 microlithography
 reticle
 photomask (also mask)
 optical lithography
 resolution
 overlay accuracy
 overlay budget
 runout
 process latitude
 negative lithography
 positive lithography
 negative resist
 positive resist
 dark-field mask
 clear-field mask
 coater/developer track system (also tracks)
 hydrophilic (also hydrated)
 hydrophobic (also dehydrated)
 dehydration bake
 hexamethyldisilazane (HMDS)

vapor prime coating
 photoresist
 contrast
 sensitivity
 exposure dose
 resist viscosity
 resist adhesion
 etch resistance
 surface tension
 conventional i-line photoresists
 novolak
 photoactive compound (PAC)
 DNQ-novolak
 deep UV (DUV)
 chemical amplification (CA)
 photoacid generator (PAG)
 spin coating
 static dispense
 dynamic dispense
 edge-bead removal (EBR)
 soft bake (also pre-bake)
 thermogravimetric analysis (TGA)

REVIEW QUESTIONS

- What is photolithography?
- Describe the difference between a reticle and a photomask.
- List the different UV wavelengths used in photolithography and the name for each wavelength used between 436 and 157 nm.
- Define resolution.

5. What is overlay accuracy, and how does this contribute to the mask overlay budget?
6. Discuss process latitude.
7. Explain the difference between negative and positive lithography.
8. Describe a clear-field mask.
9. Explain what a dark-field mask is.
10. List the eight steps of photolithography, and give a short explanation of each step.
11. What is one of the major effects of contamination on the wafer surface?
12. Explain the difference between a hydrophilic and a hydrophobic wafer surface.
13. Why is a dehydration bake done?
14. What is HMDS and what purpose does it serve?
15. Describe the most common method of applying HMDS.
16. Define a photoresist.
17. Give two purposes of a photoresist in wafer fabrication.
18. How has photoresist been improved since the early days of wafer fabrication?
19. List and describe the two major types of photoresist.
20. What is the resolution limit of negative resist? Which resist is used for submicron lithography?
21. Define contrast.
22. Explain sensitivity and discuss how this condition is related to the exposure dose.
23. Describe resist viscosity and explain why it is important.
24. Explain resist adhesion.
25. Why is etch resistance an important resist property?
26. What is surface tension, and why is it important for a photoresist?
27. Explain why storage and handling are important for resist.
28. List and describe the four components to an i-line resist.
29. What are two disadvantages of a negative photoresist?
30. What is the most common photoresist used in IC photolithography?
31. State and describe the resin used in i-line positive photoresist.
32. Describe the sensitizer used in i-line positive photoresist.
33. List the four steps to dissolution of i-line positive resist.
34. Give a reason why positive i-line resist has good resolution.
35. Why are i-line resists unacceptable at the DUV wavelengths?
36. What does chemical amplification accomplish in a resist?
37. Describe the purpose of a photoacid generator (PAG).
38. List and describe the four exposure steps for CA DUV resists.
39. What is one of the main benefits of CA DUV resists for wafer fabrication?
40. In what manner are DUV resists sensitive to contamination?
41. How is photoresist applied to a wafer?
42. List and describe the four basic steps to spin coating.
43. Explain the difference between static dispense and dynamic dispense.
44. What does resist thickness vary with? State the formula that describes this relationship.
45. Describe edge-bead removal.
46. State the four reasons for soft bake.
47. What is the ideal amount of solvent remaining in the resist after soft bake?
48. What problems would occur if the resist were not soft baked?
49. Describe the preferred method for performing soft bake.
50. What does TGA stand for and why is this analysis done?

PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS/ WEB SITES

Allied Signal
 Arch Chemicals (aka Olin)
 Ashland Specialty Chem.
 Clariant Corporation
 Dainippon Screen Mfg. Co.
 DuPont
 Eastman Chemical
 EKC Technology
 FSI International
 JSR Microelectronics, Inc.
 J.T. Baker
 Karl Suss Inc.
 Olin Microelectronics
 Rite Track
 SEMI

<http://www.electronicmaterials.com/>
<http://www.olinmicro.com/default.asp>
<http://www.ashland-act.com/>
<http://www.azresist.com/>
http://www.screen.co.jp/eed/index_E.html
<http://www.dupont.com/semiconductor/>
<http://www.eastman.com/>
<http://www.ekctech.com/ekctech.nsf>
<http://www.fsi-intl.com/>
<http://www.jsrusa.com/index2.html>
<http://www.jtbaker.com>
<http://www.suss.com/>
<http://www.olinmicro.com/>
<http://www.ritetrack.com/>
<http://www.semi.org/>

Shipley Company
 Silicon Valley Group
 TEL, Tokyo Electron Ltd.
 USHIO

<http://www.shipley.com/>
<http://www.svg.com>
<http://www.teainet.com>
<http://www.ushio.com/index2.html>

REFERENCES

1. S. Campbell, *The Science and Engineering of Micro-electronic Fabrication*, (New York: Oxford University Press, 1996), p. 152.
2. P. Castrucci, W. Henley, and W. Liebmann, "Lithography at an Inflection Point," *Solid State Technology* (November 1997): p. 127.
3. S. Campbell, *The Science and Engineering of Micro-electronic Fabrication*, p. 178.
4. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, Vol. 1, 2nd ed., *Process Technology* (Sunset Beach: Lattice Press, 2000), p. 510.
5. Ibid.
6. C. Chang and S. Sze, *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 290.
7. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, Vol. 1, *Process Technology*, 2nd ed., p. 500.
8. G. Gruetzner et al., "New Negative-Tone Photoresists Avoid Swelling and Distortion," *Solid State Technology* (January 1997): p. 79.
9. R. Dammel, *Diazonaphthoquinone-Based Resists*, (Bellingham, WA: SPIE Optical Engineering Press, 1993), p. 10.
10. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol.1, *Process Technology*, 2nd ed., p. 500.
11. H. Ito, "Deep-UV Resists: Evolution and Status," *Solid State Technology* (July 1996): p. 164.
12. D. Seeger, "Chemically Amplified Resists for Advanced Lithography: Road to Success or Detour?" *Solid State Technology* (June 1997): p. 115.
13. C. G. Willson et al., "Approaches to the Design and Radiation-Sensitive Polymeric Imaging Systems with Improved Sensitivity and Resolution," *Journal of Electrochemical Society*, vol. 133, no. 1, (Pennington, NJ: Electrochemical Society, 1986): p. 181.
14. T. Ueno, "Chemistry of Photoresist Materials," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith, (New York: Marcel Dekker, 1998), p. 451.
15. H. Ito, "Deep-UV Resists," p. 165.
16. G. Amblard, "Lithographic Evaluation of Deep UV Photoresists for 0.25 μm and 0.18 μm Technologies Design Rules," *Advances in Resist Technology and Processing XV* Proceedings of SPIE vol. 3333, February 23-25 1998, Santa Clara, CA, p. 890.
17. H. Ito, "Deep-UV Resists," p. 165-170.
18. M. Toukhy et al., "Chemically Amplified Resist Technology for I-Line Applications," *Advances in Resist Technology and Processing XV* Proceedings of SPIE vol. 3333 (Santa Clara, CA: February 23-25, 1998): p. 1212.
19. H. Ito, "Deep-UV Resists," p. 164.
20. N. Rizvi et al., "A 193 nm Excimer Laser Microstepper System," *Microlithographic Techniques in IC Fabrication*, Proceedings of the SPIE vol. 3183 (June 25-26, 1997), p. 35.
21. F. Burggraaf, "Optical Lithography to 2000 and Beyond," *Solid State Technology* (February 1999): p. 38.
22. M. Preil and H. Levinson, "Yield-Limiting Issues in Deep-UV Lithography," *Microlithography World* (Spring 1998): p. 24.
23. B. Lorefice et al., "How to Minimize Resist Usage During Spin Coating," *Semiconductor International* (June 1998): p. 179.
24. Ibid.
25. Ibid.
26. S. Campbell, *Science and Engineering of Microelectronic Fabrication*, p. 191.
27. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, *Process Technology*, 2nd ed., p. 515.
28. B. Smith, "Resist Processing," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith, (New York: Marcel Dekker, 1998), p. 529.
29. Ibid.
30. Ibid., p. 530.
31. D. Elliott, *Microlithography: Process Technology for IC Fabrication*, (New York: McGraw-Hill, 1986), p. 91.

PHOTOLITHOGRAPHY: ALIGNMENT AND EXPOSURE

Optical lithography is similar to photography with a camera. A camera uses light and lens to transfer an image of an object to a negative film. The develop process then transfers the film's image to paper from which a picture is obtained. In photolithography, the photomask and its pattern are the objects to be photographed. Optics and a light source are used to project the image of the pattern of the mask onto the resist-coated wafer. Once the resist is developed, just as with a negative film for a camera, the mask pattern appears in the resist. Subsequent processing, such as etch, creates a permanent pattern of the mask image on the wafer surface.

Many variables affect the quality of a photograph, such as the type of film, the lighting conditions, whether the object is in focus, and the type of camera lens. A wide range of variables also affect photolithography in wafer fabrication. Examples are the physical conditions of the material on the wafer surface, the type of photoresist, the resolution power of the optics, the nature of the light,

and the focusing accuracy of the system. To optimize the photolithography process for submicron critical dimensions, the production team must understand how all factors affect the final image. Team members have an important role because of their daily interaction with the equipment and product. They also should know where the wafers come from and the condition of the wafers prior to processing them in photolithography.

The alignment and exposure process represents a major equipment subsystem for modern photolithography. The wafer is first positioned within the focus range of the optical system. Wafer alignment features are aligned to similar matching features on the photomask, and the UV light is projected through the optics and the mask pattern. The mask pattern appears as light and dark features on the wafer; thus, exposing the photoresist. The numerous variables associated with alignment and exposure are reviewed in this chapter to understand their contribution toward achieving a high-quality, deep submicron patterning process.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the purpose of alignment and exposure in photolithography.
2. Describe the properties of light and exposure sources important for optical lithography.
3. State and explain the critical aspects of optics for optical lithography.
4. Explain resolution, describe its critical parameters, and discuss how it is calculated.
5. Discuss each of the five equipment eras for alignment and exposure.
6. Describe reticles, explain how they are manufactured, and discuss their use in microlithography.
7. Discuss the optical enhancement techniques for subwavelength lithography.
8. Explain how alignment is achieved in lithography.

INTRODUCTION

Modern photolithography equipment is based on *optical lithography*, which uses optics to accurately project and expose a mask pattern onto a resist-coated wafer. Basically it consists of an ultraviolet (UV) light source, an optical system, a reticle with the die pattern, an alignment system,

and a wafer covered with a light-sensitive photoresist. Optics is at the heart of photolithography's ability to pattern deep-submicron features. Photolithography is at the center of the wafer fabrication process, with wafers spending more time in this process area than any other operation (by some estimates, up to 60% of their fabrication time). Table 14.1 reviews the eight steps of photolithography and highlights the alignment and exposure step covered in this chapter.

TABLE 14.1 Eight Steps of Photolithography

Step	Chapter Covered
1. Vapor prime	13
2. Spin coat	13
3. Soft bake	13
4. Alignment and exposure	14
5. Post-exposure bake	15
6. Develop	15
7. Hard bake	15
8. Develop inspect	15

We learned in Chapter 13 about the photoresist material and its importance to successful photolithography. The critical equipment for photolithography is the step-and-repeat aligner (referred to as aligner, but commonly called stepper). Steppers align and expose successive reticle patterns by stepping from one exposure site to another on the resist-coated wafer surface. The industry has recently converted to the step-and-scan system for DUV resists, called step-and-scan technology (discussed later in this chapter).

In photolithography, wafer steppers have three basic purposes—all of which must meet the customers' specifications for accuracy and repeatability:

1. Focus and align the wafer surface to the quartz plate reticle (containing the patterns).
2. Reproduce a high-resolution reticle image on the wafer through exposure of photoresist.
3. Produce an adequate quantity of acceptable wafers per unit time to meet production requirements.

The alignment and exposure operation of the wafer stepper commences once the wafer surface has been coated with the photoresist and soft baked. Resist-coated wafers are automatically loaded onto a wafer stage in the stepper. At this stage the wafer is raised or lowered as needed to bring it into the focus range of the stepper optics. The wafer is aligned to the reticle so that the pattern can be transferred to the proper location on the wafer surface. Once the best focus and alignment are obtained, a shutter opens to allow UV light to pass from the illuminator to the reticle through a projection lens and then onto the resist-coated wafer (see Figure 14.1). The entire focusing, wafer alignment, and exposure operation is done by the stepper. Once a pattern is exposed, the stepper will step to the next location on the wafer and repeat the alignment and exposure (which explains the name step-and-repeat). The stepper is typically attached to an automated wafer track machine, which processes wafers for all other basic photolithography operations.

Importance of Alignment and Exposure

Integrated circuits are fabricated with semiconductor devices in the upper few microns of the silicon wafer, followed by successive deposition and patterning of material layers to form the circuits that interconnect the devices. Circuit designers, using computers with special design software, define the layout of the devices, metal lines, via connections, and other special circuit designs necessary for a chip to function. This is done layer by layer and structure by structure within a layer for the entire wafer. Like creating a stencil, the circuit design pattern for one or more die is transferred to a reticle, with multiple reticles needed to attain the final structure on the wafer surface (see Figure 14.2).

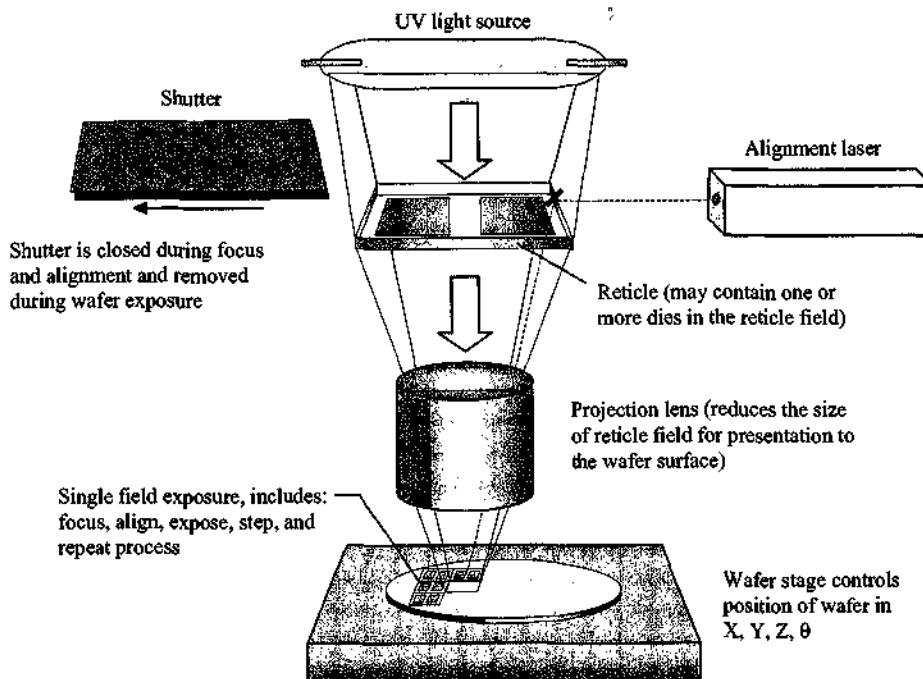


FIGURE 14.1 Reticle Pattern Transfer to Resist

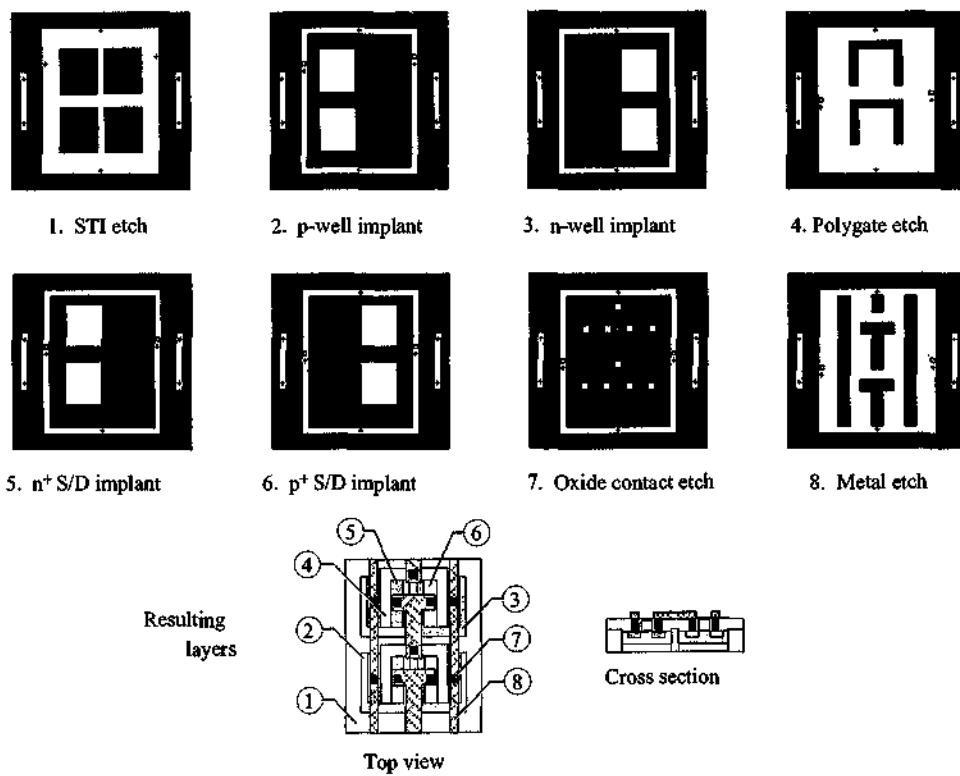


FIGURE 14.2 Layout and Dimensions of Reticle Patterns

The circuit design follows a set of design ground rules that specify the reticle pattern and its required dimensions. The design rules specify parameters such as linewidth, spacing between lines, contact size, via diameter, and the spacing between patterns defined on a reticle. Violation of design rules lead to poor production yield or unreliable product performance.

Each image defined in a reticle pattern has a particular function, such as a contact opening or metal lines. Patterns relate to a distinct material (e.g., oxide, aluminum, and so on) with dimensions and tolerances. Reticles overlay these patterns on top of one another during photolithography to form the devices and circuits on the wafer. The reticle overlay process has alignment specifications, previously referred to as the overlay budget. It is this overlay of reticle patterns onto existing features that poses the special alignment challenge for photolithography—how to accurately overlay patterns with submicron dimensions. Overlay accuracy must be met while maximizing product reliability and producing wafers in high volume.

Optical Exposure ■ During exposure, light from an illumination source passes through the aligned reticle. The reticle has opaque and transparent areas that define a pattern to be transferred to the photoresist on the wafer surface. The goal of exposure is for the reticle pattern to be accurately replicated (within specification) in the resulting image in the resist.

An aspect of exposure is, with all other factors being equal, the shorter the wavelength of the exposing light, the smaller the feature size that can be exposed. This fact has been a driving force for the continual reduction in feature sizes on wafers. Furthermore, the exposing light produces a certain amount of energy that is necessary to produce the photochemical reaction in the resist. This light energy must be uniformly distributed across the exposure field. Photolithography requires an intense exposure at the desired short wavelength to achieve the critical dimensions of today's microlithography.

OPTICAL LITHOGRAPHY

Optical lithography has traditionally been the major limiting factor to the continual reduction in microchip feature sizes since the beginning of wafer fabrication. The demise of optical lithography has been regularly predicted over the years. For instance, in 1985 it was predicted that optical lithography would be incapable of resolving a critical dimension smaller than $0.5 \mu\text{m}$.¹ It is now believed that optical lithography is capable of resolutions down to a CD of $0.1 \mu\text{m}$ and beyond. Lithography is often considered the engine that is driving the performance improvements of Moore's law.²

Photolithography for wafer fabrication has been based largely on optical lithography. The longevity of optical lithography is attributed to the basic improvements that have been made to the equipment and process. We will now review the fundamental variables in optical lithography and the basis for these ongoing improvements. At the end of Chapter 15, we will study other next-generation lithography systems that are under investigation for advanced lithography applications and that may become viable for wafer fabrication in the future: extreme UV, e-beam, X-ray, and ion-beam lithography. Nevertheless, it appears that optical lithography will continue to dominate photolithography for the near future.

Light

A light source is needed in optical lithography to project the reticle pattern on the photoresist and cause a photochemical reaction. A practical description of *light* is an electromagnetic wave that is visible to the eye. Light is also radiant energy. These two descriptions reflect the dual nature of light as a wave and a particle. Light travels in waves similar to sound waves. Since light is a wave, it can be described by wavelength (λ) and frequency (f). The relationship of the two is given by the formula shown in Figure 14.3, where v is the velocity of light.

Interference of Light Waves ■ Waves are sinusoidal in nature. Sinusoidal waves of any type (e.g., light, electrical, or sound) that have the same frequency (monochromatic) can have interference between the individual waves. For example, wave interference may be two water ripples interacting with each other and partially canceling each other out. There are two types of interference based on whether the waves are in phase or out of phase (see Figure 14.4):

Constructive interference: Two waves in phase that add to each other.

Destructive interference: Two waves out of phase that subtract from each other.

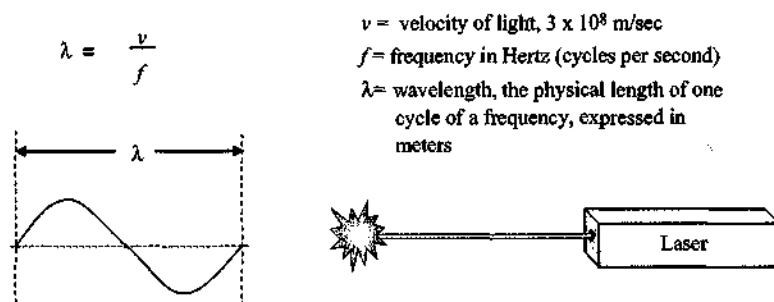


FIGURE 14.3 Light Wavelength and Frequency

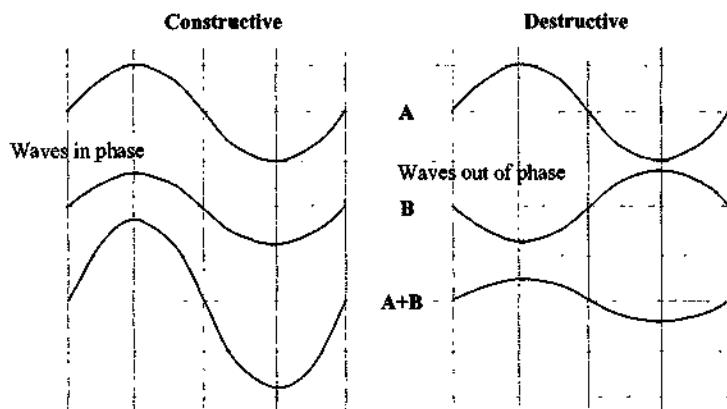


FIGURE 14.4 Wave Interference

Optical Filters. Optical filters use light interference to block unwanted incident light through either reflection or interference to obtain a light with a particular wavelength (see Figure 14.5). Optical filters will typically be made of glass and will have one or more layers of thin film coatings. The type of coating and its thickness will determine which wavelengths are blocked through destructive interference or passed through the glass.

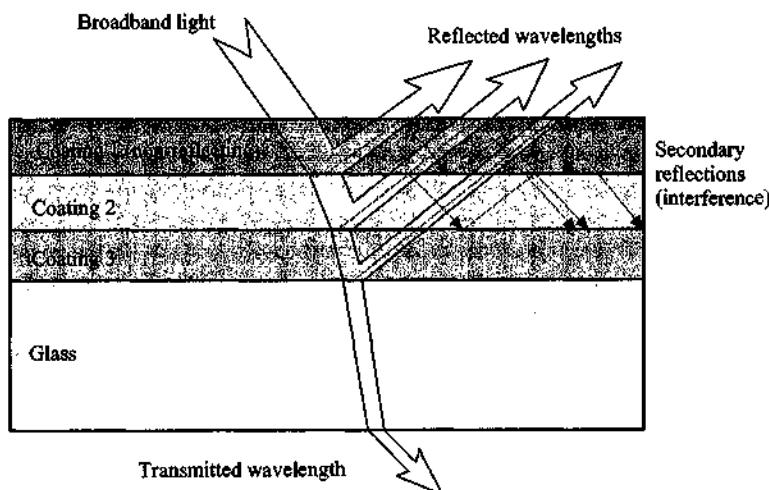


FIGURE 14.5 Optical Filtration

Electromagnetic Spectrum ■ The entire group of visible and nonvisible electromagnetic waves is called the *electromagnetic spectrum*, consisting of radiant energy that varies from very short to very long wavelengths. The visible portion of the spectrum has a wavelength between 390 nm and

780 nm. White light is composed of all the wavelengths of light in the visible spectrum. The UV spectrum ranges from 4 nm to about 450 nm, with a small overlap with the visible spectrum (see Figure 14.6). The deep ultraviolet (DUV) wavelengths are presently used for patterning critical layers in production. The vacuum ultraviolet (VUV) wavelength of 157 nm is the most likely UV successor after the DUV wavelengths of 248 nm and 193 nm. The short wavelength and high photon energy of VUV are strongly absorbed by oxygen absorption bands; therefore, operation is necessary in a vacuum or inert gas ambient (thus the name VUV). Research and development (R & D) is underway in the semiconductor industry into the future technology of using 13 nm UV wavelength, referred to as extreme UV (EUV).

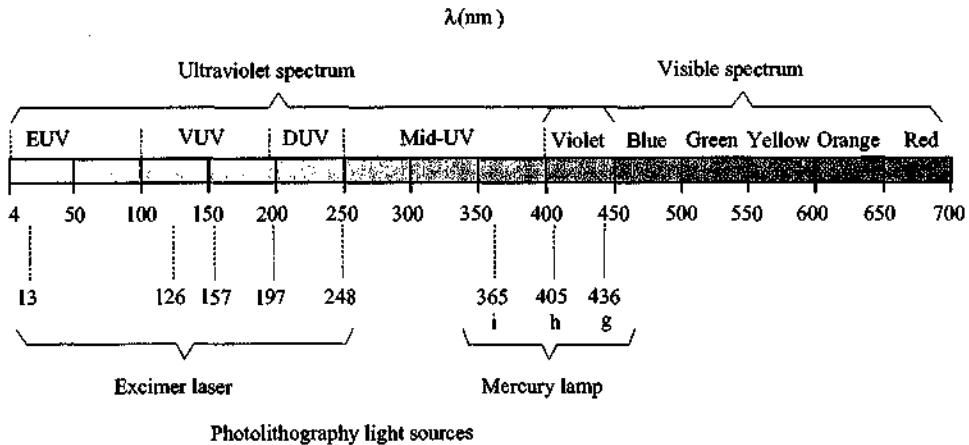


FIGURE 14.6 Ultraviolet Spectrum

Since the UV spectrum overlaps a portion of the visible spectrum, white light also contains a portion of UV (e.g., sunlight is visible and is white light, but it has UV wavelength also). Yellow light is commonly used in the photolithography area of the wafer fab because it is in the portion of the visible spectrum with little UV and, therefore, does not affect photoresists.

Exposure Sources

During exposure of the resist, photochemical transformations occur in the resist material to transfer the pattern from the mask. This is a critical step in photolithography. It must occur in the shortest amount of time and be repeatable for high-volume production of wafers.

Ultraviolet (UV) light is used to expose resist because the resist material reacts to light at this particular wavelength. Wavelength is also important because light with a smaller wavelength allows for the resolution of smaller features on the photoresist (explained in detail later in this chapter). The two types of UV exposure sources most often used today for optical lithography are:

- ◆ Mercury arc lamp
- ◆ Excimer laser

Besides these more commonly used light sources, other sources used to expose resist in advanced or special applications are X-ray, electron beam, and ion beam. These sources and their unique resist material will be discussed in Chapter 15 in sections on advanced technologies.

Mercury Arc Lamp ■ The high-pressure *mercury arc lamp* is used as the UV illumination source in all conventional i-line steppers. In this type of lamp, electrical current is passed through a tube of mercury-xenon gas to create a discharge arc. This arc emits a characteristic light spectrum, with the useful UV wavelength emissions for a mercury arc lamp between about 240 nm and 500 nm in length (see Figure 14.7).

The mercury arc lamp light spectrum has several intensity peaks. Some have been given letter names that come from the early days of spectrometry. The major intensity peaks are shown in Table 14.2.

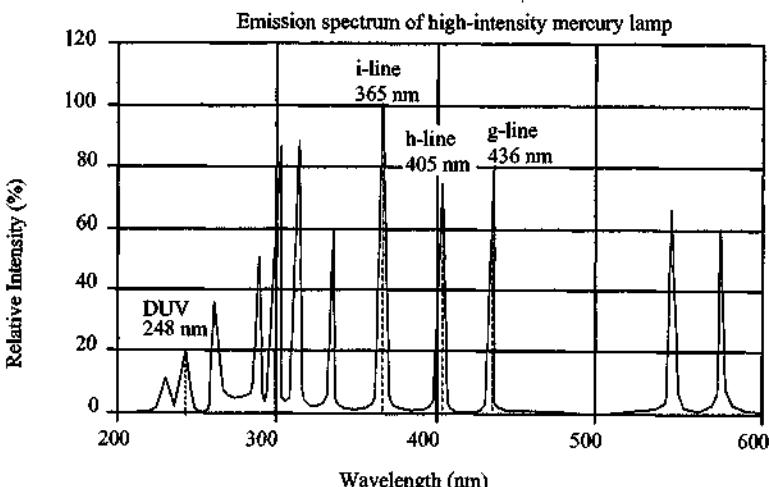


FIGURE 14.7 Emission Spectrum of Typical High Pressure Mercury Arc Lamp
(Used with permission from USHIO Specialty Lighting Products)

TABLE 14.2 Mercury Arc Lamp Intensity Peaks

UV Light Wavelength (nm)	Descriptor	CD Resolution (μm)
436	g-line	0.5
405	h-line	0.4
365	i-line	0.35
248	Deep UV (DUV)	0.25

A conventional photoresist will have a certain spectral response that corresponds to a particular UV wavelength during exposure. For instance, a DNQ-novolak i-line resist, which is used for CD feature sizes down to $0.35 \mu\text{m}$, responds to the UV i-line wavelength of 365 nm. To expose a resist to the proper UV wavelength, a set of filters is used to block the unwanted wavelengths and any infrared wavelengths. The exposure wavelength is chosen to match the critical feature size on the wafer.

An important aspect of the exposure source is the *light intensity*, which is defined by power per unit area (mW/cm^2), and measured on the surface of the resist. Another way to interpret intensity is the quantity of light per unit area, or brightness. Energy is the product of power and time. If the light intensity (which is power per unit area) is multiplied by the time of exposure, this represents the amount of exposure energy, or exposure dose, in (mJoules/cm^2 , or mJ/cm^2) received by the surface of the photoresist.

A typical i-line resist requires an exposure dose of 100 mJ/cm^2 for its exposure.³ Consider the emission spectrum shown in Figure 14.7 for a mercury arc lamp. The DUV emission at 248 nm is approximately five times less intense than the i-line emission at 365 nm. For an equivalent i-line photoresist at 248 nm, the low mercury arc lamp light intensity at the DUV wavelength would require five times more exposure time. In other words, if light intensity decreases, then the exposure time must increase by a proportional amount. This new exposure time is too long for acceptable wafer production and was a primary reason for the development of chemically amplified DUV photoresists and laser light sources with more power (see Figure 14.8 on page 374).

An undesirable parameter for the resist resin is excessive absorbance of incident radiation. With excessive resist absorption, the light intensity at the bottom of the resist is considerably less than that received at the top. This discrepancy produces image profiles with sloping sidewalls (see Figure 14.9 on page 374). To achieve straight-wall images, the resist must absorb only a small portion of the incident radiation, typically less than 20%.⁴ Minimizing resist absorbance requires optimization between the wavelength, light source dose, and the type of resist.

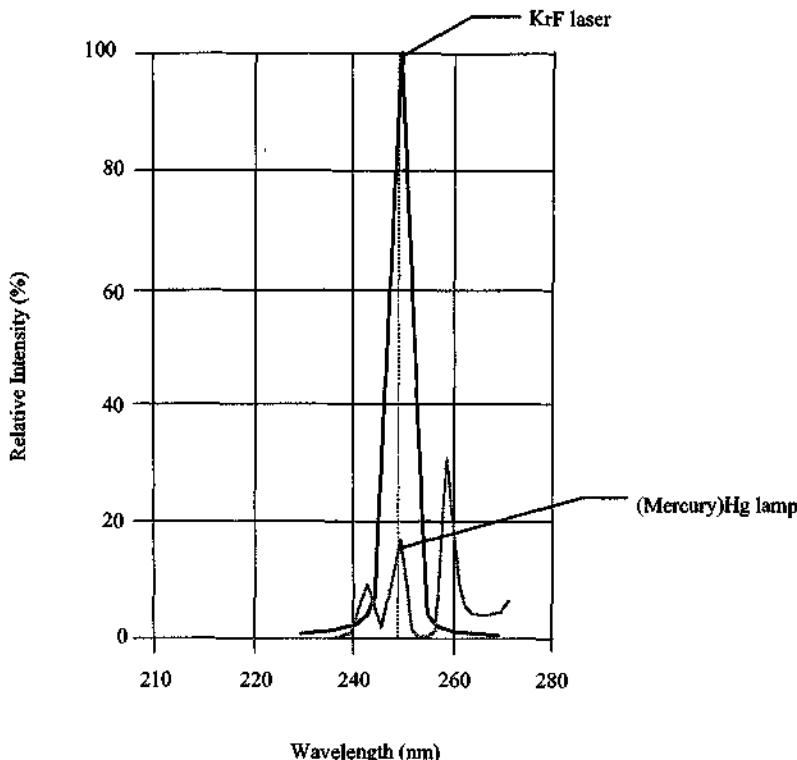


FIGURE 14.8 Spectral Emission Intensity of 248 nm Excimer Laser Versus Mercury Lamp

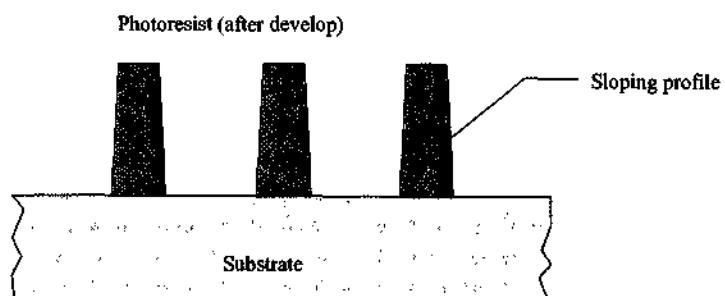


FIGURE 14.9 Excessive Resist Absorption of Incident Light

Excimer Laser ■ Laser light sources for photolithography have been available since the mid-1980s, but reliability and performance concerns delayed their implementation in wafer production until the mid-1990s. Their main benefit has been to provide more light intensity at the DUV wavelengths of 248 nm and below, since the mercury arc lamp is an inefficient emitter at these wavelengths.

The only laser light source used thus far for photolithography exposure is the *excimer laser*. An excimer is an exotic molecule formed from an atom of a noble gas and halogen, such as argon fluoride (ArF), where the molecule exists only in a quasi-stable, excited state.⁵ The word excimer actually comes from the expression *excited dimer*, which is a molecule formed from two identical atoms, such as F₂. It also is used to represent the noble gas and halogen molecules.

Most excimer lasers today contain a high-pressure mixture of two or more elements that are bound in the excited state. Laser light emission occurs when the excited state decays and the exotic molecule falls apart into its two constituent atoms. The laser maintains a condition of more molecules in the excited state than in the ground state by using a high-voltage (10 to 20 kV) pulse discharge across two flat-plate electrodes to excite a high-pressure mixture of the noble gas and halogen.⁶

The most common excimer laser used for DUV resist is krypton-fluoride (KrF) with a wavelength of 248 nm. The KrF laser typically has a power range of 10 to 20 W at a frequency of 1 kHz, which produces high-power pulses of radiant light energy to expose the resist. Table 14.3 highlights excimer laser sources for wafer fabrication photolithography.⁷ The argon-fluoride (ArF) laser, at a wavelength of 193 nm, is also envisioned to achieve DUV exposure. Notice that the fluorine (F_2) laser at the 157-nm wavelength has low output energy, which makes it less desirable for future production use because of the longer exposure time required. There has been recent improvement in the F_2 laser output energy to make it equivalent to the KrF and ArF lasers.⁸

TABLE 14.3 Excimer Laser Sources for Semiconductor Photolithography

Material	Wavelength (nm)	Maximum Output (mJ/pulse)	Frequency (pulses/sec)	Pulse Length (ns)	CD Resolution (μm)
KrF	248	300 to 1500	500	25	≤ 0.25
ArF	193	175 to 300	400	15	≤ 0.18
F_2	157	6	10	20	≤ 0.15

As seen in Table 14.3, the excimer laser produces extremely short pulses of light. The peak power of each short pulse is high and can cause damage to optical materials and lens coatings. To minimize this damage, longer laser pulse lengths are desirable and lenses are designed with features to avoid the high concentrations of light from the laser source. Another problem with early excimer lasers was excessive pulse-to-pulse energy variations. To overcome this variation, a sufficiently large number of pulses is required. Improvements to the excimer laser pulse stability continue, which reduces the number of pulses required to achieve acceptable exposure dose uniformity. With fewer pulses needed, the scanning speed of the step-and-scan system can increase, which ultimately increases productivity.⁹

The excimer laser will probably be used as an exposure source for at least another technology generation with the argon-fluoride (ArF) laser at a wavelength of 193 nm. There are no fundamental laser changes required to change from a 248-nm KrF laser to a 193-nm ArF laser. However, optical materials have undesirable absorbance and are more sensitive to laser damage from factors such as optical system heating at the 193-nm wavelength.¹⁰ This condition requires lens systems that are laser damage resistant. An industrial grade F_2 excimer laser at a wavelength of 157 nm has been demonstrated as a potential light source for 0.15 μm critical dimension.

Spatial Coherence. Recall that light is an electromagnetic wave. A light beam has *spatial coherence* when points on different waves remain in phase to one another as the waves propagate. The light waves move in unison (see Figure 14.10). A standard room light bulb has no spatial coherency (completely incoherent). Light beams from excimer lasers have a low amount of spatial coherence, which differs from the beams of conventional lasers. Spatial coherence is

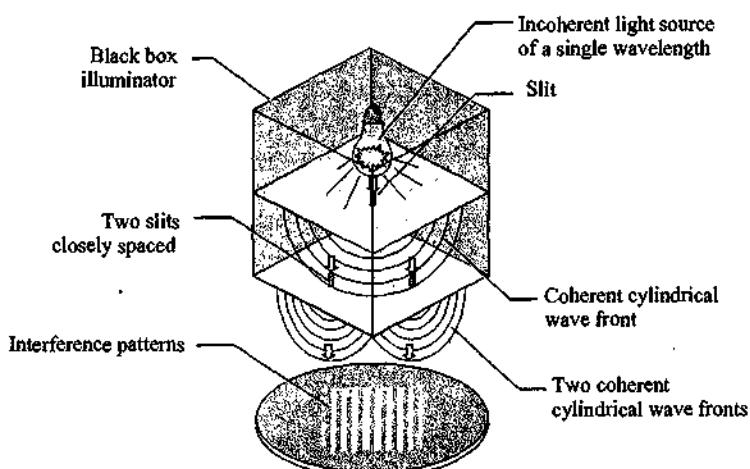


FIGURE 14.10 Spatial Coherence

controlled through optics to minimize interference patterns that can be formed in the image. If not controlled, the interference may appear as a grainy pattern of bright and dark spots on the resist, referred to as speckle.¹¹

Exposure Control. ■ A uniform dose of UV light is critical for resist exposure. The exposure dose must be repeatable from exposure to exposure and from wafer to wafer. The exposure latitude for a DUV resist is around 1% in dose variability.¹² The variation in photolithography processes, equipment, and materials requires stringent exposure control.

Exposure control in photolithography is achieved by measuring the intensity of the UV light source at the surface of the wafer with a *dose monitor*. The exposure dose is measured at different locations in the exposure field, permitting the calculation of the percent uniformity of the dose. A photodetector and an electronics circuit are used on automated steppers and step-and-scan systems to monitor and control the exposure dose through use of a shutter or by varying the speed of the scan.

Optics

Optics is the study of the physical properties and composition of light. Optics is important because lithography is based on an optical imaging process used to transfer the reticle pattern to the photoresist. The quality of the resist pattern can be limited by poor-quality optics. All photolithography equipment currently used for high-volume wafer fabrication is based on optical lithography.

Reflection of Light ■ The *law of reflection* describes the relationship between an incident light ray and the corresponding reflected ray, stating that the angle of incidence is equal to the angle of reflection (see Figure 14.11). This law is true whether the reflecting surface is rough or smooth.

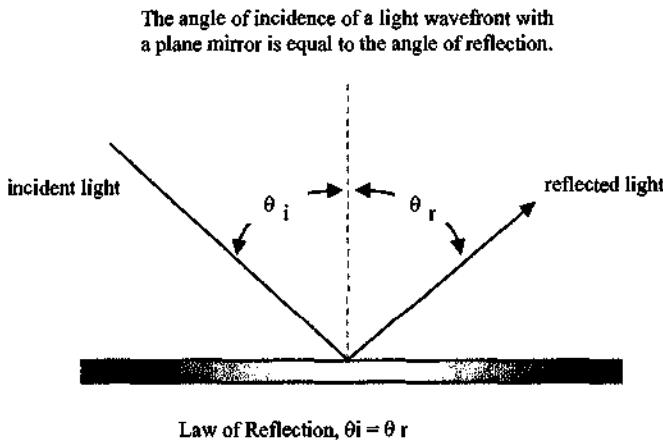


FIGURE 14.11 Law of Reflection

An example of the law of reflection is a plane mirror. A mirror reflects light from its surface and creates a virtual reversed image of an object. The image is said to be virtual because it appears to the viewer as though the object exists behind the mirror, whereas the image on the mirror surface is actually a reversed image reflection of the real object. Mirrors and reflective optics have many applications in steppers and step-and-scan systems, such as for beam orientation or for illuminators that shape and focus light (see Figure 14.12).

Refraction of Light ■ When light passes from one transparent medium into another, such as from air through a glass window, the light changes direction. The change in direction of a light ray when passing from one transparent medium into another is called *refraction*. Refraction of light is caused by the difference in the speed of light in two different media. Light traveling through a uniform medium, such as air, travels at a certain speed. As light enters into and passes through a new medium, such as glass, its speed actually decreases. This decrease occurs because the glass is optically a more dense material. The *relative index of refraction*, n , represents how much the light ray

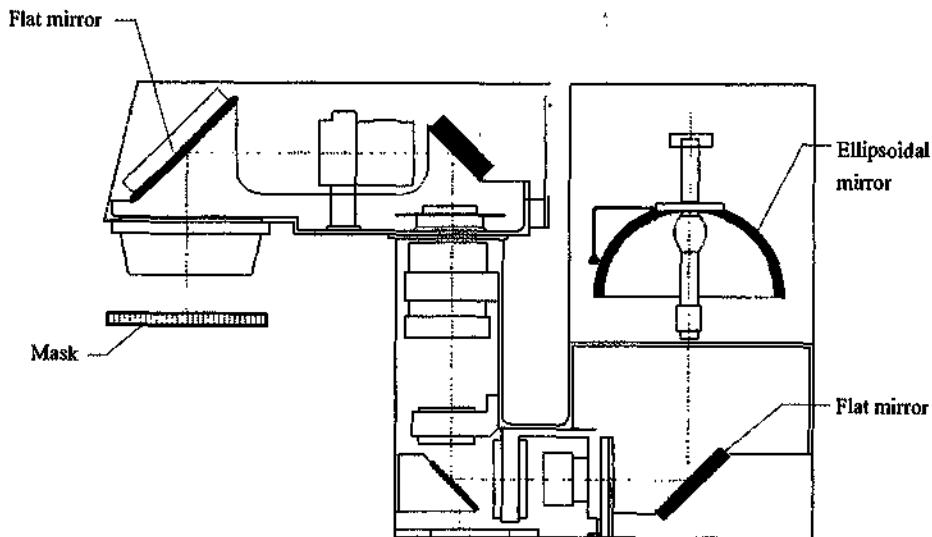


FIGURE 14.12 Application of Mirrors
(Used with permission from Canon USA, illuminator for mask aligner)

bends as it passes through the interface of the two media based on its change in velocity (see Figure 14.13). The *absolute index of refraction* compares the speed of light in a vacuum to the speed of light in the chosen medium. Therefore it is only stated for one medium and the vacuum is assumed. Examples of absolute index of refraction for different media are given in Table 14.4.¹³ The refraction of light is also dependent on the wavelength, λ , of the light ray that passes from one medium into another medium.

- Snell's Law: $\sin \theta_i = n \sin \theta_r$
- Index of refraction, $n = \sin \theta_i / \sin \theta_r$

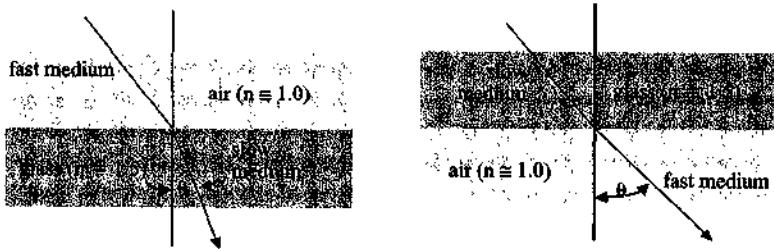


FIGURE 14.13 Refraction of Light Based on Two Mediums

TABLE 14.4 Absolute Index of Refraction for Select Materials

Material	Index of Refraction (n)
Air	1.000293
Water	1.33
Fused silica (amorphous quartz)	1.458
Diamond	2.419

Lens ■ A *lens* is defined as an optical element that refracts light from an object passing through it to form an image of the object. We will consider an individual lens, but optical systems in photolithography consists of many different lenses (refractive) and mirrors (reflective) in the optical system (see Figure 14.14.) Lenses are important in the optical system of photolithography equipment because of the need to project the image of the reticle pattern onto the photoresist. This projection must be done with the proper resolution, dimensional control, and alignment. Lenses and refractive optics in photolithography steppers and step-and-scan systems play a critical role in achieving the resolution needed to achieve the CD linewidths in modern wafer fabrication.

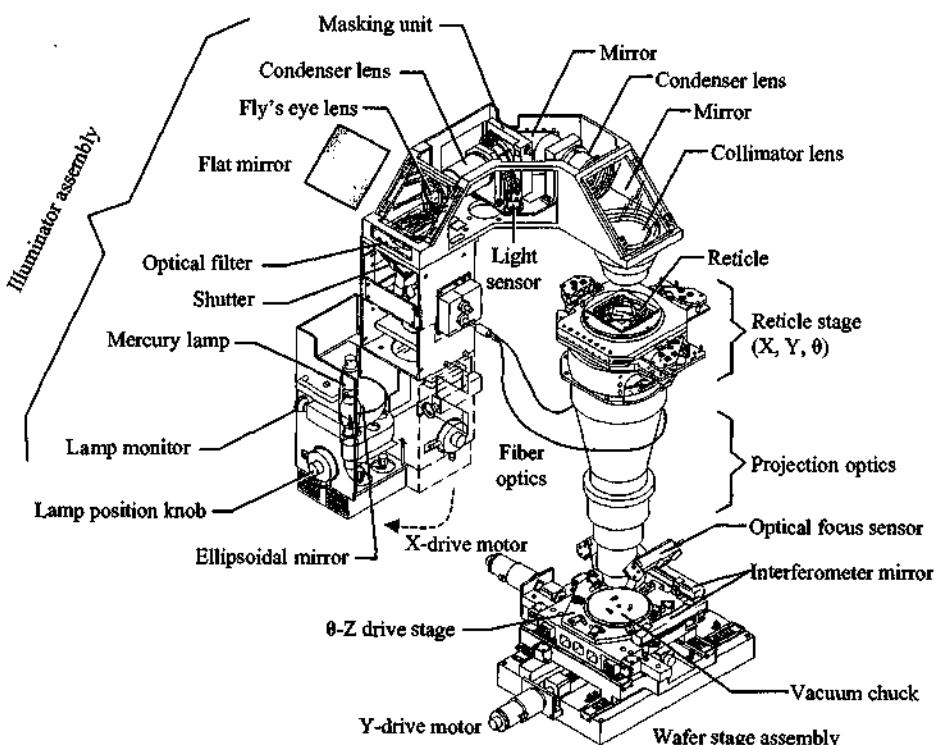


FIGURE 14.14 Optical System of Lenses
(Used with permission from Canon USA, FPA-2000 II exposure system)

A lens has either one or both surfaces as spherical-shaped, which defines whether the lens is converging (convex) or diverging (concave). Lenses refract light in such a way that light rays converge (see Figure 14.15) or diverge (see Figure 14.16) on a principal focus. The principal focus, or *focal point*, is where the light rays are refracted and will converge—the image carried by the light is sharp and in focus. The distance from the center of the lens to the focal point is called the *focal length*.

Lens Material. Lenses have traditionally been made of glass. The lens material is an important variable as the exposure wavelength is continually reduced. For 248-nm wavelength DUV, a suitable lens material is fused silica. It has less light absorption at the DUV wavelengths. At the 193-nm DUV and 157-nm VUV wavelength, calcium fluoride (CaF_2) is being investigated as a possible candidate for lens material.

Traditional optical materials such as glass have greater absorbance and are more sensitive to laser damage at the 248-nm KrF wavelength.¹⁴ Absorption causes a loss in exposure power and induces heat in the optics, which leads to refractive index changes and imaging problems. Thermal effects can also cause focus changes due to optical system heating from DUV laser light sources.

Another concern regarding lens materials is that a laser beam can create compaction damage. *Lens compaction* is a structural rearrangement of the lens material that causes densification of the lens material (see Figure 14.17). Compaction occurs in lens materials, including fused silica. It is not well understood, but it occurs due to the total cumulative laser beam exposure and peak power density.

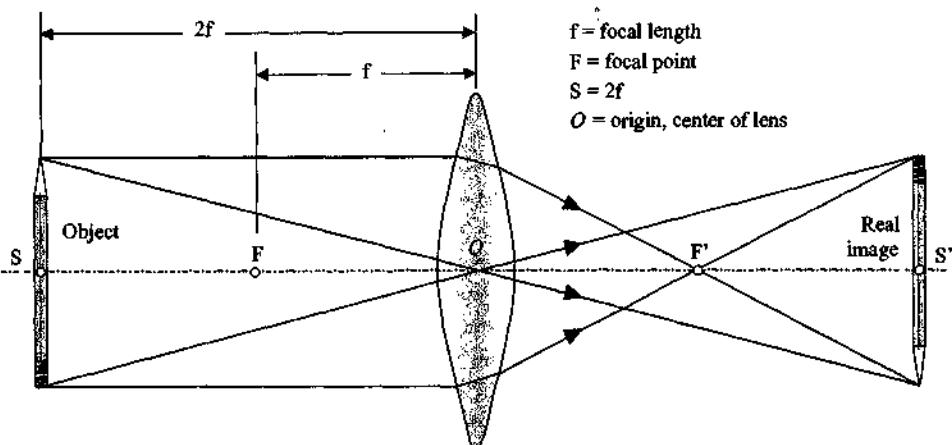


FIGURE 14.15 Converging Lens with Focal Point

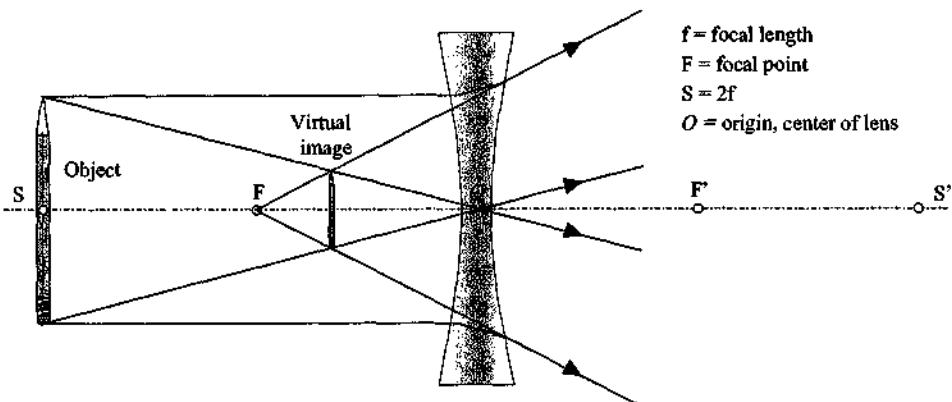


FIGURE 14.16 Diverging Lens with Focal Point

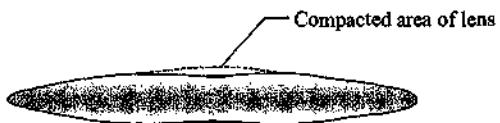


FIGURE 14.17 Laser-Induced Lens Compaction

It can increase the lens material's refractive index in the region traversed by the laser beam. If the index of refraction changes even slightly, then there could be a change in the light wave path, which would lead to loss of image quality. The control of compaction for DUV wavelengths is being researched.

Lenses used in photolithography optical systems are of the finest quality materials and workmanship, but they are not perfect. Deviations from the ideal lens behavior that result from design, fabrication, or usage flaws are referred to as *aberrations*. There are different types of aberrations, such as inaccurate lens surfaces. For the most part, the lens system designer designs the optical system to minimize individual lens aberrations.¹⁵

Diffraction ■ Light travels in straight lines. When light passes through a narrow opening or past a sharp edge, interference patterns occur along the edge of the opening. The effect is a fuzzy image rather than the expected sharp edge that occurs between light and shadow (see Figure 14.18 on page 380). The light appears to bend around the slit edges. This phenomenon is referred to as *diffraction*.

- Light travels in straight lines.
- Diffraction occurs when light hits edges of objects.
- Diffraction bands, or interference patterns, occur when light waves pass through narrow slits.

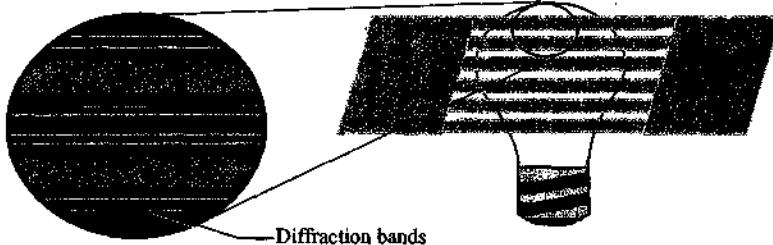


FIGURE 14.18 Interference Pattern from Light Diffraction at Small Opening

You can experience diffraction by putting two fingers close together and looking through them toward a light. The edge of your fingers is not sharp. Diffraction of light passing through a small opening causes an unexpected light intensity profile. In essence, regions outside the boundary are illuminated and the light intensity profile does not have a sharp edge. The central spot is bright and bounded by bands of decreasing intensity which are called *diffraction orders*. The amount of diffraction depends on the width of the opening and the wavelength of the light.

Light diffraction is a concern in photolithography because of the extremely small patterns of sharp edges and narrow spaces on reticles. Light during exposure must pass through these patterns (see Figure 14.19). Diffraction patterns rob exposure energy and scatter it, leading to exposure of unwanted areas of the photoresist. The problem is worse in small holes, such as small contact openings of 200 nm. The interference patterns caused by diffraction can make small contact holes and small lines difficult to print.

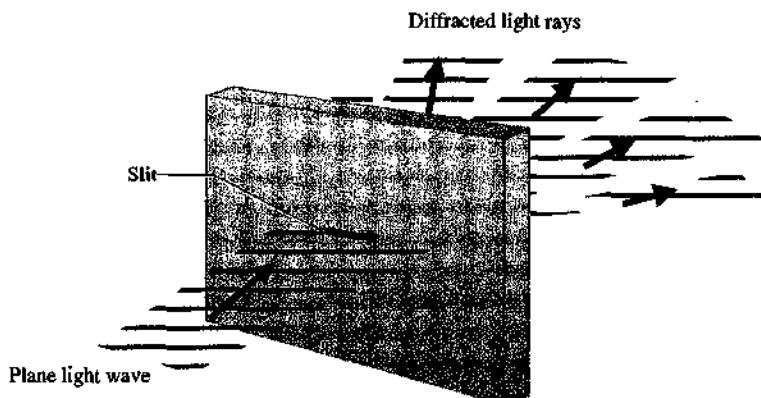


FIGURE 14.19 Diffraction in a Reticle Pattern

Numerical Aperture ■ A lens is able to capture some diffracted light (see Figure 14.20). The ability of a lens to collect diffracted light is referred to as the *numerical aperture (NA)* of the lens. For a given lens, the NA is a measure of how much diffracted light the lens can accept and image by converging the diffracted light to a single point.

The numerical aperture is defined by the following formula:

$$NA = (n) \sin \theta_m \approx (n) \frac{\text{radius of lens}}{\text{focal length of lens}}$$

Where, n = index of refraction of the image medium ($n \approx 1$ for air)

θ_m = angle between the optical principal axis and the marginal ray at the edge of the lens

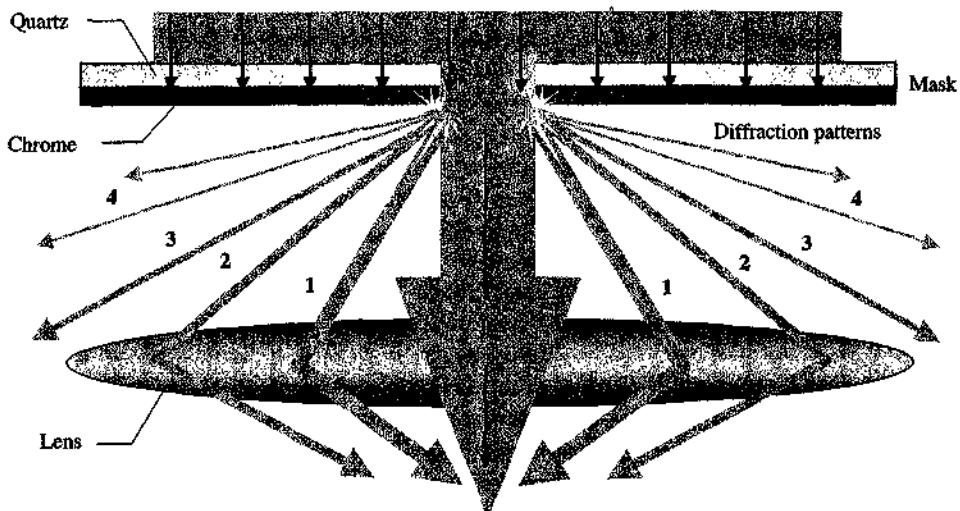


FIGURE 14.20 Lens Capturing Diffracted Light

Note that the value for $\sin \theta_m$ can be approximated by using its trigonometric relationship. This approximation highlights how increasing the radius of the lens will increase the NA and capture more diffracted light. With an increased NA, more of the diffracted light can be converged to a single point for imaging (see Figure 14.21). However, increasing the radius of the lens to increase the NA also means more complicated and costly optical systems.

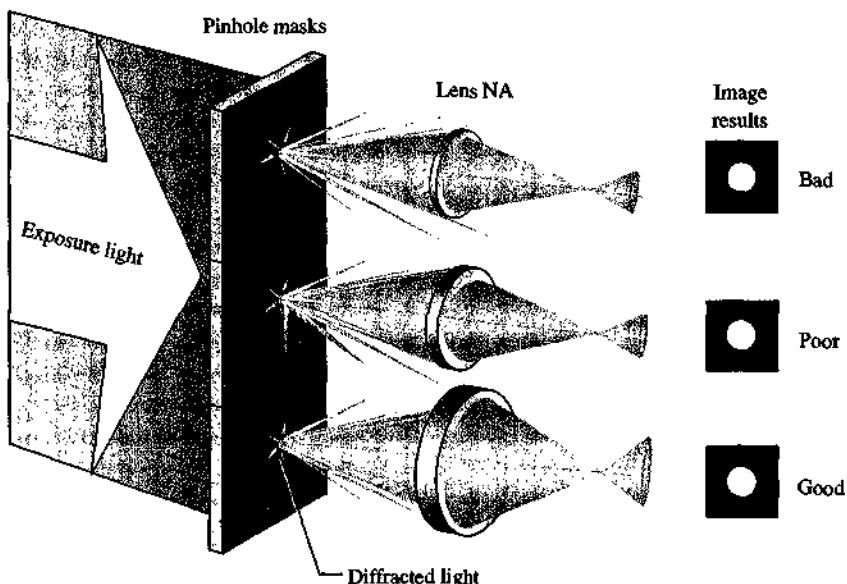


FIGURE 14.21 Effect of Numerical Aperture on Imaging

NA values are typically specified for photolithography steppers and step-and-scan systems. NA has improved considerably over the years. For an air or vacuum medium, the NA value must be less than or equal to 1. Some typical NA values for photolithography equipment are provided in Table 14.5.

TABLE 14.5 Typical NA Values for Photolithography Tools

Type of Equipment	NA Value
Scanning projection aligner with mirrors (1970s technology)	0.25
Step-and-repeat	0.60 to 0.68
Step-and-scan	0.60 to 0.68

Antireflective Coatings ■ Exposure light passes through a reticle to pattern the resist. Below the resist is the underlying layer that will ultimately be etched and patterned. If this underlying film is reflective, as with metal and polysilicon layers, then light rays reflect off this film and potentially damage the adjacent resist. This damage can adversely affect CD control. The two primary light reflectivity problems are reflective notching and standing waves. *Reflective notching* occurs when vertical surfaces on the sides of etched structures reflect light into the resist where exposure is not intended (see Figure 14.22).

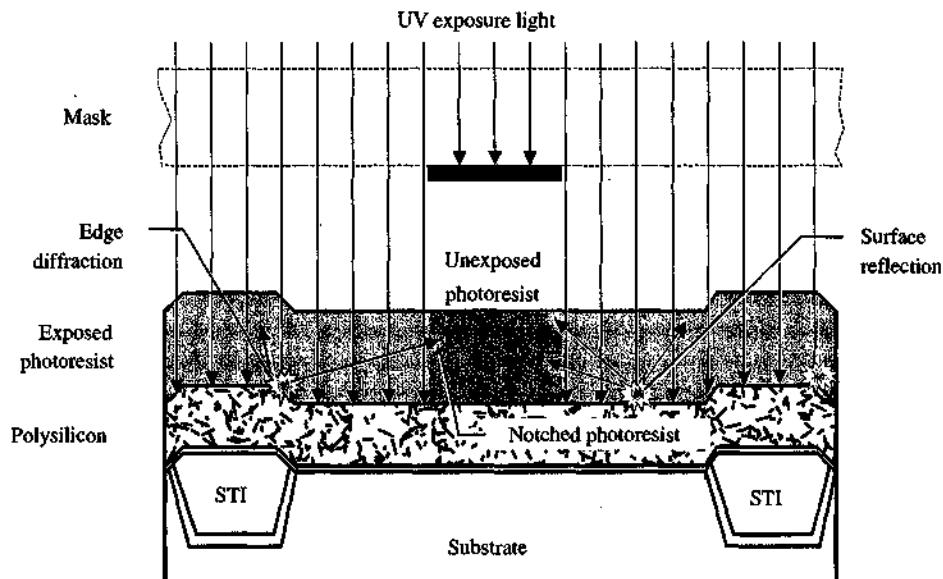
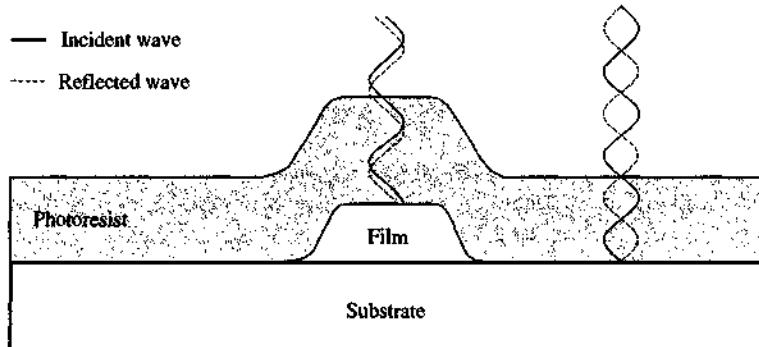


FIGURE 14.22 Photoresist Reflective Notching Due to Light Reflections

Standing Waves. An example of wave reflection and interference in photolithography is the phenomenon of standing waves. If a layer of photoresist is applied to a reflective wafer surface and exposed to monochromatic light, then the incident light wave striking the resist passes through the resist layer and is reflected from the surface of the wafer (see Figure 14.23). *Standing waves* represent interference between the incident light waves and reflected light waves, which causes nonuniform exposure along the thickness of the photoresist film. The occurrence of standing waves is more pronounced with DUV resists because many wafer surfaces (e.g., oxide, nitride, and polysilicon) are more reflective at the shorter DUV wavelengths.¹⁶ After exposure, the sides of the resist features have striations of overexposed and underexposed areas. Standing waves essentially degrade the resolution of the image in the resist.

Using an *antireflective coating* (ARC) applied directly to the reflective substrate surface reduces standing wave effects in photoresist (see Figure 14.24). ARCs reduce unintended light reflection by suppressing the exposure light, with the latest ARCs capable of suppressing 99%



Standing waves cause nonuniform exposure along the thickness of the photoresist film.

FIGURE 14.23 Incident and Reflected Light Wave Interference in Photoresist



Effect of Standing Waves in Photoresist
(Photo courtesy of Grant Willson's research group at the University of Texas at Austin)

of the substrate-reflected light. They are deposited on the wafer as a thin layer, typically from 200 to 2000 Å, depending on the type of ARC and material used.¹⁷ Dyes can also be added to the resist to help prevent light wave interference. In addition, a post-exposure bake (PEB) between exposure and development can reduce the extent of standing wave striations in conventional i-line resists. The PEB redistributes the photoactive compound (PAC) in the resist and allows for straighter resist sidewall profiles by reducing standing waves.

There are two basic types of ARCs: bottom antireflective coating beneath the photoresist to reduce substrate reflections, and top antireflective coating deposited over the resist to reduce secondary reflections from the resist surface. The bottom ARC has emerged as the most effective method in reducing reflections and problems such as standing waves and, therefore, will be the focus of our discussion.

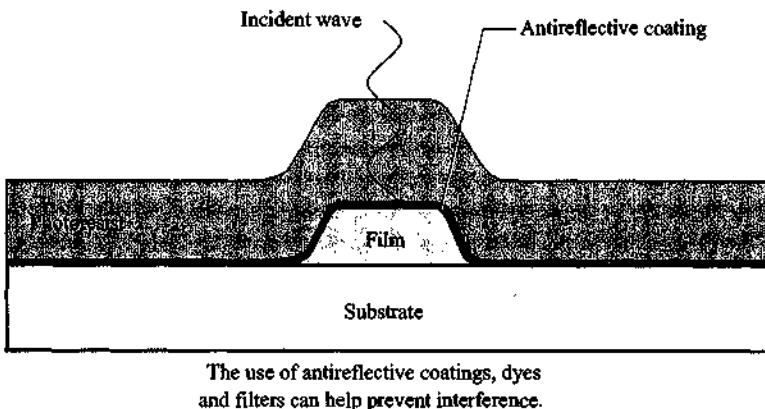


FIGURE 14.24 Antireflective Coating to Prevent Standing Waves

Bottom ARCs. A *bottom antireflective coating (BARC)* is used to suppress unintended light reflection from a reflective layer that is below the resist, as shown in Figure 14.25 on page 384. The BARC material is an organic or inorganic dielectric material that is applied to the wafer before the photoresist.

Organic ARCs reduce reflection by absorbing light and are typically spin-coated on the wafer in the same manner as the photoresist. Inorganic ARCs are deposited by plasma-enhanced chemical vapor deposition (PECVD). Inorganic ARCs do not absorb light but instead work by phase-shift cancellation of specific wavelengths based on the refractive index, film thickness, and other parameters (see Figure 14.26 on page 384). Successful phase cancellation of light requires very tight control of process parameters, such as a thickness tolerance for the BARC of 15 Å.¹⁸ TiN, used as a diffusion barrier for interconnect metal, is also a good antireflective coating. However, the reflectivity of materials changes for shorter wavelengths, making interference effects more difficult to control.

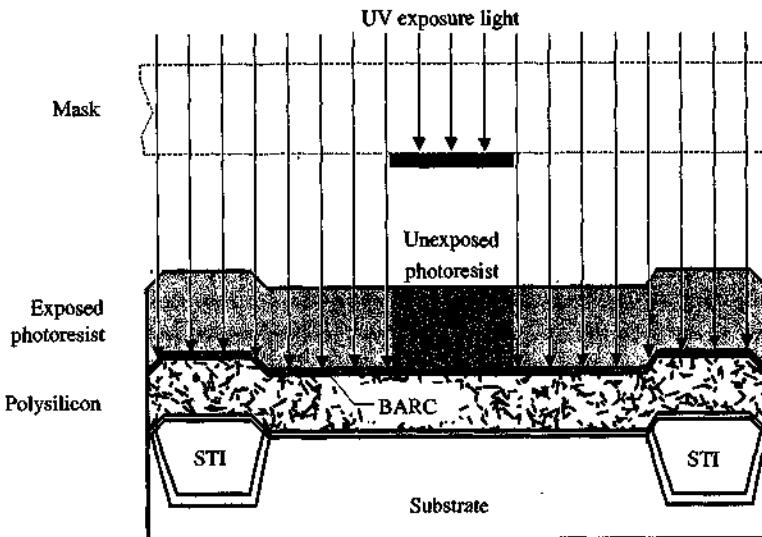


FIGURE 14.25 Light Suppression with Bottom Antireflective Coating

One factor in selecting an ARC is its ability to be removed after the completion of the photolithography process step. In some cases organic ARCs (primarily top ARCs) are aqueous-based and relatively easy to remove by rinsing during the development step. Inorganic ARCs are more difficult to remove, especially if their chemistry is similar to the underlying layer. This ARC layer is sometimes left on the wafer surface and becomes a part of the device.

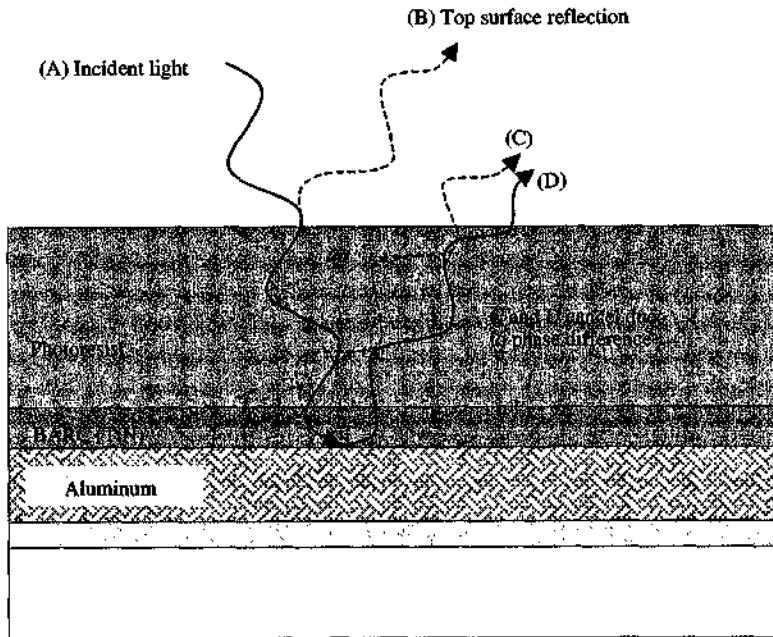


FIGURE 14.26 BARC Phase-Shift Cancellation of Light

Top ARCs. The *top antireflective coating (TARC)* reduces reflection at the interface between the resist surface and air (see Figure 14.27). TARC materials do not absorb light, but instead act as a transparent thin-film interference layer that uses destructive interference between light rays to eliminate reflectance.¹⁹

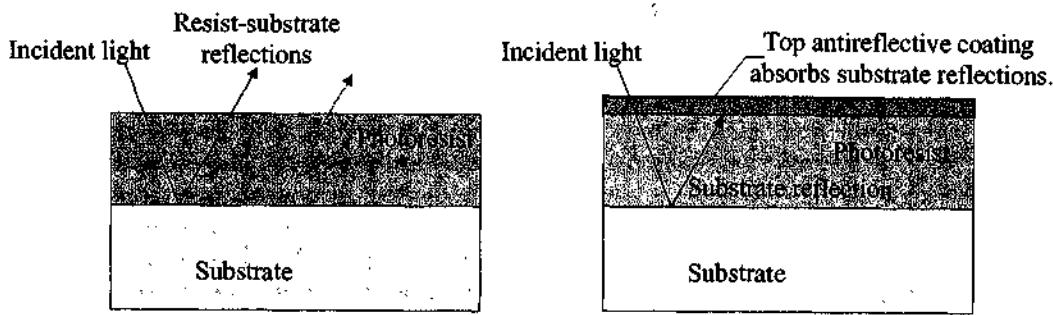


FIGURE 14.27 Top Antireflective Coating

Resolution

In photolithography, *resolution* is defined as the ability to discretely discern pairs of closely spaced features on the wafer (e.g., equal lines and spaces). This quality is shown in Figure 14.28. Resolution is critical given the high packing density for wafer features in advanced IC semiconductor manufacturing. Resolution is an important parameter for any optical system and is critical for photolithography because of the need to print extremely small feature sizes for wafer fabrication.

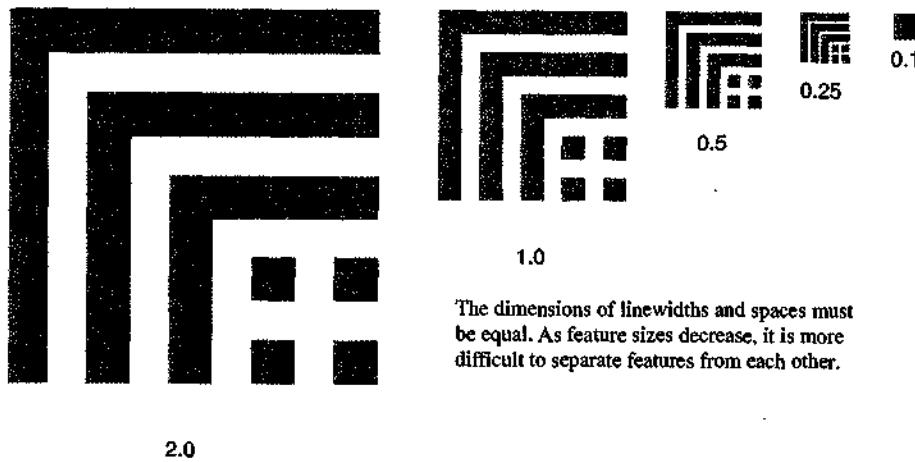


FIGURE 14.28 Resolution of Features

The formula for resolution, R, is:

$$R = \frac{k\lambda}{NA}$$

Where, k = factor that represents specific applications, with a range between 0.6 to 0.8

λ = wavelength of the light source

NA = numerical aperture of the exposure system

This formula shows that there are three parameters in photolithography that affect the ability to resolve fine geometry patterns in photoresist:

1. wavelength, λ
2. numerical aperture, NA
3. process factor, k

It should now be obvious why decreasing the wavelength, λ, of the exposing light source is important for improving resolution. A decrease in λ will improve the resolution capability of the lithography system. In other words, smaller wavelength is better. The other important parameter for

improving resolution is increasing the NA of the projection lens. Lenses with a high NA also have a high resolution. However, recall that to increase NA also requires a larger diameter lens, which becomes costly.

The third parameter, k , represents process factors of the optical system and can affect resolution. However, there are practical limitations to decreasing k much below 0.6. There are some resolution enhancement techniques available, such as phase-shift masks (PSM) and optical proximity correction (OPC) that are becoming important for reducing k to improve pattern resolution. These techniques are discussed later in this chapter.

Calculating Resolution ■ When given the wavelength, λ , numerical aperture, NA, and process factor, k , it is possible to calculate the predicted resolution, R , of an optical system (see Figure 14.29). Consider the following example:

$$\begin{aligned}\lambda &= 193 \text{ nm} \\ \text{NA} &= 0.60 \\ k &= 0.6 \\ R &= \frac{k\lambda}{\text{NA}} = \frac{(0.6)(193 \text{ nm})}{0.6} = 193 \text{ nm}\end{aligned}$$

For this optical system, the resolution of the smallest printable pattern is predicted to be 193 nm. If the wavelength of the exposure source is decreased, then the resolution will decrease. As the NA increases, the resolution will decrease. However, as explained in the following section, there is a price to pay for increasing the NA of the system, and that is decreased depth of focus.

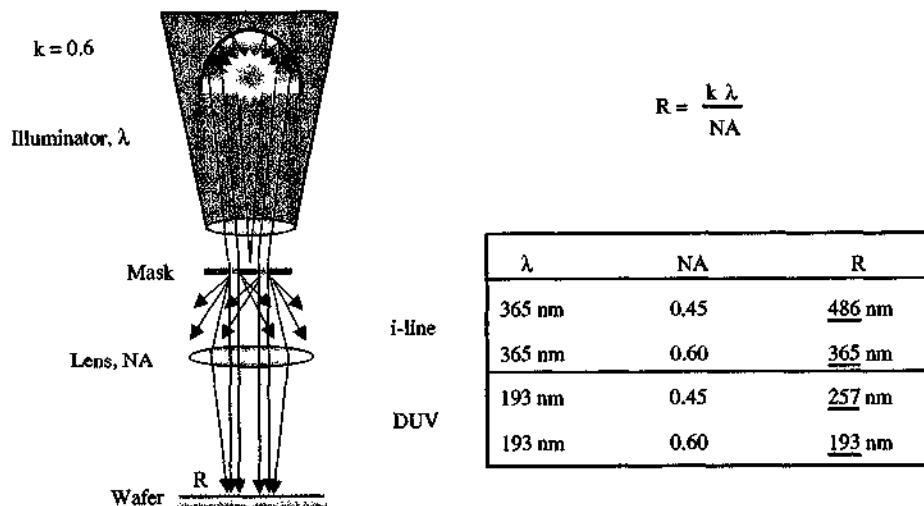


FIGURE 14.29 Calculating Resolution for a Given λ , NA, and k

Depth of Focus ■ The range around the focal point over which the image is continuously in focus is called the *depth of focus*, or *DOF* (see Figure 14.30). The *center of focus (COF)* is the point from the center of the lens where the best imaging occurs. The depth of focus is the range above and below the center of focus where the exposure energy is relatively constant. The COF may not be exactly at the middle of the resist layer, while the DOF should extend beyond the upper and lower surface of the resist coating. The actual usable DOF of any exposure system should be confirmed through tests conducted with the relevant processing parameters and environmental conditions. The goal is to find and maintain the best focus across the wafer and from wafer to wafer.

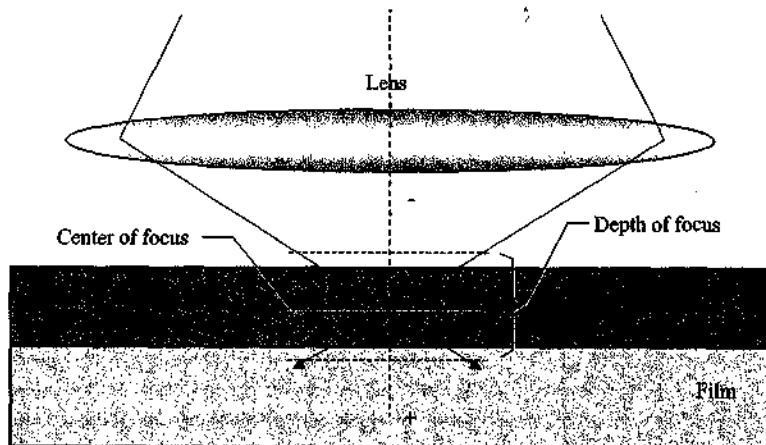


FIGURE 14.30 Depth of Focus (DOF)

The equation to describe the DOF is:

$$\text{DOF} = \frac{\lambda}{2(\text{NA})^2}$$

Where,

λ = wavelength of the exposing light

NA = numerical aperture of the optical system

Depth of focus, also known as *depth of field (DOF)*, is commonly used in photography work to indicate the acceptable focus range of a camera's lens. Have you ever seen a picture of someone in which the background is out of focus? The photograph was taken with a camera of reduced DOF. Whereas a depth of focus of 30 cm for handheld cameras is common, in semiconductor photolithography DOF values of 1 μm or less are the norm.

The DOF is calculated using the λ of the illumination source and the NA of the projection lens. As the numerical aperture of a lens increases, more optical detail is captured and the resolution capability of the system increases. The significance of the DOF equation is that if resolution is increased, then the depth of focus will decrease (see Figure 14.31). Improving patterning resolution is imperative for submicron feature sizes. However, the resulting decrease in the DOF severely reduces the process latitude of the optical system.

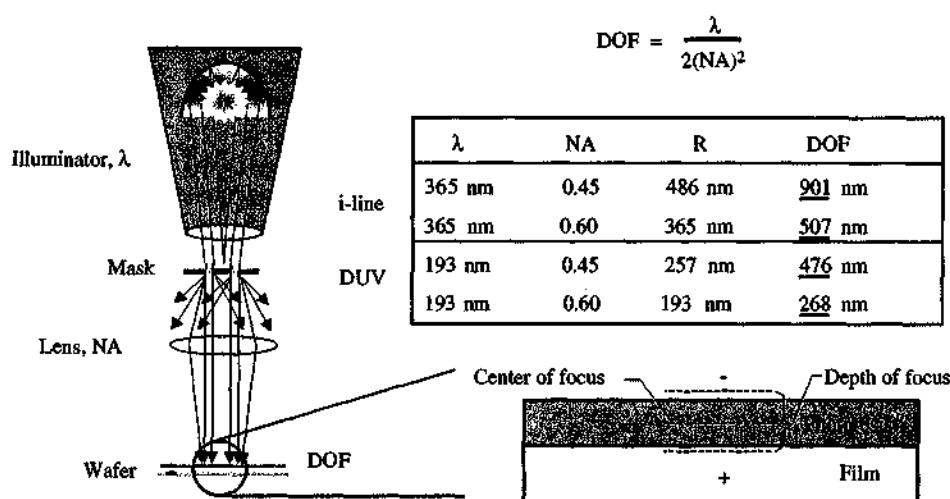


FIGURE 14.31 Resolution Versus Depth of Field for Varying NA

Resolution Versus Depth of Focus ■ In optical lithography, two critical parameters for the image quality are resolution and depth of focus. The semiconductor industry has always been challenged by the competing goals of achieving better resolution to pattern critical dimensions while simultaneously maintaining acceptable depth of focus. Any aspect of the photolithography tool or process that interferes with the ability of the wafer to lay flat, stable, and parallel to the focal plane of the lens will affect the focus quality of the exposure process. If the image plane moves away from the plane of best focus in the resist, image quality suffers. Imagine a stepper having a $\text{DOF} = 1 \mu\text{m}$. Suppose a human hair (diameter $\approx 100 \mu\text{m}$) were to fall between the wafer stage and the wafer. The wafer would be lifted roughly $99 \mu\text{m}$ beyond the focal point of the stepper's lens. The tool would not be able to achieve focus nor be able to successfully align or expose the wafer to the reticle. Photolithography tools have introduced techniques in their equipment that compensate for factors such as poor wafer flatness and equipment vibration but cannot compensate for gross conditions of contamination and damaged parts.

The wafer surface has traditionally been nonplanar due to surface topography introduced during deposition and etching steps. In the 1980s, this wafer surface topography limited resolution because the optical system was not able to adequately focus on all wafer surfaces. Many different techniques were introduced to try to reduce the surface topography during this era.

Surface Planarity. The most important technique to planarize (make flat) wafer surfaces is a planarization process known as chemical mechanical planarization (CMP). CMP was introduced into wafer fabrication in the 1990s and has made a significant contribution toward planarizing the wafer surface to meet reduced depth of focus requirements (which permits increased pattern resolution). CMP has been a major factor in reducing nonplanar wafer surface topography for $0.25\text{-}\mu\text{m}$ critical dimensions and below. The significance of CMP for photolithography is that it makes possible the reduction of the depth of focus to achieve a higher pattern resolution.

PHOTOLITHOGRAPHY EQUIPMENT

The history of photolithography since the early days of wafer fabrication can be separated into five major equipment eras. Each era is roughly based on the type of equipment needed to achieve the CD resolution necessary at the time. The five microlithography equipment eras are:

- ◆ Contact aligner
- ◆ Proximity aligner
- ◆ Scanning projection aligner (scanner)
- ◆ Step-and-repeat aligner (stepper)
- ◆ Step-and-scan system

Contact Aligner

The *contact aligner* was the primary method of photolithography during the SSI era until the early 1970s. It was used in production mode for linewidth dimensions of about $5 \mu\text{m}$ and above. Although linewidths of $0.4 \mu\text{m}$ are possible, the contact aligner is not widely used today.

The mask for a contact aligner has the complete array of all die patterns to be photographed on the wafer surface. The wafer is coated with photoresist and mounted on a stage that has manual knobs to control left, right, and rotational position (a stage is a generic name for a positioning device with x , y , and rotational, or θ , positioning). The mask and wafer are both viewed simultaneously through a microscope having split vision optics (see Figure 14.32). The mask pattern is then aligned to the pattern on the wafer through manual operator control of the stage position.

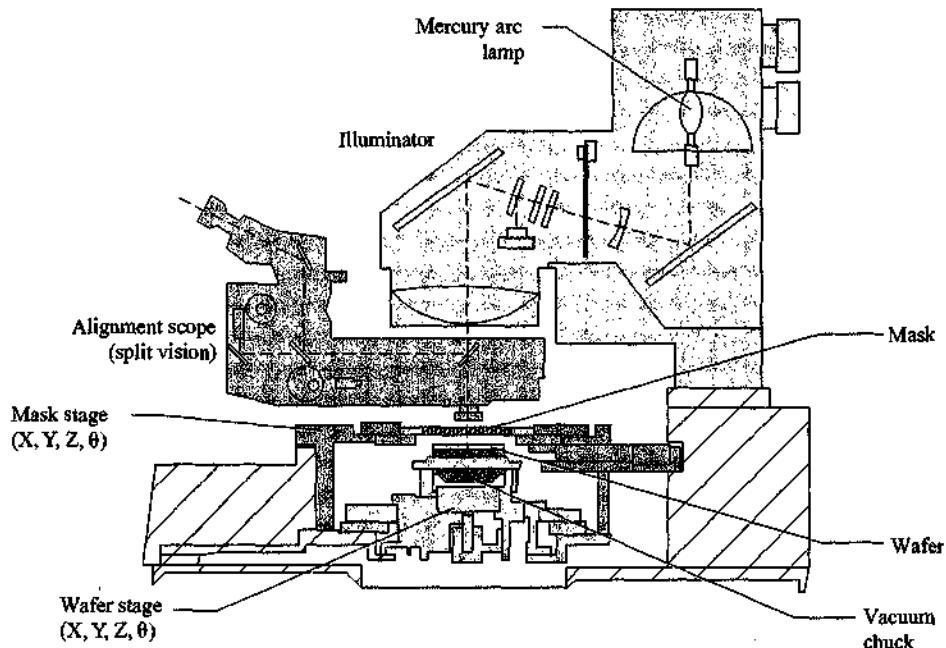


FIGURE 14.32 Contact/Proximity Aligner System
(Used with permission from Canon USA)

Once the mask is aligned to the wafer, then the mask is brought into direct contact with the resist coating on the wafer surface, which is why the equipment is called a contact aligner. At this time, the wafer and mask are exposed to ultraviolet (UV) rays. The UV rays pass through the transparent portions of the mask and the mask pattern is transferred to the resist.

Contact aligner systems were operator dependent and prone to contamination because the mask was placed in direct contact with the resist. Particulate contamination damaged the soft resist layer, the mask, or both, requiring the mask to be replaced every 5 to 25 operations. There were also resolution problems in the immediate area of any particle. As wafer sizes increased, there were overlay accuracy problems from attempting to pattern the entire wafer with one mask because alignment tolerances had to be held over wider areas.

Contact aligner printing actually can produce good image resolution on the wafer surface because the mask pattern and wafer are as close to each other as possible. This proximity reduces image distortion. However, contact alignment is very operator dependent, which introduces problems with repeatability and control.

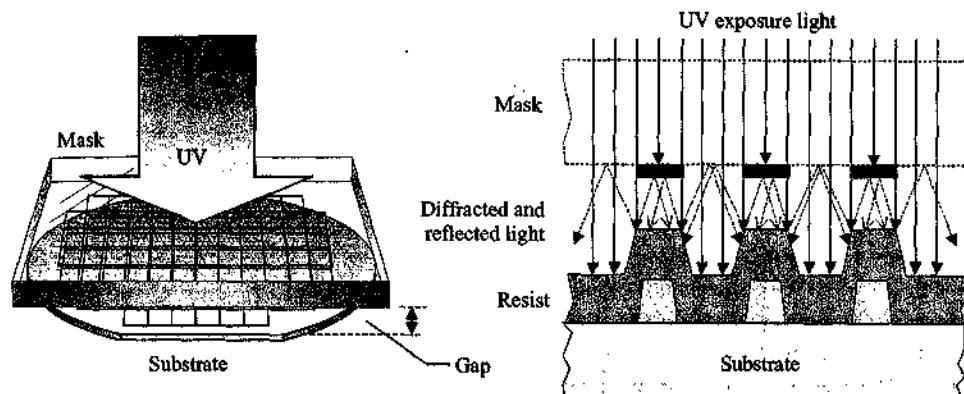
Proximity Aligner

The *proximity aligner* evolved from the contact aligner and was common in the early 1970s during the SSI and early MSI era. These aligners are still used today in low-volume laboratory applications or older wafer fabs for discrete components where retooling is not economically feasible. They are suitable for linewidth dimensions of 2 to 4 μm , depending on factors such as the reflectivity of the substrate surface.

In proximity alignment the mask, which continues to transfer the entire wafer pattern, does not make direct contact with the resist. It is positioned in close contact to the resist surface, with a gap of approximately 2.5 to 25 μm between the mask and the resist on the wafer surface. The light originating from the source is collimated, which means the light rays are forced to be parallel with each other.

The proximity aligner was an attempt to alleviate the contamination problems associated with the contact aligner by creating a gap between the resist surface and mask that would not necessarily trap particles. Even though the magnitude of the gap was controlled, the performance of the proximity aligner was reduced because of the light scattering as the UV rays passed through the transparent regions of the mask and the air (see Figure 14.33 on page 390). This condition impaired

the system's resolution capability, which was a major problem due to the need to reduce the linewidth critical dimension.



Diffraction of light on edges results in reflections from underside of mask causing undesirable resist exposure.

FIGURE 14.33 Edge Diffraction and Surface Reflectivity on Proximity Aligner

Scanning Projection Aligner

The industry recognized the need to move away from any form of contact or near contact alignment system due to its contamination problems, edge diffraction, resolution limitations, and operator dependency. The development of the *scanning projection aligner* (also referred to as *scanner*) in the early 1970s attempted to address these problems. Scanning projection aligners were the dominant lithography exposure tool in the late 1970s and early 1980s.²⁰ These aligners are still used today in older wafer fabs. They are suitable for noncritical layers containing linewidths greater than 1 μm .

The *scanning projection aligner* concept is to project a full wafer mask with a 1:1 image onto the wafer surface using a mirror system (i.e., based on reflective optics). Since the mask is 1X, there is no magnification or reduction involved and the pattern on the mask has the same dimensions as the pattern on the wafer.

The UV light rays are focused onto the wafer through a narrow slit, allowing a uniform source of light (see Figure 14.34). The mask and resist-coated wafer are mounted on a scanning carriage and moved in unison across the narrow beam of UV light to expose the resist on the wafer. As the scanning motion takes place, the image of the mask is eventually printed on the wafer surface.

A major challenge of scanning projection aligners was making a good 1X mask that contained all the chips on the wafer. If the chip had submicron feature sizes, then the mask also had submicron dimensions. As submicron feature sizes were introduced, this photolithography approach made it difficult for the mask maker to have no defects in the mask.

Step-and-Repeat Aligner (Stepper)

The mainstay of the 1990s microlithography equipment for wafer fabrication is the *step-and-repeat aligner* (commonly referred to as a *stepper*). Steppers have their distinct name because the tool projects only one exposure field (which may be one or more chips on the wafer), and then steps to the next location on the wafer to repeat the exposure. Steppers were first commercialized in the early 1980s.²¹ Though not universally accepted at first, optical lithography steppers have dominated IC fabrication since the later 1980s, primarily for applications with critical dimensions down to 0.35 μm (conventional i-line photoresist) and some 0.25 μm (DUV photoresist).

A stepper uses a reticle, which contains the pattern in an exposure field corresponding to one or more die. A mask is not used in a stepper since a mask contains the entire die matrix. The optical projection exposure system of steppers uses refractive optics to project the reticle image onto the wafer (see Figure 14.35).

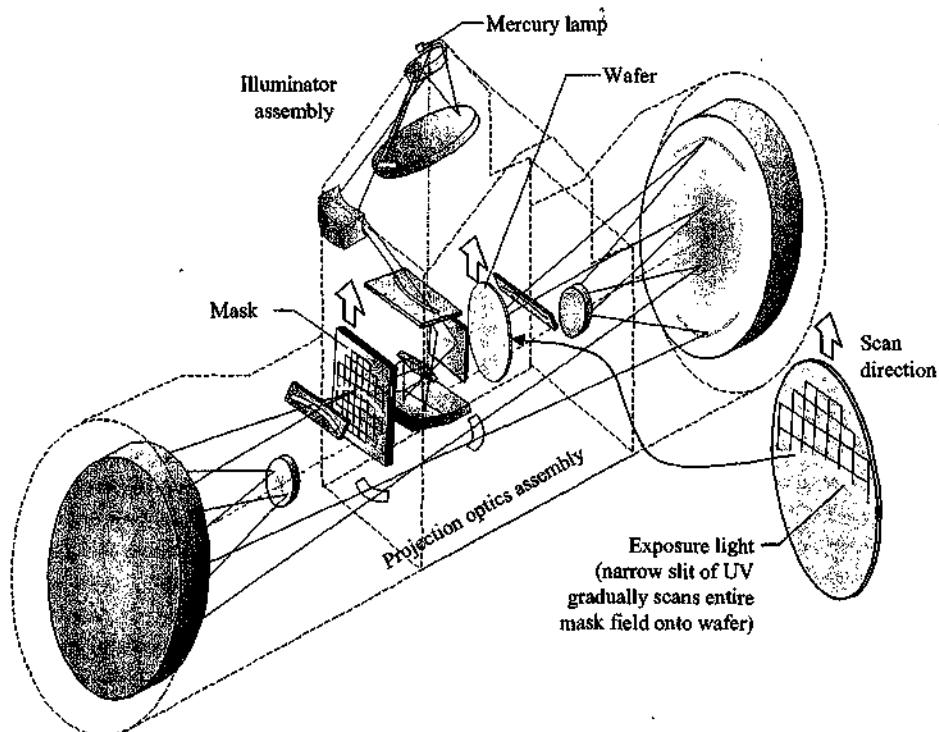


FIGURE 14.34 Scanning Projection Aligner
(Redrawn with permission from Silicon Valley Group Lithography Systems,
Perkin-Elmer 500 Micralign)

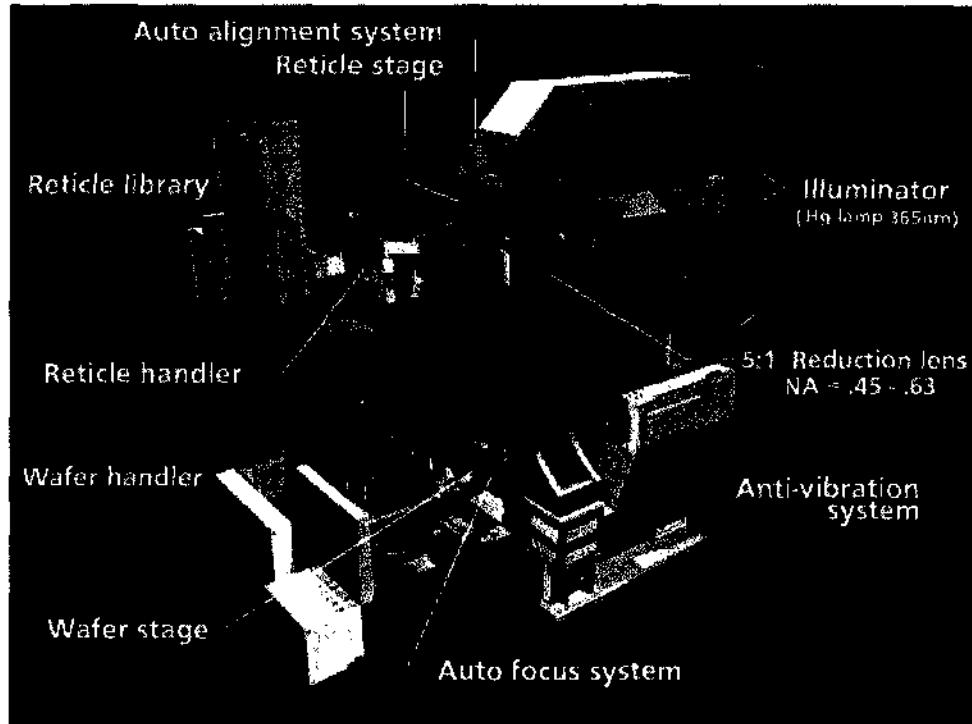


FIGURE 14.35 Step-and-Repeat Aligner (Stepper)
(Used with permission from Canon USA, FPA-3000 i5)

An advantage of optical steppers is their ability to use a reduction lens. Traditionally, i-line stepper reticles are sized 4X, 5X, or 10X larger than the actual image to be patterned (initially steppers used 10X reticles and then later 5X and 4X). To further explain the purpose of a reduction lens, a stepper with a 5X reticle requires a 5:1 reduction lens to transfer the correct image size to the wafer surface. This demagnification factor makes it easier to fabricate the reticle because the features on the reticle are five times larger than the final image on the wafer. At least one stepper manufacturer produces a stepper that provides a 1:1 projected image. The advantage of a nonreduction stepper is a low-cost stepper that can be used on noncritical patterning layers.

At each step in the exposure process, the stepper will focus the wafer and the reticle to the projection lens, align the wafer to the reticle, expose the resist with UV light that passes through the transparent regions of the reticle, and then step to the next location on the wafer to repeat the entire sequence. By following this process, the stepper will ultimately transfer the full die array onto the wafer in a sequence of exposure steps (see Figure 14.36). Because the stepper exposes only a small portion of the wafer at one time (e.g., typically a field size of one die on the wafer for a larger microprocessor die), compensations for variations in wafer flatness and geometry can be easily performed.

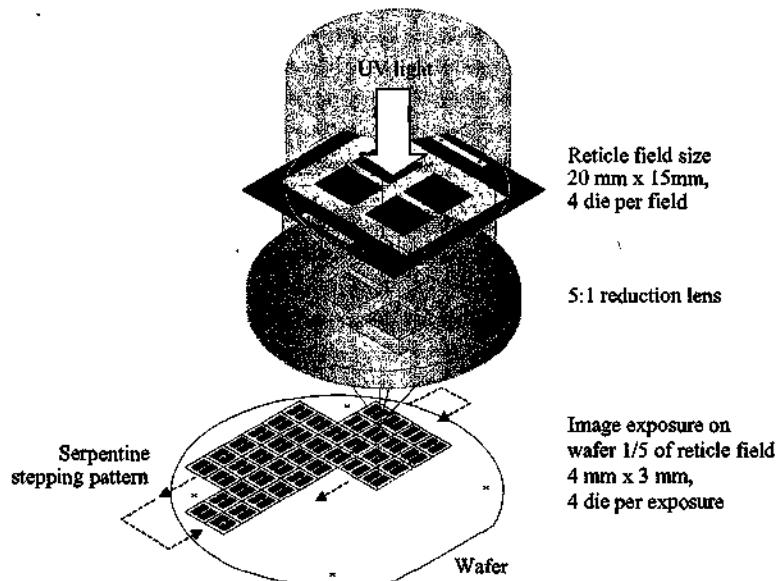


FIGURE 14.36 Stepper Exposure Field

Steppers have used conventional mercury arc lamp illumination sources (for g-line of 436 nm, h-line of 405 nm, and i-line of 365 nm) with a CD down to $0.35\text{ }\mu\text{m}$. To obtain a 248 nm DUV wavelength source, the mercury arc lamp source is replaced with a KrF (krypton-fluoride) excimer laser. This equipment permits patterning a $0.25\text{-}\mu\text{m}$ CD. Typically, DUV stepper technology is used for patterning critical layers while conventional i-line exposure methods are used for noncritical layers. This mix-and-match approach to lithography is used to reduce production costs.

The growth in the physical size of semiconductor chips along with smaller critical dimensions creates a need for a larger exposure field size and improved lens optics for steppers. This in turn dictates more complex lithographic lens design and fabrication, with the cost of lens systems alone in steppers exceeding \$1 million. This cost limits a conventional step-and-repeat exposure field to $22 \times 22\text{ mm}$. To overcome these problems, an evolutionary stepper known as step-and-scan has become predominant for DUV photolithography at technology nodes of $0.25\text{ }\mu\text{m}$ and below.

Step-and-Scan System

A recent development in lithographic exposure equipment employs a technique referred to as a *step-and-scan system*. The step-and-scan optical lithography system is a hybrid tool that combines the technology from scanning projection aligners and step-and-repeat steppers by using a reduction lens to scan the image of a large exposure field onto a portion of the wafer. A focused slit of light is scanned simultaneously across the reticle and wafer (see Figure 14.37). One standard exposure field size for this type of aligner is 25×33 mm for a six-inch-square reticle size. Once the scan and pattern transfer is completed, then the wafer is stepped to the next exposure field and the process is repeated.

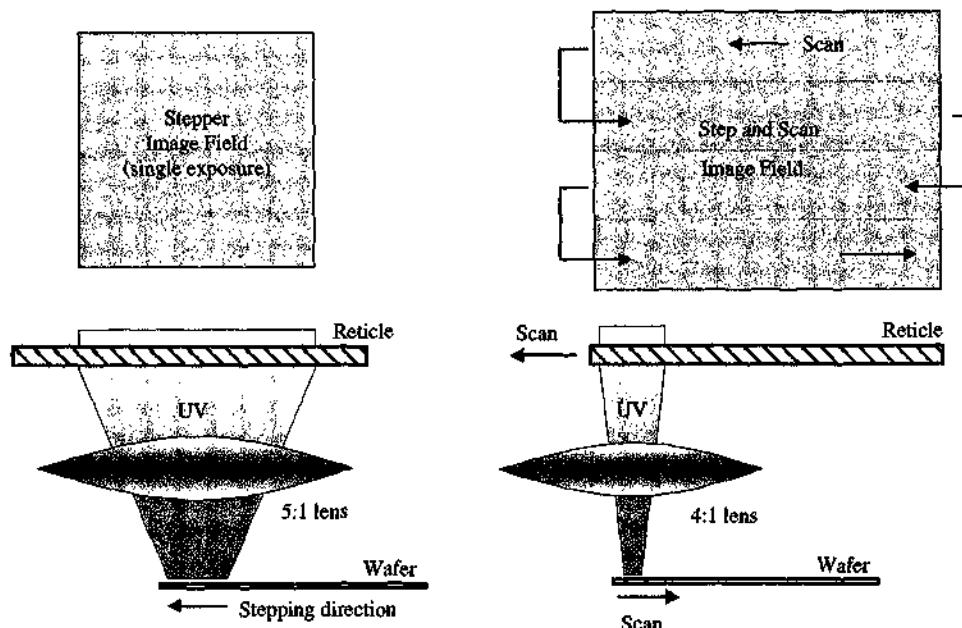


FIGURE 14.37 Wafer Exposure Field for Step-and-Scan
(Used with permission from ASM Lithography)

A benefit from using a step-and-scan system to expose wafers is the increased exposure field for larger chip sizes. The lens field only has to be a narrow slit, as in the older full-wafer scanning projection aligners. It scans a reduction reticle (typically a 4X) through a small, well-corrected image field of 26×33 mm before stepping to the next location.²² The optical system can now be designed for this much smaller field, permitting a smaller lens system (see Figure 14.38 on page 394). Another major advantage of a larger field size is the opportunity to put more die on a reticle, thus exposing more die in a single exposure.

Another significant step-and-scan system advantage is the ability to adjust focus throughout a scan (called on-the-fly focus), permitting compensation for lens defects and changes in wafer flatness. This improved control of focus during a scan yields improved CD uniformity control across the exposure field.

The major challenge with step-and-scan systems is the increased demand on mechanical tolerance control due to the motion of the stage controls for the reticle and wafer. Whereas steppers only had to move the wafer rapidly to a new position, a step-and-scan system has to precisely move both the wafer and reticle simultaneously and in opposite directions. These scanning and stepping motions are performed while holding position tolerances to within a few tens of nanometers during the scan.

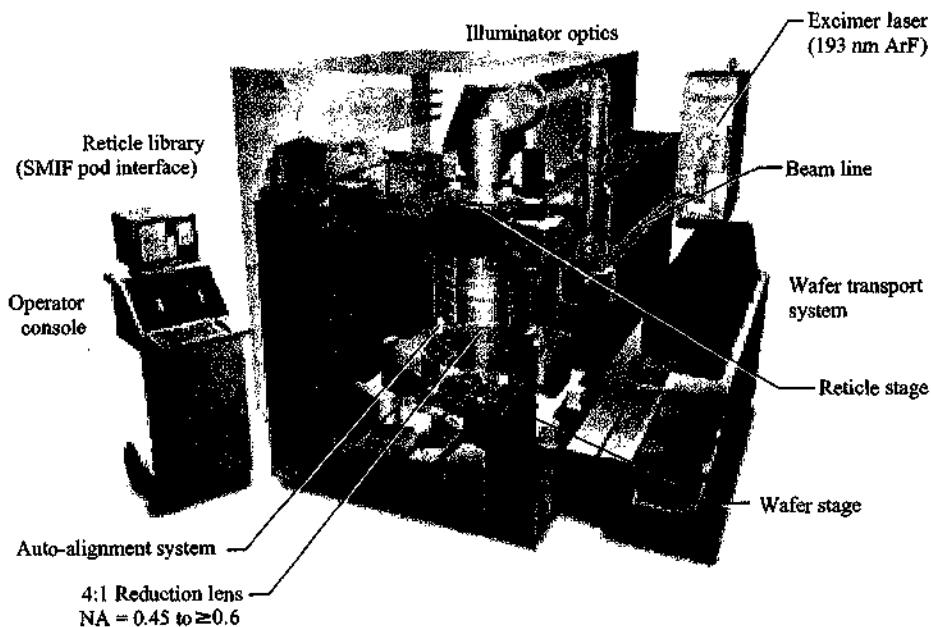


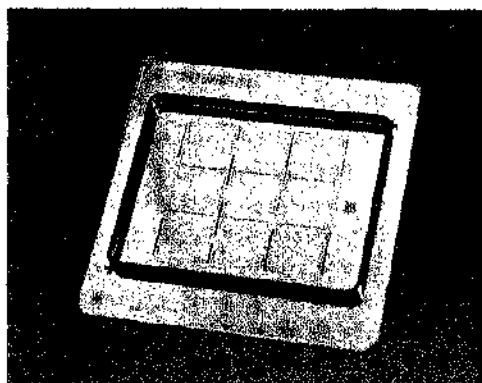
FIGURE 14.38 Step-and-Scan Exposure System
(Used with permission from ASM Lithography, PAS 5500)

Reticles

A *reticle* is a transparent plate that has a pattern image that will be transferred to the photoresist coating on the wafer. Two terms are commonly used in wafer fabrication: reticle and mask. A reticle contains the pattern image for only part of the wafer (e.g., 4 die) that must be stepped and repeated across the entire substrate. Reticles are used for step-and-repeat steppers and step-and-scan systems found in wafer fabrication. A *photomask*, or *mask*, contains the pattern image for a complete wafer die array and the pattern is transferred in a single exposure (1:1 image transfer).²³ Masks are used for older proximity printers and scanning alignment projectors. Table 14.6 compares reticles versus masks to provide insight into why the photolithography industry changed from mask to reticle.

TABLE 14.6 Comparison of Reticle versus Mask

Parameter	Reticle Number of Exposures	Mask Single Exposure
Critical Dimension	Easier to pattern submicron dimensions on wafer due to larger pattern size on reticle (e.g., 4:1, 5:1).	Difficult to pattern submicron dimensions on mask and wafer without reduction optics.
Exposure Field	Small exposure field that requires step-and-repeat process.	Exposure field is entire wafer.
Mask Technology	Optical reduction permits larger reticle dimensions—easier to print.	Mask has same critical dimensions as wafer—more difficult to print.
Throughput	Requires sophisticated automation to step-and-repeat across wafer.	Potentially higher (not always true if equipment is not automated).
Die alignment and focus	Adjusts for individual die alignment and focus.	Global wafer alignment, but no individual die alignment and focus.
Defect density	Improved yield but no reticle defect permitted. Reticle defects are repeated for each field exposure.	Defects are not repeated multiple times on a wafer.
Surface flatness	Stepper compensates during initial global prealignment measurements or during die-by-die exposures.	No compensation, except for overall global focus and alignment.



Photolithography Reticle (Photo courtesy of Advanced Micro Devices)

A reticle must be perfectly manufactured. All wafer circuit features ultimately come from patterns on the reticle; therefore, the quality of the reticle plays a key role in achieving high-quality imaging during submicron photolithography. If reticle pattern defects such as distortion and incorrect image placement are not detected, they will be reproduced in the resist on the wafer. Reticles undergo extensive automated testing for defects and particles once they are fabricated.

Reticle Materials ■ The primary reticle substrate material for submicron lithography is fused silica. It is always used for DUV lithography because it has high optical transmission in the DUV portion of the spectrum (248 and 193 nm). Fused silica is the most expensive material for reticles and has very low thermal expansion. A low thermal expansion means the reticle is dimensionally stable with changes in temperature. Because of this, there is only minor thermal run-out, which is an increase in a dimension due to a change in the thermal environment. Other desirable properties for reticle material are high optical transmission and no defects on the surface or within the material.

The most common opaque material deposited on the reticle substrate is a thin layer of chromium (referred to as chrome). This chrome layer will undergo patterning to create the necessary design shapes to define the circuitry for the wafer layer (e.g., holes, lines, pads, and so on). The chrome thickness is usually less than 1,000 Å and is sputter deposited. There sometimes is an antireflective layer of chromium oxide (about 200 Å thick) placed on the chrome.

Reticle Reduction and Size ■ Reticles are used in stepper and step-and-scan system applications that require a reduction lens to reduce overlay accuracy during patterning. A stepper commonly uses a reticle reduction ratio of 5:1 or 4:1, whereas a step-and-scan system uses a reticle reduction ratio of 4:1. Each reticle layer usually contains the necessary pattern for one or more die. The small field exposure size on steppers and step-and-scan systems permits close tolerance control during reticle alignment.

Most reticles are currently 6 × 6 inches (152 mm) square, although 5 × 5 inches is still common. Reticles are usually 0.090" to 0.250" thick. A six-inch reticle with a 4X reduction defines a usable field size at the wafer surface, which on current photolithography step-and-scan equipment is about 25 × 25 mm (see Figure 14.39 on page 396). Multiple exposures are required as the reticle is stepped over the wafer. The required exposure field size is forecasted to increase to accommodate the increasing size of DRAM and microprocessor chips. The industry is investigating a changeover to a nine-inch reticle to replace the current six-inch reticles. This will increase the number of die exposed per field, which permits fewer steps in the stepper or step-and-scan to complete the exposure of a complete wafer.²⁴ However, there is substantial cost to implementing this change, which may negate the benefit.

Reticle Fabrication ■ A common method for creating a pattern on a reticle is by the use of an electron beam (e-beam). This technique uses direct writing of the pattern from electronically stored original patterns. There is an immense amount of data involved in this process, and the total time to write a pattern on a reticle can be several hours. The basic steps for patterning a reticle are similar to those for a wafer.

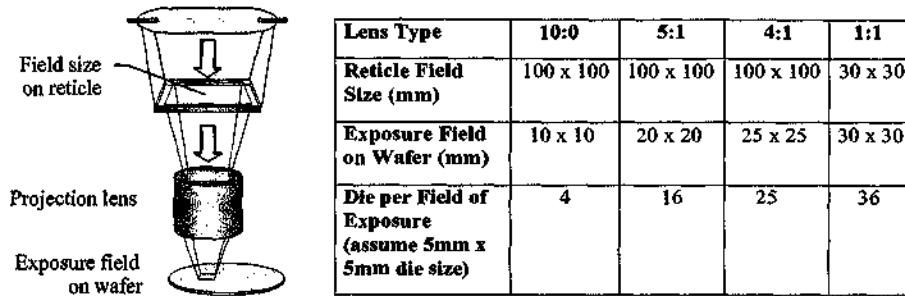
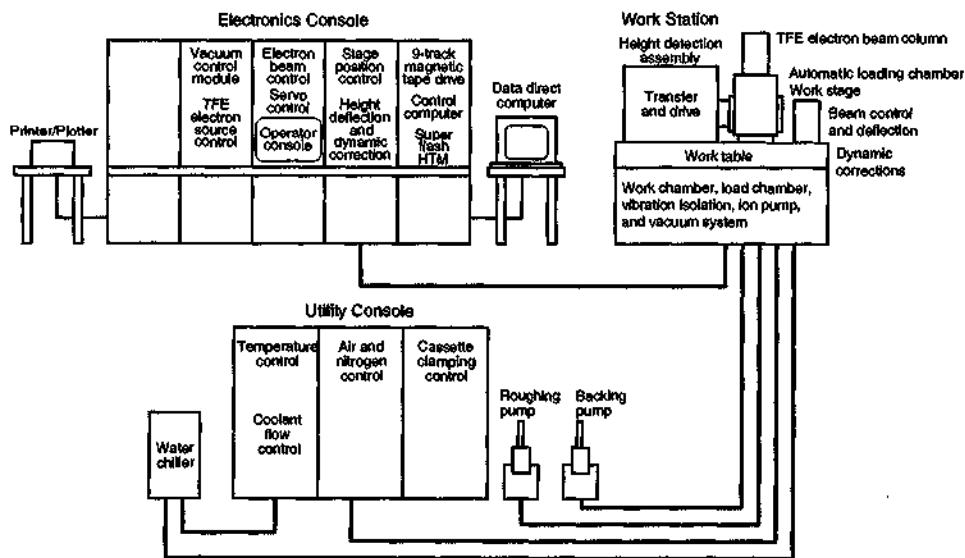


FIGURE 14.39 Comparison of Reticle Reduction Versus Exposure Field

E-Beam Lithography. The direct-write method of *electron beam (e-beam) lithography* is capable of transferring a high-resolution pattern directly to the reticle surface. In e-beam lithography, an electron source produces many electrons that are accelerated and focused in the shape of a beam toward the reticle (see Figure 14.40). The electron beam is focused either magnetically or electrostatically and is scanned in the desired pattern across a special e-beam resist on the reticle surface. The beam can be scanned across the entire reticle (raster scan) or scanned only over the printed areas (vector scan) to eventually transfer the pattern to the reticle.

To apply the e-beam resist on the reticle, the reticle is cleaned and the chrome film is spin coated with a suitable photoresist, then soft baked. The standard e-beam resist is a positive-tone poly(butene 1 sulfone), or PBS. However, this resist is not suitable for submicron linewidths. Alternative resists are being developed, such as chemically amplified resists.²⁵ Advanced e-beam mask writers are capable of patterning 0.36- μm minimum feature sizes on a reticle.²⁶

After exposing and developing the resist, the final patterned surface is etched into the chrome film with a wet- or dry-etch process (advanced reticle fabrication uses dry etch). Recall that the mask features are four or five times larger than the wafer images (due to lens reduction in steppers and step-and-scan tools). However, the dimensional tolerances are very tight on mask features in order to produce acceptable critical dimensions on wafers.

FIGURE 14.40 Principles of Electron Beam Lithography
(Used with permission from Etec Systems, Inc., MEBES 4500 System)

E-beam replaced optical resist patterning on reticles because it has a shorter wavelength and higher exposure speed compared to UV sources. These qualities produce better dimensional control with increased production throughput.²⁷ E-beam is a reliable method with high resolution and flexibility, but speed and system complexity are disadvantages for its use to pattern wafers.²⁸ It does have applications in very specialized wafers, such as application-specific integrated circuits (ASICs) that benefit from no reticle to achieve a quick turnaround time from development to production.

Sources of Reticle Damage ■ Reticles on advanced steppers and step-and-scan systems are moved into and out of the tool with automated storage and handling. A reticle must be meticulously clean for it to create perfect images during its many exposures to pattern a wafer layer. Recall that a stepper will step-and-repeat one image over the entire wafer. If there is even one foreign particle on a reticle circuit image, then it is repeated at every site on the wafer.

There are actually many possible sources of damage during use of a reticle, such as dropping the reticle, scratches on the surface, electrostatic discharge (ESD), and particles of dirt. ESD causes a problem if the reticle is handled by an improperly grounded technician. This condition could potentially discharge a small surge of current through the micron-sized chromium lines on the reticle surface, melting a circuit line and destroying the pattern.

If an airborne particle of dirt lands on a critical region of the reticle, this could damage the circuit and create an imaging defect. A solution to particulate contamination on reticles is to protect the surface with a thin, optically transparent membrane known as a *pellicle*. The pellicle is tightly stretched on a sealed frame about 5 to 10 mm above the mask surface. No dirt particles are able to reach the mask surface. If a dust particle lands on the pellicle, it is far removed from the focal plane and is invisible to the projection optics (see Figure 14.41).

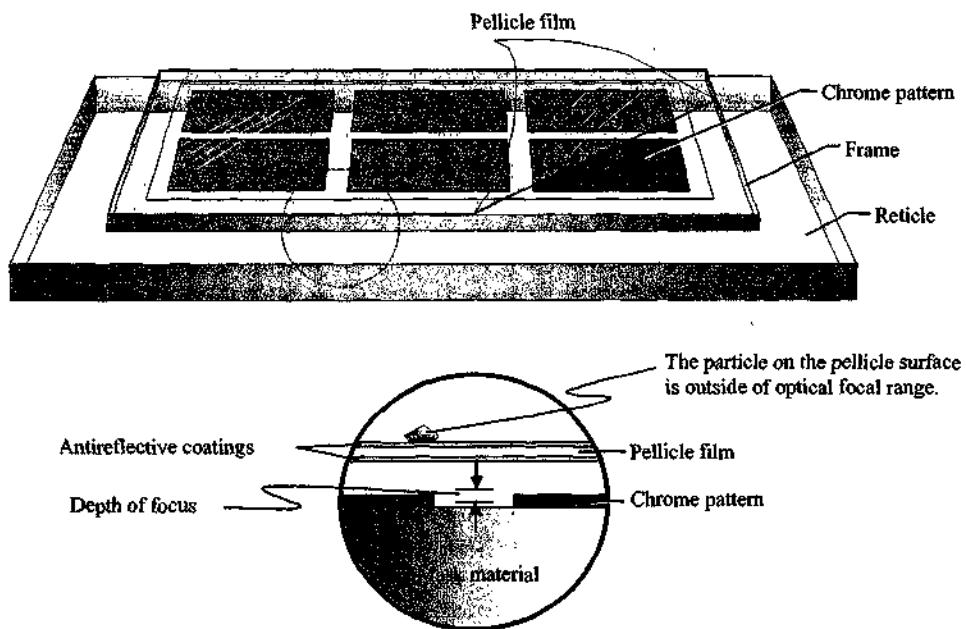


FIGURE 14.41 Pellicle on a Reticle

The pellicle material is transparent to the exposing light energy. There are different materials and thickness used for pellicles, such as nitrocellulose acetate that is typically 0.7 μm thick, or a 12 μm thick Mylar fluorocarbon material. The pellicle is fragile; therefore, any action that would scratch the mask surface will rupture the pellicle membrane (which alerts the user to a potential source of reticle damage). When used for deep UV exposure wavelengths, the optical transparency of the pellicle must be carefully evaluated.

Optical Enhancement Techniques

As wafer critical dimensions decrease to feature sizes of $0.15\text{ }\mu\text{m}$ and below, factors such as diffraction and light scattering prevent reticles from effectively transferring a replica of the pattern to the wafer. Optical enhancement techniques are being used on the reticle to improve the quality and definition of the image. This has become an important area of optical lithography known as *subwavelength lithography*, which permits a wafer to be patterned with a resolution slightly below the light exposure wavelength.

Phase-Shift Mask (PSM) ■ *Phase-shift mask (PSM)* is a method developed in 1982 to overcome problems associated with light diffraction through small openings patterned on the reticle. With PSM, the reticle is modified with an additional transparent layer so that alternating clear regions cause the light to be phase-shifted 180° (see Figure 14.42). As we have learned about interference between out-of-phase waves, there is now destructive interference. The light diffracted into the nominally dark area on the left will encounter destructive interference with the light diffracted from the right clear area. The light diffraction underneath the opaque area is reduced. Phase-shifting techniques on reticles improve image contrast, and has become a critical factor for microlithography at CDs of $0.18\text{ }\mu\text{m}$ and below. There are many approaches to PSM that use the same basic concept, but they are all based on the principle of destructive interference.

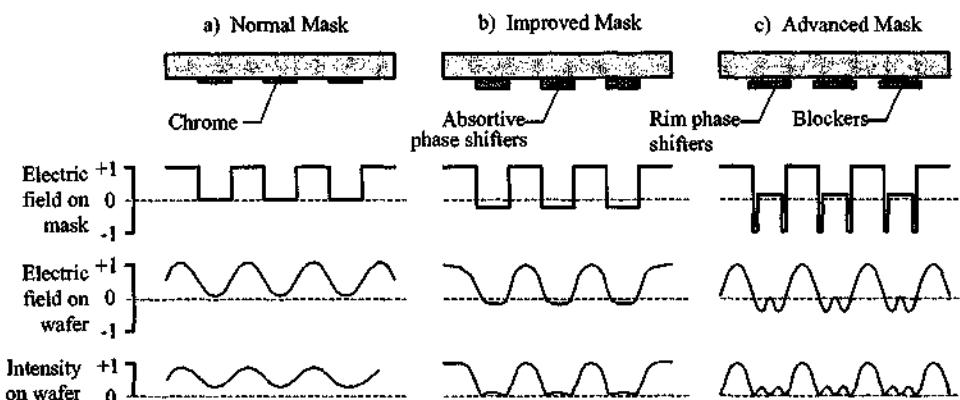


FIGURE 14.42 Phase-Shifting Mask
(Reprinted from the January 1992 edition of *Solid State Technology*, copyright 1992 by PennWell Publishing Company)

Optical Proximity Correction (OPC) ■ Uniformity of feature size dimensions is critical in high-performance ICs, especially in the transistor gate region where varying linewidths will affect the speed of the device. Because of optical proximity effects due to light diffraction and interference between closely spaced features on the reticle, the widths of lines in the lithographic image are affected by other nearby features (see Figure 14.43). Tightly grouped lines will print with a different size than isolated lines, although both have the same linewidth dimension on the reticle. This particular example is referred to as iso-dense bias.

It is possible to introduce selective image size biases (alterations) into the reticle pattern to compensate for optical proximity effects. This is called *optical proximity correction (OPC)*. Computer algorithms are available to the reticle designer for generating optical proximity corrections on the reticle for smaller feature sizes. However, producing reticles with this level of control is a challenge in manufacturing, especially since CD feature sizes are already extremely small. As the critical dimension is reduced to less than $0.18\text{ }\mu\text{m}$, aggressive scaling of dimensions will require increased use of OPC, which will make reticle fabrication more complicated.²⁹

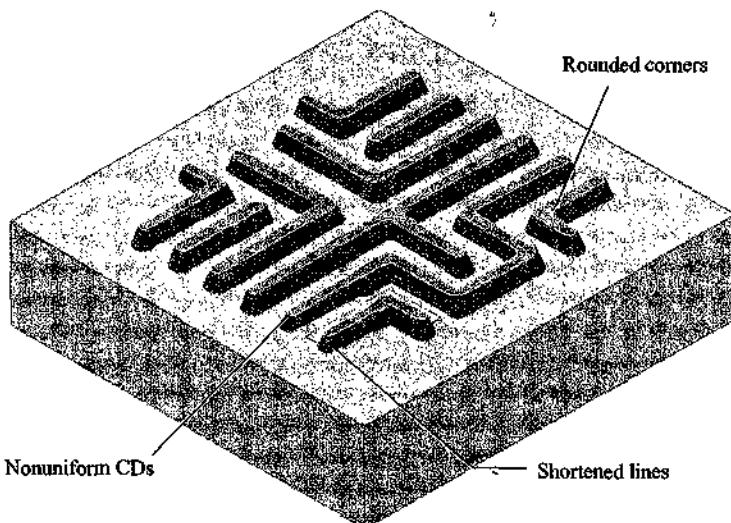


FIGURE 14.43 Optical Proximity Effects

Off-Axis Illumination ■ The conventional form of illumination for lithographic lenses has the exposing light centered on the centerline of the projection optics (see Figure 14.44). However, light diffraction creates a problem as the reticle openings are made smaller and smaller for submicron features sizes, to the point that the imaging lens will not transfer the pattern. A solution is to have the incident light strike the mask at an angle in order to align the diffraction fringes with the lens, which produces a symmetrical intensity profile that corresponds to the reticle pattern. This is referred to as *off-axis illumination* or (*OAI*). This technique reduces the resolution limit and increases the depth of focus for imaging.

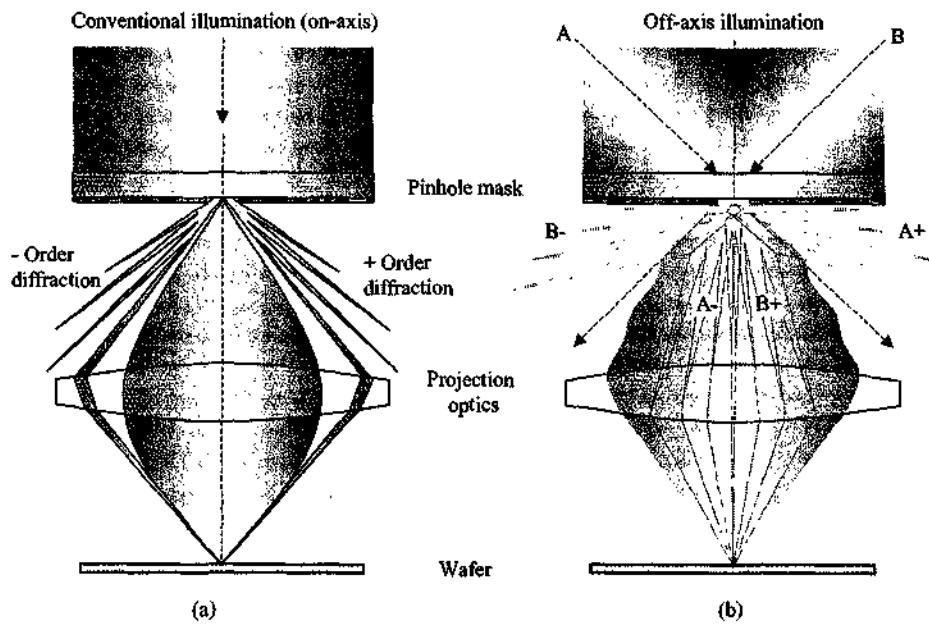


FIGURE 14.44 Off-Axis Illumination

Bias ■ *Print bias* is the difference between a feature dimension on a reticle and the same printed dimension on the wafer. The amount of print bias may be different for varying sized features. An example of print bias is a square contact with dimensions near the optical system resolution that prints with rounded corners. One way to address this problem is by adding serifs to the features on the reticle pattern, which help square features such as contact corners during printing on the wafer (see Figure 14.45 on page 400).³⁰

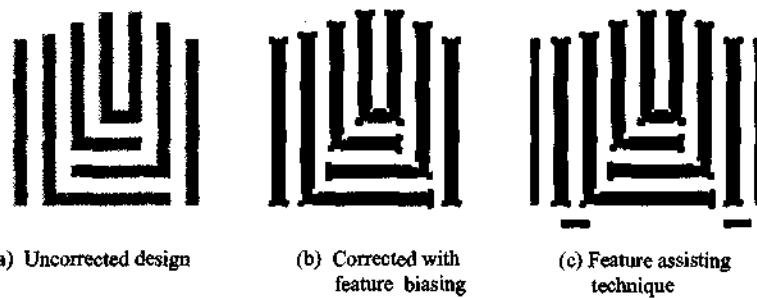


FIGURE 14.45 Serifs to Minimize Rounding of Contact Corners
(Redrawn from "Optical/Laser Microlithography," SPIE Proceeding VIII 2440 (February 1995)

Alignment

The alignment process begins with the proper alignment of the reticle to a fixed reference mark on the body of the stepper or step-and-scan system. Once the reticle is aligned to the body of the exposure tool, then the wafer stage positioning is measured with respect to the reticle. This positioning provides any baseline correction data that the alignment software will use to compensate for variations in reticle characteristics. This process is referred to as *baseline compensation (BLC)*.

To successfully pattern the wafer, the wafer pattern must be correctly aligned to the existing patterns on the reticle pattern. The IC will function properly only if each successive projected image is correctly matched to the wafer patterns. To this end, *alignment* is the process of determining the position, orientation, and distortion of the patterns already on the wafer and then placing them in correct relation to the projected image from the reticle. Alignment should be fast, repeatable, accurate, and precise. The outcome of the alignment process, or how accurately each successive pattern is matched to the previous layer, is known as *registration*.³¹

Overlay accuracy (also known as registration) is the measure of the alignment system's ability to overlay the reticle pattern onto the wafer pattern. Overlay budget describes the maximum relative displacement between a patterned layer and the previously defined layer (see Figure 14.46). In general, the overlay budget is about one-third of the critical dimension. For 0.15 μm design rules, the overlay budget is expected to be 50 nm.³²

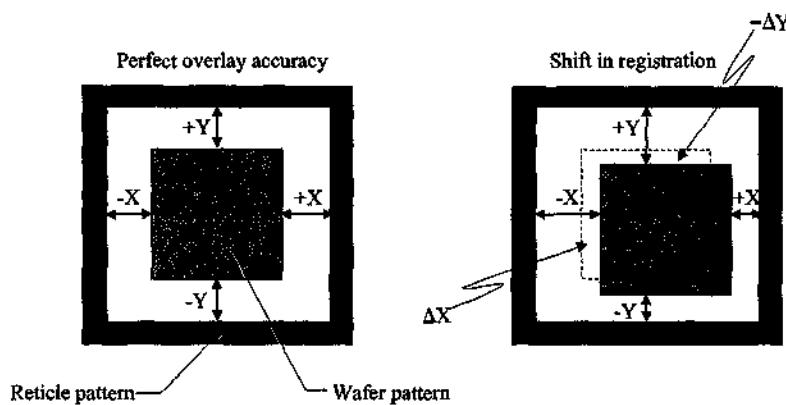


FIGURE 14.46 Overlay Budget

For steppers and step-and-scan systems, each reticle pattern is aligned and exposed at multiple locations as the aligner steps across the wafer. Each field corresponds to the reticle pattern of a single large chip or several smaller chips. A *grid* is the particular path that the photo tool follows to step across the wafer and expose the individual fields (see Figure 14.47).

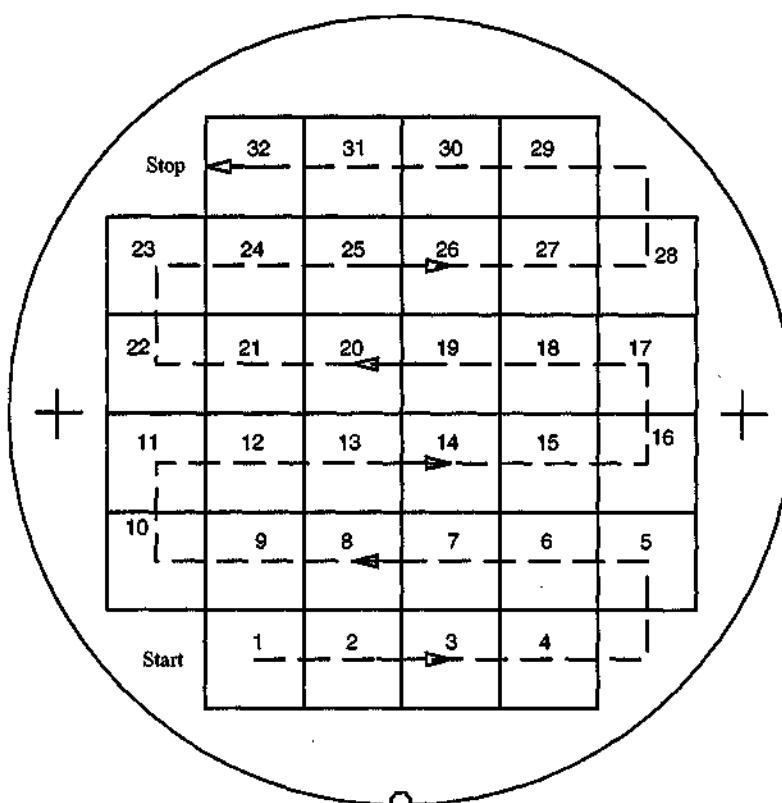


FIGURE 14.47 Grid of Exposure Fields on Wafer

Steppers and step-and-scan systems have a sophisticated automatic alignment system that detects the position and orientation of both the wafer and the reticle and then aligns the wafer to the reticle before exposure (see Figure 14.48 on page 402). The tool's alignment system includes alignment marks on both the reticle and wafer, an alignment detection system, and the electromechanical positioning devices such as motors and drive mechanisms. The goal of alignment detection is to determine as rapidly as possible the position of the different alignment marks in the exposure tool coordinate system. Alignment software algorithms within the tool's process controller calculate offset values and the direction of wafer stage movements necessary to bring the wafer within the overlay budget specified for the tool.

Alignment Marks ■ *Alignment marks* are visible patterns placed on the reticle and the wafer to determine their position and orientation. Marks, also referred to as targets or fiducial marks, may be one or more lines on the reticle, which then become trenches when printed on the wafer. Marks may also be a shape on the reticle which overlays on wafer marks (see Figure 14.49 on page 402). Reticle alignment (RA) marks appear on both the left and right sides of the reticle. The RA marks are aligned to fiducial marks mounted on the stepper's body. Global alignment (GA) marks are printed on the left and right sides of the wafer during first mask exposure. GA marks are used for making coarse alignments of each wafer during subsequent layers. Fine alignment (FA) marks are printed during each field exposure. FA marks are used for making final alignment adjustments between each wafer exposure site and the reticle. The shape and location of marks varies depending on the equipment manufacturer. Once an alignment mark is aligned, it is assumed that the remainder of the reticle patterns is also correctly aligned.

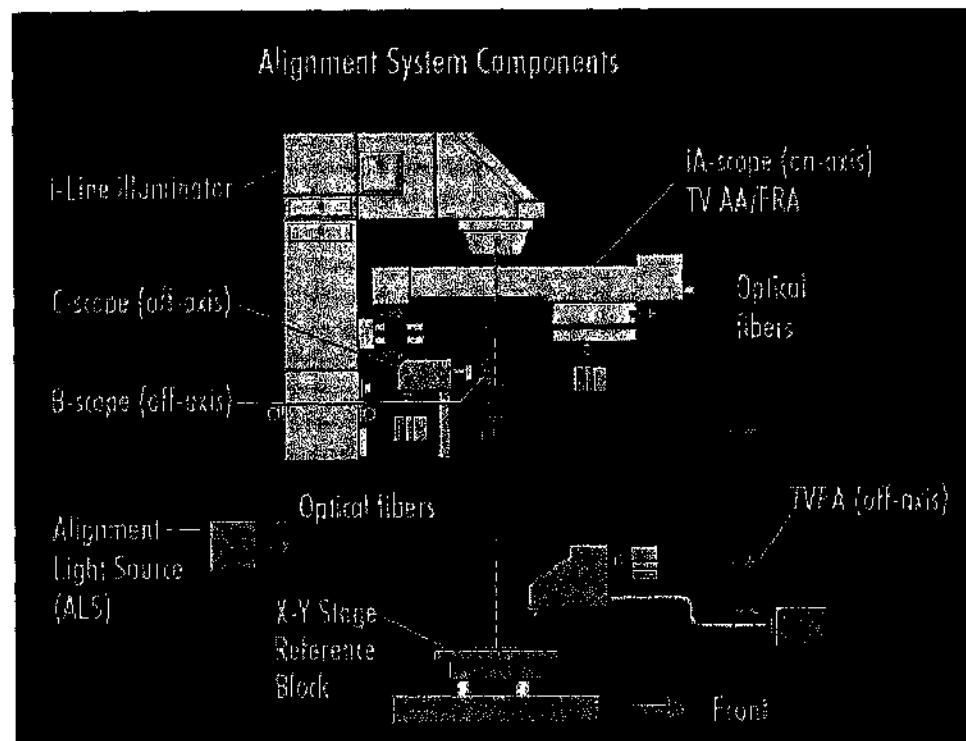


FIGURE 14.48 Step-and-Repeat Alignment System
(Used with permission from Canon USA, FPA-2000 II)

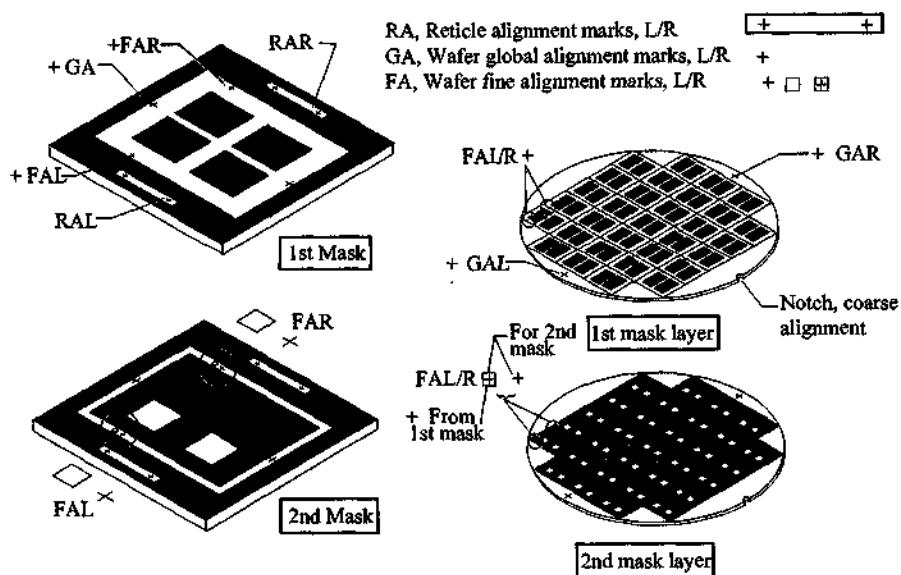


FIGURE 14.49 Alignment Marks

The computer-controlled automatic alignment system identifies the alignment marks. An alignment illumination system projects light through the reticle marks and onto the wafer surface. A common light source is the stepper UV light, with filtering to block unwanted wavelengths that could activate the resist. Another source is a laser such as a helium-neon laser. This laser emits light at 633 nm, which will not activate the resist.

Light illuminates the alignment marks and then photo detectors are used to optically detect the reticle and wafer targets. The alignment illumination system can use the main stepper's projection optics to illuminate the marks (known as on-axis or through-the-lens), or use its own alignment

optics (known as off-axis) (see Figure 14.50). There are many different types of alignment methods in use. Laser beams are also used to precisely control the levelness and position of the X-Y stage that holds the wafer chuck. Laser interferometry is used to measure the position of the wafer stage at all times. Once the position data are measured, the data are fed into the system computer with a software interface to the electromechanical system used to make the correct adjustments to align the wafer to the reticle.

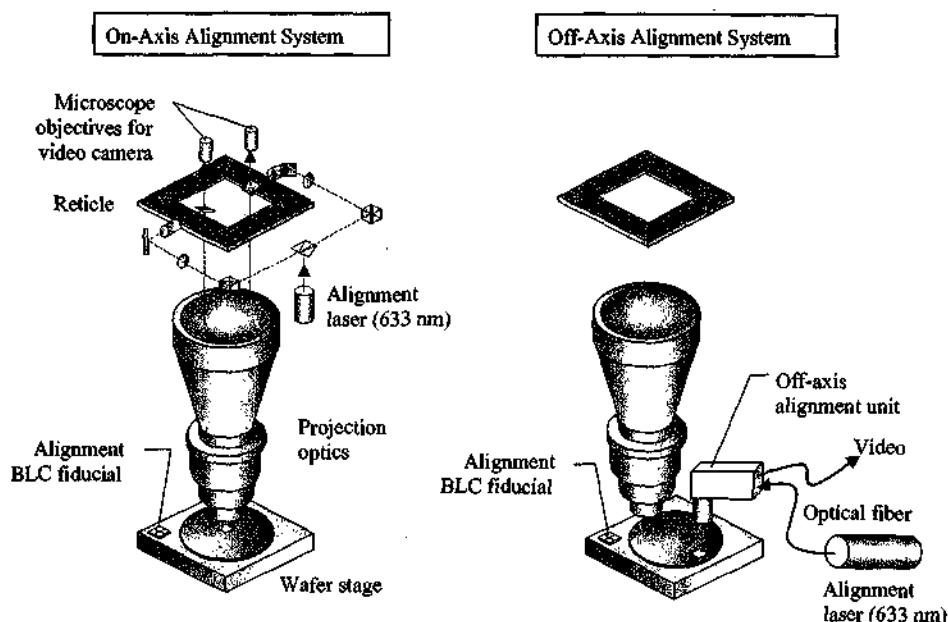


FIGURE 14.50 On-Axis Versus Off-Axis Alignment System
(Used with permission from Canon USA, redrawn after FPA-2000 II schematics)

Step-and-Scan Alignment Marks. An improvement for a step-and-scan system is that multiple alignment marks can be placed in each exposure field, permitting alignment to occur with each field exposure. This means that position parameters such as wafer rotation and X-Y alignment can be adjusted at each scan site, which improves overlay accuracy and reduces the overlay budget.

Alignment Mark Deterioration. The increasing number of film layers on high-performance ICs require greater flatness from one layer to the next. This flatness is achieved by using chemical mechanical planarization (CMP) to planarize the wafer (see Chapter 18). However, CMP also deteriorates alignment marks on the wafer, which makes them more difficult to detect.³³ Stepper and step-and-scan equipment manufacturers have developed alternative approaches to align the wafer. In one case, a laser beam is used to detect a phase change in the beam reflection off the alignment mark relative to a reference signal. Slight phase changes are relatively easy to detect, even with layers that are too flat to measure with the standard light reflection system.

Types of Alignment ■ Each reticle must be aligned to the exposure tool's body when the reticle is mounted on the reticle stage. The reticle alignment marks are illuminated by a laser beam (an example is HeNe) through a fixed referenced mark. Once aligned, the wafer stage is then aligned to the reticle.

The wafer is initially prealigned by holding it in a chuck at the wafer stage and locating the notch or flat. For a *first mask exposure*, the wafer flat or notch is the only alignment feature on the wafer. The blank wafer is “blindly” exposed to the first reticle layer without any further alignment. The only consideration is for ensuring that the lens is properly focused to the wafer prior to each exposure.

For subsequent exposures the wafer and reticle position data are measured with respect to a coordinate system defined for the exposure tool and are then used in a global or field-by-field manner to perform alignment. *Global alignment* (or coarse alignment) uses several marks to align all the exposure fields on the wafer. The goal is to quickly align the entire wafer relative to the reticle.

Field-by-field alignment data (also called fine alignment) is used to align a single field. The stepper steps to each die on the grid, focuses, aligns, and then exposes. This method is the slowest but achieves excellent registration, if necessary, for a critical layer. The specific details of stepper alignment systems are unique to each manufacturer due to the extreme tolerance control required to align the photomask and wafer pattern to within a few nanometers.

Environmental Conditions

The environmental conditions for photolithography steppers and step-and-scan systems are critical for high-yield wafer fabrication. Minor variations in environmental conditions can lead to device defects. Photolithography tools have an enclosed environmental chamber to control conditions such as temperature, humidity, vibration, atmospheric pressure, and particle contamination.

Temperature ■ It is critical to control temperature during photolithography. Temperature can adversely affect all aspects of aligners: mask stage, optical elements, light sources, wafer stage, and alignment systems. Temperatures are typically controlled to within a tenth of a degree Celsius. For this reason, the high-temperature illuminator and heat generating power supplies are generally located away from the main body of the alignment and exposure tool.

Humidity ■ Humidity is carefully controlled inside the fab and within the photolithography environmental chamber. Excessive humidity can affect the resist adhesion to the wafer. Humidity also affects the density of the air, which can adversely affect light passing through it; thus producing adverse effects on interferometer positioning, lens NA, and focusing.

Vibration ■ The occurrence of vibration would create problems for positioning, alignment, focusing, and exposure. Problems could be misregistration, misalignment, defocus, and nonuniform exposure. There are different methods to minimize vibration. In some cases, floors that support the lithography equipment are isolated by large shock absorbers from the other areas of the fab. The photo tools also have their own design features to isolate vibration, such as pneumatic isolation devices and momentum absorption structures.

Atmospheric Pressure ■ Atmospheric pressure changes can affect the refractive index of air in the projection optics and laser interferometers used for stage positioning of steppers. This condition leads to nonuniform CDs and poor registration. For this reason, equipment manufacturers usually include pressure sensors to monitor the atmospheric pressure inside the environmental chamber. The pressure measurement data is fed into a computer for monitoring and control. In some cases, the lens elements are sealed in an airtight housing with a constant air flow and pressure inside the housing.

Particle Contamination ■ Photo tools are designed so that a class 1 or better clean environment is maintained inside the tool chamber. All materials and hardware used in the equipment construction are selected for their low particle generation. Lubricants are avoided. If they are required, then a low vapor pressure lubricant is specified to minimize outgassing. Wafer and reticle handling is done by mechanisms that hold the product from underneath, with automated handling systems used on modern equipment. Air solenoids are vented externally from the tool to avoid scattering particles inside the tool.

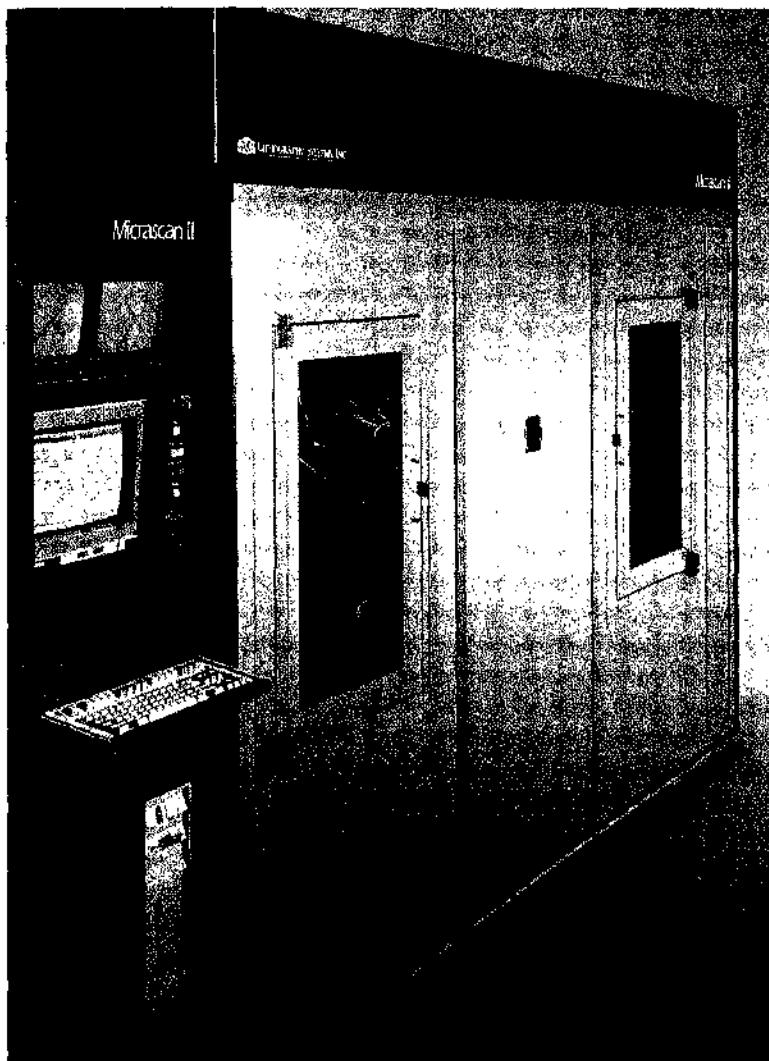
Comparison of Photo Tools

Steppers and step-and-scan systems based on optical lithography are the photolithography work-horses in semiconductor manufacturing for VLSI/ULSI deep submicron feature sizes. A comparison of the various features of steppers and step-and-scan systems is listed in Table 14.7.

TABLE 14.7* Comparison of Photolithography Tools

Model	Wave-length (nm)	Stepper/ Step-and scan	Illumination	Exposure Field Size (mm)	N/A	Resolution (μm)	Overlay Accuracy (nm)
ASML							
PAS5500/22	365	Stepper	Mercury arc	22 × 27.4	0.4	0.70	≤70
PAS5500/60B	365	Stepper	Mercury arc	18 × 24.2	0.54	0.45	≤70
PAS5500/200B	365	Stepper	Mercury arc	22 × 27.4	0.48 to 0.60	0.32	≤50
PAS5500/300B	248	Stepper	KrF laser	22 × 27.4	0.40 to 0.57	0.25	≤45
PAS5500/500	248	Step-and-scan	KrF laser	26 × 34	0.40 to 0.63	0.22	≤45
PAS5500/900	193	Step-and-scan	ArF laser	26 × 33	0.45 to >0.6	0.13 to 0.15	
Canon							
FPA-3000i5	365	Stepper	Mercury arc	22 × 22	0.45 to 0.63	0.35	≤80
FPA-3000EX3	248	Stepper	KrF laser	22 × 22	0.40 to 0.60	0.25	≤80
FPA-4000ES1	248	Step-and-scan	KrF laser	25 × 33	0.40 to 0.63	0.25	≤70
FPA-5000ES2		Step-and-scan		26 × 33	0.68	0.18	
ISI Lithography							
XLS 7500	365	Stepper	Mercury arc	20 × 20	0.55	0.5	≤90
XLS 7800	248	Stepper	KrF laser	22 × 22	0.35 to 0.53	0.35	≤80
Isis 2021	365	Stepper	Mercury arc	22 × 22	0.35 to 0.60	≤0.35	≤60
Isis 302k	248	Stepper	KrF laser	22 × 22	0.35 to 0.55	≤0.25	≤50
ArF MicroStep	193	Stepper	ArF laser	1.5 × 1.5	0.40 to 0.60	≤0.18	
Nikon							
i12D	365	Stepper	Mercury arc	22 × 22	Var. to 0.63	≤0.35	≤55
EX12B	248	Stepper	KrF laser	22 × 22	Var. to 0.55	≤0.28	≤55
S201A	248	Step-and-scan	KrF laser	25 × 33	Var. to 0.60	≤0.25	≤50
SVGL							
MS II+	248	Step-and-scan	KrF laser	22 × 33.5	0.50	0.3	≤70
MS III	248	Step-and-scan	KrF laser	26 × 33.5	0.40 to 0.60	0.25	≤55
MS III+	193	Step-and-scan	ArF laser	26 × 34	0.68	0.18	
Ultratech							
2244I	355 to 375	Stepper		22 × 44	0.32	0.75	≤350
Saturn	355 to 375	Stepper		22 × 44	0.365	0.65	≤250

*Adapted from R. DeJule, "Wafer Stepper Trends," *Semiconductor International* (February 1997), p. 88.



Step-and-Scan System
(Photo courtesy of Silicon Valley Group Lithography Systems)

MIX AND MATCH

Most attention for advancements in wafer fabrication technology is given to leading-edge areas like chemically amplified DUV resists capable of deep submicron CD linewidth control. However, recognize that there can be up to 20 to 30 layers in a high-performance IC. CDs and overlay accuracy are critical quality measures on layers with sub-quarter micron structures. For the subcritical layers, novolak-based DNQ resists with i-line steppers are acceptable and cost-justified.

The total cost of owning and operating equipment for wafer production is referred to as the *cost of ownership*, or *COO*. COO costs include the up-front capital expenditures plus the hidden costs from areas such as maintenance, training, downtime, and reduced process latitude. There is a high COO for operating a high-resolution DUV resist on a subcritical layer where i-line DNQ-novolak resist would suffice. Chip manufacturers have learned it is not cost-effective to use an advanced tool and resist system for patterning subcritical layers.

A *mix-and-match* approach to photolithography matches the photolithography resist and equipment technology to the criticality of the wafer layer. For instance, analysis of a 0.22 μm DRAM device demonstrated that a 0.28 μm i-line stepper with a 70 nm overlay accuracy could image 13 noncritical layers of the total 20 layers on the device. The remaining 7 layers were imaged with a leading-edge DUV step-and-scan system that could achieve the 0.22 μm resolution. In this study, it was estimated that the cost of the i-line product averaged about \$5 per layer, vs. \$7.50 for the DUV layers.³⁴ This is the benefit of the mix and match strategy for photolithography technology—to reduce cost of ownership.

ALIGNMENT AND EXPOSURE QUALITY MEASURES

Typical quality measures for alignment and exposure are provided in Table 14.8.

TABLE 14.8 Key Quality Measures for Exposure

Quality Parameter	Types of Defects	Remarks
1. Focus-exposure dose.	A. Incorrect focus-exposure for system. Conduct focus-exposure optimization test while measuring CD linewidth.	<ul style="list-style-type: none"> Verify uniform and optimum exposure from illumination source. Make CD measurements of line versus a series of exposure dose values for a nominal focal position (e.g., 30% from top surface). With nominal focal position, find optimum dose for producing CD. Modify focal position and conduct CD measurements. Wide range of acceptable dose exposure at optimum focal position. Verify the bulk resist meets all quality parameters.
2. Light intensity of illumination source.	A. Non-uniform light intensity in the exposure field.	<ul style="list-style-type: none"> Check light intensity for specified energy and uniformity at several locations on a wafer. Most aligners have built-in photodetector (measures in mW/cm^2). Evaluate resist to ensure it is not outgassing and condensing on optical elements. This will degrade lens transmission and field uniformity (important for DUV resists).*
3. Reticle alignment in stepper or step-and-scan tool.	A. Reticle alignment targets will not align properly with wafer alignment targets.	<ul style="list-style-type: none"> Verify appropriate process recipe is loaded for the specific mask layer. Verify appropriate wafers and reticle are loaded for a specific job. Check rotation of reticle on reticle stage or wafer on wafer stage due to vacuum leak at chuck or electro-mechanical problem with stage. Problem with internal optics of aligner. Possible cause is temperature or pressure change that affects the NA of the lens.
4. Pattern resolution.	A. Poor resolution for CDs on wafer: linewidths and holes do not meet specification.	<p>Resolution is often a focus problem:</p> <ul style="list-style-type: none"> Perform focus-exposure test. Check environment (temperature, pressure). Wafer is not flat on chuck, possibly due to backside contamination or chuck problem. Look for possible incoming process-related problems or improper process parameters or reticle. Look for optics problems (e.g., lens aberrations).

Quality Parameter	Types of Defects	Remarks
5. Reticle quality.	<p>The following are reticle defects:</p> <ul style="list-style-type: none"> A. Dirt or scratches on reticle. B. Pattern defects on reticles: <ul style="list-style-type: none"> • Break in line. • Bridge between features. • Missing geometry. • Opaque spot from isolated chrome. • Pinhole in chrome line. C. Glass fracture. D. Lifted chrome (poor adhesion). E. Reticle plate flatness. 	<ul style="list-style-type: none"> • Scratches can remove chrome and cause defect in resist. • A break in line extends completely across the chrome feature. • Bridge will join two chrome features across a clear space on the reticle. • Missing geometry is a pattern not on reticle, such as a missing contact. • Opaque spot is an area of chrome that should not be on reticle. • Pinhole is a hole in a chrome pattern. • Reticles should not have fractures and be controlled for flatness and warpage.

*O. Nalamasu et al., "Single-Layer Resist Design for 193 nm Lithography," *Solid State Technology* (May 1999), p. 29.

ALIGNMENT AND EXPOSURE TROUBLESHOOTING

Troubleshooting problems encountered during alignment and exposure are shown in Table 14.9.

TABLE 14.9 Common Alignment and Exposure Troubleshooting Problems

Problem	Probable Cause	Corrective Action
1. Excessive overlay error.*	<p>A. Incorrect alignment system measurement of reticle and wafer alignment marks</p>	<p>Possible measurement error sources are:</p> <ul style="list-style-type: none"> • Verify correct process recipe and reticle is used for a specific mask layer. • Verify the calibration and stability of the alignment system to determine the position of the alignment marks. • Verify calibration of alignment mark-to-pattern relationship, including thermal and/or mechanical effects. • If error is within one tool, then optical distortion is probably not source of problem. If error is tool-to-tool, check difference in optical distortion of two different projection optics.
	<p>B. Problem with reticle.</p>	<ul style="list-style-type: none"> • Verify there is no problem with reticle mounting and/or reticle heating that could change alignment mark-to-pattern position. • Check there is no particulate contamination on alignment marks that makes the alignment system incorrectly determine the mark position.

Quality Parameter	Types of Defects	Remarks
	C. Error in wafer or reticle stage that holds and positions wafer and reticle.	<ul style="list-style-type: none"> • Wafer and/or reticle stage has errors in position and rotation during exposure that contributes to overlay errors. • Excessive vibration of wafer or reticle stage. Tools have built-in vibration and shock isolation. • Unacceptable wafer or reticle stage heating that distorts wafer or reticle. • Chucking errors that vacuum clamp wafers differently and cause wafer distortion.
	D. Problem with projection optics.	<ul style="list-style-type: none"> • Calibration error in the lens magnification adjustment causes pattern mismatches. • Focus errors and/or unacceptable field flatness that causes image distortion or shifts. • Unacceptable heating causes optics distortion.
2. Drift in KrF laser parameters. ^{**}	<p>A. Laser properties that can change are:</p> <ul style="list-style-type: none"> • Laser spectral bandwidth and energy distribution. • Wavelength stability. • Output energy and repetition rate. • Pulse-to-pulse energy stability. 	<p>All laser measurements and calibrations should be performed after specialized training by the supplier, including:</p> <ul style="list-style-type: none"> • Use special wavemeter to measure wavelength. Drift in wavelength affects focus at the wafer plane. • Assess background optical noise levels and electronic offsets to determine impact on energy distribution. • Measure bandwidth of laser output using procedure specified by supplier.

*G. Gallatin, "Alignment and overlay," *Microlithography, Micromachining and Microfabrication* vol. 1, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 318.

**P. Das and U. Sengupta, "Krypton Fluoride Excimer Laser for Advanced Microlithography," *Microlithography Science and Technology* ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 299.

SUMMARY

Modern photolithography is based on optical lithography. Alignment and exposure are critical in lithography to meet critical submicron resolution requirements. There have been five photolithography equipment eras. The step-and-repeat and step-and-scan tools are the preferred optical lithography tools used today for advanced IC fabrication. The continued success of optical lithography is based on improvements to all subsystems. Light is important to project the pattern on the resist. UV light at certain energy peaks creates the photochemical reaction in the resist. UV exposure sources are the mercury arc lamp for conventional lithography and the excimer laser for deep UV. Reflection and refraction are important parameters for manipulating the light through lens systems for image projection. Diffraction describes how the light bends as it passes through the narrow patterns on the reticle. The ability of a lens to capture light is the numerical

aperture (NA). Special antireflective coatings are placed on the wafer surface to diffuse reflected light and avoid damage to the light-sensitive resist. Resolution is the smallest feature printable on the wafer and is limited by wavelength, NA, and process factors. Reducing pattern resolution also leads to reduced depth of focus, which requires a planar wafer surface. This is achieved by chemical mechanical planarization (CMP). Quartz reticles contain the pattern to be imaged, using a 4X or 5X reduction optics to simplify reticle fabrication with e-beam lithography. Optical enhancement techniques such as phase-shift mask and optical proximity correction permit subwavelength lithography. Alignment of the reticle to the wafer is critical for meeting overlay budget requirements. Environmental conditions must be tightly controlled for optimum lithography. A mix-and-match approach uses conventional lithography for noncritical layers and DUV lithography for critical layers.

KEY TERMS

optical lithography
 light
 optical filters
 electromagnetic spectrum
 mercury arc lamp
 light intensity
 excimer laser
 spatial coherence
 dose monitor
 optics
 law of reflection
 refraction
 relative index of refraction, n
 absolute index of refraction
 lens
 focal point
 focal length
 lens compaction
 aberrations
 diffraction
 numerical aperture (NA)
 reflective notching
 standing waves
 antireflective coating (ARC)
 bottom antireflective coating (BARC)
 top antireflective coating (TARC)
 resolution
 depth of focus (DOF)

depth of field
 center of focus (COF)
 contact aligner
 proximity aligner
 scanning projection aligner (scanner)
 step-and-repeat aligner (stepper)
 step-and-scan system
 reticle
 photomask or mask
 electron beam (e-beam)
 pellicle
 subwavelength lithography
 phase-shift mask (PSM)
 optical proximity correction (OPC)
 off-axis illumination (OAI)
 print bias
 baseline compensation (BLC)
 alignment
 registration
 overlay accuracy
 overlay budget
 grid
 alignment marks
 first mask exposure
 global alignment
 field-by-field alignment
 cost of ownership (COO)
 mix and match

REVIEW QUESTIONS

1. Describe optical lithography.
2. What are the three basic purposes of a stepper?
3. Explain why design ground rules are important for alignment.
4. What is the alignment challenge for photolithography?
5. Describe the relationship between light exposure wavelength and image resolution.
6. Define light. Why is it needed in optical lithography?
7. List and explain the two types of light wave interference. What is an optical filter?
8. What is the electromagnetic spectrum, and what is the UV range?
9. List and describe the two UV exposure sources used in optical lithography.
10. What is light intensity and exposure dose, and how are they related? Why is exposure dose important?
11. What happens to resist sidewalls if there is excessive resist light absorbance?
12. Describe how an excimer laser functions.
13. What excimer laser is used as a 248 nm light source? As a 193 nm light source?
14. What is spatial coherence? Why is this controlled in photolithography?
15. What is a typical dose exposure latitude for a DUV resist? How is this measured?
16. What is optics, and why is it important in lithography?
17. State the law of reflection and give an example.
18. What is refraction? Explain the relative index of refraction and the absolute index of refraction.
19. What is a lens? Explain the focal point and focal length.
20. What lens material is used at the DUV exposure wavelength? Why?

21. Explain how lens compaction occurs and what problem it creates.
22. Describe lens aberration.
23. What is diffraction? Why is it a concern in optical lithography?
24. What is numerical aperture (NA)? State its formula, including the approximate formula.
25. What happens to NA if the radius of the lens is increased?
26. List and explain the two primary problems of light reflection from the wafer surface.
27. What is an antireflective coating, and how does it reduce standing waves?
28. State the formula for resolution. What three lithography parameters affect resolution?
29. What happens to resolution if the light wavelength decreases? If NA goes up?
30. Calculate the resolution of a scanner if $\lambda = 248$ nm, NA = 0.65, and k = 0.6.
31. Define depth of focus and center of focus. Write the equation to calculate DOF.
32. What happens to depth of focus as resolution increases?
33. List the two important parameters for image quality in optical lithography.
34. Why is surface planarity important for pattern resolution?
35. Explain the contact aligner. Does it use a mask or reticle?
36. Explain how a proximity aligner works. What problem did it attempt to resolve?
37. Explain how a scanning projection aligner works. What problems did the scanning projection aligner try to resolve?
38. Explain the basic functions of a step-and-repeat stepper.
39. What benefits are gained from step-and-scan technology for lithography?
40. Define a reticle. What is the difference between a reticle and a photomask? What reticle reduction is usually used in a stepper? What reduction is used in a step-and-scan system?
41. What material is used to make a reticle? What opaque material is patterned on a reticle?
42. What is a typical usable field size for a 4:1 reticle?
43. What lithography technology is used to pattern reticles? Why?
44. Explain why a pellicle is used on reticles.
45. Explain phase-shift mask (PSM).
46. Discuss optical proximity correction (OPC).
47. How does off-axis illumination increase the pattern resolution?
48. What is print bias and what is one method to correct it?
49. Explain baseline compensation for alignment.
50. What is alignment? What is registration?
51. Define overlay accuracy. What is the overlay budget? About how much of the critical dimension is the overlay budget?
52. What is a grid?
53. What are alignment marks? Describe RA, GA, and FA alignment marks.
54. How does a step-and-scan system improve alignment?
55. Explain global alignment and field-by-field alignment.
56. List and discuss five environment conditions that must be controlled in steppers and step-and-scan systems.
57. Explain the mix-and-match approach to photolithography. Why is it beneficial?

PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES

ASML
 Canon Semiconductor
 Charles Evans and Associates
 Cymer Inc.
 Dupont Semiconductor Products
 ETEC Systems Inc.
 KLA-Tencor
 Karl Suss
 Nikon
 SEMI
 Silicon Valley Group
 SPIE, International Society of Optical Engineering
 Ultratech Stepper
 USHIO

<http://www.asml.com/>
<http://www.usa.canon.com/indtech/semicondeq/>
<http://www.cea.com>
<http://www.cymer.com/>
<http://www.dupont.com/semiconductor/lith.html>
http://www.etec.com/semiprod_frame.html
http://www.kla-tencor.com/product/photo_frame.html
<http://www.suss.com/>
<http://www.nikon.com/>
<http://www.semi.org/>
<http://www.svg.com/>
<http://www.spie.org/>
<http://www.ultratechstepper.com/>
<http://www.ushio.com/index2.html>

REFERENCES

1. H. Levison and W. Arnold, "Optical Lithography," *Microlithography, Micromachining and Microfabrication* vol. I, ed. P. Rai-Choudhury (Bellingham, WA: SPIE, 1997): p. 13.
2. K. Derbyshire, "Issues in Advanced Lithography," *Solid State Technology* (May 1997): p. 133.
3. M. Hibbs, "System Overview of Optical Steppers and Scanners," *Microlithography Science and Technology*, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 16.
4. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies* (Boston: Kluwer Academic Publishers, 1995), p. 212.
5. M. Hibbs, "System Overview of Optical Steppers and Scanners," p. 18.
6. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 174.
7. S. Campbell, *The Science and Engineering of Microelectronic Fabrication* (New York: Oxford University Press, 1996), p. 162.
8. R. DeJule, "Extending Optical Lithography to 157 nm?" *Semiconductor International*, (February 1997): p. 84.
9. M. Hibbs, "System Overview of Optical Steppers and Scanners," p. 20.
10. J. Shamaly, "A Full-Field ArF Exposure Tool for 180-nm Lithography," *Microlithography World* (Summer 1996): p. 23.
11. M. Hibbs, "System Overview of Optical Steppers and Scanners," p. 20.
12. C. Cromer et al., "Improved Dose Metrology in Optical Lithography," *Solid State Technology* (April 1996): p. 76.
13. E. Hecht, *Optics*, 2nd ed. (Reading, MA: Addison-Wesley Publishing Co., 1990). p. 143.
14. J. Shamaly, "Full-Field ArF Exposure Tool," p. 23.
15. C. Mack, "Optical Lithography Modeling," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 118.
16. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 206.
17. C. Bencher et al., "Dielectric Antireflective Coatings for DUV Lithography," *Solid State Technology* (March 1997): p. 109.
18. Ibid.
19. B. Smith and M. Hanratty, "Optical Lithography Modeling," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 587.
20. W. Arnold, "Is a Scanner Better than a Stepper?" *Solid State Technology* (March 1999): p. 77.
21. M. Hibbs, "System Overview of Optical Steppers and Scanners," p. 9.
22. W. Arnold, "Is a Scanner Better than a Stepper?" p. 77.
23. S. Wolf and R. Tauber, *Process Technology*, vol. 1, *Silicon Processing for the VLSI Era* (Sunset Beach: Lattice Press, 1986), p. 476.
24. R. Singh, S. Vu, and J. Souze, "Nine-Inch Reticles: An Analysis," *Solid State Technology* (October 1998): p. 83.
25. R. DeJule, "Resists for Next-Generation Masks," *Lithography Technology News, Semiconductor International* (October 1998): p. 46.
26. "Rapid Innovation in Spite of Slow Market," *Solid State Technology* (September 1998): p. 42.
27. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, p. 176.
28. M. Sasago, "Aggressive Optical Lithography," *Microlithography World* (Autumn 1998): p. 5.
29. R. DeJule, "Resists for Next-Generation Masks," p. 46.
30. H. Levison and W. Arnold, "Optical Lithography," *Microlithography, Micromachining and Microfabrication* vol. I, ed. P. Rai-Choudhury (Bellingham, WA: SPIE, 1997), p. 69.
31. G. Gallatin, "Alignment and Overlay," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 318.
32. R. DeJule, "Lithography," *Semiconductor International* (September 1998): p. 50.
33. R. DeJule, "Wafer Stepper Trends," p. 88.
34. R. DeJule, "More Productivity at Subcritical Layers," *Semiconductor International* (April 1998): p. 50.

PHOTOLITHOGRAPHY: PHOTORESIST DEVELOPMENT AND ADVANCED LITHOGRAPHY

At this stage of the photolithography process, wafers have been coated with resist and exposed to UV light to produce a photochemical reaction in specified areas of the resist. The resist contains a precise coded image of the reticle pattern and is ready for the development step of photolithography. Photoresist development is where the three-dimensional pattern is physically produced in the resist. This step determines whether the resist image is an accurate replication of the reticle pattern. In deep submicron wafer fabrication, success at this step is critical for subsequent processing.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain why and how a post-exposure bake is done for conventional and chemically amplified DUV resist.
2. Describe the negative and positive resist development process for conventional and chemically amplified DUV resist.
3. List and discuss the two most common resist development methods and the critical development parameters.
4. State why a hard bake is done after resist development.
5. Explain the benefits of a post-develop inspection.
6. List and describe the four different alternatives for advanced lithography, including the challenges for introducing each alternative into production.
7. Describe and give the benefit for the advanced resist process of top surface imaging.

The advanced lithography section summarizes the next generation lithography tools. Fundamental research into alternative lithography tools is necessary so that the semiconductor industry will be ready to replace optical lithography when it has reached its resolution limit. Optical lithography appears to be the optimum patterning process for at least the next several technology nodes below 0.18 micron.

INTRODUCTION

Following the sequential steps of the eight-step photolithography process, wafers have completed alignment and exposure. The reticle pattern has been aligned to each die on individual wafers and exposed to UV light for image transfer to the resist. Wafers with exposed resist leave the stepper or step-and-scan system and return to the automated track system. The next processing step is often a post-exposure bake and then wafers are ready for development. Photoresist development is the chemical dissolution of the soluble resist in the developer chemical. Development removes those areas of the resist rendered soluble by the exposure step. The insoluble resist remaining on the wafer should be an exact copy of the pattern on the reticle (for a positive-tone resist). Photochemistry renders the exposed regions soluble for positive-tone resists and insoluble for negative resists.

The goal of resist development is to accurately replicate the reticle pattern in the resist while maintaining acceptable resist adhesion. The ability to develop resist with deep submicron geometries is essential for advanced ICs. Resist development and inspection is the intermediate pattern transfer step before the wafer moves into another workbay to undergo etch or ion implantation. An improperly developed resist will cause a high reject rate, whereas a properly developed resist is the foundation for a high-yield process. The quality of the resist pattern determines the success of the subsequent process steps.

In early days of wafer fabrication, resist development was a stand-alone process step, with its own tool and workstation. Cassettes of wafers were manually moved from the exposure tool to the develop operation. The manual processing and lack of tool controls produced excessive variability that is unacceptable for submicron lithography. In modern wafer fabrication automated wafer tracks have integrated the development operation into a comprehensive process of photolithography. This represents steps one through seven of the eight-step photolithography process.

The eight steps to photolithography are in Table 15.1. The highlighted steps are covered in this chapter.

TABLE 15.1 Eight Basic Steps of Photolithography

Step	Chapter Covered
1. Vapor prime	13
2. Spin coat	13
3. Soft bake	13
4. Alignment and exposure	14
5. Post-exposure bake	15
6. Develop	15
7. Hard bake	15
8. Develop inspect	15

The procedure used to develop resist depends on whether it is positive-tone or negative-tone, and whether it is a conventional i-line resist (DNQ-novolak) or chemically amplified deep UV (CA DUV) resist. Positive resists are the norm for submicron lithography. Noncritical layers typically use conventional i-line resist (making up the major share of photolithography processing), while critical layers with CDs of 0.25 μm and below use DUV resists.

Advanced Lithography

Optical lithography has thrived while contributing to the continual reduction of device feature sizes. It is now expected by some industry experts that optical lithography will continue as the main lithography process until after the arrival of the 0.1 μm technology node sometime around 2006.¹ The reason optical lithography has continued to succeed in wafer fabrication is the ongoing introduction of improvements in equipment, process, and design. We reviewed new exposure techniques in Chapter 14, such as optical proximity correction (OPC) and phase-shift masks (PSM). At the end of this chapter we will review next generation microlithography technologies.

POST-EXPOSURE BAKE

After the wafer with exposed resist exits the exposure system, it enters the wafer track system and undergoes a short *post-exposure bake (PEB)* step. A thermal PEB is necessary for chemically amplified DUV resists to catalyze (initiate) critical resist chemical reactions. For conventional i-line resists based on DNQ chemistry, PEB is done to improve adhesion and reduce standing waves. Resist manufacturers include recommended time and temperature specifications for PEB in their product literature.

DUV Post-Exposure Bake (PEB)

It is during the post-exposure bake (PEB) that the exposed regions of a chemically amplified DUV resist become soluble in the developer. Recall that for a chemically amplified DUV resist, a protecting chemical (e.g., t-BOC) makes the resist insoluble in the developer. During UV exposure, a photoacid generator (PAG) generates an acid in the exposed regions. To make the exposed resist soluble to the developer, the post-exposure bake (PEB) heats the resist, which causes the acid-catalyzed deprotection reaction. The acid removes the protecting group from the resin and the exposed resist is now soluble in the developer solution (see Appendix E).

PEB is an important step in resist processing for chemically amplified DUV resists. The requirement for PEB to catalyze the chemical deprotection reaction must be considered nearly as critical as exposure in terms of control, uniformity, and latitude requirements.²

Temperature Uniformity ■ The temperature uniformity and duration of PEB for DUV resist is important. Excessive variations will affect the kinetics of the acid catalytic reaction in the resist. Most often, PEB is done on a hot plate in an automated wafer track system. The actual temperature and time that the coated wafer is heated by the hot plate depends on the type of resist used. A representative PEB temperature and time is 90 to 130°C for 1 to 2 minutes. The PEB temperature is usually 10 to 15°C hotter than the soft bake. The hot plate temperature variation and the bake latitude of the particular DUV resist is critical and can affect the amount of CD variation in the resist during the develop step. For commercially available DUV resist systems, a representative post-exposure bake latitude for CD variation is about 5 nm /°C.³ To reduce CD inconsistencies, hot plates are set at 130°C and controlled within ± 0.1°C.⁴

PEB Delay ■ Early DUV resists were very sensitive to the delay time from UV exposure to PEB. A delay greater than a few minutes permitted the acid to neutralize due to amine contamination from the ambient air. The neutralization occurred on the top surface of the resist, creating a thin, less-soluble inhibition layer that led to the formation of a “T-top” after development (see Figure 15.1). More recent DUV resists permit a delay of up to about 30 minutes. However, it is still desirable to develop the resist as soon as possible after exposure to minimize any potential chemical reactions that can inhibit the development process.

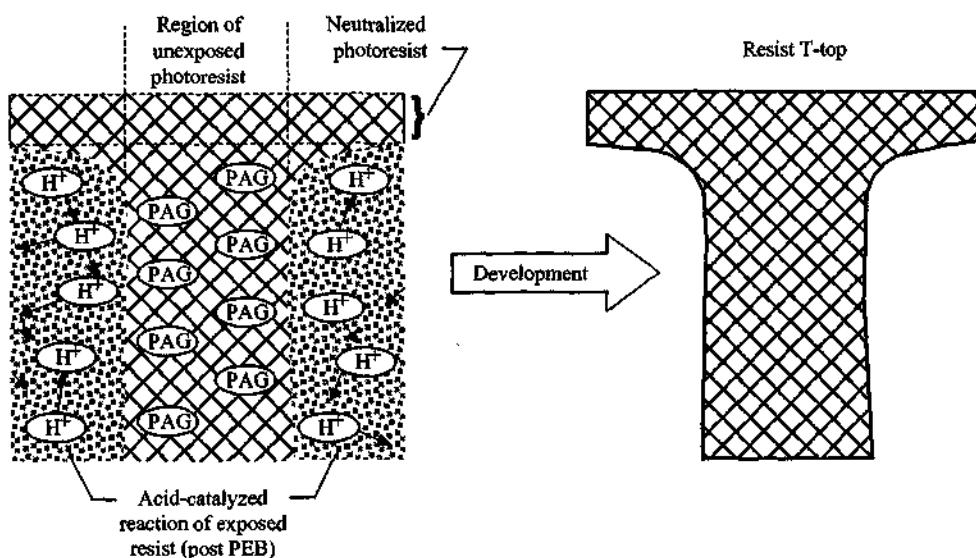


FIGURE 15.1 Amine Contamination of DUV Resist Leading to T-top Formation

Conventional I-Line PEB

It is common practice to perform a PEB for conventional DNQ-novolak i-line resist. The PEB extends some of the same benefits gained by the soft-bake step that occurred before exposure. PEB helps drive out remaining solvents that are left in the resist, reducing solvent content from 7 to 4% (before exposure) to about 5 to 2%.⁵ The most significant benefit from PEB for i-line resist is a reduction in the standing wave defect from reflective surfaces during exposure.

Recall from Chapter 14 that standing waves result from light exposure of a resist that is coated on a reflective substrate. Coherent light reflects off the substrate and interference produces a nonuniform intensity in the resist film that appears as standing waves in the sidewall. The PEB minimizes the standing wave effect because the increased temperature causes the PAC sensitizer in the resist to diffuse through the novolak polymer matrix, essentially producing an averaging effect across the standing wave boundary (see Figure 15.2).

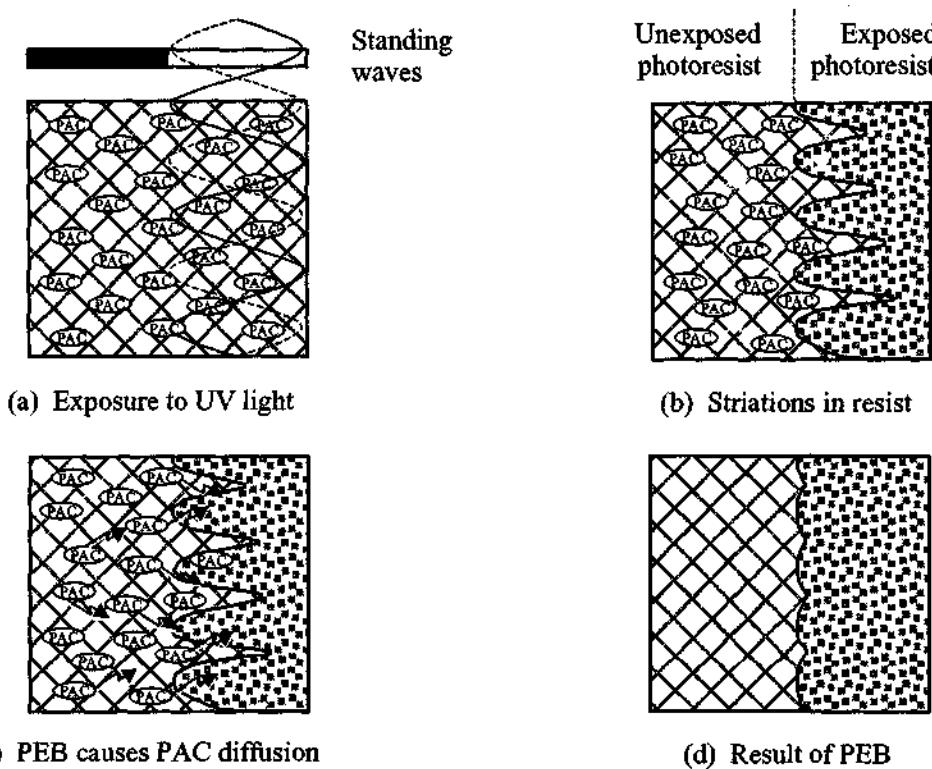


FIGURE 15.2 Reduction of Standing Wave Effect Due to PEB

DEVELOP

Photoresist development (develop) uses a liquid chemical developer to dissolve the soluble regions of the resist that were formed during the mask exposure. The primary goal of resist development is to accurately replicate the reticle pattern in the resist material. The emphasis is on producing CD features that meet the required specifications. If the CDs meet the specifications, then all other features are assumed acceptable since the CD is the most difficult structure to develop.

Resist patterning problems occur if the development process is not properly controlled (see Figure 15.3 on page 417). These resist problems can negatively affect production yield, showing up as defects in the subsequent etch process. The three primary types of development problems are underdevelopment, incomplete development, and overdevelopment. The photoresist lines in Figure 15.3 illustrate these types of problems as compared to a normal developed line. An underdeveloped line appears wider than a normal line and will have sloping sidewalls. The incomplete developed line has residual resist on the substrate that should have been removed during develop. Overdevelopment removes too much resist causing features to appear narrower and poorly-defined.

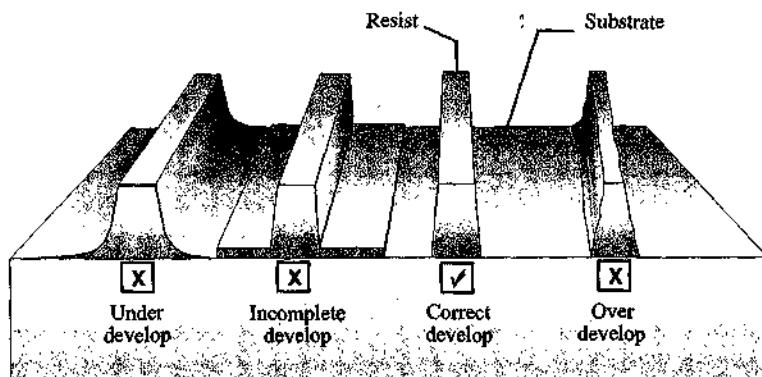


FIGURE 15.3 Photoresist Development Problems

Negative Resist

Negative resist is crosslinked (hardened) by exposure to UV light. This makes the exposed resist nonsoluble in the developer solution (see Figure 15.4).

Very little chemical reaction is necessary for *negative resist development* in the developer solution. This process consists mainly of a solvent wash of the unexposed resist, which is not crosslinked and therefore soft and soluble. The developer is typically an organic solvent such as xylene that is sprayed on the resist while the wafer is spinning on a vacuum chuck. Developer spray may be followed by another organic solvent sprayed on the wafer to stop the develop process.

An organic solvent rinse cycle is then introduced to clean the wafer and remove partially crosslinked pieces of resist. In addition, rinse removes any remaining developer to ensure the development process stops. Rinse typically consists of an organic solvent such as n-butylacetate or an alcohol or trichloroethylene. There is finally a spin-dry step to dry the wafer.

A significant problem with negative resist is swelling and distortion of the crosslinked exposed resist due to absorption of the developer solution during the wash. Because of this, the side-walls of remaining negative resist on the wafer become jagged and swollen. This absorption is a primary reason why traditional negative resists are not suitable for geometries below about 2 μm . There is ongoing research into improving negative resists to minimize swelling.

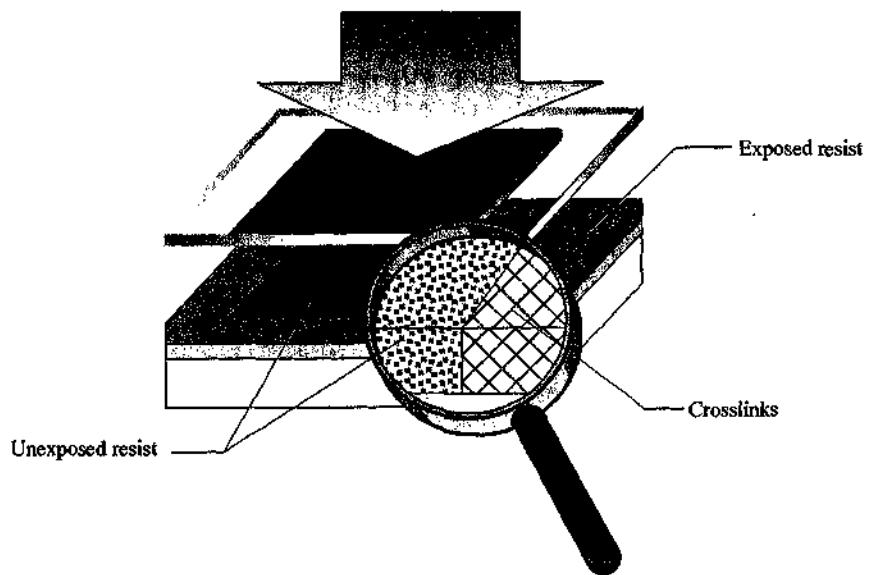


FIGURE 15.4 Negative Resist Crosslinking

Positive Resist

Positive-tone resists are the most common resists used in submicron wafer fabrication due to improved linewidth resolution. The two general types of positive resists used in wafer microlithography are conventional DNQ i-line resists (noncritical layers down to 0.35- μm linewidth) and chemically amplified (CA) DUV resists (critical layers with 0.25- μm linewidth and below). Both of these types of resists use a phenolic-based resin. Conventional i-line resists consist of a novolak resin, while CA DUV resists commonly use a PHS resin. The type of resin is important for resist development because this is the material that the developer will remove in order to pattern the resist. Although these two resists are very different chemically (see Chapter 13), it is generally true that phenolic resin is soluble in a base solution.

Developer ■ *Positive resist development* involves a chemical reaction between the developing solution and the resist to dissolve the exposed resist (see Figure 15.5). The rate at which a developer dissolves the resist is termed the *dissolution rate* (also referred to as the speed of the developer). A fast dissolution rate is desirable for productivity, but too fast a rate can also be bad for resist performance. Developers also have selectivity. High *developer selectivity* means the developer reacts quickly with the exposed resist (fast removal rate) relative to the slow reaction with the unexposed resist (slow removal rate). A developer with high selectivity produces sharper and cleaner resist sidewalls, which is desirable for high-density patterning.

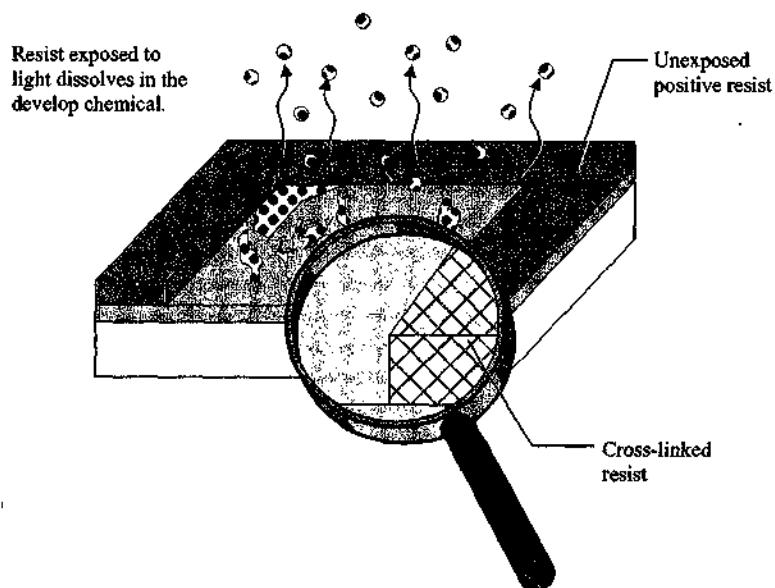


FIGURE 15.5 Development of Positive Resist

The chemical reaction of developer with positive resists is significantly different than the solvent wash needed to dissolve negative-tone resists. Positive resist developer solution is based on a strong alkaline diluted with water (aqueous) developer. Early developer solutions were an alkaline-water mixture of sodium hydroxide or potassium hydroxide. However, both of these solutions contain mobile ionic contaminants (MICs), which are unacceptable for processing high-performance IC circuits sensitive to corrosion. The most common developer today for positive-tone resists is tetramethyl-ammonium hydroxide (*TMAH*). This developer has a very low concentration of metal ions, particularly sodium ions, to avoid introducing MICs on the wafer surface. An advantage of aqueous developers is that the rinse only requires DI water.

Standardized TMAH developer formulations are common in the industry. These developers are more economical and they improve developer repeatability for better quality control. TMAH developer concentrations in the range of 0.2 to 0.3 normality (N) allow sufficient resist selectivity with high contrast between exposed and unexposed resist regions. *Normality* refers to the exact chemical makeup of a particular developer (or any chemical solution); it has a direct effect on developer

dissolution rates. A 0.26 N TMAH solution is becoming a standard for resist processing.⁶ The objective is to have a high selectivity toward removing the exposed portions of the resist while leaving the unexposed resist intact on the wafer.

Exposed positive resist dissolves into the TMAH developer. The unexposed resist does not dissolve into nor absorb any of the developer. For this reason positive unexposed resist does not swell as with the negative resist. The lack of swelling makes positive resist desirable for producing submicron geometries.

TMAH requires careful control of the developer strength as measured with pH. It is also important to control the developer solution temperature and DI water rinse temperature. The dissolution rate varies as a function of developer temperature, with a result of faster resist dissolution occurring at lower developer temperatures.⁷ Small ppm levels of surfactants are often added to the TMAH developer. A surfactant reduces surfaces tension and increases wetting, which makes a liquid flow across the surface better. This condition improves the dissolution of small geometry features such as contacts. Surfactants also minimize any residual film left on the resist surface.

TMAH developer chemistry for positive resist is highly caustic, meaning it destroys or erodes by chemical action. Alkaline chemicals erode metal as well as human tissue. Proper safety precautions (e.g., chemical protective clothing and eyewear) must be used by technicians when working around develop stations, such as on the automated wafer tracks.

Conventional I-Line Resists. Recall from Chapter 13 that positive i-line resists create carboxylic acid during UV exposure. When TMAH developer is applied to exposed resist, a chemical reaction occurs that neutralizes the carboxylic acid by the alkaline TMAH developer. The neutralized carboxylic acid of the exposed resist rapidly goes into solution. The unexposed (hard) resist does not react with the developer and remains intact on the substrate surface.

Chemically Amplified DUV Resists. To review, chemically amplified (CA) DUV resists often consist of phenolic resin in the form of PHS (poly[hydroxystyrene]). PHS has a protection group (such as t-BOC) that makes the PHS insoluble in the base developer. The PAG (photoacid generator) in the CA DUV generates an acid during exposure. During post-exposure bake, this acid removes (or “cleaves”) the protection group away from PHS in exposed regions. The PHS is now highly soluble in the base developer.

The developer does not actually react with the PHS during the development operation. This is because there are hydroxyl OH groups that exist in a “corkscrew” configuration along the PHS polymer chain. These OH groups create effective diffusion paths and bring about extremely high dissolution rates for the PHS in the basic developer solution.⁸

Development Methods

Static immersion of a cassette of wafers into developer solution was an early method for resist development. However, it is no longer used today in high-volume production. Batch production is not suitable for automated wafer tracks that efficiently transport single wafers simultaneously between photolithography operations. Static immersion consumes a large quantity of developer chemicals. Furthermore, batch immersion is not conducive to producing uniform development for high-density ICs across large diameter wafers.

The two most common development techniques used to remove exposed resist on spin-coated wafers today are:

- ◆ Continuous spray development
- ◆ Puddle development

Continuous Spray Development ■ The dissolution of exposed resist with a *continuous spray develop* tool and solution is a similar process to resist dispense systems. In the past, batch development was done using spray systems similar to a spin-rinse-dryer. Development can also be done in a wafer track system after the wafer has completed post-exposure bake. A single wafer is positioned on a vacuum chuck and spun at a slow speed (e.g., 100 to 500 RPM) while one or more nozzles dispense developer on the resist-coated wafer surface (see Figure 15.6). The developer is dispensed in a fine mist, with some processes using ultrasonic atomization to allow for low-velocity dispersion. A low-velocity exit minimizes adiabatic (constant heat transfer) cooling effects during dispense, where the

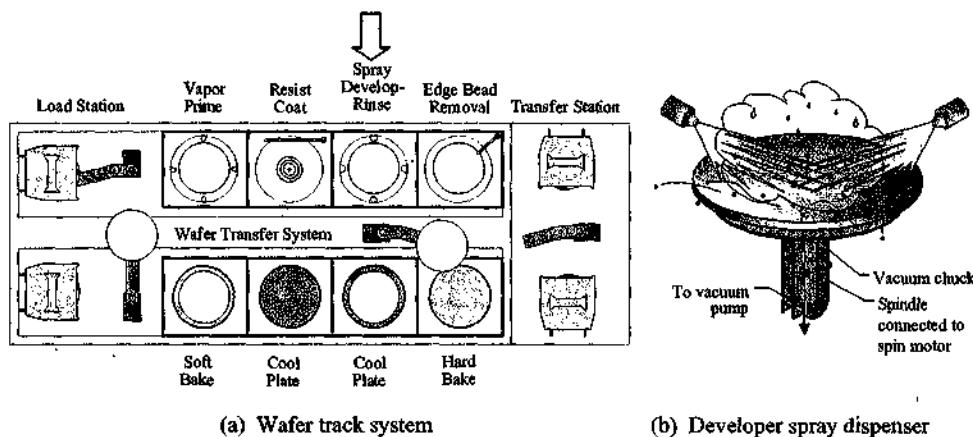


FIGURE 15.6 Resist Development with Continuous Spray

temperature of the developer drops due to its expansion from a high pressure region to a low pressure region. Early nozzle designs required a heating system for the developer to minimize the cooling effect.

The nozzle spray pattern and speed of the wafer rotation are critical variables to achieve repeatability in the resist dissolution rate and uniformity across the wafer. In recent years the spray process method has been largely replaced by puddle development because the latter method allows more process latitude for these variables.

Puddle Development ■ The *puddle develop* approach uses the same basic equipment as the spray method to develop the resist. A small amount of developer is dispensed onto the wafer and forms a puddle (see Figure 15.7). Adequate developer volume is needed to form a puddle meniscus over the entire wafer. Excessive developer is avoided to minimize backside wafer wetting. The wafer can be stationary or slowly rotating on a heated chuck.

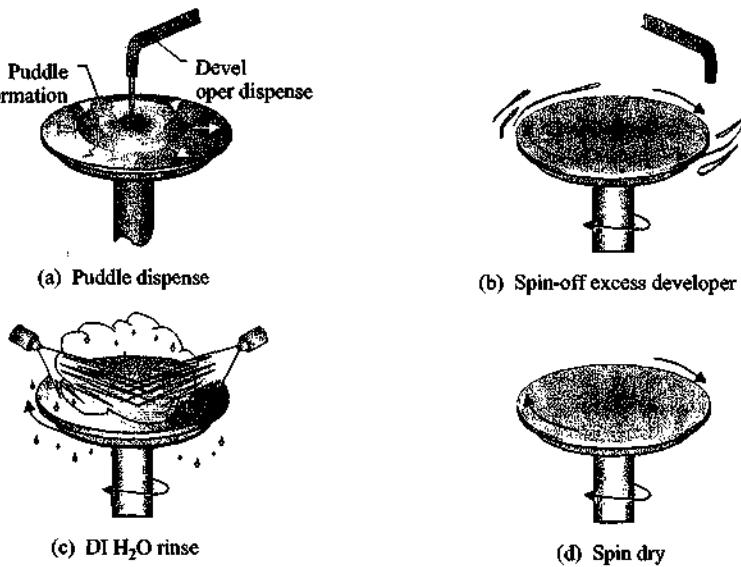


FIGURE 15.7 Puddle Resist Development

There are variations as to whether the wafer is static or rotating after the initial developer is puddled on the wafer. In all cases, the developer must be left on the resist for sufficient time to allow the soluble resist areas to become completely dissolved. As an example, a multiple-puddle method is used where the first puddle is left on the wafer for a predetermined time (such as 10 to 30 seconds, depending on the type of developer). It is then spun off and a new puddle is dispensed and

left on the wafer for a defined time. This second puddle replenishes the developer chemicals and rejuvenates the chemical reaction between the developer and the resist. Some manufacturers also spray the developer onto the wafer during the second puddle application.

After the resist has dissolved in developer, the wafer is rinsed with DI water and spun dry. It is important to rinse all remaining chemicals after development from both the topside and backside of the wafer.

For optimal performance, the flow of developer is kept low to reduce variations in development rate at the edge of the wafer. This reduction is an advantage of puddle development over spray because a minimal amount of developer chemical is used. At the same time, adequate developer volume is needed to achieve full and uniform coverage over the entire wafer. Puddle development introduces fresh chemicals for every wafer, which improves wafer-to-wafer uniformity. The puddle development method minimizes temperature gradients and permits control of the variables affecting uniformity of single wafer resist development.

Resist Development Parameters

Critical parameters must be controlled during the development process. These parameters are:

- ◆ Developer temperature
- ◆ Normality
- ◆ Developer time
- ◆ Rinse
- ◆ Developer volume
- ◆ Exhaust flow
- ◆ Wafer chuck

Developer Temperature ■ The optimum temperature for developer mixtures is between 15 to 25°C. The temperature must be maintained to within $\pm 1^\circ\text{C}$ or better from the optimal setpoint once it has been determined. The temperature of the developer has a direct effect on resist dissolution rate. As previously noted, faster resist dissolution for positive resists occurs at lower developer temperatures. For solvent developers (negative resist), the dissolution rate increases with an increase in temperature.

Developer Time ■ Developer continues to react with resist until it is removed by a rinse process. This condition can lead to overdevelopment of the resist and cause unacceptable CD variations. Off-line analysis of CD resist linewidth can be done on a sample of wafers from a batch to determine optimum develop time. However, this sampling gives only limited information and is costly.

An in-situ *dissolution rate monitor* (DRM) is used in automated wafer track systems for real-time monitoring of the development process.⁹ A DRM uses interferometric signal data collected by an optical sensor that measures the phase difference of light reflected off the baked-on resist film surface and the wafer surface underneath. As the resist dissolves, the phase difference between the two surfaces decreases to zero at the breakthrough point for endpoint (EP) detection. Monitoring each wafer with a DRM provides more useful information about the develop process and permits more accurate control of CDs.

Developer Volume ■ The amount of developer volume dispensed on the wafer is critical for successful resist development. The use of inadequate developer volume results in *scumming*, which is a residue film left on the wafer surface if it has been underdeveloped (such as from inadequate developer volume) or improperly rinsed. Excessive developer is undesirable because the goal is to reduce chemical usage in the fab. Excessive developer is a cost issue, but this condition will not necessarily ruin the resist on the wafer.

Normality ■ *Normality* (*N*) is a concentration unit that refers to the quantity of substances (stoichiometry) in a solution. For a particular developer, normality represents the exact chemical makeup, which in turn indicates how strong or weak the developer is. Normality has a direct effect on resist dissolution rates. It is primarily altered by the addition of moisture, so the chemical distribution system must be carefully controlled to minimize leaks.

Rinse ■ Rinsing is typically done with DI water, which is sprayed onto the wafer, and then the wafer is spun dry. Rinse serves to stop the development process and also remove developer from the wafer surface.

Exhaust Flow ■ The develop module is enclosed on the wafer track machine. The exhaust flow is critical so as not to disturb the developer that is applied in spray form. Too low an exhaust can leave residual developer mist in the developer module, which builds up and reacts with subsequent wafers.

Wafer Chuck ■ The wafer must be held perfectly flat on the wafer chuck during the static or spin development cycle. This placement ensures coverage uniformity during puddle development.

HARD BAKE

A post-development thermal bake, referred to as *hard bake*, is done to evaporate any residual solvent and harden the resist. This action improves the resist's adhesion to the substrate and prepares it for subsequent processing, such as becoming more resistant to etch. Hard bake also removes any residual developer and any water.

Since all light exposure is complete, the baking temperature for hard bake can be elevated toward the solvent boiling point, effectively evaporating the solvent and achieving maximum resist densification. For DNQ-novolak resist, there is also the need to evaporate any remaining DNQ to avoid nitrogen diffusion during subsequent high-energy processes. The nitrogen causes localized resist popping in the densified resist, which disperses resist particles on the wafer surface.¹⁰

The starting point for the hard bake temperature is the recommended setting determined by the resist manufacturer. After that, the process can be fine-tuned to meet the adhesion and dimensional control necessary for the product. Nominal hard bake temperatures are 130°C for positive resist and 150°C for negative resist. Hard bake is typically done on an in-line hot plate in a wafer track system, or in an in-line oven. Resist is a material that softens and flows when sufficiently heated (see Figure 15.8). Higher hard bake temperatures could cause the resist to flow slightly, which can deform the pattern.

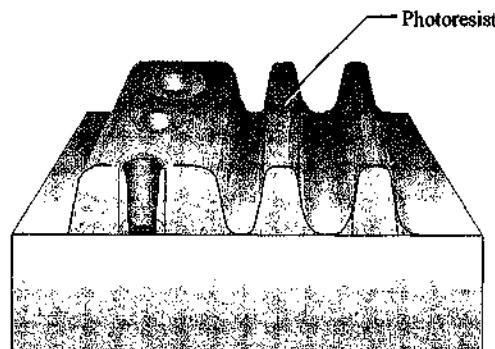


FIGURE 15.8 Softened Resist Flow at High Temperature

Resist Hardening With Deep UV. Resist hardening of DNQ-novolak resist is also achieved by exposing it to deep UV. This exposure causes the positive resist resin to crosslink and form a thin surface crust, which increases the resist thermal stability. The crosslinked resist can now withstand thermal processes up to 210°C without significant resist flow.¹¹ This condition is beneficial for subsequent processing such as plasma etching and ion implant (see Chapters 16 and 17, respectively), where temperatures can rise up to 125 to 200°C. If the resist is not sufficiently hardened, it risks flowing, which reduces the resolution of the original pattern.

DEVELOP INSPECT

The imaged pattern in the resist undergoes a *post-develop inspection* to find defects. This step is necessary to identify and remove defective wafers before they continue to the following processes of etch or implant. A wafer with a defective resist pattern that is etched or implanted will become scrap. The post-develop inspection is used to characterize the photo process, providing information to the photolithography production team for corrective action.



Automated Inspect Tool for Develop Inspect
(Photo courtesy of Advanced Micro Devices)

Most post-develop defects are relatively large and belong to a wide range of defect types (see the Quality Measure section at the end of this chapter). The defects may occur during all previous photolithography steps plus during the processing before lithography. Post-develop inspection requires a flexible system. Traditionally, this inspection has been a manual, operator-intensive inspection process with an optical microscope. Depending on the criticality of the resist layer, the inspection may be done on a sample basis. Automated inspection equipment for post-develop inspection is becoming more common in advanced fabs, especially for deep submicron lithography where the defects are not detectable with optical microscopes.

There are two possible outcomes for a wafer that fails post-develop inspection. If the wafer has a problem from a previous operation that makes it unacceptable, then the wafer is scrapped. If the problem is associated specifically with the quality of the pattern in the resist film, then the wafer is reworked. Wafer rework is done by stripping the photoresist off the wafer surface and inserting the wafer back into the photolithography process (see Figure 15.9).

This is one of the few operations in the fabrication process where rework of rejected parts is possible. The percent of wafers reworked is monitored and action plans are put in place for rejects. The goal is zero defects, while many fabs run typically less than 2% rejects. If the rework rate is greater than about 4%, then there is a quality problem that requires corrective action.

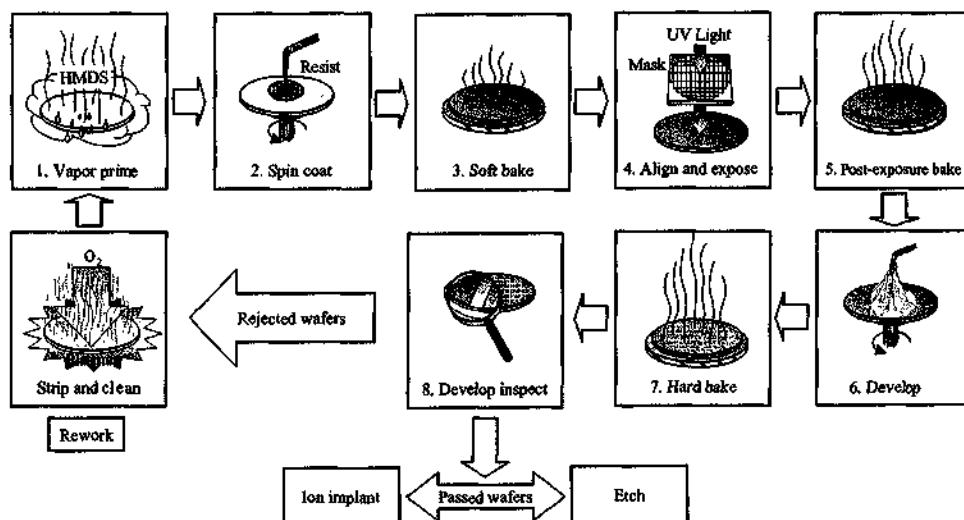


FIGURE 15.9 Develop Inspect Rework Flow

ADVANCED LITHOGRAPHY

The resolution capability of optical lithography has been extended by many equipment and process improvements, especially in the areas highlighted in Table 15.2. These improvements are the basis for a recent development in optical lithography referred to as subwavelength printing. *Subwavelength lithography* images CDs that are significantly less than the light source wavelength. An example is a 248 nm DUV light source used to print CD linewidths at 0.2 μm (200 nm) or below. Recent research activities have demonstrated a 50 nm CD pattern using a 248 nm DUV light source, chemically amplified DUV resist, and phase-shift mask. Subwavelength printing further extends the versatility of optical lithography while lowering its cost of ownership.

TABLE 15.2 Photolithography Improvements

- | |
|--|
| 1. Reduction in wavelength of the UV light source. |
| 2. Increase in numerical aperture of optical lithography tools. |
| 3. Chemically amplified DUV resists |
| 4. Resolution enhancement techniques (e.g., phase-shift masks and optical proximity correction). |
| 5. Wafer planarization (chemical mechanical planarization, or CMP) to reduce surface topography. |
| 6. Advances in photolithography equipment (e.g., stepper and step-and-scan). |

Next-Generation Lithography

Predicting the actual resolution limit of optical lithography has been a futile exercise for many years. There is some future limit where the extension of optical lithography will not be feasible. Wafer patterning must then transition to an alternative lithography process referred to as *next-generation lithography*. Major research activity is underway to investigate the type of next generation lithography tool that will replace optical lithography. Advanced study of this nature requires many years of research and development effort and sustained funding. There are four major lithography technologies considered as possible successors to optical lithography:¹²

- ◆ Extreme UV (EUV)
- ◆ Ion projection lithography (IPL)
- ◆ SCALPEL
- ◆ X-ray

Each of these technologies has demonstrated patterning capability down to 70 nm or below. At the time of this writing, International SEMATECH has selected EUV and SCALPEL for research funding as the next generation lithography tool.

Extreme UV (EUV) ■ *Extreme UV* lithography builds on the production knowledge of optical lithography. It uses a laser-produced plasma source to produce UV wavelengths of ~ 13 nm and is expected to print images down to about 30 nm.¹³ The source operates in a vacuum environment to produce the EUV radiation, which is collected by condenser optics and formed into a beam (see Figure 15.10 on page 425). The beam is reflected off a reflective reticle that scans the pattern through the beam. An all-reflective (meaning mirrors instead of lens) 4X projection optics will image the UV beam onto a resist-coated wafer. The wafer is scanned at one-fourth the reticle speed in the opposite direction.

There are a number of challenges for bringing EUV into wafer production.¹⁴ The precision optics will be difficult to achieve given the stringent requirements for manufacturing the high-quality surfaces. Mirror reflectivities will be optimized by using precision multilayer coatings on the surfaces. The reflective reticle is four times larger than the feature size; therefore, a 100 nm CD has a minimum reticle feature size of 400 nm (at 4X demagnification). The reticle consists of a thin metal absorber patterned on a multilayer coating that is patterned with conventional electron beam mask tools. The short wavelengths and penetration depths will require the use of top surface imaging resists (see the following section) or bi-layer resists with a very thin top layer. For alignment, the total overlay budget is approximately 35 nm for devices at the 0.1 μm design rules. All equipment will be in a vacuum environment.

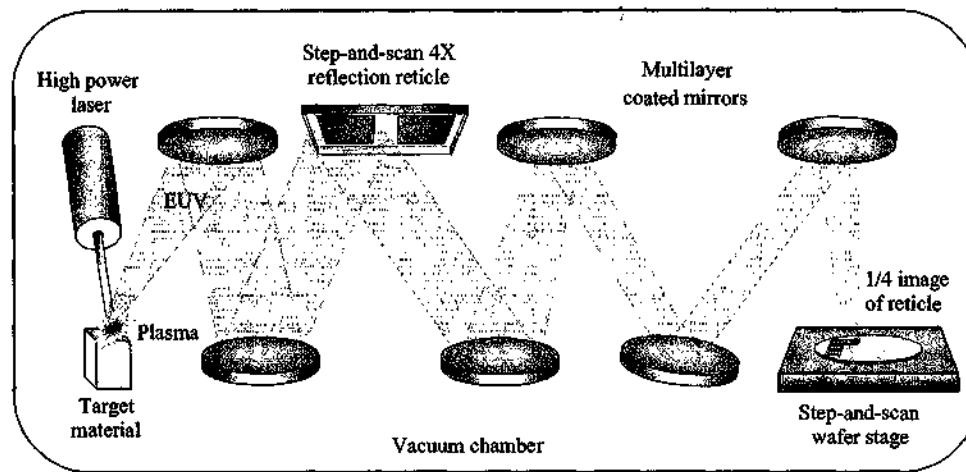


FIGURE 15.10 Concept for Extreme Ultraviolet Lithography
Redrawn from International SEMATECH's Next Generation Lithography Workshop brochure

SCALPEL ■ Feasibility studies are in progress for an electron beam imaging approach referred to as *SCALPEL* (SCattering with Angular Limitation Projection Electron Beam Lithography).¹⁵ *SCALPEL*, in development since the late 1980s, uses an established electron beam source instead of a light source to image a wafer pattern (see Figure 15.11). A multilayer membrane mask is used that absorbs few electrons. The pattern is formed as the electron beam passes through a high atomic number layer in the mask that scatters the electrons for a high-contrast image at the wafer plane.

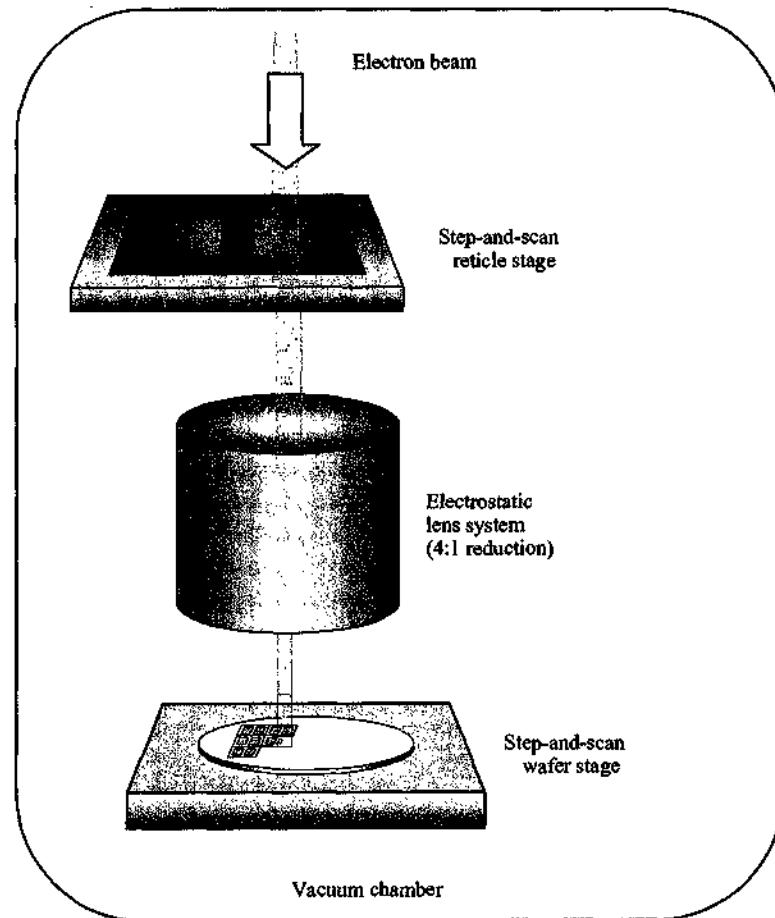


FIGURE 15.11 Concept of SCALPEL
Redrawn from International SEMATECH's Next Generation Litihography Workshop brochure

Increased beam current permits faster exposure but there is also a problem with beam blur caused by space charge effects at higher beam currents. Beam blur occurs when like-charged particles repel one another and cause the image to blur. The mask for SCALPEL is a 4X mask that does not require complicated resolution enhancement techniques, nor does the system need expensive optics.

SCALPEL prints linearly, using a step-and-scan writing strategy that produces stripes of exposed resist. To expose an entire chip, several stripes are stitched together per chip. This approach will require good overlay accuracy so that the stitching does not exceed CD error budgets. The first commercial SCALPEL tools are targeted for 2002.

Ion Projection Lithography (IPL) ■ *Ion projection lithography (IPL)* is a technique in early development that uses ion beams to expose resists, either through a mask or by serially writing on the resist with a finely focused beam (see Figure 15.12). If a mask is used, then a stitched approach is needed, using a broad beam to create small subfields of exposed resist on the wafer surface. IPL uses multielectrode electrostatic optics to direct hydrogen or helium ions to the wafer. Ions transfer their energy more efficiently to the resist than electron beams because ions have a larger mass. There are also fewer secondary electrons produced by ion bombardment and those produced have very low energy, which makes for less backscatter. Backscatter creates proximity effects that determine the limit to the minimum feature size for wafer patterning.¹⁶ Very high resolution is achievable from ion projection lithography, with 50 nm feature sizes already demonstrated in a research environment.

X-Ray ■ *X-ray lithography* is an established technology for imaging patterns on wafers with CDs below 100 nm (0.1 μm). An X-ray source projects X-rays onto a special mask that forms a pattern

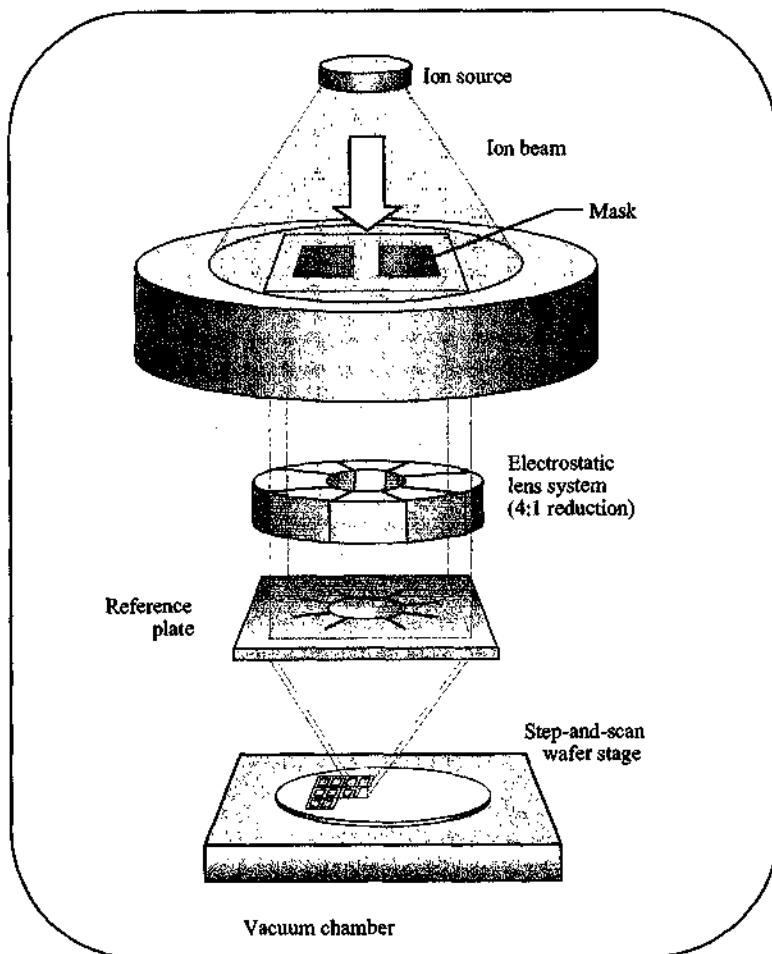


FIGURE 15.12 Ion Projection Lithography
Redrawn from International SEMATECH's Next Generation Lithography Workshop brochure

on a resist-coated wafer. This technique is not widely used for wafer fabrication, primarily due to its high cost of ownership in comparison to optical lithography.¹⁷ One major chipmaker has used X-ray lithography for patterning commercial chips throughout the 1990s, including advanced microprocessors on 200-mm diameter silicon wafers. The system components for X-ray lithography are: (1) a mask that has a pattern of X-ray absorbing materials on a material that transmits X-rays, (2) an X-ray source, and (3) an X-ray resist.

The X-rays used in lithography are known as soft X-rays. They operate in a region of the electromagnetic spectrum from about 0.1 nm to about 10 nm (see Figure 15.13).¹⁸ These differ from the familiar X-rays used in medicine, which have a shorter wavelength and are referred to as hard X-rays.

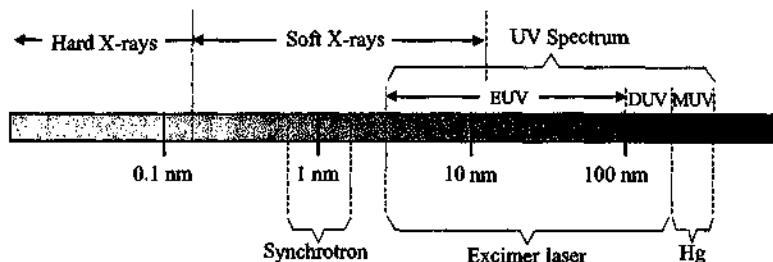


FIGURE 15.13 X-ray Spectrum

The most common X-ray source for lithography is known as a synchrotron (an electron storage ring). High-energy electrons are forced into closed curved paths by magnetic fields and made to accelerate which causes them to emit radiation. This action produces X-ray radiation that is very intense and reasonably collimated (parallel). A projection system based on special mirrors will project the X-rays onto the mask and resist coated wafer. Each synchrotron ring provides radiation for multiple X-ray steppers (e.g., 15 or more).

A special X-ray photomask is used to define the pattern. Because X-ray lithography has such a short wavelength, there are no diffraction interference effects from the mask, producing wide process latitude. An X-ray mask has a membrane substrate (e.g., polyimide) that transmits X-rays and a patterned material that absorbs X-rays, such as gold, tungsten, or tantalum (see Figure 15.14). However, the ability to fabricate masks is one of the many challenges for X-ray lithography. X-ray patterning uses a 1X mask that has the same dimensions as the wafer CDs. This feature eliminates the benefit from having a 4X mask with four times larger feature sizes than the wafer. Mask makers

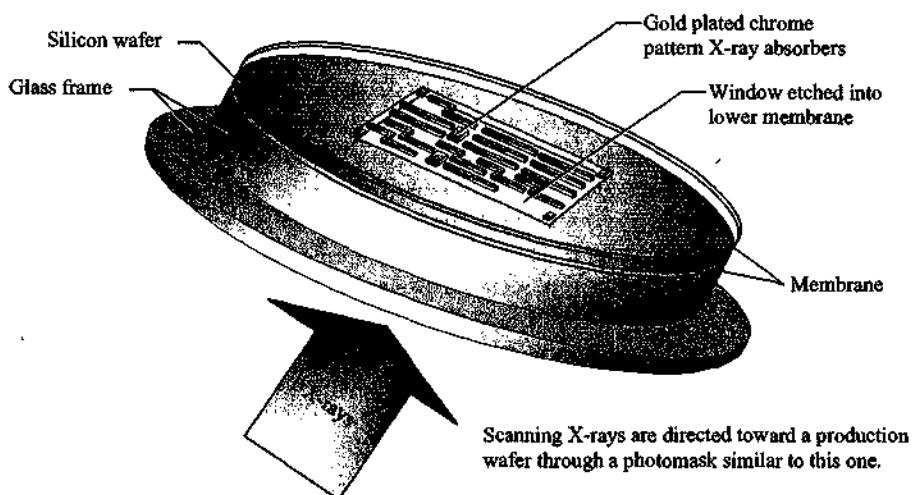


FIGURE 15.14 Concept of X-ray Photomask

Redrawn from K. Nalcamura, *Lithography, ULSI Technology*, ed. by C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 314.

are challenged to control pattern placement and CDs and minimize mask defects on X-ray masks with existing e-beam mask writers.¹⁹

Advanced Resist Processing

Photoresist has undergone two basic turning points in semiconductor manufacturing: (1) changing to positive resist, and (2) using chemically amplified deep UV resist (see Figure 15.15). Because resist is the medium that transfers the reticle pattern to the wafer surface material, it is a critical contributor to next-generation lithography.

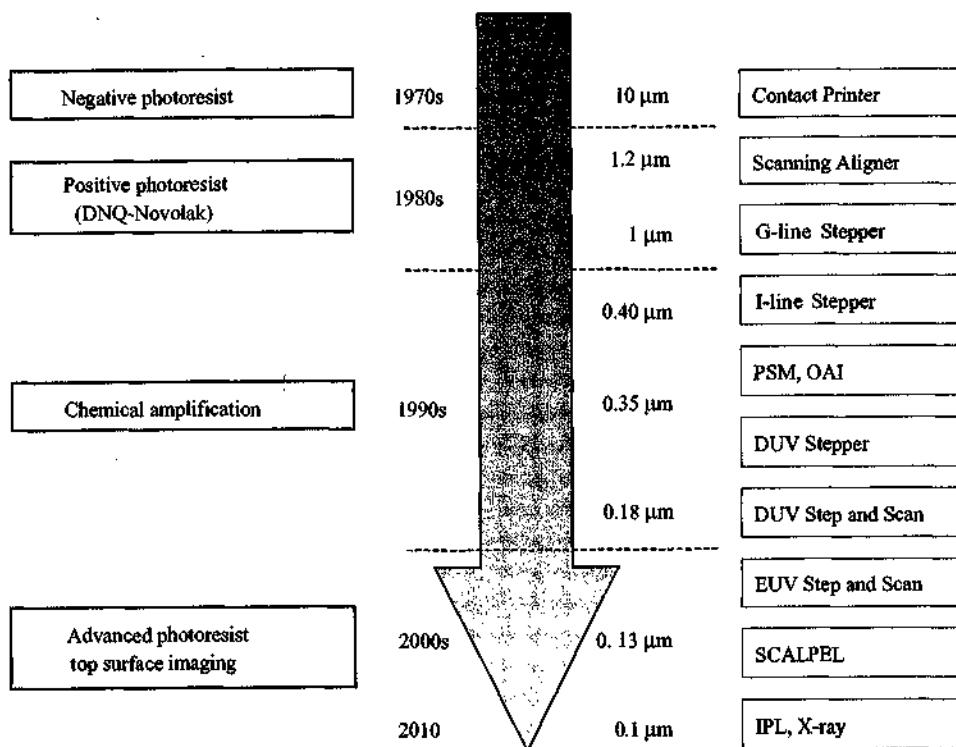


FIGURE 15.15 Development Trends of Photoresist and Lithography

An alternative resist technology that is being evaluated for sub-0.1 μm feature sizes is top-surface imaging resist. In *top-surface imaging*, the photoresist is imaged only at its top surface. This feature action provides a wider process latitude for focus issues than that found on typical resist systems since conventional resist must have its entire resist thickness in focus. At the same time, top-surface imaging permits resist materials to be opaque to the exposure light.²⁰ This feature could be beneficial since 248 nm DUV resist materials become opaque when the UV wavelength is reduced to 193 nm.

DESIRE Process ■ There are different ways to achieve surface imaging on resists. One method is through a process referred to as DESIRE (diffusion-enhanced silylated resist), which uses *silylation* to selectively place a thin layer of silicon on a resist. This process effectively transfers the reticle pattern to the resist-coated wafer in the form of a silicon pattern on the top surface of the resist.

The silicon is applied by first exposing the spin-coated resist to UV light per the standard process (see Figure 15.16 on page 429). The unexposed resist has the PAC (photoactive compound) sensitizer that is a diffusion inhibitor, while the exposed resist is more susceptible to silicon diffusion. There is often a presilylation bake to crosslink unexposed resist and increase the contrast between the exposed and unexposed region. The silicon is then incorporated into the resist by treating it with a gas-phase silylating agent (such as hexamethyldisilazane, or HMDS) at elevated

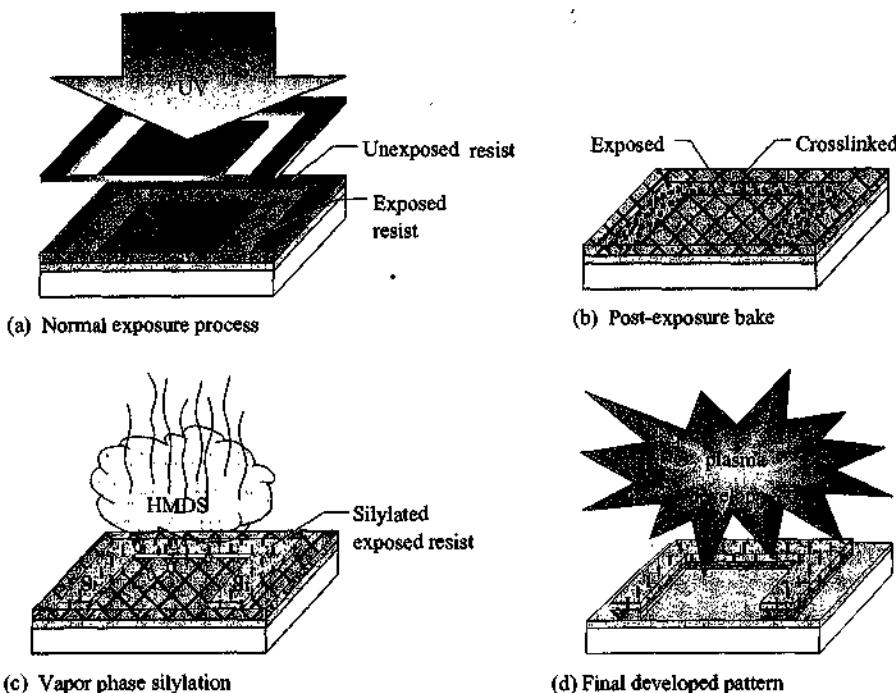


FIGURE 15.16 Top-Surface Imaging

temperatures. The exposed areas of the resist selectively bond chemically to the silicon to a depth of 100 to 200 nm.²¹ The silicon-bonded areas of the resist represent the pattern transfer from the reticle to the substrate.

The final step is to dry-develop the wafers in oxygen plasma to remove the unprotected resist. During this plasma etch the silicon is converted into silicon dioxide which forms a thin protective mask that stops the etch process in the resist directly underneath the SiO₂. This is just one example of many silylation-based processes. They all essentially use a single-layer patterning technique that depends on selectively incorporating a silicon-containing compound into the photoresist after the exposure step.

DEVELOP QUALITY MEASURES

There are different types of defects found at post-develop inspect. The quality measures are shown in Table 15.3.

TABLE 15.3 Key Measures at Post-Develop Inspection

Quality Parameter	Types of Defects	Remarks
1. Critical dimensions.	A. Wider CDs than normal.	<ul style="list-style-type: none"> • Improper stepper focus. • Not enough time or energy during exposure. • Not enough develop time or weak developer solution. • Incorrect process recipes during exposure or develop steps.

Quality Parameter	Types of Defects	Remarks
	B. Narrower CDs than normal.	<ul style="list-style-type: none"> Excessive time or energy during exposure. Excessive develop time or strength of developer solution. Incorrect process recipes during exposure or develop steps.
2. Contamination.	A. Particles and foreign contamination on the resist surface.	<ul style="list-style-type: none"> Equipment needs to be kept clean, with special focus on the track equipment. Inadequate cleaning and rinsing of wafers. Developer chemicals and rinse water require point-of-use filters to remove contaminants.
3. Surface defects.	A. Scratches in the resist surface.	<ul style="list-style-type: none"> Wafer handling errors or tool misadjustments related to cassette indexers and robotic handling systems.
	B. Particles, spots and stains.	<ul style="list-style-type: none"> Chamber exhaust flow, dispenser alignment, dispense pressure, wafer leveling, splashbacks, drips, spin speed, are all possible contributors.
	C. Missing resist, extra resist, or scumming.	<ul style="list-style-type: none"> Incorrect puddle time. Incorrect dispense volume or position. Improper rinsing after the develop process. Improper or uneven baking.
	D. Striations in photoresist along the sidewalls of the resist features.	<ul style="list-style-type: none"> Standing waves or reflective notches (unacceptable CD variation). Improper or no antireflective coating (ARC) was used.
4. Overlay registration.	A. Improper alignment or overlaying of one layer over a previous layer.	<ul style="list-style-type: none"> This is not a problem caused by the develop process. It is more likely a stepper-induced problem. Wrong process recipe or reticle was used. Poor temperature and humidity control.

DEVELOP TROUBLESHOOTING

Various troubleshooting problems encountered during develop are shown in Table 15.4.

TABLE 15.4 Common Develop Troubleshooting Problems

Problem	Probable Cause	Corrective Actions
1. Linewidths and holes do not meet CD requirements.	A. Underdeveloped or underexposed positive resist across entire wafer.	<ul style="list-style-type: none"> • Ensure correct stepper process recipe was used. • Check for insufficient exposure time and energy settings. • Check stepper illuminator system. • Verify dose meter (light integrator) is functioning properly. • Ensure correct recipe on developer was used. • Check for insufficient puddle time and quantity settings. • Check developer equipment. • Check bake temperatures.
	B. Overdeveloped or overexposed positive resist across entire wafer.	<ul style="list-style-type: none"> • Ensure correct stepper process recipe was used. • Check for excessive exposure time and energy settings. • Check stepper illuminator system. • Verify dose meter (light integrator) is functioning properly. • Ensure correct recipe on developer was used. • Check for excessive puddle time and quantity settings. • Check developer equipment. • Check bake temperatures.
	C. No measurable CDs.	<ul style="list-style-type: none"> • Check exposure or develop operation. • Check reticle or process recipe. • Wafer may have skipped coat, expose, post-exposure bake, or develop step. • Rework wafer.
2. Resist scumming.	A. Resist residue remains on the wafer after the develop operation is completed.	<ul style="list-style-type: none"> • Check develop equipment process recipe. • Verify puddle time and puddle quantity are correct. • Check the rinser operation. • Check bake oven times and temperatures. • Rework wafer.

Problem	Probable Cause	Corrective Actions
3. Contamination and defects.	A. Possible causes may include the chemicals, rinse water, and process chamber.	<ul style="list-style-type: none"> Verify puddle time and puddle quantity are correct. Clean the develop process chamber, then recheck a test wafer for further contamination. Check, and if necessary, replace line filters for the developer and rinse water.
	B. Misting or backsplashing from the dispenser can cause contamination.	<ul style="list-style-type: none"> Check chamber exhaust level. Check alignment of dispenser relative to the wafer. Check for drips from dispenser assembly. Rework wafer.
4. Collapsing of resist pattern after develop.	A. High aspect ratio ($>5:1$) will lead to collapsing of resist lines.*	<ul style="list-style-type: none"> Verify that the resist is not excessively thick, since this will increase the aspect ratio and make the resist more likely to collapse. Check that the resist has proper adhesion to the wafer. The problem resolution may require a material or process change (e.g., more rigid resist).
5. Unacceptable CD variation at top of CA DUV resist profile.	A. Amine contamination of resist after exposure.	<ul style="list-style-type: none"> Check integrity of environmental chamber filtering system. Check to see if coated-wafers may have been exposed to external chemical contamination. Asses whether resist is best selection for length of delay (newer resists withstand a longer time before PEB).

*J. Yu, et al., "Analysis of Resist Pattern Collapse and Optimization of DUV Process for Patterning Sub-0.20 mm Gate Line," *Advances in Resist Technology and Processing XV, Proceedings of SPIE*, Vol. 333, (Bellingham, WA: SPIE, 1998); p. 880.

SUMMARY

Resist development removes those areas of the resist rendered soluble by the exposure step. After exposure, a post-exposure bake (PEB) is performed for chemically amplified DUV resists to catalyze critical resist chemical reactions. The temperature uniformity, duration, and time delay for the PEB of DUV resist is important. For conventional DNQ-novolak resist, a PEB drives out additional solvent for better adhesion and reduces the standing wave effect. Negative resist development is mainly a solvent wash of unexposed resist. The exposed negative resist tends to swell during development, which limits its use in submicron lithography. Positive resist development involves a chemical reaction between the developer and the exposed resist, with parameters of develop speed and selectivity. The most common positive-tone developer is TMAH. The two most common development methods are continuous spray and puddle, which require the control of critical parameters for an optimum process. Hard bake is a post-development thermal bake

used to drive out any residual solvent and harden the resist. A post-develop inspection is performed to characterize defects for corrective action and remove defective wafers from the process before etch or implant.

Lithography improvements have produced sub-wavelength lithography, where patterning achieves a CD below the exposure light wavelength. Next-generation lithography is evaluating the lithography technology for eventually replacing optical lithography. There are four primary alternatives: Extreme UV (EUV), SCALPEL, ion projection lithography (IPL), and X-ray. EUV uses a UV wavelength of about 13 nm to achieve a pattern resolution of 30 nm. SCALPEL uses an established electron beam for lithography. IPL uses ion beams to expose resist. X-ray lithography projects X-rays with a wavelength from about 0.1 to 10 nm onto a special mask to pattern a resist. Top-surface imaging is a technique used to image only the top surface of the resist for reduced depth-of-focus and increased resolution.

KEY TERMS

resist development
post-exposure bake (PEB)
negative resist development
positive resist development
dissolution rate
developer selectivity
continuous spray develop
puddle develop
dissolution rate monitor (DRM)
scumming
normality

hard bake
post-develop inspection
subwavelength lithography
next-generation lithography
extreme UV (EUV)
SCALPEL
ion projection lithography (IPL)
X-ray lithography
top surface imaging
silylation

REVIEW QUESTIONS

1. Explain resist development. What is its goal?
2. Why is a post-exposure bake done for chemically amplified DUV resist? Be specific about deprotection.
3. Why is temperature uniformity important for PEB?
4. Describe the benefits of a post-exposure bake for an i-line conventional resist.
5. Explain negative resist development. What is the primary problem for submicron patterning?
6. Why is positive-tone resist the most commonly used resist?
7. What is the developer dissolution rate? Is it desirable for the rate to be high or low?
8. Explain resist selectivity and whether it should be high or low.
9. What is the name for the most common positive-tone developer?
10. Why is a surfactant added to the developer?
11. Explain how the TMAH developer functions for a conventional i-line resist.
12. Is there a chemical reaction between PHS and the developer for chemically amplified DUV resists?
13. List the two methods of resist development.
14. Explain continuous spray development.
15. Describe puddle development.
16. List the seven resist development parameters.
17. Explain why hard bake is done.
18. Describe UV resist hardening.
19. Why is a post-develop inspection performed?
20. What is subwavelength lithography?
21. List four alternative lithography methods under evaluation for next-generation lithography.
22. Explain extreme UV (EUV).
23. Describe SCALPEL lithography.
24. Discuss ion projection lithography (IPL).
25. Explain X-ray lithography. What is one of the major challenges for X-ray lithography?
26. Describe top-surface imaging. Why is silylation used in this process?

PHOTOLITHOGRAPHY MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES

Allied Signal
Applied Materials
Arch Chemicals (aka Olin)
Ashland Specialty Chem.
ASML
Canon Semiconductor
Charles Evans and Associates
Clariant Corporation
Cymer Inc.
DuPont
Eastman Chemical

<http://www.electronicmaterials.com/>
<http://www.appliedmaterials.com/products/>
<http://www.olinmicro.com/default.asp>
<http://www.ashland-act.com/>
<http://www.asml.com/>
<http://www.usa.canon.com/indtech/semicondeq/>
<http://www.cea.com>
<http://www.azresist.com/>
<http://www.cymer.com/>
<http://www.dupont.com/semiconductor/>
<http://www.eastman.com/>

EKC Technology	http://www.ekctech.com/ekctech.nsf
ETEC Systems Inc.	http://www.etec.com/sempiprod_frame.html
FSI International	http://www.fsi-intl.com/
International SEMATECH	http://www.semtech.org/
JSR Microelectronics, Inc.	http://www.jsrusa.com/index2.html
J.T. Baker	http://www.jtbaker.com/
Karl Suss Inc.	http://www.suss.com/
Lucent Technologies	http://www.bell-labs.com/project/SCALPEL/
MICRO Magazine	http://www.micromagazine.com/
Nikon	http://www.nikon.com/
Olin Microelectronics	http://www.olinmicro.com/
Photonics Inc.	http://www.photonics.com/
SEMI	http://www.semi.org/
Semiconductor International	http://www.semiconductor.net/
Shipley Company	http://www.shipley.com/
Silicon Valley Group	http://www.svg.com/
Solid State Technology	http://sst.pennet.com/home/home.cfm
SPIE	http://www.spie.org/
TEL, Tokyo Electron Ltd.	http://www.teainet.com
Ultratech Stepper	http://www.ultratechstepper.com/

REFERENCES

1. P. Burggraaf, "Optical Lithography to 2000 and Beyond," *Solid State Technology* (February 1999): p. 31.
2. B. Smith, "Resist Processing," *Microlithography: Science and Technology*, ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 542.
3. D. Seeger, "Chemically Amplified Resists for Advanced Lithography: Road to Success or Detour?" *Solid State Technology* (June 1997): p. 115.
4. A. Braun, "Track Systems Meet Throughput and Productivity Challenges," *Semiconductor International* (February 1998): p. 63.
5. B. Smith, "Resist Processing," p. 541.
6. B. Smith, "Resist Processing," p. 551.
7. C. Mack et al., "New Model for the Effect of Developer Temperature on Photoresist Dissolution," *Advances in Resist Technology and Processing XV* Proceedings of SPIE vol. 3333, (Bellingham, WA: SPIE, 1998): p. 1218.
8. B. Smith, "Resist Processing," p. 548.
9. D. Velikov et al., "Endpoint Detector Monitors Photoresist Develop Process," *Semiconductor International* (August 1996): p. 144.
10. B. Smith, "Resist Processing," p. 561.
11. B. Smith, "Resist Processing," p. 562.
12. R. DeJule, "Next-Generation Lithography Tools: The Choices Narrow," *Semiconductor International* (March 1999): p. 48.
13. Ibid.
14. Challenges summarized from A. Hawryluk, N. Ceglio, and D. Markle, "EUV Lithography," *Microlithography World* (Summer 1997): pp. 18–20.
15. R. DeJule, "Next-Generation Lithography Tools," p. 48.
16. S. Ghandhi, *VLS Fabrication Principles: Silicon and Gallium Arsenide*, 2nd ed., (New York: Wiley, 1994), p. 693.
17. T. Ueno and J. Sheats, "X-ray Lithography," ed. J. Sheats and B. Smith (New York: Marcel Dekker, 1998), p. 403.
18. T. Ueno and J. Sheats, "X-ray Lithography," p. 405.
19. R. DeJule, "Next-Generation Lithography Tools," p. 48.
20. S. Postnikov et al., "Top Surface Imaging Through Silylation," *Advances in Resist Technology and Processing XV* Proceedings of SPIE vol. 3333, (Bellingham, WA: SPIE, 1998): p. 997.
21. K. Nakamura, "Lithography," *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 294.

CHAPTER 16

ETCH

A wafer is fabricated into many functional microchips by creating electronic devices in the silicon and then sequentially depositing layers of dielectric and conductive materials to interconnect the devices. This is the concept of planar technology for wafer fabrication, used since the early days of semiconductor manufacturing.

In general, interconnect materials are deposited on the wafer surface and then selectively removed to form the circuit patterns that were defined by photolithography. This selective material removal process, known as etch, follows post-develop inspection. It is critical that etch be performed correctly or else the chip will not function. More importantly, once material is removed by etch, it is difficult to correct a mistake. Improperly etched wafers are most likely scrapped and discarded at a loss to the company.

Etch requirements depend on the type of feature being fabricated, such as the aluminum alloy metal stack, a polysilicon gate, a silicon trench for isolation, or a

dielectric via. IC structures are complex, with a range of materials that require different etch parameters. Shrinking feature sizes drive tighter etch dimensional control and make inspections more difficult.

An example of an etch application is aluminum. The traditional metallization process deposits an aluminum alloy layer on the wafer surface. This action is followed by photolithography and etch to form the interconnects. The different metal layers are electrically connected by previously-formed tungsten plugs in interlayer dielectric (ILD) vias.

With the introduction of damascene processing for copper metallurgy, the metallization process becomes a dielectric etch to form a trench in the ILD. Copper is blanket-deposited into the dielectric patterns and polished back to the ILD height using chemical mechanical planarization. With damascene processing, there will be less emphasis on metal etch and more importance placed on dielectric etch.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. List and discuss nine important etch parameters.
2. Explain dry etch, including its advantages, and discuss how etching action takes place.
3. List and describe the equipment systems for seven dry plasma etch reactors.

4. Explain the benefits of high-density plasma (HDP) etch, and discuss the four types of HDP reactors.
5. Give an application example for dielectric, silicon, and metal dry etch.
6. Discuss wet etch and its applications.
7. Explain how photoresist is removed.
8. Discuss etch inspection and its related important quality measures.

INTRODUCTION

Etch is the process of selectively removing unneeded material from the wafer surface by using either chemical or physical means. The fundamental goal of etching is to accurately reproduce the mask features on the resist-coated wafer. The patterned resist layer is not attacked significantly by the etchant. This mask layer is used to protect specific regions of the wafer surface while permitting selective etching through openings in the photoresist layer (see Figure 16.1 on page 436). The etch process follows photolithography in the general CMOS process flow (see Figure 16.2 on page 436). In this manner, etch can be viewed as the last major pattern transfer process step necessary to replicate the desired pattern on the wafer surface.

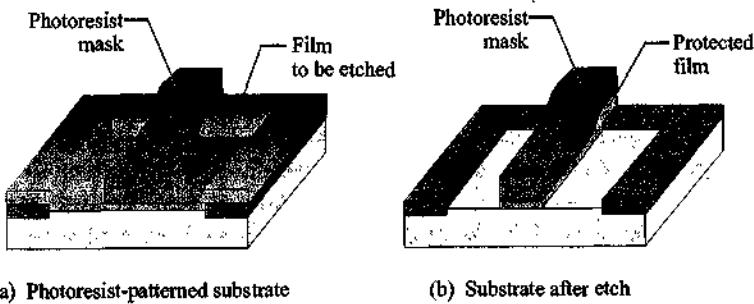
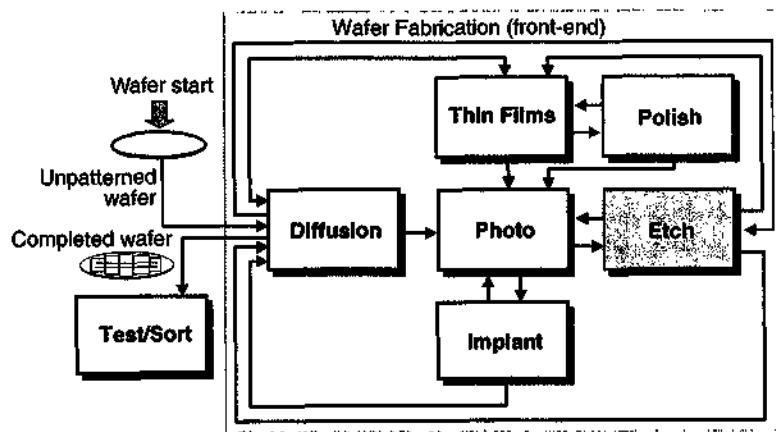


FIGURE 16.1 Applications for Wafer Etch in CMOS Technology

FIGURE 16.2 General CMOS Process Flow
(Used with permission from Advanced Micro Devices)

Etch Processes

Two basic types of etch processes are used in semiconductor manufacturing: dry etch and wet etch. *Dry etch* exposes the wafer surface to a plasma created in the gaseous state. The plasma passes through the openings in the patterned resist and interacts physically or chemically (or both) with the wafer to remove the surface material. Dry etching is the primary method used for etching devices with submicron geometries. Thus, dry etching will receive the most coverage in this text. In *wet etch*, liquid chemicals such as acids, bases, and solvents are used to chemically remove wafer surface material. Wet etch is generally applicable only for larger geometries ($>3\text{ }\mu\text{m}$). It is still used to remove some layers from the wafer and to remove dry etch residues.

Dry etch can also be categorized by the type of material removed. The three major material categories for etching are metal etch, dielectric etch, and silicon etch. *Dielectric etch* is used for applications with dielectric material, such as silicon dioxide. The formation of contact and via structures require a dielectric etch to form openings in the ILD. This is a challenge for etching high-aspect ratio openings (ratio of height to width for an opening). *Silicon etch* (including polysilicon) is used for applications requiring silicon removal, such as etching polysilicon transistor gates and silicon trench capacitors. *Metal etch* is primarily used for removal of aluminum alloy stacks to form interconnect wiring on metal layers. At the time of this writing, there is no acceptable method for etching copper metal with submicron feature sizes (which is an important reason for the introduction of dual-damascene processing into the wafer fab).

Etch can also be categorized as either patterned or unpatterned etching. *Patterned etching* uses a masking layer (patterned photoresist) to define areas of the surface material that are to be etched. Only selected portions of this wafer surface layer are removed during the etch process. Patterned etching is used to form the many different features on the wafer surface, including gates, metal interconnects, vias, contact holes, and trenches. *Unpatterned etching, etchback, or stripping*, occurs when the entire wafer surface is etched with no mask present. This etching process is used to strip masking layers (e.g., STI nitride strip and the titanium strip after the salicide process used

to form spacers for transistor implants). Etchback is used when it is desirable to reduce the overall thickness of a specific layer of film (e.g., when planarizing a surface to reduce topographical features). Photoresist is another example of a material stripped off the wafer. In summary, patterned and unpatterned etch processes can be performed using either dry-etch or wet-etch techniques.

ETCH PARAMETERS

Etch has specific requirements it must meet in order to replicate the mask pattern on the wafer surface material. Important parameters for etch are:

- ◆ Etch rate
- ◆ Etch profile
- ◆ Etch bias
- ◆ Selectivity
- ◆ Uniformity
- ◆ Residues
- ◆ Polymer formation
- ◆ Plasma-induced damage
- ◆ Particle contamination and defects

Etch Rate

Etch rate is the speed at which material is removed from the wafer surface during etching (see Figure 16.3). It is usually measured in Å/minute. The depth of the etched opening is known as the *step height*. It is desirable to have a high etch rate in order to keep wafer throughput high. This feature is even more important when using single-wafer processing in cluster tools. The etch rate is determined by process and equipment variables such as the type of material being etched, reactor configuration, gases used for the etch, and the process parameter settings. The etch rate is calculated by the following formula:

$$\text{Etch Rate} = \frac{\Delta T}{t} \quad (\text{Å}/\text{minute})$$

Where, ΔT = amount of material removed (Å or μm)
 t = time elapsed during etch (typically minutes)

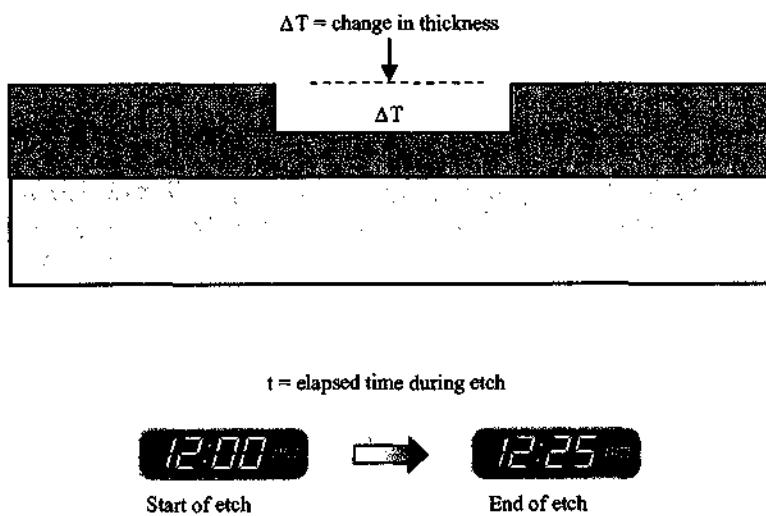


FIGURE 16.3 Etch Rate

Etch rate is generally proportional to the concentration of the etchant. Factors such as wafer surface geometry can affect the wafer-to-wafer etch rates. Wafers with significant surface area for etching will deplete the etchant concentration and etch slower, whereas wafers with small surface areas for etching will etch faster. This condition is referred to as *loading effects*. The decrease in etch rate is caused by the plasma etching reactions consuming most of the available etchant species in the gas phase. The change in etch rate due to loading effects is a primary reason why effective endpoint detection is critical (see the following section).

Etch Profile

Etch profile refers to the shape of the sidewall of the etched feature. There are two basic etch sidewall profiles: isotropic and anisotropic. An *isotropic etch profile* etches at the same rate in all directions (laterally and vertically), leading to undercutting of the etched material under the mask (see Figure 16.4). This action results in an undesirable loss of linewidth. Wet chemical etching is usually isotropic in nature, which is the primary reason why wet etching is not used for selective patterned etching of submicron devices. Some dry plasma systems are also capable of providing an isotropic etch profile. There are instances where isotropic etching may be desirable depending on the specific needs of the material being etched and the requirements of subsequent processing steps.

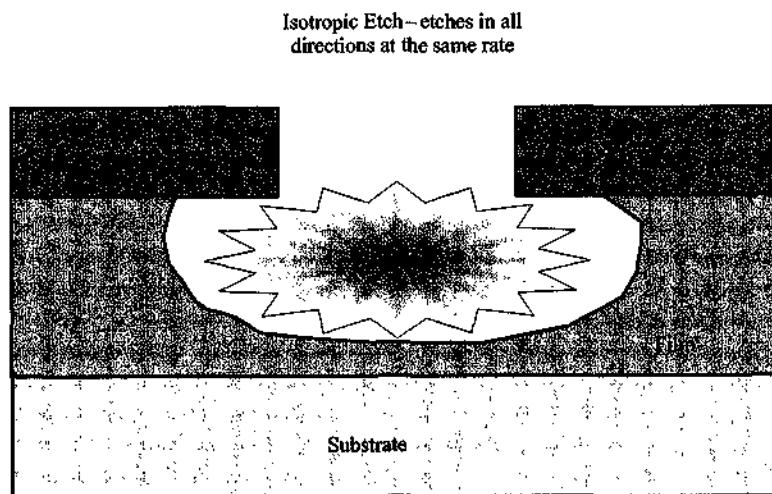


FIGURE 16.4 Wet Chemical Isotropic Etch

The desirable etch sidewall profile for submicron feature sizes is an *anisotropic profile*, where the rate of etching is in only one direction perpendicular to the wafer surface (see Figure 16.5). There is very little lateral etching activity. This leaves vertical sidewalls, permitting a higher packing density of etched features on the chip. Anisotropic etch is critical for the patterning of submicron devices with small linewidths and features. Advanced IC applications usually require 88 to 89° vertical sidewall profiles. Anisotropic etch is achieved with most dry plasma etching. Table 16.1 shows the profile for wet etch and the various profiles achievable with dry etch.

The amount of anisotropic etch can be moderate (slight sidewall angle) or highly anisotropic (vertical sidewalls). Etch profile refers to the shape of the etched film wall. A vertical profile is the result of a highly anisotropic etch.

With smaller geometries, the etch profiles have higher aspect ratios. It is difficult to get etchant chemicals in and reaction by-products out of the high-aspect ratio openings. To overcome this, it is desirable to have *directionality* to drive the plasma into the high-aspect ratio openings. If plasma ions are directional (perpendicular to the wafer surface), then only the surface is bombarded, not the feature's sidewalls. This action forces etchant chemicals into high-aspect ratio openings with little undercutting. For advanced ICs with sub-0.25 μm CDs, directionality is achieved with a high-density plasma source capable of generating enough etchant species to achieve acceptable etch rates.

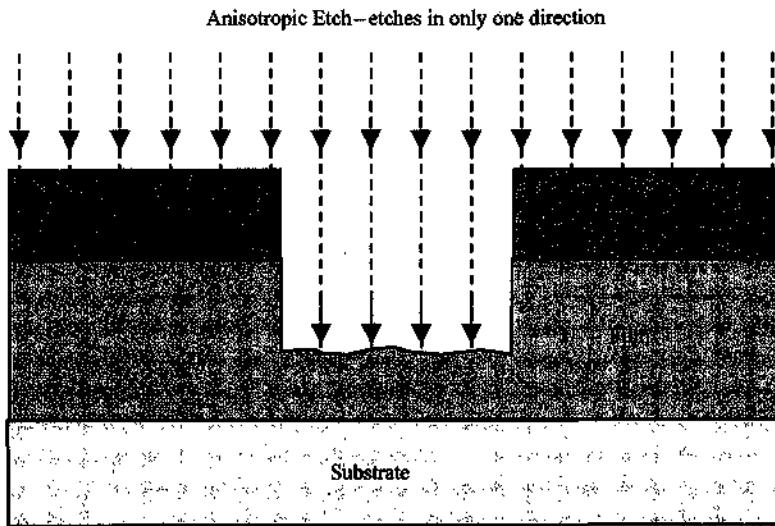


FIGURE 16.5 Anisotropic Etch with Vertical Etch Profile

TABLE 16.1 Sidewall Profiles for Wet Etch Versus Dry Etch

Type of Etch Wet Etch	Sidewall Profile Isotropic	Diagram
Dry Etch	Isotropic (depending on equipment and parameters)	
	Anisotropic (depending on equipment and parameters)	
	Anisotropic—Taper	
	Silicon Trench	

Etch Bias

Etch bias is a measure of the change in linewidth or space of a critical dimension (CD) after performing an etch process (see Figure 16.6 on page 440). It is usually caused by undercutting (see Figure 16.7 on page 440), but can also be the result of an etch profile. Undercutting occurs when

the etch process removes excessive material below the mask, causing the top surface of the etched film to be recessed from the resist edge. The formula to calculate etch bias is:

$$\text{Etch bias} = W_b - W_a$$

Where, W_b = the original linewidth in photoresist before etch.

W_a = the final linewidth of the etched material after resist removal.

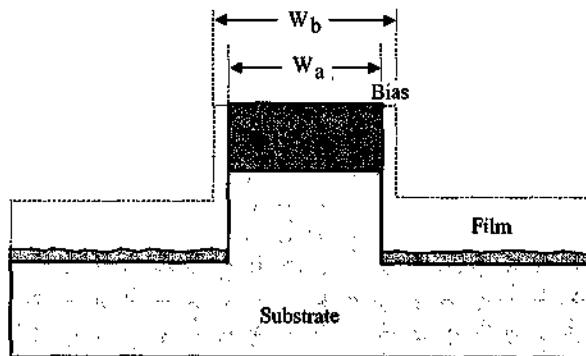


FIGURE 16.6 Etch Bias

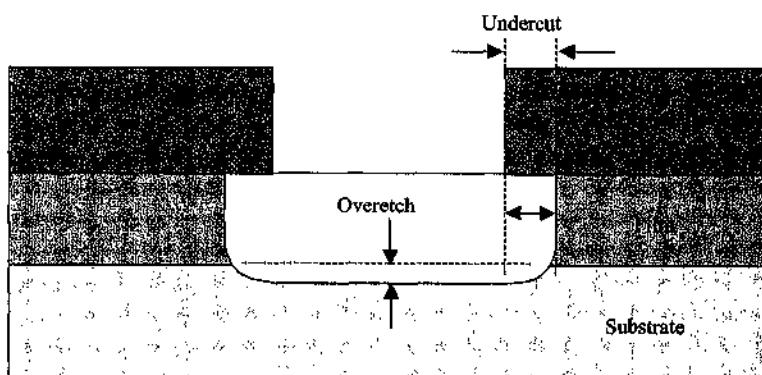


FIGURE 16.7 Etching Undercut and Slope

Selectivity

Selectivity represents how much faster one film etches than another film under the same etch conditions. It is defined as the etch rate of the material being etched relative to the etch rate of another material (see Figure 16.8). High selectivity means that etching only occurs on the desired layer. A high selectivity etch process does not etch the underlying film (etching stops at the right depth) and the protective photoresist is not etched. Shrinking geometries require thinner layers of resist. High selectivity is necessary in most advanced processes to ensure critical dimension and profile control. Specifically, the smaller the critical dimension then the higher the selectivity must be.

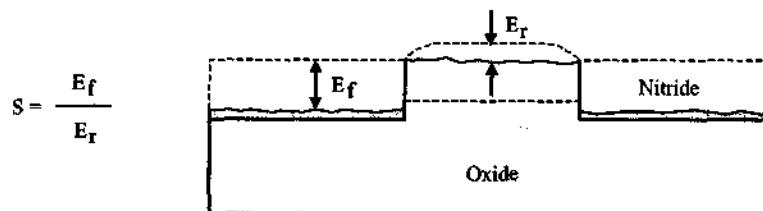


FIGURE 16.8 Etch Selectivity

The selectivity, S_R , for the film undergoing etch and the mask layer (e.g., photoresist) can be calculated from the following formula,

$$S_R = \frac{E_f}{E_r}$$

Where, E_f = the etch rate of the film undergoing etch
 E_r = the etch rate of the masking layer (e.g., photoresist)

Based on this formula, selectivity is often expressed as a ratio. A process with poor selectivity may be 1:1 (meaning the film being etched is removed as quickly as the photoresist mask), whereas a good selectivity could be 100:1. Interpret this as the etched film is removed 100 times faster than the film not being etched (e.g., photoresist).

Dry etching frequently does not provide adequate etch selectivity to the layer underneath. In this case, a plasma reactor should be equipped with an endpoint detection system that signals when it is time to stop the etch process with minimal overetching. The endpoint detector notifies the etch equipment controller when the underlying film is beginning to be exposed in order to stop the etch process.

Uniformity

Etch uniformity is a measure of the capability of the process to etch evenly across the entire surface area of the wafer, across the entire wafer lot, and from lot to lot. Uniformity is closely related to selectivity, since nonuniformity results in additional overetch. Maintaining uniformity across the wafer surface is key to ensuring consistent manufacturing performance. The challenge is that etching must be uniform in different types of wafer surface pattern density, such as densely populated wafer areas, large open spaces, and within high-aspect ratio features. Some problems in uniformity occur because etch rates and profiles depend on feature size and pattern density.¹ The etch rate is slower in small openings, to the point that etching can actually stop in small geometries with high aspect ratios. For instance, silicon trenches with a high-aspect ratio opening etch slower than trenches with a small-aspect ratio opening (see Figure 16.9). This phenomenon is known as *aspect ratio dependent etching (ARDE)*, also referred to as *microloading*. The objective is to minimize ARDE across the wafer surface in order to improve uniformity.

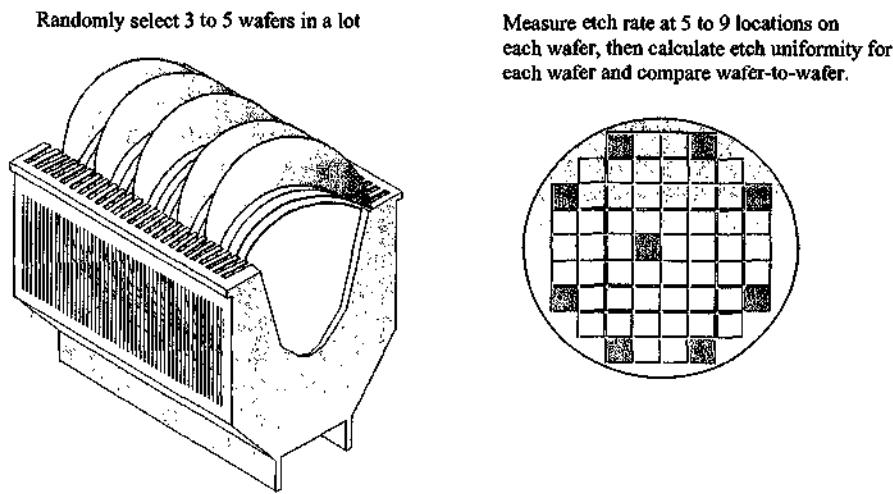


FIGURE 16.9 Etch Uniformity

Residues

Etch residue is the unwanted material remaining on the wafer surface after etch. It often coats the walls of the chamber and the bottom of the feature being etched. It can occur for reasons such as

contaminants in the etched film, improperly chosen etch chemistry (e.g., etching too fast), contaminants in the chamber, and nonuniform dopant distribution in the film. There are different names for the residues left after etch, including stringers, veils, crowns, and fences. Stringers are small residues of the etched material that are not totally removed, are electrically active, and can form an undesirable connecting short between features. Residues are a source of wafer contamination for IC fabrication and can cause problems during resist stripping. An overetch is sometimes done at the end of the etch process to remove residues. In some cases the etch residues can be removed by a resist strip process or by wet chemical etching.

Polymer Formation

A *polymer formation* is sometimes intentionally deposited on the sidewalls of the etch feature to form an etch-resistant film that prevents lateral etching (see Figure 16.10). This produces highly anisotropic features because it blocks the etching of the sidewall, thus increasing the etch directionality. The result is better control of critical dimensions of patterned structures. The polymers come from photoresist carbon converted into polymers during etching and combines with etching gases (i.e., C_2F_4) and etch by-products to form this sidewall polymer.² The need for sidewall polymers depends on the type of etching gas used.

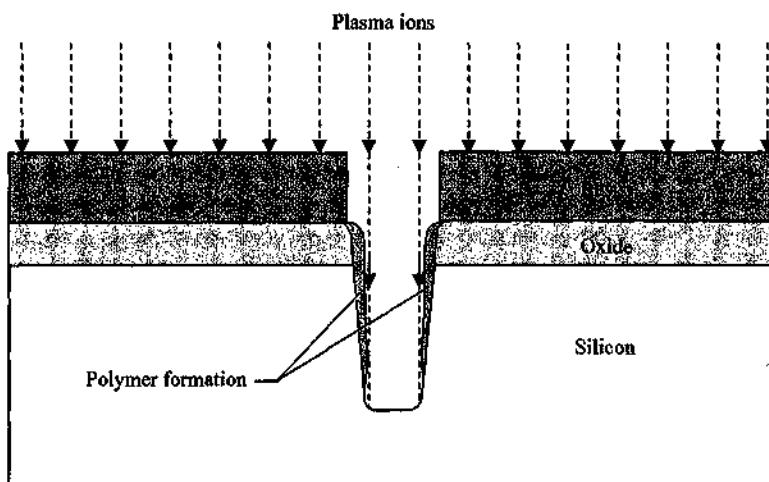


FIGURE 16.10 Polymer Sidewall Passivation for Increased Anisotropy

These sidewall polymers are complex, containing etchant elements and reaction by-products such as aluminum, titanium from barrier layers, oxides, and other inorganic materials. The polymer chains have strong carbon-fluorine bonds that are difficult to oxidize and remove.³ However, the polymers must be removed after the etch process or device yield and reliability is affected. The cleaning of the sidewall often requires special gas chemistry for a plasma stripping process or possibly a wet clean using strong solvents followed by DI-water rinsing (see the following section).

Unfortunately, an undesirable side effect of polymer deposition is that the internal parts of the process chamber also become coated with polymer. Etch process chambers require periodic cleaning to remove the polymer and to replace parts that can not be cleaned.

Plasma-Induced Damage

Plasma consists of energized ions, electrons, and excited molecules, which can cause *plasma-induced damage* to sensitive devices on the wafer surface. A major type of damage is nonuniform plasma that creates trapped charges on the gate electrode of a transistor, causing a breakdown of the thin gate oxide.⁴ Plasma becomes nonuniform due to poor equipment design or from operating the plasma reactor outside of the optimum process window. Another type of device damage is from energetic ion bombardment to gate oxide that is directly exposed to the ions. This damage could occur at the edge of a gate electrode during etching. Plasma damage is sometimes removed from the wafer through anneals or chemical wet etching.

Particle Contamination

Wafer damage from plasma can also come from particle contamination generated by the plasma near the wafer surface. Studies have shown that particles are trapped near the plasma/sheath interface because of the electrical potential difference.⁵ When the plasma is turned off, these particles fall onto the wafer surface. Plasma based on fluorine gas chemistries produces fewer particles than chlorine or bromine gas chemistries because fluorine generates etch by-products with a higher vapor pressure. Control of particle contamination is done by an optimized tool design, proper tool operation and shutdown, and use of the appropriate gas chemistry for the film being etched.

DRY ETCH

Dry etch is the primary etching method used to remove surface material in semiconductor manufacturing. The goal of dry etch is to reproduce the image of a mask on the wafer surface with a high degree of integrity. The advantages to using dry etch over wet etch are listed in Table 16.2.

TABLE 16.2 Advantages of Dry Etch Over Wet Etch

Description of Advantages
1. Etch profile is anisotropic with excellent control of sidewall profiles.
2. Good CD control.
3. Minimal resist lifting or adhesion problems.
4. Good etch uniformity within wafer, from wafer-to-wafer, and from lot-to-lot.
5. Lower chemical costs for usage and disposal.

There are disadvantages to using dry etch. The primary disadvantages are poor selectivity to the underlying layer, risk for device damage from plasma, and expensive equipment.

In the dry etching process, a low-pressure plasma discharge is used to remove material in small feature sizes of integrated circuits (the creation of a plasma was discussed in Chapter 8). The plasma interacts with the wafer surface to cause etching action and the subsequent removal of the surface material. The major actions that occur for plasma etching of a substrate material are shown in Figure 16.11.⁶

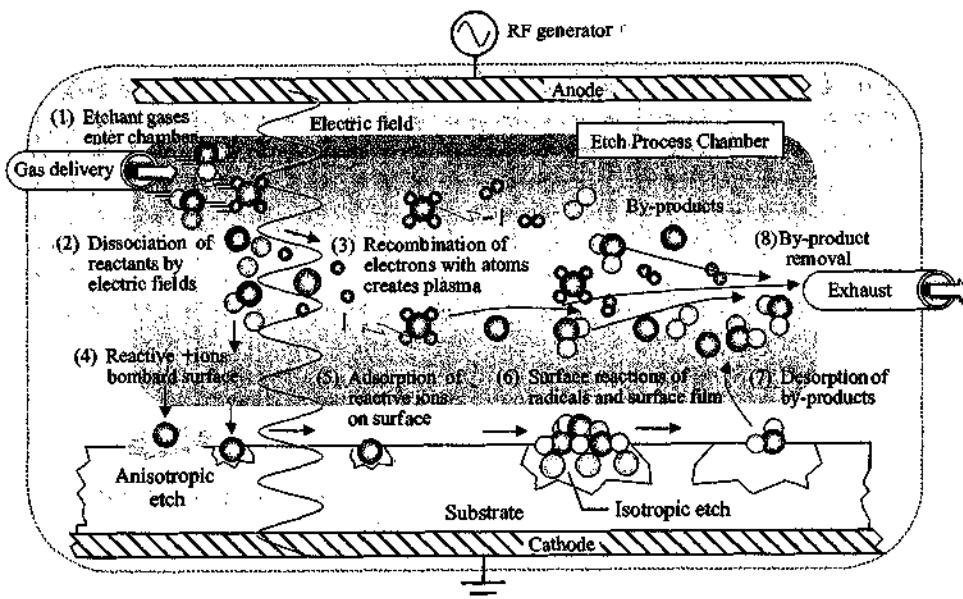


FIGURE 16.11 Plasma Etch Process of a Silicon Wafer

Etching Action

Etching action in dry etch systems is achieved by either a chemical or a physical technique or a combination chemical/physical technique (see Figure 16.12). In a purely *chemical mechanism*, the plasma creates reactive species (free radicals and reactive atoms) that chemically react with the materials on the wafer surface. The gases introduced into the chamber (typically containing chlorine and/or fluorine) are carefully selected to attain high selectivity (that is, to minimize the chemical reaction with the photoresist mask and underlying wafer layer). The chemical etching component of the plasma produces poor CD control because of its isotropic profile. Volatile by-products of the reaction are removed by the low-pressure pumping system.

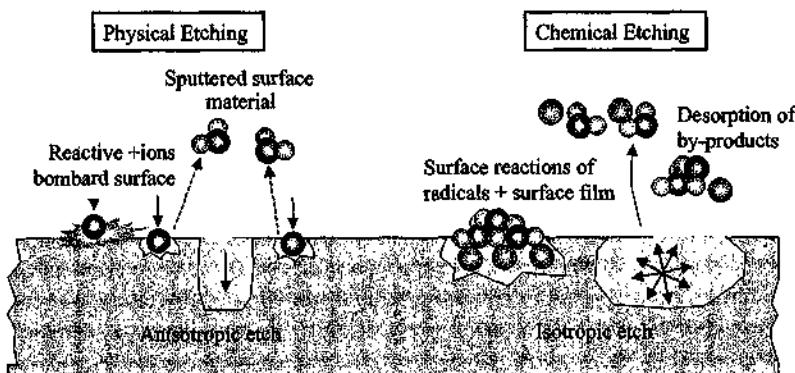


FIGURE 16.12 Chemical and Physical Dry Etch Mechanisms

To achieve etching action with a *physical mechanism*, the plasma provides energetic species (bombarding positive ions) that are accelerated toward the wafer surface by strong electric fields. The ions physically remove the unprotected wafer surface material by a sputter-etch action. Typically a nonreacting gas such as argon (Ar) is used. A benefit to this mechanical etching approach is the strong directionality of the etch, making it possible to achieve highly anisotropic profiles (near-vertical sidewalls) for good CD control. Sputter etching produces a high etch rate; however, it does have poor selectivity. Another problem is that the species removed by sputtering are not volatile and may redeposit back on the wafer, causing particulate and chemical contamination.

There is also a *combined physical and chemical mechanism* where ion bombardment improves the chemical etching action. The etch profile is varied from isotropic to anisotropic by adjusting the plasma conditions and gas composition. Combined physical and chemical etch produces good CD control with fair selectivity and is often preferred for most dry etch processes. Table 16.3 summarizes the different etch parameters for chemical, physical, and combined chemical/physical etching. The advantages from each type depend on the objectives of the etch process.

Note that dry etch systems may be designed to operate either as isotropic or anisotropic etchers depending on the direction of the RF electric field relative to the wafer surface. This means that positive ion sputtering can occur respectively on the wafer surface or on the edge of the wafer. If the field is perpendicular to the wafer surface, then the etching is done by a combination of heavy positive ion sputtering and some radical chemical reactions. If the field is parallel to the surface of the wafer, very little physical sputtering occurs, so the etching is done primarily by chemical reactions between radicals and the surface material.

Potential Distribution

The *plasma potential distribution* in the glow discharge region of a plasma has a strong effect on the etching capability of the system. This is because the amount of particle energy that bombards the etched surface depends on the electrical potential distribution. Figure 16.13 shows a reactor glow discharge between two electrodes. The power electrode has RF power applied, while the ground electrode is at ground potential. The potential of the plasma is positive relative to the

grounded electrode (which is also connected to the reactor walls, making them grounded as well). The plasma region has the most positive potential in the system.

TABLE 16.3 Chemical Versus Physical Dry Plasma Etching

Etch Parameter	Physical etch (RF field perpendicular to wafer surface)	Physical etch (RF field parallel to wafer surface)	Chemical etch	Combined Physical and Chemical
Etch Mechanism	Physical ion sputtering	Radicals in plasma reacting with wafer surface*	Radicals in liquid reacting with wafer surface	In dry etch, etching includes ion sputtering and radicals reacting with wafer surface
Sidewall Profile	Anisotropic	Isotropic	Isotropic	Isotropic to anisotropic
Selectivity	Poor/difficult to increase (1:1)	Fair/good (5:1 to 100:1)	Good/excellent (up to 500:1)	Fair/good (5:1 to 100:1)
Etch Rate	High	Moderate	Low	Moderate
CD Control	Fair/good	Poor	Poor to nonexistent	Good/excellent

*Used primarily for stripping and etchback operations.

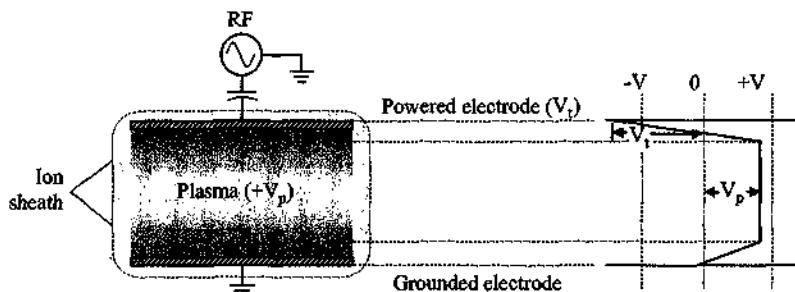


FIGURE 16.13 Schematic View of Reactor Glow Discharge with Potential Distribution

The powered electrode in an etch reactor develops a negative self-bias (DC self-bias or DC bias) voltage relative to ground because fast-moving electrons leave the plasma to strike the electrode. After a certain amount of negative charge, electrons are repelled from the electrode, thus creating the dark space (ion sheath) region with positive ionic charge. The magnitude of the powered electrode self-bias voltage depends on the amplitude and frequency of the RF voltage applied to the electrodes.⁷ If the electrodes are of similar area, the potential difference across the dark space of each electrode will be the same. Since the powered electrode develops a negative self-bias, the plasma must assume a positive potential to produce a potential of equivalent magnitude at the grounded electrode. In essence, the plasma forms a compensating positive potential.

There are etch trends that are affected by parameters of the etch process. If the RF frequency is reduced, then ions efficiently cross the plasma dark space in a small fraction of an RF cycle. This action increases the ion energy and etch rate. The amount of ion bombardment depends on the size of the electrodes. For asymmetrical electrode size, if the area of the powered electrode is small, then the positive plasma potential is small. This condition results in a larger dark space potential at the powered electrode and a high-energy ion bombardment of the powered electrode surface.⁸ Some basic trends for etching process parameters are provided in Table 16.4 on page 446.

TABLE 16.4 Effects of Changing Plasma Etch Parameters

Increase (\uparrow) or Decrease (\downarrow) in Etch Control Parameters	Ion Energy	DC Bias	Etch Rate	Selectivity	Physical Etch
RF Frequency	\uparrow \downarrow	\downarrow	\downarrow	\uparrow	\downarrow
	\downarrow \uparrow	\uparrow	\uparrow	\downarrow	\uparrow
RF Power	\uparrow \downarrow	\uparrow	\uparrow	\downarrow	\uparrow
	\downarrow \uparrow	\downarrow	\downarrow	\uparrow	\downarrow
DC Bias	\uparrow \downarrow	\uparrow	\uparrow	\downarrow	\uparrow
	\downarrow \uparrow	\downarrow	\downarrow	\uparrow	\downarrow
Electrode Size	\uparrow \downarrow	\downarrow	\downarrow	\uparrow	\downarrow
	\downarrow \uparrow	\uparrow	\uparrow	\downarrow	\downarrow

PLASMA ETCH REACTORS

The basic components of a plasma dry etch system include: a reaction chamber where the etching takes place, an RF power supply to ignite the plasma, a gas flow control system, and a vacuum system to remove etch by-products and gases. The etch system includes sensors, gas flow control units, and an endpoint detector. A wide range of gas chemistries is used in etch (see the following section), but in general, fluorine etches SiO₂; chlorine and fluorine etch aluminum; chlorine, fluorine, and bromine etch silicon; and oxygen removes photoresist. Different parameters controlled in dry plasma etching are: vacuum operation, gas mixture, gas flow rate, temperature, RF power, and the wafer position relative to the plasma. The interaction of these different parameters is a function of the dry etch process controller.

Prior to the 1980s, most plasma etch equipment used a barrel reactor designed for batch processing (many wafers processed simultaneously). This type of reactor is no longer common in advanced IC production and is only used in noncritical applications. The current trend in semiconductor manufacturing is for single-wafer processing in integrated cluster tools. This technique achieves manufacturing efficiency by reducing the batch size for single-piece flow. Integrated cluster tools also reduce exposure to contamination between process steps and achieve better wafer uniformity because the reactor parameters are optimized for one wafer. The most important development in dry etch systems for sub-0.25 μm geometries is the high-density plasma reactor.

The different types of dry plasma reactors are:

- ◆ Barrel plasma etcher
- ◆ Parallel plate (planar) reactor
- ◆ Downstream etch systems
- ◆ Triode planar reactor
- ◆ Ion beam milling
- ◆ Reactive ion etch (RIE)
- ◆ High-density plasma etchers

Barrel Plasma Etcher

The *barrel reactor* is a cylindrical design with almost pure chemical isotropic etching at a pressure of about 0.1 to 1 torr (see Figure 16.14). Wafers are mounted vertically in a quartz boat with a small separation between wafers. The RF power is applied by placing electrodes on both sides of the cylinder. There is typically a perforated metal cylindrical etch tunnel that confines the plasma to the outer region between the etch tunnel and chamber wall. The wafers are placed parallel to the electric field to minimize physical etching. The etchant species in the plasma diffuse to the etch tunnel, while the energetic ions and electrons of the plasma do not enter this region. The etching is almost purely chemical with isotropic etching and high selectivity.⁹ There is minimal plasma-induced

damage because there is no physical sputtering on wafers. The barrel plasma reactor has been used primarily for stripping photoresist from wafers. Oxygen is the primary reactant for stripping photoresist.

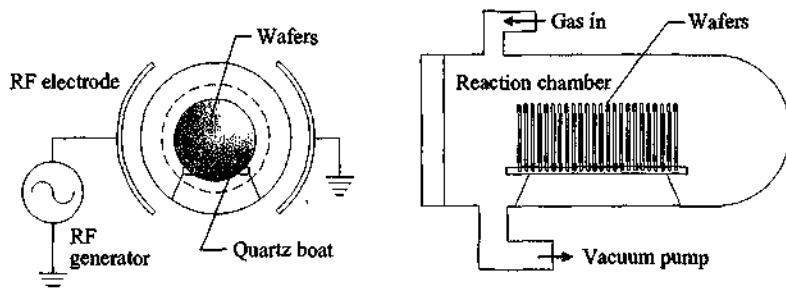


FIGURE 16.14 Typical Barrel Reactor Configuration

Parallel Plate (Planar) Reactor

The *parallel plate (planar) reactor* has two parallel plates that are symmetrical in size and position in the reactor (see Figure 16.15). A wafer can be placed backside down on the grounded electrode cathode with the RF signal applied to the upper electrode. This is the plasma etch mode, with energetic ion bombardment since the plasma potential is always above ground potential. If a wafer is placed directly on the RF-powered electrode, the wafer is in direct contact with the plasma and energetic ions. This contact can result in high-energy ion bombardment and is said to be in the reactive ion etch mode. Physical and chemical etch mechanisms occur in both the plasma etch mode and the reactive ion etch mode. However, energies of the bombarding ions are about ten times higher in the reactive ion etch mode.¹⁰ A reactive gas is needed to form the plasma, such as fluorine (F_2) or oxygen (O_2). The reactor operates at a pressure of about 0.1 to 1 torr, with high RF power to control the etch rate. The planar etcher was one of the early reactors. Single-wafer parallel plate etchers are commonly used today.

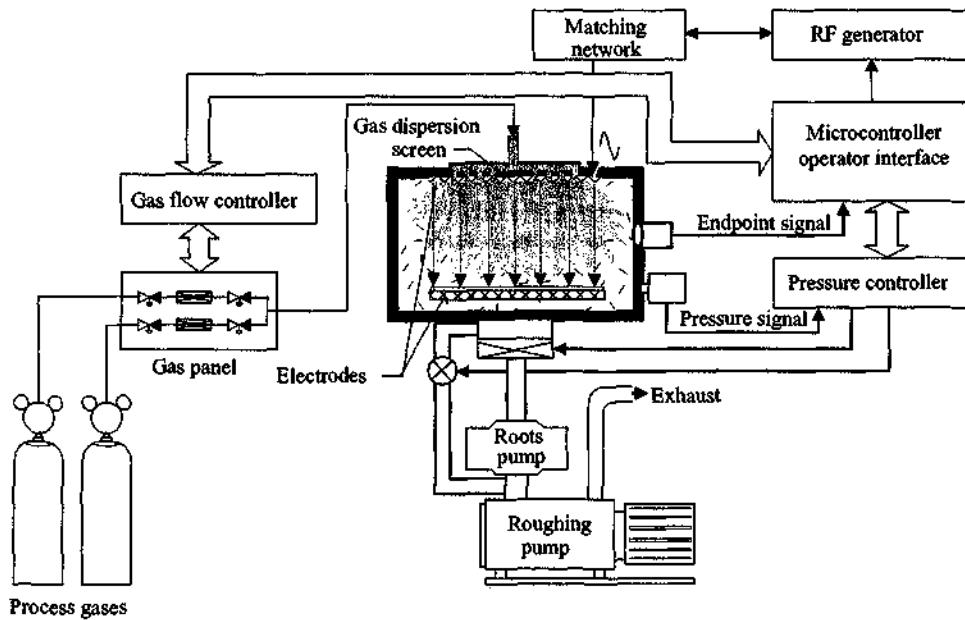


FIGURE 16.15 Parallel Plate Plasma Etching

Downstream Etch Systems

Repeated exposure to ion bombardment at the wafer surface increases the probability of device damage. A means of reducing damage to the wafer surface as well as heat buildup due to ion bombardment is to locate the wafer etch region away from the plasma in a *downstream reactor*. The plasma is formed in a separate source at a pressure of about 0.1 to 1 torr, transferred to the process chamber, and uniformly distributed over the heated wafer surface (see Figure 16.16). Since there are no ions to create directional etching, downstream reactors employ chemical etching, are isotropic, and are often used to remove resists or other noncritical layers (see the following section). A microwave source (2.45 GHz) for exciting the plasma for downstream etching is common because it produces the maximum concentration of atomic oxygen and the lowest concentration of ionic oxygen, which minimizes device damage and yields a high strip rate.

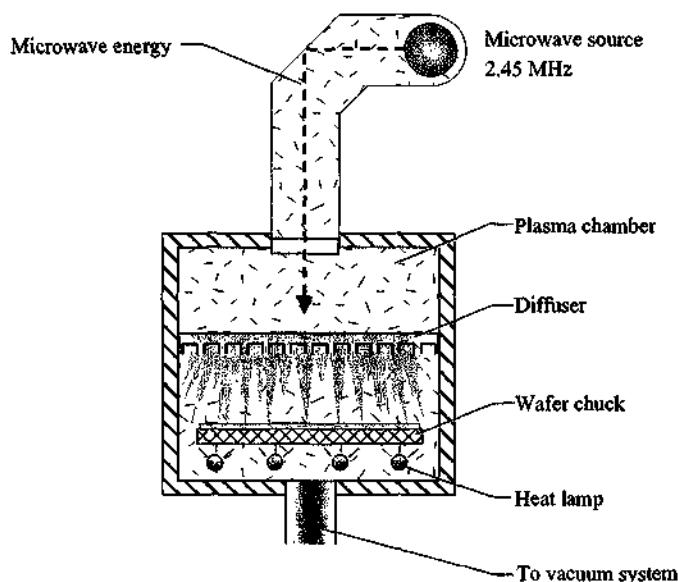


FIGURE 16.16 Schematic of a Downstream Reactor

There are variations of the downstream etching concept used with different etch systems discussed in the following section. The objective is to minimize wafer exposure to plasma ions in order to reduce or eliminate plasma-induced damage.

Triode Planar Reactor

The *triode planar reactor* adds a third electrode to attain control over the amount of ion bombardment. Figure 16.17 shows a setup with two power supplies. The inductively-coupled RF generates the plasma to create the ions and reactive species at a pressure of about 10^{-3} torr. The low-frequency generator controls the ion bombardment. A typical use is in single-crystal silicon trench etching.

Ion Beam Milling

Ion beam milling, also called *ion beam etching (IBE)*, has a physical etch mechanism with a strongly directional plasma. It is capable of anisotropic etching of small features. The plasma is commonly generated using an RF inductively-coupled source or a microwave source. Fast-moving electrons are emitted by a hot filament. Argon atoms enter the plasma chamber through a diffuser screen. An electromagnet surrounds the plasma chamber. The magnetic field causes electrons to travel in circular paths. This cyclical motion produces a high number of collisions with argon atoms, which results in a high number of positive argon ions. Positive argon ions are drawn out of the plasma source with grid electrodes and formed into a high-density beam using a set of collimated electrodes (see Figure 16.18). A high-voltage accelerator grid boosts ion energy as high as 2.5 keV.

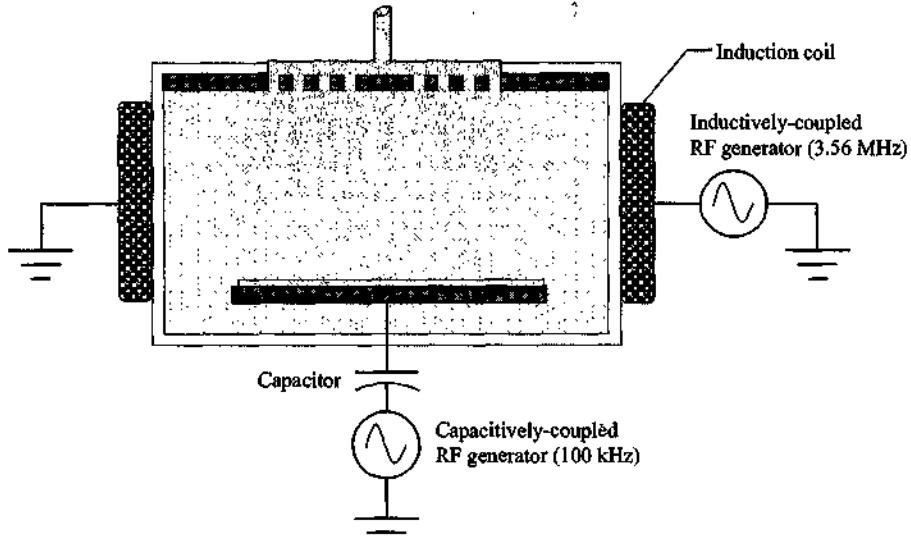


FIGURE 16.17 Triode Planar Reactor

A neutralizing filament emits electrons to recombine with argon atoms to prevent charging of the wafer with the positive ions. Ion beam etchers operate with argon (Ar) gas in the low-pressure range of 10^{-4} torr, which is lower than that commonly used for high-density plasma etching. Ion beam etching is used to etch difficult materials, such as gold, platinum, and copper.¹¹ The wafer can be tilted to produce variable sidewall geometries. The major problems inhibiting widespread use of ion beam etchers in semiconductor processing are low selectivity (usually below 3:1) and low etch rates, which lead to low wafer throughput.

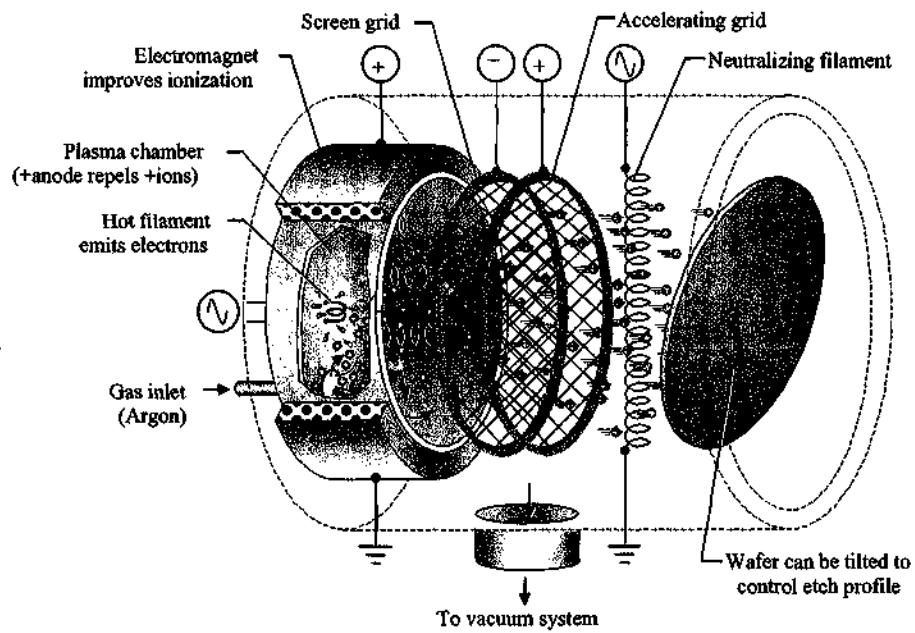


FIGURE 16.18 General Schematic of Ion Beam Etcher
Redrawn from *Advanced Semiconductor Fabrication Handbook*, Integrated Circuit Engineering Corp., pp. 8-12.

Reactive Ion Etch (RIE)

Reactive ion etch (RIE) is a technique for removing material from the wafer surface with both a reactive chemical process and a physical process using ion bombardment. RIE is similar to the standard parallel-plate plasma etcher, except the wafers are placed on the RF-powered electrode (cathode) and the powered electrode size is greatly reduced relative to the grounded electrode size (see Figure 16.19). In this manner a DC self-bias develops on the cathode and the wafers acquire a large voltage difference with respect to the plasma. This condition creates directionality for the ionized species moving toward the wafer, creating improved anisotropic sidewalls for features. There is no sputtering on the anode. The pressure is relatively low at <0.1 torr.

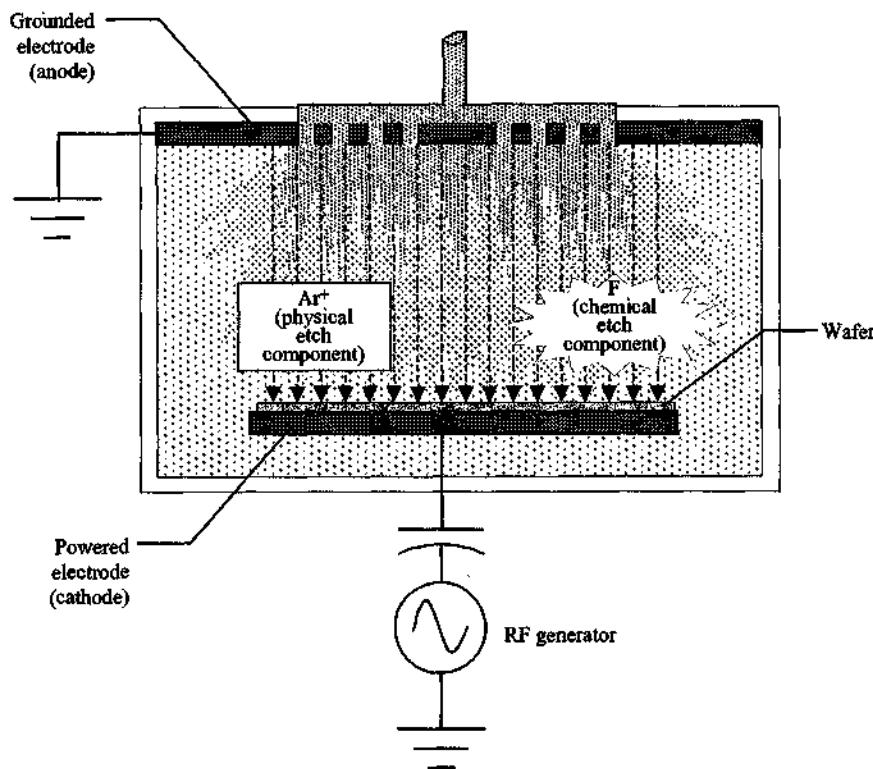


FIGURE 16.19 Parallel-Plate RIE Reactor

High-Density Plasma Etchers

The most predominant dry etching methodology in use for critical layers of advanced ICs is *high-density plasma etch* with single-wafer processing in a cluster tool.¹² Standard plasma systems previously used in wafer fabrication operated in a vacuum regime of a few hundred mtorr, where it is relatively straightforward to create a plasma (such as by applying an RF signal across two parallel plates). However, for 0.25-μm geometries and below, it has become too difficult to get etchant ions into and etch by-products out of high-aspect ratio features. The etching action slows down and will actually stop altogether at the bottom and lower sides of the feature profile.

The solution is to lower the system pressure to about 1 to 10 mtorr to increase the mean free path lengths of gas molecules and ions. This condition effectively reduces collisions that cause loss of profile control. However, a drawback to reduced pressure is the lower etch rate due to the rapid reduction in ion density as pressure decreases. To overcome this concern, a high-density plasma is needed to generate enough ions for an acceptable etch rate with a lower pressure. High-density refers to the number of active species in the plasma relative to conventional plasma at the same process pressure. In conventional plasma, the degree of ionization is typically on the order of 0.01% to 0.1%. High-density plasma more efficiently couples input power with the plasma, resulting in greater dissociation of etch species to achieve a degree of ionization as high as 10%. This technique yields highly directional, low-energy ions that produce anisotropic etching in high aspect ratio openings.



High-Density Plasma Etcher
(Photo courtesy of Applied Materials, Inc.)

High-density plasma etchers generally immerse the plasma in a magnetic field. The reasons for using a magnetic field during plasma etch are: (1) plasma generation is more efficient for creating highly-directional, low-energy ions entering high-aspect ratio openings, yet causes less wafer damage; (2) plasma is denser, with more reactive species and charged particles that increases the etch rate; and (3) the DC bias on the wafer can be reduced, which leads to less wafer bombardment (or damage).

Electron Cyclotron Resonance (ECR) ■ The *electron cyclotron resonance (ECR)* reactor was one of the earliest high-density plasma reactors to be developed commercially, first introduced in the early 1980s. It is still used in today's modern wafer fabs for etching 0.25- μm feature sizes and below. The ECR reactor produces very dense plasmas at operating pressures of about 1 to 10 mtorr. ECR etching uses microwave excitation (2.45 GHz) in the presence of a magnetic field to generate high-density plasma (see Figure 16.20 on page 452). A key feature of the ECR reactor is a magnetic field parallel to the direction of reactant flow that causes free electrons to move in a spiral path due to the magnetic force. An efficient transfer of energy occurs from the electric field to the plasma electrons when the electron orbit frequency (known as electron cyclotron resonance) equals the frequency of the applied electric microwave field. This resonance condition increases the probability of electron collisions, which generates a very dense plasma to create a large flux of ions. The reactive ions move toward the wafer surface and react with the surface layer to cause etching.

A low-power RF bias (13.56 MHz) or DC bias can be applied to the electrode holding the wafer to control the energy of the ions striking the surface. This action permits the ECR to operate as a combined chemical and physical etch process and produces an anisotropic etch profile. The main drawback to the ECR reactor for semiconductor manufacturing is its equipment complexity.¹³ A variation of the ECR is a system with pairs of magnets and microwave antennas distributed around a central reactor, known as distributed ECR (DECR).

Inductively-Coupled Plasma (ICP) ■ Another high-density, low-pressure etch reactor that has plasma decoupled from the wafer is the *inductively-coupled plasma (ICP)* reactor. This reactor is less complicated and less expensive than the ECR and widely used in the United States. The ICP generates plasma by means of a spiral coil separated from the plasma by a dielectric plate or quartz tube (see Figure 16.21 on page 452).¹⁴ The wafer is located away from the coil so it is not affected

by its electromagnetic field. The wafer can be biased to have both chemical and physical etching. This reactor can achieve anisotropic sidewall profiles in high-aspect ratio openings.

Another inductively-coupled reactor that generates a high-density plasma is the *helicon wave*. This system receives power from an RF signal (13.56 MHz) that is inductively coupled into the plasma from a double-loop antenna located outside a quartz source tube.

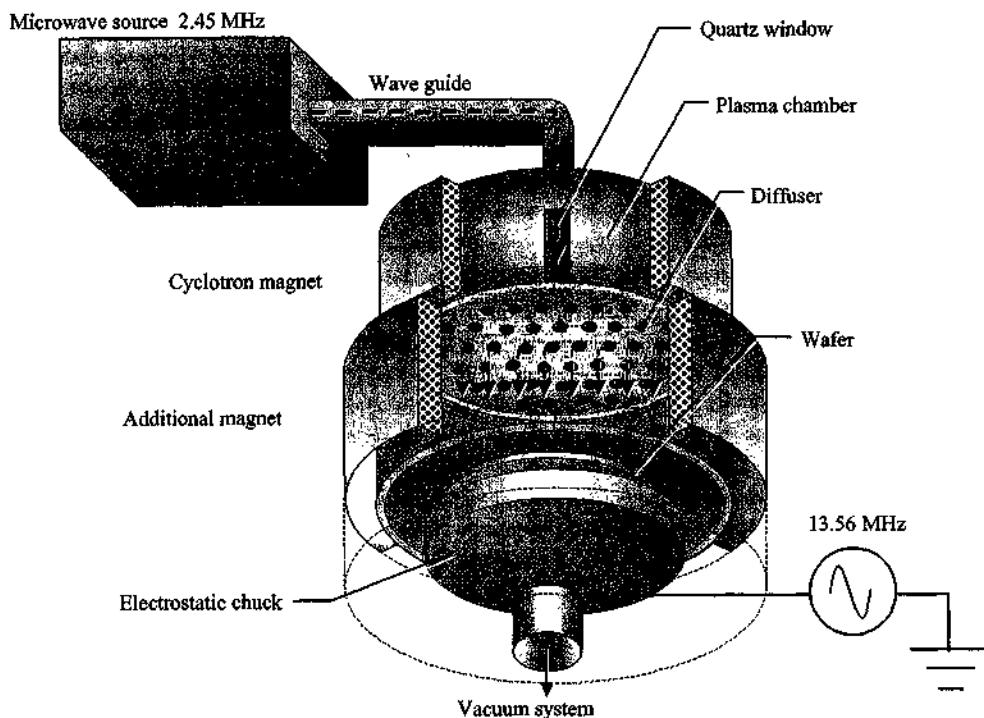


FIGURE 16.20 Schematic of Electron Cyclotron Reactor
Redrawn from Y. Lii, "Etching," *ULSI Technology*, ed. by C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 349.

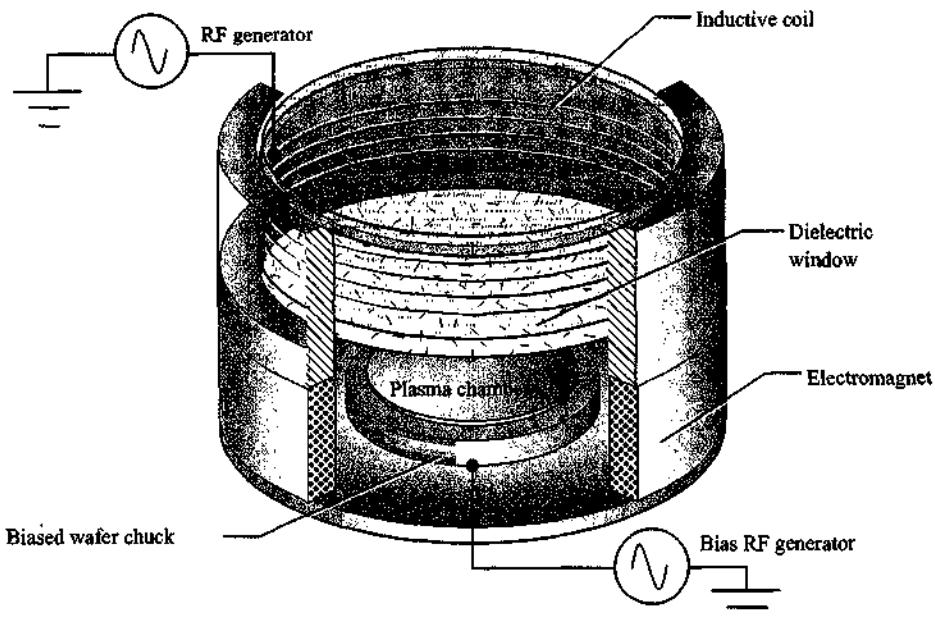


FIGURE 16.21 Inductively-Coupled Plasma Etch
Redrawn from Y. Lii, "Etching," *ULSI Technology*, ed. by C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 351.

Dual Plasma Source (DPS) ■ Figure 16.22 is a schematic of a *dual plasma source (DPS)* etch chamber (also referred to as decoupled plasma source). It has four major components: a source power unit, an upper chamber, a lower chamber with the wafer, and a movable electrode.¹⁵ As with the previous high-density plasma systems, there are two sources of RF power used. The source power unit has an inductive coil for transferring RF power to the plasma in order to generate the reactive ions and neutral species. This is referred to as the source power. The source power unit has a temperature control system. RF power is also supplied to the wafer electrode to bias the wafer substrate and is referred to as the bias power. The wafer cathode is movable in the vertical direction. Only the upper chamber is exposed to the plasma and process gases, which keeps the lower chamber clean and makes it much easier to maintain.

A key aspect of the DPS plasma is the decoupling of the source plasma power from the bias power. This arrangement permits greater control over the ion density and ion energy, which results in a larger process window for physical and chemical etching. The result is improved critical dimension control and less etch residue.

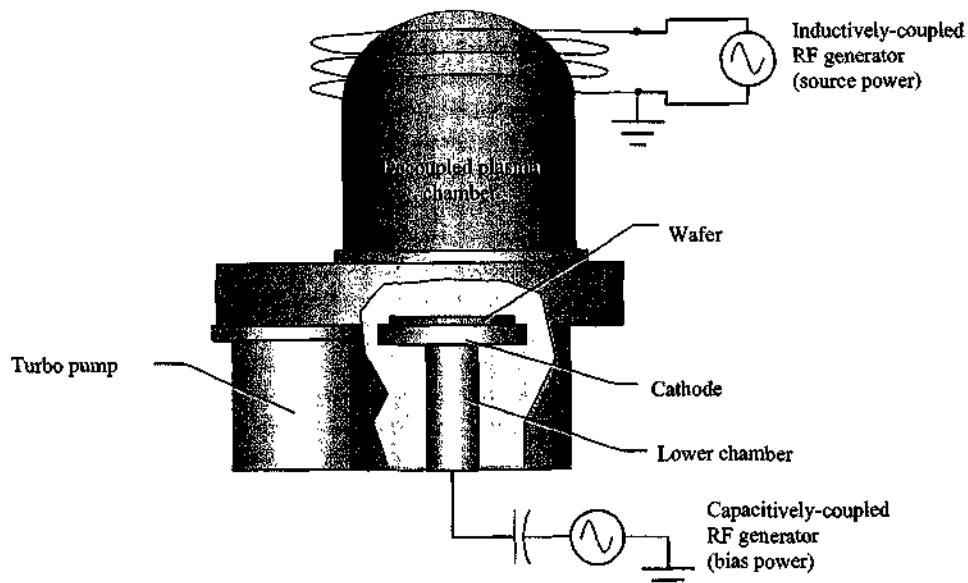


FIGURE 16.22 Dual Plasma Source (DPS)

Redrawn from Y. Ye et al., *Proceedings of Plasma Processing XI*, Vol. 96-12, ed. G. Mathod and M. Meyyappan (Pennington, NJ: The Electromechanical Society, 1996): p. 222.

Magnetically Enhanced RIE (MERIE) ■ The *magnetically enhanced RIE (MERIE)* reactor (also referred to as a *magnetron*) is a combined physical and chemical etch system. It is similar to the RIE reactor except that now there is a magnetic field that holds the plasma away from the chamber walls and increases the electron and ion concentration near the wafer to create a high-density plasma (see Figure 16.23 on page 454). The magnetic field can be rotated electrically by its three-phase AC power supply or physically rotated in a dipole ring magnet system. Confining the plasma with the magnetic field creates a high-density plasma and permits lower pressure, effectively maintaining etch directionality and uniformity, especially when etching high-aspect ratio features.

Etch System Review

Shrinking feature sizes and the introduction of new wafer materials place stringent demands on etch performance. Dry etch systems use physical or chemical or a combination of these two mechanisms to etch material. Some equipment is anisotropic while other equipment is purely isotropic. One can appreciate the wide variety of etch equipment that is available for wafer manufacturing. The capability and control of a particular etch system is critical for successful wafer fabrication. Table 16.5 on page 454 summarizes the important characteristics for each equipment system.

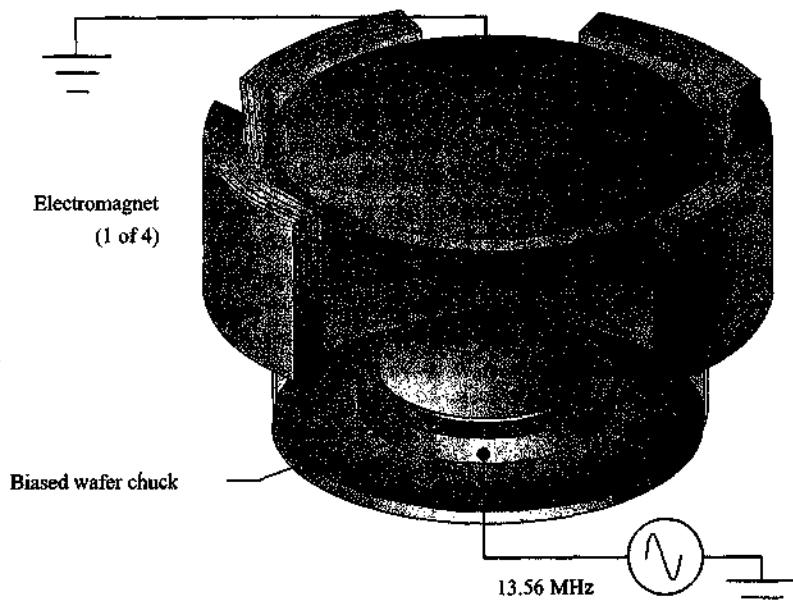


FIGURE 16.23 Magnetically Enhanced Reactive Ion Etch (MERIE)
Redrawn from *Wet/Dry Etch* (College Station, TX: Texas Engineering Extension Service, 1996), p. 165.

TABLE 16.5 Dry Etcher Configurations

Configuration	Etch Mechanism	Pressure (Torr)	Arrangement	High-Density Plasma	Biasing	Bias Source	Profile
Barrel	Chemical	10^{-1} to 1	Coil or electrodes outside vessel	No	In cassette (bulk)	RF	Isotropic
Parallel Plate (Planar)	Physical and Chemical	10^{-1} to 1	Planar diode (two electrodes)	No	On powered electrode (anode or grounded electrode)	RF	Anisotropic and isotropic
Downstream Plasma	Chemical	10^{-1} to 1	Coil or electrodes outside vessel	No	In cassette (bulk) downstream of plasma	RF or Microwave	Isotropic
Triode Planar	Physical	10^{-3}	Triode (three electrodes)	No	On powered electrode		Anisotropic
Ion Beam Milling	Physical and Chemical	10^{-4}	Planar triode	No	On powered electrode (anode)		Anisotropic
Reactive Ion Etch (RIE)	Physical	<0.1	Planar or cylindrical diode	No	On cathode		Anisotropic
Electron Cyclotron Resonance (ECR)	Physical	10^{-4} to 10^{-3} (low)	Magnetic field in parallel with plasma flow	Yes	On cathode	RF or DC	Anisotropic
Distributed ECR	Physical	(low)	Magnets distributed around central plasma	Yes	On cathode	RF or DC	Anisotropic

Configuration	Etch Mechanism	Pressure (Torr)	Arrangement	High-Density Plasma	Biasing	Bias Source	Profile
Inductively-Coupled Plasma (ICP)	Physical	(low)	Spiral coil separated from plasma by dielectric plate	Yes	On cathode	RF or DC	Anisotropic
Dual Plasma Source	Physical	(low)	Independent plasma and wafer biasing	Yes	On cathode	RF or DC	Anisotropic
Magnetically Enhanced RIE (MERIE)	Physical	(low)	Planar diode with magnetic field confining plasma	Yes	On cathode	RF or DC	Anisotropic

Endpoint Detection

Dry etch differs from wet etch in that it usually does not have good etch selectivity to the underlying layer. For this reason, a form of *endpoint detection* is required to monitor the etch process and stop etching to minimize overetching of the underlying layer. Endpoint detection systems measure different parameters, such as a change in the etch rate, the types of etch products removed from the etch process, or a change in the active reactants in the gas discharge (see Figure 16.24). A method used for endpoint detection is optical emission spectroscopy. This measurement tool is integrated into the etch chamber for real-time monitoring of the etch process.

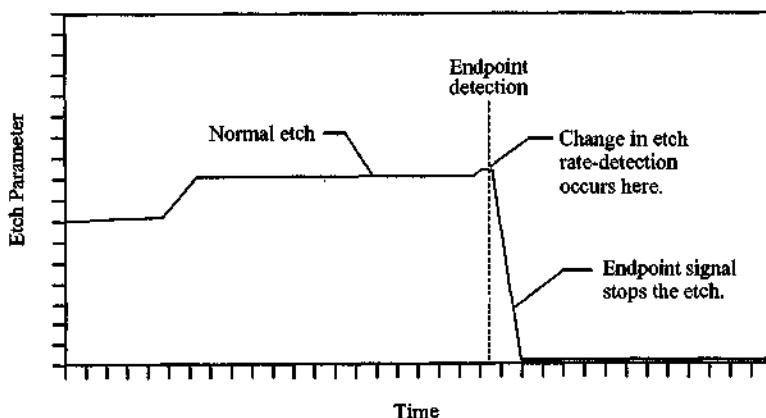


FIGURE 16.24 Endpoint Detection for Plasma Etching

Optical Emission Spectroscopy ■ An excited species emits light at a wavelength that corresponds to a specific material (see Table 16.6 on page 456).¹⁶ The light emitted by excited atoms and molecules in a gas discharge can be analyzed with optical emission spectroscopy to identify the element. The emitted light is passed through a detector with a filter to let light of a specific wavelength pass through that is used to identify the etched materials. The emission intensity is directly related to the relative concentration of a species in a plasma. In this manner the endpoint detector can determine when the etch process has removed the desired material and proceeded into the underlying layer. Optical emission spectroscopy is the most common method for etch endpoint detection because it is easy to implement with high sensitivity.¹⁷

Optical emission can also be used to perform etch reactor diagnostics. If a system has leaks to the atmosphere, the presence of nitrogen is identified. Water moisture can be detected, which could be caused by inadequate vacuum pumpdown after chamber cleans. Absolute concentrations of species can not be obtained by optical emission spectroscopy.

TABLE 16.6 Characteristic Wavelengths of Excited Species in Plasma Etch

Material	Emitting Species of Etchant Gas	Some Products	Wavelength (nm)
Silicon	CF ₄ /O ₂	SiF	440; 777
	Cl ₂	SiCl	287
SiO ₂	CHF ₃	CO	484
Aluminum	Cl ₂	Al	391; 394; 396
	BCl ₃	AlCl	261
Photoresist	O ₂	CO	484
		OH	309
		H	656
Nitrogen	N ₂ (used as a purge gas prior to and after etch)	N ₂	337
		NO	248

Vacuum for Etch Chambers

The vacuum system in an etch process is critical because it affects plasma parameters associated with gas flow and pressure. A typical high-density plasma vacuum system for etch will achieve a chamber pressure in the 1 mtorr range or less with gas flows as high as 800 sccm at the wafer surface. The vacuum components must be capable of handling the corrosive by-products produced during etching. The pumps used for an etch process are typically a turbopump, a roots-type blower and a dry backing pump. Typically the turbopump is located close to the chamber to maximize the pump speed at the wafer surface, while the other pumps are located in a service bay or sub fab area.¹⁸ Additional dry pumps are used to evacuate the loadlock and transfer chamber of an integrated cluster tool.

The fluorine, chlorine, and bromine gas chemistries used in etch (see the following section) are effective at removing the wafer surface films because of their reactivity. Unused reactants as well as reaction by-products are removed at a high rate. The high reactivity of etch by-products creates corrosion problems for vacuum system components, requiring special designs such as nitrogen purge of bearings and magnetically levitated (maglev) pumps that only rely on bearings during start-up and shutdown.¹⁹

DRY ETCH APPLICATIONS

There are many different types of plasma dry etching applications needed to fabricate integrated circuits. These applications are found in all materials used in the fab. For our purposes, we will review dry etch by the type of material to be etched: dielectric, silicon, and metal. The move to smaller critical dimensions, higher aspect ratio openings, and new materials in wafer fabrication creates challenges in etch processing for all three material types. Optimizing etch conditions leads to competing objectives for the production team. In general, the requirements for successful dry etch are:

1. High selectivity to avoid etching materials that are not to be etched (primarily photoresist and underlying materials).
2. Fast etch rate to achieve an acceptable throughput of wafers.
3. Good sidewall profile control.
4. Good etch uniformity across the wafer.
5. Low device damage.
6. Wide process latitude for manufacturing.

Critical etch parameters are determined through optimization activities for each specific dry etch application. Some of these parameters are shown in Figure 16.25. Note that in many cases optimization is being done through process equipment modeling using computer software. This is due to the high cost of conducting prototype hardware testing with actual wafers.

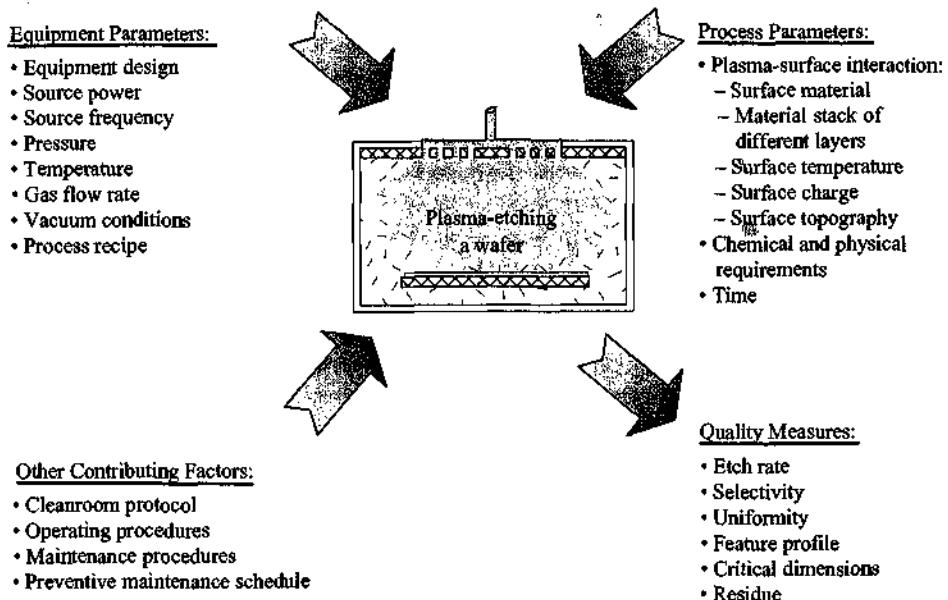


FIGURE 16.25 Dry Etch Critical Parameters

Table F.1 in Appendix F lists the typical gas chemistries used for plasma etching. Dielectric and silicon materials are usually etched with fluorine-based gas etchants such as CF_4 . Aluminum etching is typically done with chlorine-based gas chemistries such as BCl_3 .

Dielectric Dry Etch

Dielectric dry etch is the most complicated etch process for 200 mm wafers and will remain the biggest etch challenge for 300 mm wafers.²⁰ The challenges for oxide etching are tighter process specifications because of shrinking feature sizes, new trench etch processes for interlayer dielectrics (ILD) used in the dual-damascene approach for copper metallization, and the need to etch new ILD low- k materials.

Oxide ■ Common reasons for etching oxide are for forming contact holes and vias. These features are critical applications, with requirements such as etching high-aspect ratio openings in the oxide. Aspect ratios for DRAM applications are expected to be at 6:1 for 0.18 μm feature sizes. The required selectivity to underlying silicon and polycide is about 50:1.²¹ There are new oxide etch applications such as the dual-damascene structure, with new trench requirements and high-aspect ratio etch. There are also noncritical oxide etch applications such as a low-aspect ratio via etch.

The oxide plasma etching process is commonly based on fluorocarbon chemistry. Fluorocarbon is a fluorinated hydrocarbon. It has one or more of the hydrogen atoms in a hydrocarbon molecule replaced by fluorine (F). In this manner, many gases are sources of fluorine, such as CF_4 , C_3F_8 , C_4F_8 , CHF_3 , NF_3 , and SiF_4 . Common gases are tetrafluoromethane (CF_4), which has a high etch rate but poor selectivity with polysilicon, and trifluoromethane (CHF_3), which has a high rate of polymer formation. Fluorocarbon gases are chemically stable in their ground state (non-plasma state) and will not attack silicon or its oxide because their bonds are stronger than SiF bonds. There are also buffer gases such as Ar and He added to the etch chemistry. Argon has a relatively high mass that is used for physical etching (sputtering). Helium has a low mass and is used to diminish the etchant gas concentration (referred to as a diluent) to enhance plasma uniformity.

The desired fluorocarbon feed gas, which may have a primary gas component such as CF_4 , is fed into the plasma and dissociates (see Figure 16.26 on page 458). This reaction forms many different types of reactive ions (radicals) and neutrals.²² Examples of reactive species are CF_3^+ , CF_2^+ , CHF^+ , HF, and F. It is the radical F species that attacks the oxide and causes etching. The F reactive ions easily form volatile by-products, which are then pumped out of the etch reactor by the vacuum system. The CF_x radicals also serve to passivate the sidewall surfaces with a polymer. A higher carbon-to-fluorine ratio generally means more polymer formation, lower etch rate, and higher oxide-to-silicon

selectivity. In some cases, a silicon source is used above the plasma to getter excess fluorine atoms, thereby increasing the ratio of carbon to fluorine.

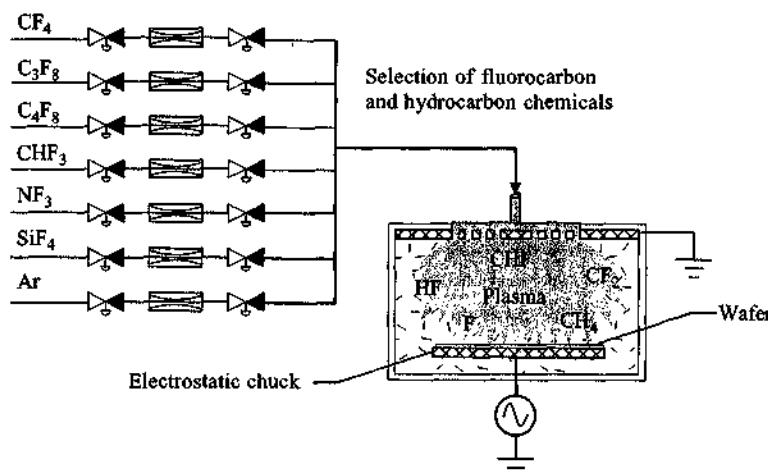


FIGURE 16.26 Oxide Etch Reactor

Underlying Material Selectivity. One of the major challenges for oxide etching is obtaining high selectivity to the underlying material, usually silicon, nitride, or an antireflective coating. For example, high selectivity to silicon is critical during contact etch through the Li oxide dielectric to avoid etching into the source/drain region. High selectivity to silicon is required when etching oxide sidewall spacers on the gate structure (see Chapter 9). For via etch, high selectivity to TiN, W, or Al is a requirement.

One method of achieving selectivity to silicon is by adding oxygen to the gas chemistry to control the selectivity between oxide and silicon. Small concentrations of O₂ improve the etch rate of both oxide and silicon. At greater concentrations of up to about 20% O₂, there is more rapid etching of oxide over silicon, improving the selectivity to minimize etching of the underlying silicon. Another way to improve selectivity is to add hydrogen to the gas mixture as the etch rate of silicon decreases until it reaches almost zero for a H₂ volume of about 40%. At the same time, the oxide etch rate is not affected by H₂ at concentrations below 40%.²³

Selectivity during oxide etch can be maintained to the underlying material by use of a hard “etch stop” masking layer of silicon nitride, as with etching the contact hole (see Figure 16.27). This action requires good oxide etch selectivity to the nitride layer. A hard mask approach to selectivity adds processing steps and is done only when necessary.

Example: Silicon nitride, SiN₃, serves as etch-stop during Li oxide etch
Note: The numbers show the order of the five operations.

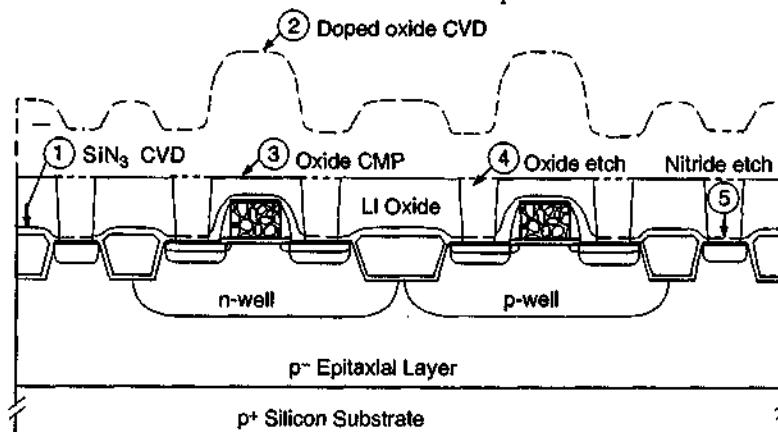


FIGURE 16.27 Etch Stop Hard Mask Layer

Oxide/silicon selectivity is also achieved by forming a passivating polymer layer on top of the silicon as part of the etch process. The polymer residue is introduced into the gas chemistry to inhibit lateral silicon etching. The polymer tends to form more easily on silicon rather than SiO_2 , which is thought to occur because the carbon combines with the oxygen from the SiO_2 to form volatile CO or CO_2 , which is then pumped away.²⁴ This approach to selectivity permits continued etching of the oxide while inhibiting the silicon etch. Unfortunately the polymer also deposits on the chamber surfaces during etch and becomes a source of particulate contamination, requiring frequent cleaning of the etch system.

Photoresist Selectivity. Achieving high selectivity to photoresist during oxide etch is important for avoiding tapered sidewalls. The photoresist defines the pattern that is being etched. In the case of contacts and vias, millions of holes are being etched simultaneously, each requiring the removal of a precise amount of surface material, often at different depths (see Figure 16.28). One factor that reduces photoresist selectivity (an undesirable situation) is the effective formation of aggressive fluorine atoms in a high-density plasma. Free fluorine will etch the organic photoresist. This lowers selectivity from about 10:1 from conventional plasma etchers to a range of 4:1 to 7:1 in the high-density plasma tools.²⁵ In addition, the need to etch through antireflective coating layers, which lengthens the etch time, further reduces the resist thickness. In general, deep UV resists are less resistant than i-line resists to plasma processing, which also lowers resist selectivity.

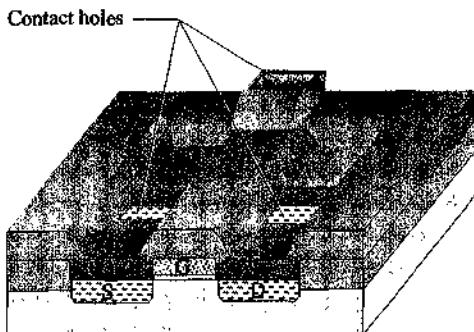


FIGURE 16.28 Contact Etching to Varying Depths

Sidewall Profile. Contact windows in the local interconnect (LI) oxide dielectric layer generally have dimensions equivalent to the smallest feature size with a high-aspect ratio. For this type of application, a high degree of anisotropy with a vertical sidewall profile is needed. The most significant factor is the highly directional ion bombardment of high-density plasma. A factor in obtaining optimum sidewall profile is the amount of resist selectivity. Low resist selectivity can lead to tapered sidewalls in an anisotropic etch process because of gradual photoresist erosion during the etch process.

Silicon Nitride ■ There are two basic types of silicon nitride used in wafer fabrication. One type is deposited with LPCVD at 700° to 800°C, which produces a stoichiometric compound of Si_3N_4 film. The other nitride film is deposited by plasma-enhanced CVD at <350°C that produces a film having lower density.²⁶ The etch rate is faster for the PECVD film because of its lower density.

Silicon nitride is etched by different gas chemistries. A common primary gas is carbon tetrafluoride (CF_4) that is mixed with O_2 and N_2 . Increasing the amount of O_2/N_2 dilutes the concentration of fluorine species and lowers the etch rate of any underlying oxide. A high etch rate of 1200 Å/min and a high selectivity to oxide would be about 20:1 for a nitride film deposited with LPCVD.²⁷ Selectivity to oxide is important in an application with a thin pad oxide used as an etch-stop, requiring the etch process to quickly slow down when oxide is reached. Other possible primary gases used for nitride etch are SiF_4 , NF_3 , CHF_3 , and C_2F_6 .

Silicon Dry Etch

Plasma dry etch of silicon is a critical process in wafer fabrication. The two major layers for plasma etching are (1) polysilicon gate length formation for the MOS gate structure, and (2) single-crystal

silicon trench creation used for either device isolation or capacitor structures in dynamic random-access memories (DRAM).

Polysilicon etch gas chemistries have traditionally been fluorine-based gases, including CF_4 , CF_4/O_2 , SF_6 , $\text{C}_2\text{F}_6/\text{O}_2$, and NF_3 . The fluorine atoms produce a very fast etch reaction but with an isotropic profile and average resist selectivity. The sidewall profile is improved by reducing the fluorine atoms and increasing the ion energy, but this also reduces the poly/oxide selectivity. The energy of the bombarding ions must be low enough to avoid sputtering away an underlying oxide layer.

To overcome these difficulties, polysilicon plasma etch gases are often based on chlorine, bromine, or chlorine/bromine chemistries. Chlorine (Cl_2) produces an anisotropic silicon sidewall profile with good oxide selectivity (>10 selectivity for poly/oxide and poly/nitride). Bromine gas chemistry, as either Br_2 or HBr , etches silicon anisotropically with very high selectivity (>100) to oxide and nitride. This quality is important for applications such as poly gate etching over the gate oxide.²⁸ Bromine gas chemistry also has better selectivity to photoresist than Cl_2 . Another silicon etch chemistry is a blend of chlorine and bromine, such as HBr and Cl_2 with the addition of O_2 . Adding O_2 increases the etch rate and the selectivity to oxide. Polymers form during etching with chlorine, and fluorine gas chemistries to deposit on sidewalls and contribute to sidewall profile control. SiF_4 , SiCl_4 , and SiBr_4 are some volatile etch by-products for fluorine, chlorine, and bromine gas chemistries.

Poly Gate Etch ■ In MOS devices, doped LPCVD polysilicon is used as the gate conductor material. The linewidth of the doped polysilicon conductor defines the gate length in the active device and therefore affects transistor performance (see Figure 16.29). For this reason, CD control is critical. The poly gate etch process must have high selectivity to the underlying gate oxide with excellent uniformity and repeatability. A high degree of anisotropy is also required since the polysilicon line also serves as an implant barrier during the source/drain doping step. Sloped sidewalls would yield partial doping of the implant species under the polysilicon gate structure.

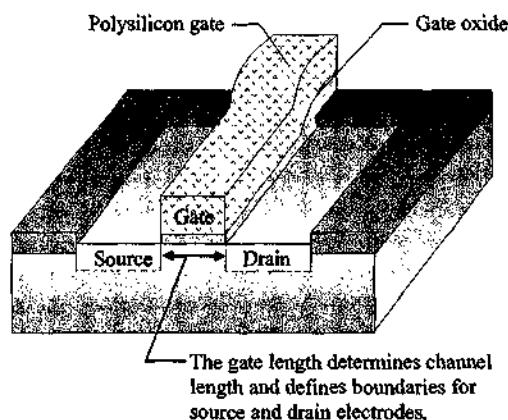


FIGURE 16.29 Polysilicon Conductor Length

Etching polysilicon (and silicon in general) is often a three-step process. This permits optimization for either anisotropy or selectivity during the different etch steps. These three steps are:

1. The first step is a *breakthrough step* that removes the native oxide, hard mask (e.g., SiON) and surface contaminants to achieve a uniform etch (this minimizes etching surface defects from contaminants that act as a micro-mask).
2. This is followed by a *main-etch step* to endpoint. This step removes most of the polysilicon film without damaging the gate oxide and achieves the desired anisotropic sidewall profile.
3. The final step is an *overetch step* that removes the remaining residues and stringers of polysilicon while maintaining high selectivity to the gate oxide. This etch process should be done while avoiding any microtrenching (formation of small trenches) of the gate oxide around the periphery of the polysilicon (see Figure 16.30).

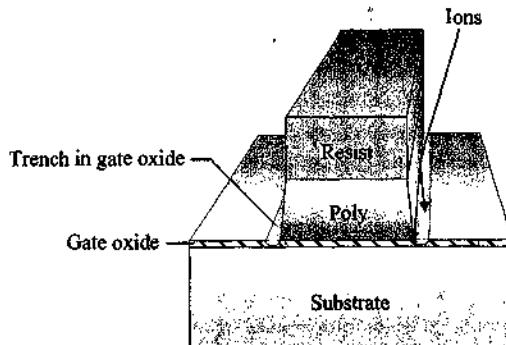


FIGURE 16.30 Undesirable Microtrenching During Polysilicon Gate Etching

The polysilicon gates are difficult structures to etch—requiring care and precision in the etch process. Devices with a $0.15\text{ }\mu\text{m}$ -feature size will have a projected gate oxide thickness of 20 to 30 Å (equivalent thickness of 6 to 10 monolayers of oxide atoms). The process specifications for selectivity of polysilicon etch to oxide typically require no more than a 5 Å loss of oxide thickness (about 1.5 monolayers) across a $70,686\text{ mm}^2$ surface area on a 300 mm wafer. This will require development of a selectivity $>150:1$ for the main polysilicon etch process to prevent punchthrough of the gate oxide and an overetch selectivity of $>250:1$ to remove etch residue and stringers.²⁹

Polysilicon etch gas chemistries have traditionally been fluorine-based gases, including CF_4 , CF_4/O_2 , SF_6 , $\text{C}_2\text{F}_6/\text{O}_2$, and NF_3 . The fluorine atoms are instrumental in the etching of silicon, however, this chemical-type of etch results in etch with an isotropic profile. This profile is improved by reducing the fluorine atoms and increasing the ion energy, but this also reduces the polysilicon-to- SiO_2 selectivity. The energy of the bombarding ions must be low enough to avoid sputtering away the gate oxide. To overcome these difficulties, polysilicon plasma etch gases often contain both chlorine and fluorine. Chlorine produces an anisotropic polysilicon sidewall profile with good selectivity to SiO_2 . Another common gas chemistry for etching polysilicon is bromine because it gives higher etch selectivity of polysilicon to gate oxide than chlorine. Volatile etch by-products for fluorine, chlorine, and bromine gas chemistries are SiF_4 , SiCl_4 , and SiBr_4 , all of which are evacuated from the chamber after etching.

Single-Crystal Silicon Etch ■ Single-crystal silicon is etched to form trenches for applications such as device isolation or for the fabrication of vertical capacitors in high-density DRAM ICs. Silicon trench fabrication permits a decrease in the surface area needed for these technologies, which is important as feature sizes continue to decrease on advanced ICs. For STI device isolation in both bipolar and MOS technology, the shallow silicon trenches (depths less than $1\text{ }\mu\text{m}$ to several μm s in depth) are filled with oxide dielectric. Capacitors are fabricated by etching a deep trench (a depth greater than $5\text{ }\mu\text{m}$), oxidizing the sidewalls, and then filling the trench with polysilicon. The vertical capacitor trench design is advantageous because it takes up less surface area and provides higher capacitance values.

Silicon Trench Etching. Silicon trench etching requires precise dimensional control for each trench on the IC. This is a challenge, considering there are millions of capacitor trenches on a high-density IC, with each trench required to have identical smooth, nearly vertical sidewalls, the correct depth, and rounded corners at the bottom and top of the trench. A multistep etch process (as outlined above) is needed, with a modification for the final step where the bottom of the trench is rounded to remove any silicon damage (see Figure 16.31 on page 462).

Sidewall passivation formed by adding carbon to the gas is often used to protect the trench sidewalls from lateral etch attack. The sidewall shape is a function of the wafer temperature—increasing the temperature results in less deposition of sidewall passivation and more lateral etching. Wafer temperature can be controlled by backside cooling of the vacuum chuck with helium.³⁰

For plasma dry etching of shallow trenches, fluorine gas chemistry is sometimes used because it has a high silicon etch rate with adequate selectivity to the photoresist that is used as the mask material. For deep trenches (e.g., several microns depth), a chlorine-based or bromine-based gas chemistry is often used because this gas chemistry has a high silicon etch rate and high selectivity

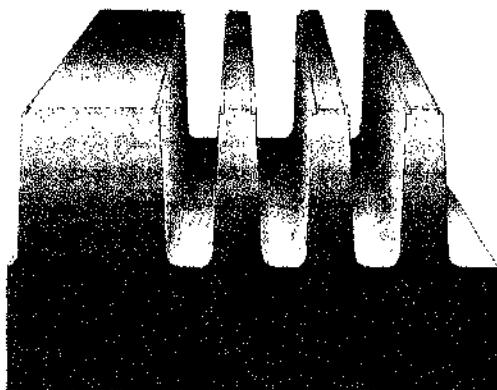


FIGURE 16.31 Silicon Trench Etching

to the oxide mask. Bromine gas chemistry is becoming common in high-density plasma reactors because it is an aggressive etchant that does not require extensive use of carbons for sidewall passivation (which reduces contamination problems). The primary problem with bromine gas chemistry is that it is extremely corrosive to the gas delivery system and reactor. The gas suppliers and equipment manufacturers are addressing this issue.

Metal Dry Etch

A major application in metal etch is aluminum alloy for interconnect wiring. Recall from Chapter 12 that copper alloyed with aluminum reduces electromigration and hillock formation in aluminum, while silicon is added to aluminum to minimize spiking at contact interfaces. Silicon poses no additional etch resistance. Tungsten, barrier metals, and the contact metal are also common metals etched during wafer fabrication. However, copper metal is not dry etched because it does not easily form volatile compounds, which makes plasma etching difficult. Due to the density of interconnect lines on the critical layers of ULSI wafers, an anisotropic dry etch is necessary to attain the desired edge profile and narrow lines. The major requirements for metal etching are:³¹

1. High etch rates (>1000 nm/min).
2. High selectivity to the masking layer (>4:1), interlayer dielectric (>20:1), and the underlying layers.
3. High uniformity with excellent CD control and no microloading (<8% at any location on the wafer).
4. No device damage from plasma-induced electrical charging.
5. Low residue contamination (e.g., copper silicon residue, developer attack, and surface defects).
6. Fast resist strip, often in a dedicated cluster tool chamber, with no residual contamination.
7. No corrosion.

Aluminum and Metal Stacks ■ A chlorine-based chemistry process is commonly used for aluminum etch. Pure chlorine etches aluminum isotropically. To achieve a directional etch process with anisotropic etching, sidewall passivation is needed by adding polymers to the etch gas, such as with CHF_3 or carbon obtained from the photoresist. BCl_3 is also used for aluminum etch, which has heavy molecules for a physical etch and good sidewall profile control. A small amount of N_2 is often added to the gas chemistry to minimize microloading and assist sidewall passivation. Aluminum etching with fluorine-based gas chemistry is unacceptable because of the low vapor pressure of the etch by-product, nonvolatile AlF_3 , which means that the etch rate is extremely limited. Bromine-based chemistry has been investigated for metal etch.

A challenge in metal etch for interconnect wiring is the complexity of the multilayer stack that is common in VLSI/ULSI technology (see Figure 16.32). There is often an antireflective coating (ARC) of TiN or other materials and an underlying adhesion barrier (e.g., Ti), which complicates the etch process. To remove ARC films, possible etch gas chemistries are CCl_4/O_2 or SF_6/Cl_2 . TiN films are removed in CCl_4/N_2 or Cl_2/Ar or $\text{BCl}_3/\text{Cl}_2/\text{CHF}_3$.³² Another factor is that aluminum

oxidizes almost instantly when exposed to air and the resulting aluminum oxide (Al_2O_3) inhibits the chlorine etch reactions.

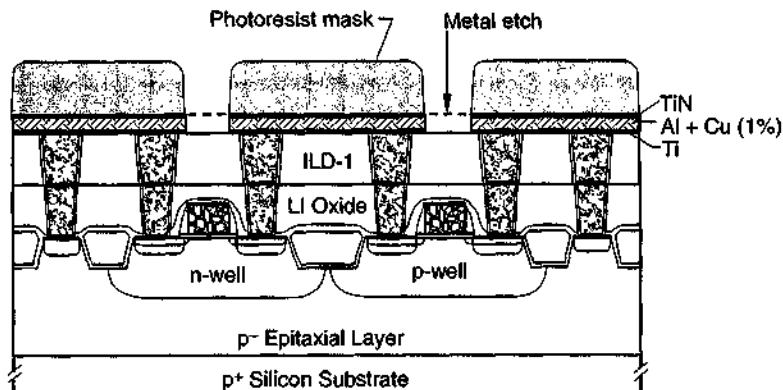


FIGURE 16.32 Metal Stack for VLSI/ULSI Integration

To etch a metal stack structure, multiple steps are used. This includes the first step to remove the native oxide layer (e.g., Al_2O_3 for aluminum) on the surface. The typical steps for etching a metal stack are:

1. Breakthrough step to remove native oxide.
2. ARC layer etch (may be combined with above step).
3. Main etch step of aluminum.
4. Overetch step to remove residue. It may be a continuation of the main etch step.
5. Barrier layer etch.
6. Optional residue removal process to prevent corrosion.
7. Resist removal (see section below).

Thorough corrosion control is essential to device performance after metal etching. Any remaining corrosive by-products of the etch process must be quickly neutralized or removed from the wafer surface. For aluminum etching, the major corrosive by-products are AlCl_3 or AlBr_3 . These by-products react with water to form highly corrosive HCl or HBr , which are damaging acids to aluminum. Control of water vapor and oxygen content is critical in the etch process, which is best achieved in single-wafer etch reactors with a loadlock chamber to isolate the wafers from atmospheric contaminants and moisture. Another approach is to use the resist strip process to remove corrosive compounds while removing the resist, thus reducing their potential for corrosion.³³ The need to minimize the time delay and exposure to moisture are reasons for an etcher to have an integrated photoresist removal chamber.

Tungsten ■ Tungsten (W) is an important metal commonly used for via fill in a multilevel metal structure (see Chapter 12). Tungsten etch can be done with fluorine- and chlorine-based chemistry. However, fluorine-based chemistry, such as SF_6 and CF_4 , has poor oxide selectivity. Chlorine-based gas chemistry has good oxide selectivity. N_2 is often added to the gas mixture to give better selectivity to the photoresist mask. Sometimes an O_2 forming gas is added to reduce carbon deposits. A chlorine gas such as Cl_2 or CCl_4 can be used to etch tungsten with improved anisotropy and selectivity.

Tungsten Etchback. *Tungsten etchback* is used as one step in the formation of tungsten plugs. Via openings are first etched in the SiO_2 interlayer dielectric (ILD). The W is blanket-deposited (CVD) into the via openings on top of a TiN barrier layer. This operation is followed by a dry plasma etchback step to remove the excess blanket layer and finish with a tungsten-filled via (see Figure 16.33 on page 464). The etch involves a two-step process. In the first step, 90% of the tungsten is etched at a high rate with excellent uniformity. In the second step, the etch rate is reduced and a gas chemistry is used with a high selectivity to the TiN barrier layer.³⁴ The etch

reduction serves to decrease gas pressure and wafer temperature to reduce loading effects that may cause a recess in the W plug. Anisotropic etching is not needed, but minimal residue and plug loss are important. This etchback process has been largely replaced for advanced wafer fabrication by chemical mechanical planarization (CMP) to remove excess W and planarize the top surface of the contacts and vias (see Chapter 18).

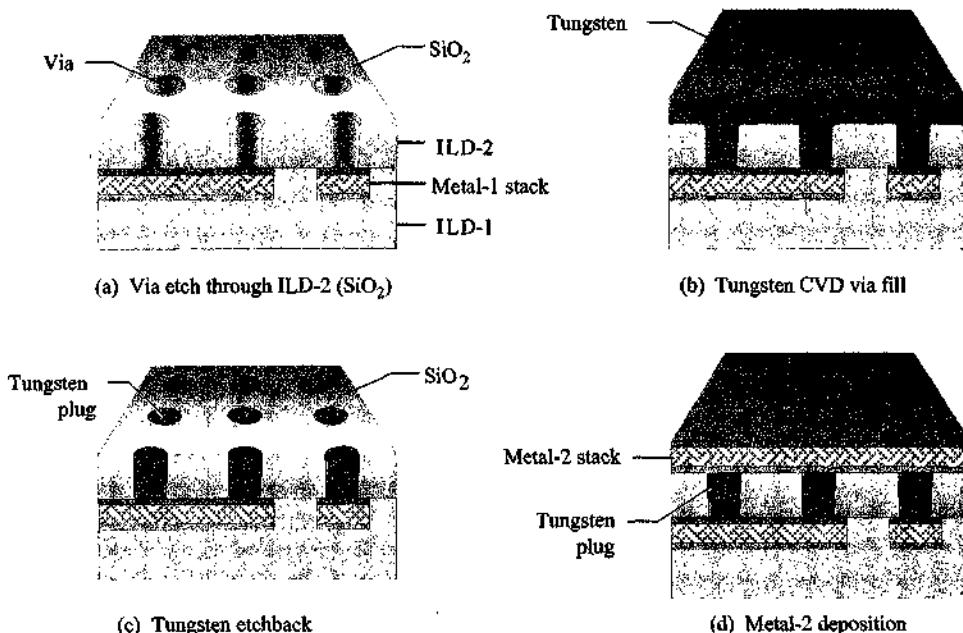


FIGURE 16.33 Tungsten Etchback

Contact Metal Etch ■ Refractory metals alloyed with silicon form the silicides common in wafer fabrication, including CoSi_2 , WSi_2 , TaSi_2 , and TiSi_2 (see Chapter 12). A polycide is a composite of doped polysilicon and a refractory metal. Silicon dioxide does not form an alloy with the refractory metals, and it is the unreacted metal that must be removed during the contact metal etch process. Contact metal etch is critical in MOS device fabrication because the dimensional control will affect the channel length of the device. Contact metal etch with dry plasma can be done with fluorinated and chlorinated gas chemistries. Fluorine-based gas chemistry, such as NF_3 and SF_6 , is used because of its good dimensional control with increased etch rate. Since the contact formation is a self-aligning process, no photoresist mask is required for this etch process.

WET ETCH

Wet etching has been associated with wafer fabrication since the beginning of the semiconductor industry. It has been replaced to a large degree by dry etching but still plays an important role in oxide cleaning, residue removal, the stripping of surface layers, and etching for larger geometry applications. Wet cleaning (see Chapter 6) can actually be considered a wet etch process. An example is the residue layer and damaged silicon layer formed by reactive ion etching of a contact hole. The contact residue is removed by downstream microwave ashing followed by wet etching as a form of cleaning to remove residues and silicon damage.³⁵ The benefits from wet etching over dry etching are good selectivity to the underlying film, no risk of plasma damage to devices, and use of simple equipment. Basic wet-etch parameters (see Table 16.7) must be controlled during the etching process. These parameters are applicable to all wet etching applications.

Wet etching is done by exposing a batch of wafers, usually 25 but sometimes up to 50 wafers, to an appropriate acid bath, either by immersion or spray. Immersion is the simplest method, while spray etching requires less volume of chemicals and is faster than immersion. Wet etching produces an isotropic sidewall with substantial undercutting of the film underneath the edge of the mask material, making it unsuitable for feature sizes $< 3 \mu\text{m}$. Isotropic etching with undercutting is the main

TABLE 16.7 Wet-Etch Parameters

Parameter	Explanation	Difficulty to Control
Concentration	Solution concentration (e.g., ratio of $\text{NH}_4\text{F:HF}$ to etch oxide).	Most difficult parameter to control because the bath concentration is continually changing.
Time	Time of wafer immersion in the wet chemical bath.	Relatively easy to control.
Temperature	Temperature of wet chemical bath.	Relatively easy to control.
Agitation	Agitation of the solution bath.	Moderately difficult to properly control.
Number of Runs	Solution must be replaced after specified number of runs to reduce particles and to ensure proper solution strength.	Relatively easy to control

reason why the majority of semiconductor etching is done with dry plasma etch. Other disadvantages to wet etch are the chemical safety concerns for wet baths, its tendency to cause resist lifting and bubbles, the difficulty in controlling the bath parameters for uniformity, and the high cost of chemical disposal.

Types of Wet Etch

Wet etching can provide an alternative for high etch selectivity that might not be available with dry etching. In addition, wet etching eliminates plasma damage. As previously stated, wet etch has been widely replaced by plasma dry etching in advanced IC wafer fabrication. Materials such as silicon and aluminum are nearly always dry etched for submicron fabrication. General safety information for wet etch chemicals is provided in Appendix A.

Wet Oxide Etch ■ Oxide can be wet etched by hydrofluoric acid (HF). Selective removal of oxide is done by spraying or immersing the wafer in a dilute solution of HF that is frequently buffered with ammonium fluoride (NH_4F), known as a *buffered oxide etch (BOE)* or *buffered HF (BHF)*. A chemical buffer is a solution that resists changes in pH when small amounts of strong acid or base are added. Buffering the HF with NH_4F provides a well-controlled etch solution that slows and stabilizes the etch and does not appreciably attack the photoresist. The BOE is rinsed from the wafer using DI water and quick dump or cascade overflow rinse equipment. The quick dump rinse is faster and uses less DI water, but cascade rinsing can significantly reduce particulate contamination levels.

SiO_2 is an amorphous material and etches equally well in all directions when exposed to BOE. If a 1 μm thick oxide layer is etched, there is also a lateral etch of 1 μm under the mask material. This lateral etching and the subsequent undercutting is undesirable from a dimensional standpoint, limiting the density of lines and spaces that can be achieved.

The oxide etch rate also depends on whether it is thermal or deposited oxide (see Table 16.8 on page 466). Since dry grown oxide is denser than wet grown oxide, dry grown oxide etches at a slower rate. In addition, doped oxides etch differently and usually faster than undoped oxides.

Wet Chemical Strips ■ Due to the high selectivity of wet etch, *wet chemical strips* are sometimes used to remove surface layers, including photoresist (see the following section) and masking layers. Silicon nitride (Si_3N_4) is widely used in wafer fabrication as a mask material in the isolation technologies of shallow trench isolation (STI), local oxidation of silicon (LOCOS), and self-aligned contact structure. The removal of the nitride masking layer is often done by a wet chemical strip with hot phosphoric acid (H_3PO_4).³⁶ The wet acid bath solution is typically maintained at around 160°C and has the desirable property of high selectivity to exposed oxide layers. Control of the hot phosphoric nitride removal bath is difficult and it is done as a timed operation (no endpoint detection) with the use of monitor wafers. There is usually some oxynitride formation on the exposed nitride layer that requires a short hydrofluoric acid (HF) deglaze before the nitride removal. If the oxynitride layer is not removed, then there is a risk of nonuniform nitride removal. Another material often

TABLE 16.8* Approximate Oxide Etch Rates in BHF^a Solution at 25°C

Type of Oxide	Density (g/cm ³)	Etch Rate (nm/s)
Dry grown	2.24 to 2.27	1
Wet grown	2.18 to 2.21	1.5
CVD deposited	<2.00	1.5 ^b to 5 ^c
Sputtered	<2.00	10 to 20

a) 10 parts of 454 g NH₄F in 680 ml H₂O and one part 48% HF

b) Annealed at approximately 1000°C for 10 minutes

c) Not annealed

*B. El-Kareh, *Fundamentals of Semiconductor Processing Technology* (Boston: Kluwer Academic, 1995), p. 277.

removed by wet strip is the titanium layer deposited to form the titanium silicide at the contacts. Ammonium hydroxide (NH₄OH) and hydrogen peroxide (H₂O₂) are diluted with DI water for this process.

HISTORICAL PERSPECTIVE

Etch technology has undergone many changes throughout the years of wafer fabrication. The initial barrel plasma etcher was a simple tool with limited operator controls. The modern plasma etcher generates a high-density plasma, has independent RF power supplies for plasma generation and wafer bias, endpoint detection, gas pressure and flow control, and is integrated with software control of the etch parameters. Table 16.9 details the advances for polysilicon etching from wet etch technology through reactive ion etch (RIE) and up to modern high-density plasma etch tools for subquarter micron geometries.

PHOTORESIST REMOVAL

One of the last steps after etching is the removal of the photoresist mask. The photoresist serves as the pattern transfer medium from the reticle to the wafer surface and blocks films from being etched or ions from being implanted. Once etch or implantation is complete, the resist serves no further purpose on the wafer surface and must be completely removed. In addition, any trace residues from the etch process are also removed.

Photoresist stripping is the general term that describes the wet removal of the resist. Stripping is sometimes used for resist residues that are difficult to remove. In most applications, wet resist stripping is not cost effective because of the required handling and disposal of wet chemicals. Furthermore, if the resist has previously undergone dry etch processing, its top surface could be hardened from exposure to fluorinated and chlorinated gas chemistry. This renders the resist insoluble in most wet chemistry stripping solutions. In this case, plasma stripping is required to remove at least the top resist layer.

A major challenge for photoresist removal is that resists are designed and processed to achieve maximum adhesion to the wafer surface. As discussed in Chapter 13, this is required for the resist to withstand the rigors of the etch process and ion implantation. Maximum resist adhesion creates an obvious challenge for photoresist removal. Another challenge is the need for a high resist strip rate to attain a higher wafer throughput. However, a high strip rate generally tends to leave more resist residues, which places more burden on the removal process to effectively remove all resist and residues.

Plasma Ashing

Plasma ashing is the dry removal of resist with oxygen, and is the dominant technique for bulk resist removal.³⁷ The first plasma ashers in the 1970s used barrel reactor technology, with oxygen

TABLE 16.9* Polysilicon Etch Technology Evolution

Geometry Requirements	Time Frame and Reactor Design	Chemistries	Strengths	Limitations and Problems	Controls
4 to 5 μm , isotropic etch	Pre-1977: wet etch	HF/HNO ₃ buffered with acetic acid or H ₂ O	Batch process	Resist lift; bath aging; temperature sensitive	Operator judgement for endpoint
3 μm	1977: barrel etcher	CF ₄ /O ₂	Batch process	Nonuniformity; isotropic etch; large undercut	Manometer and timer
2 μm	1981: single wafer etch	CF ₄ O ₂	Single wafer; individual etch endpoint, improvement in repeatability	Low oxide selectivity; isotropic process	Endpoint detection
1.5 μm	1982: single wafer RIE	SF ₆ /Freon 11, SF ₆ /He	MFCs; independent pressure and gas flow control, improvement in repeatability	Low oxide selectivity; profile control	MFCs; separate gas flow and pressure control
To 0.5 μm	1983: variable gap; load-locked	CCl ₄ /He, Cl ₂ /He, Cl ₂ /HBr	Load-locked chamber; variable gap; improvement in repeatability	Microloading in high-aspect ratios; profile control	Control of electrode gap; computer controls
To 0.25 μm and below	1991: inductively-coupled plasma (ICP)	Cl ₂ , HBr	High-density plasma; low pressure; simple gas mixtures; improvement in repeatability	Complex tool; many variables	Independent RF control for plasma generation and wafer bias

*Adapted from C. Almgren, "The Role of RF Measurements in Plasma Etching," *Semiconductor International*, (August 1997): p. 100.

plasmas to strip the resist. Barrel reactors have been widely replaced with downstream plasma reactors that strip the resist with an oxygen plasma in single-wafer chambers.

Asher Overview ■ A photoresist *asher* removes the resist layer by reacting atomic oxygen atoms with the resist material in a plasma environment (see Figure 16.34 on page 468). Atomic oxygen (O) is created by using microwave or RF energy to dissociate molecular oxygen (O₂). There are usually forming gases such as N₂ or H₂ added to improve the ash performance and enhance polymer residue removal.³⁸ Thus, a typical ashing gas chemistry would be O₂/N₂. Recall from Chapter 13 that resist is basically a hydrocarbon polymer. The atomic oxygen atoms rapidly react with the resist material to create volatile carbon monoxide (CO), carbon dioxide (CO₂), and water (H₂O) as the main by-products. The ashing by-products are pumped away by the vacuum system.

Plasma Damage ■ A concern during the ashing process is plasma damage to devices on the wafer surface from ion bombardment or wafer charging. As gate oxides are scaled thinner, this concern is more critical due to the sensitivity of thin oxides to plasma degradation. Wafer damage from plasma has largely been remedied by using single-wafer downstream ashers that position the wafer away from the damaging plasma ions, allowing only reactive chemical species to reach the wafer. Microwave (2.45 GHz) frequencies are common for source plasma generation in downstream systems. This is because a microwave glow discharge produces more atomic oxygen reactant and a lower proportion of ionic species. This yields lower risk of ion-induced damage.

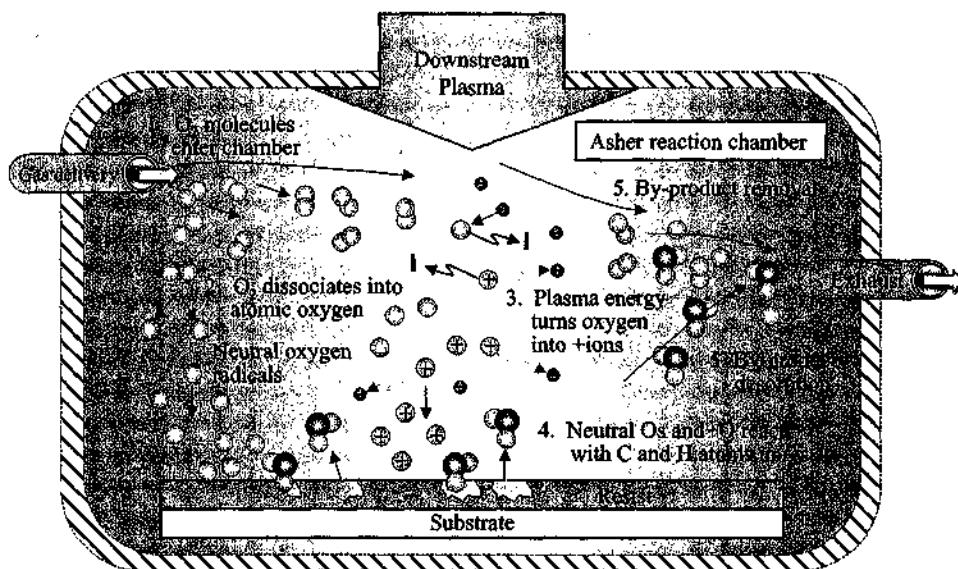


FIGURE 16.34 Atomic Oxygen Reaction with Resist in Asher

Residue Removal ■ Standard ashing has evolved to include the removal of post-etch residues such as sidewall polymers and via veils (see Figure 16.35). As previously discussed, these residues are complex and can contain plasma etching and ashing by-products such as aluminum, titanium, oxides, and silicon. A complicating factor is that ashing may occur at an elevated temperature (e.g., 200°C), which can harden residues and make them harder to remove. If not removed, residues are a source of particles and contaminants that increase the wafer defect density. To fully remove etch residues, especially the inorganic residue materials (e.g., silicon oxides, metal oxides, aluminum, and so on), some ashers are based on alternative plasma chemistries to oxygen, such as NO or N₂O. Other ash systems add a small amount of fluorine to the gas chemistry in the form of CF₄ or NF₃ to more effectively remove residues containing oxide and silicon by making them more water soluble.³⁹ The ash process is typically followed by a DI water rinse to remove residual particles on the wafer surface.⁴⁰ In some cases a strong sulfuric acid (H₂SO₄) mixture with hydrogen peroxide (such as the piranha mixture discussed in Chapter 6) is used to thoroughly clean the remaining stripped photoresist.

Wet cleaning for residue removal still is found in backend processes, particularly for the removal of sidewall passivating residues that have high-density inorganic contaminants. Dry ashing alone is not sufficient to strip and clean residues because of the multitude of inorganic materials present that are not volatized by the plasma. Until recently, wet chemistry residue strippers were commonly based on a hydroxylamine compound. However, there is a corrosion risk to metal and barrier layers from this stripper and new hydroxylamine-free formulations are in use based on a chelation chemistry to strip residues (a chelate is a complex ion that involves metal cations).⁴¹ This new formulation dissolves inorganic residues at low temperature with no corrosion.

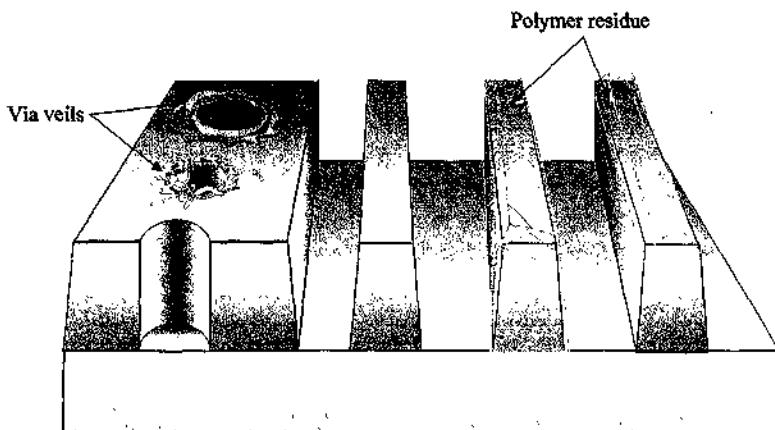


FIGURE 16.35 Post-Etch Via Veil Residue

ETCH INSPECTION

The final step in the etch process is an *etch inspection* to verify quality measures. This inspection occurs on the patterned wafer after all etching and photoresist stripping is completed. It has traditionally been done with a manual microscope inspection that used incident white or ultraviolet light to inspect for defects such as stains and large particulate contamination. Manual microscopes have been largely replaced by automatic inspection systems, especially on critical layers with deep sub-micron patterning. Advanced metrology instruments are able to automatically inspect patterned wafers for patterned defects and distortions (see Chapter 7). One of the most important inspections that occurs at final inspection is the automated measurement to verify critical dimensions (CD) on a particular mask level. The quality of the etch process is also verified by inspecting for etch problems such as overetching, underetching and undercutting.

Etch inspection is similar to the inspection at develop inspect, except that there is no photoresist and defects usually cannot be reworked. Almost any defect, if numerous enough, will cause the wafer to be scrapped. An exception to this rule is surface particle contamination, which can be cleaned. Representative quality measurements for etch final inspection are described in Table 16.10.

ETCH INSPECTION QUALITY MEASURES

Typical quality measures for etch inspection are provided in Table 16.10.

TABLE 16.10 Quality Measures for Etch Final Inspection

Quality Parameter	Types of Defects	Remarks
1. Critical dimension bias.	A. Linewidth change: excessive difference between photoresist linewidth and final feature linewidth after etch.	<ul style="list-style-type: none"> CD bias for linewidth change is measured by comparing pre-etch linewidth in photoresist with post-etch linewidth of the same feature on the wafer. Measurements can be done with an SEM. Excessive CD bias requires optimization of the etch process for undercutting or slope. The photoresist profile has an effect on CD bias. Vertical resist profile produces best CD bias.
2. Metal corrosion.	A. Corrosion or attack of metal film after etch.	<ul style="list-style-type: none"> Corrosion often results from residual HCl on the wafer that is exposed to water vapor in the air. Corrosion may be visible as small bubbles along the side of metal lines. This defect is looked for by optical microscope or SEM.
3. Sidewall contaminants after etching.	A. Residual sidewall passivants that remain after etch, including remaining photoresist.	<ul style="list-style-type: none"> Minimize formation by optimizing the pre-metal etch photoresist hard bake. Residues are usually removed by post-etch solvents or dilute buffered HF.*
	B. Contaminants backspattered onto the sidewall of the metal line or a via hole.	<ul style="list-style-type: none"> Backspattering can occur due to polymer used to passivate sidewall, leaving a "veil" residue (thin overhang along metal line or via sidewall).
4. Loading effects.	A. Microscopic nonuniformity of etch process.	<ul style="list-style-type: none"> Etch rate between narrow openings is slower than open-field areas due to reduced density of reactive radicals in small space. Balance pressure and power of etch process.**

Quality Parameter	Types of Defects	Remarks
5. Shorts after metal etch.	A. Bridging of metal lines after etch leading to electrical short.	<ul style="list-style-type: none"> Reduce pattern density effects from microloading.
6. Excessive post-etch residue.	<p>A. The following types of residue may exist after etch:</p> <ul style="list-style-type: none"> Stringers (small strings of residue). Veils (thin residue overhang). Crowns Rails Corrosion after metal etch. 	<p>The causes of residue after etch are:</p> <ul style="list-style-type: none"> Nonuniform etch process. Films deposited over topography. Nonuniform dopant distribution in film. Contaminants in the film (in addition to intentional alloys such as Cu in Si). Contaminants in gas or chamber. Check supply filters or clean system. Incorrect process parameter, such as high etch rate.

*K. Mautz, *Optimization of Single Wafer and Batch Metal Etch Manufacturing Processes* vol. 96-12 (Pennington, NJ: The Electrochemical Society, 1996), p. 283.

**S. Gonzales, J. Quijada, and G. Grivna, *Submicron Metal Etch Integration Study* vol. 2875 (Bellingham, WA: SPIE, 1996), p. 302.

DRY ETCH TROUBLESHOOTING

Problems associated with dry etch equipment and process are provided in Table 16.11.

TABLE 16.11 Common Dry Etch Troubleshooting Problems

Problem	Probable Cause	Corrective Actions
1. Incorrect etch rate.	<p>A. Change in RF power.</p> <p>B. Incorrect temperature.</p> <p>C. Problem with pressure.</p> <p>D. Endpoint detection not functioning properly.</p> <p>E. Improper wafer spacing.</p> <p>F. Improper gas-flow dynamics.</p> <p>G. Improper maintenance.</p> <p>H. Incorrect process recipe.</p>	<ul style="list-style-type: none"> Check and troubleshoot RF generator and matching unit. Check backside wafer cooling system. Calibrate vacuum gauges (e.g., capacitive manometer) and pressure control system. Check endpoint detection system. Check wafer-to-electrode gap. Verify gas distribution system. Perform chamber wet clean. Verify process recipe and parameters.
2. Inadequate selectivity.	<p>A. Etch rate too high.</p> <p>B. Improper gas flow or pressure.</p> <p>C. Endpoint detection problem.</p> <p>D. Wrong wafer temperature.</p> <p>E. Incorrect process recipe.</p>	<ul style="list-style-type: none"> Verify etch rate. Calibrate MFCs and vacuum gauges. Check/calibrate endpoint detection. Check wafer cooling system. Confirm process recipe and parameters.
3. Improper sidewall profile angle.	<p>A. Contamination on sidewall.</p> <p>B. Temperature of wafer.</p> <p>C. System pressure.</p> <p>D. Incorrect process recipe (misprocess).</p>	<ul style="list-style-type: none"> Check for polymer buildup in chamber. Backside contamination of wafer causing nonuniform heating. Check/calibrate MFCs and perform leak test to check for contamination.

Problem	Probable Cause	Corrective Actions
4. Etch nonuniformity across wafer.	A. Depletion of etchant gas concentration due to ARDE. B. Improper gas flow. C. Temperature of wafer. D. Improperly positioned wafer in chamber. E. Chamber configuration. F. Improper film thickness. G. Improper maintenance.	<ul style="list-style-type: none"> • Verify acceptable design for dense and nondense areas of wafer. • Check/calibrate gas distribution system. • Check thermocouples and wafer cooling. • Check robotics, wafer handling system, and vacuum chuck. • Check reactor plate spacing. • Measure and verify film thickness. • Perform chamber wet clean.
5. Plasma damage.	A. Nonuniform plasma. B. Excessive ion bombardment of gate oxide. C. Excessive RF power. D. Improper maintenance.	<ul style="list-style-type: none"> • Poorly designed or maintained plasma equipment. • Suboptimum process conditions. • Check recipe and RF generator.
6. Particle Contamination.	A. Leak/contamination from gas lines. B. Tool operation. C. Improper gas chemistry.	<ul style="list-style-type: none"> • Leaks or faulty MFCs. • Improper tool shutdown, operation, or maintenance. • Wrong process recipe. • Perform wet clean.
7. Metal Corrosion.	A. Moisture. B. Gas flow. C. Contaminants from etch process. D. Wrong maintenance procedure.	<ul style="list-style-type: none"> • Excessive time delay for post-etch residue cleanup. • Check MFCs for correct process gases. • Control time to resist strip. • Check maintenance procedure.

SUMMARY

Etch selectively removes material from the wafer surface by chemical or physical means. Dry etch uses a plasma, whereas wet etch uses liquid chemicals. Etch is generally divided into categories including dielectric, silicon, and metal etch. There are nine important parameters: etch rate, etch profile, etch bias, selectivity, uniformity, residues, polymer formation, plasma-induced damage, and particle contamination. Plasma etch has many advantages over wet etch and is the most common etch process. Sidewall profiles are generally divided into isotropic and anisotropic. Anisotropic, or vertical sidewalls, are produced by dry plasma etching. Oxide etch requires high selectivity to the underlying material. Silicon etch is used for critical

applications such as polysilicon gate formation and trench formation. A main application of metal etch is aluminum interconnect wiring, with the need to etch through complex metal stacks. Wet etch has been largely replaced by dry etch for advanced IC fabrication but is still used for oxide removal, wet cleans, and stripping. The removal of photoresist at the completion of etch or ion implant is known as photoresist ashing. An ashing uses a downstream process to minimize plasma-induced damage and is usually followed by a wet etch to remove residues. Etch inspection is the last operation in the etch process and is necessary to ensure defects are identified and corrected.

KEY TERMS

etch	combined physical and chemical mechanism
dry etch	plasma potential distribution
wet etch	barrel reactor
dielectric etch	parallel plate (planar) reactor
silicon etch	downstream reactor
metal etch	triode planar reactor
patterned etching	ion beam milling or ion beam etching (IBE)
unpatterned etching (etchback or stripping)	reactive ion etch (RIE)
etch rate	high-density plasma etch
step height	electron cyclotron resonance (ECR)
loading effects	inductively coupled plasma (ICP)
etch profile	dual plasma source (DPS)
isotropic etch profile	magnetically enhanced RIE (MERIE) or magnetron
anisotropic etch profile	endpoint detection
directionality	breakthrough step
etch bias	main-etch step
selectivity	overetch step
etch uniformity	tungsten etchback
aspect-ratio-dependent etching (ARDE)	buffered oxide etch (BOE) or buffered HF (BHF)
microloading	wet chemical strips
etch residue	photoresist stripping
polymer formation	plasma ashing
plasma-induced damage	photoresist ashing
chemical mechanism	etch inspection
physical mechanism	

REVIEW QUESTIONS

1. Define etch. What is the goal of etching?
2. What are the two types of etch processes? Give a short description of each type.
3. List the three major material categories for dry etch.
4. Explain the difference between patterned and unpatterned etching.
5. List nine important parameters for etch.
6. Define etch rate and state its formula. Why is it desirable to have a high etch rate?
7. Explain loading effects and how this is related to the etch rate.
8. Describe isotropic and anisotropic etch profiles and what are the desirable and undesirable aspects of each profile.
9. Is a dry etch profile isotropic, anisotropic or both? What about a wet etch profile?
10. What is directionality and why is it desirable in etching?
11. What is etch bias and what is it caused by? State and explain the etch bias formula.
12. Define selectivity. Does dry etching have good or poor selectivity? What does high selectivity mean? State and explain the selectivity formula.
13. What is etch uniformity? What is the challenge for etch uniformity? Explain ARDE and discuss how it is related to etch uniformity. What is another name for ARDE?
14. Discuss etch residues, why they occur, and how they are removed.
15. Why is a polymer formation deposited sometimes on feature sidewalls during etch? What is an undesirable side effect of polymer formation?
16. What is plasma-induced damage from etching and what problems occur from this damage?
17. Explain how plasma can cause particle contamination on the wafer surface.
18. What is the goal of dry etch? List the advantages of dry etch over wet etch. What are the disadvantages of dry etch?
19. List the three ways to achieve etching action during dry etch.

20. Explain the chemical mechanism and the physical mechanism for achieving etching action.
21. Describe the combined physical and chemical mechanism for dry etch.
22. Describe the plasma potential distribution and why it is important for etch.
23. What are the etch results if the RF frequency is decreased? RF power is decreased? DC bias is increased? Electrode size is decreased?
24. Describe the basic components of a plasma dry etch system. What gas chemistries are used for silicon dioxide, aluminum, silicon and photoresist?
25. Describe the barrel plasma etcher.
26. Describe the parallel plate (planar) reactor.
27. Why is downstream etching beneficial? Describe a downstream reactor.
28. How does a triode planar reactor function?
29. Explain ion beam milling. For what materials is it used?
30. Describe reactive ion etch (RIE).
31. Explain the principle for how high-density plasma improves etching of high-aspect ratio gaps.
32. Give three reasons for using a magnetic field with a high-density plasma etcher.
33. Describe electron cyclotron resonance (ECR).
34. Explain inductively-coupled plasma (ICP).
35. Describe dual plasma source (DPS).
36. Discuss the MERIE technology for high-density plasma.
37. What is endpoint detection and why is it necessary for dry etch? What is the most common type of endpoint detection?
38. Describe optical emission spectroscopy for endpoint detection.
39. List six requirements for successful dry etch.
40. List three challenges for oxide etching.
41. Oxide plasma etching is usually based on what gas chemistry? Give a specific gas chemistry common for oxide etch.
42. Give an example why oxide etch selectivity to the underlying material is important. Provide two methods for achieving oxide etch selectivity to silicon.
43. What is one factor that reduces photoresist selectivity? Is this desirable or undesirable?
44. Explain the effect of low resist selectivity on contact via etch in the Li oxide.
45. Is the etch rate faster for silicon nitride deposited with PECVD or LPCVD? Why?
46. What gas chemistries are usually used for etching polysilicon and why has these chemistries replaced fluorine chemistries?
47. Give three requirements for polysilicon gate etch.
48. List and describe the three steps for etching polysilicon.
49. Discuss the polysilicon etch gas chemistries.
50. How are silicon trench sidewalls protected during plasma etch from lateral etch attack?
51. What gas chemistry is sometimes used for shallow trench etching? Deep trench etching? Why are these gas chemistries used?
52. List the seven major requirements for metal etching.
53. List the seven steps for etching a metal stack.
54. What control is necessary after metal etching?
55. What gas chemistry has good oxide selectivity during tungsten dry etch?
56. Describe the tungsten etchback process.
57. What is the main reason why dry etch has largely replaced wet etch?
58. List four wet etch parameters, explain each parameter and state how difficult it is to control.
59. How is oxide wet etched? What are the two names of the HF solution buffered with ammonium fluoride?
60. What is the wet etch rate difference for dry grown versus wet grown oxide?
61. Describe the silicon nitride wet chemical strip.
62. What is photoresist stripping?
63. What is plasma ashing? What is the purpose of a photoresist ash?
64. What is a concern from the ashing process? How is this concern resolved?
65. How are post-etch residues removed?
66. What is an important aspect of performing an etch inspection?

ETCH EQUIPMENT SUPPLIERS' WEB SITES

Applied Materials
 Eaton Corporation
 Gasonics International
 Hitachi
 International SEMATECH
 Lam Research Corp.
 Leybold-Inficon
 MRC, Materials Research Corp.
 Plasmos
 SEMI
 Tegal Corporation
 TEL, Tokyo Electron Ltd.

<http://www.appliedmaterials.com/products/>
<http://www.semiconductor.eaton.com/>
<http://www.gasonics.com/>
<http://www.hitachi.com/semiequipment/products.html>
<http://www.sematech.org>
<http://www.lamrc.com/>
<http://www.leyboldinficon.com/>
<http://www.materialsresearch.com/>
<http://www.plasmos.com/>
<http://www.semi.org/>
<http://www.tegal.com>
<http://www.teainet.com>

REFERENCES

1. Y. Lii, "Etching," *ULSI Technology*, ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 342.
2. P. Singer, "New Frontiers in Plasma Etching," *Semiconductor International* (July 1996), p. 153.
3. R. DeJule, "Managing Etch and Implant Residue," *Semiconductor International* (August 1997): p. 60.
4. C. Gabriel, "Measuring and Controlling Gate Oxide Damage from Plasma Processing," *Semiconductor International* (July 1997): p. 151.
5. Y. Lii, "Etching," p. 364.
6. Adapted from G. Oehrlein and J. Rembetski, "Plasma-Based Dry Etching Techniques in the Silicon Integrated Circuit Technology," *IBM Journal of Research & Development* vol. 36, no. 2 (Armonk, NY: March 1992): p. 140.
7. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era—Vol. 1, Process Technology*, (Sunset Beach: Lattice Press, 1986), p. 545.
8. Ibid.
9. Ibid., p. 569.
10. Ibid., p. 700.
11. P. Singer, "New Frontiers in Plasma Etching," *Semiconductor International* (July 1996): p. 153.
12. Ibid., p. 154.
13. Y. Lii, "Etching," p. 349.
14. Ibid.
15. Y. Ye et al., "0.35-Micron and Sub-0.35 Micron Metal Stack Etch in a DPS Chamber—DPS Chamber and Process Characterization," *Proceedings of the Eleventh International Symposium on Plasma Processing* vol. 96-12 ed. G. Mathad and M. Meyyappan (Pennington, NJ: The Electrochemical Society, 1996), p. 222.
16. G. Oehrlein, "Reactive Ion Etching," *Handbook of Plasma Processing Technology* ed. S. Possnagel, J. Cuomo, and W. Westwood (Park Ridge, NJ: Noyes Publishing, 1990).
17. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era, Vol. 1—Process Technology* p. 697.
18. J. Baliga, "Vacuum Pump Designs Adjust to Harsher Conditions," *Semiconductor International* (October 1997): p. 87.
19. P. Singer, "Vacuum Pumping in Etch and CVD," *Semiconductor International* (September 1995): p. 78.
20. S. Tandon, "Challenges for 300 mm Plasma Etch System Development," *Semiconductor International* (March 1998): p. 78.
21. Ibid.
22. P. Singer, "The Many Challenges of Oxide Etching," *Semiconductor International* (June 1997): p. 110.
23. B. El-Kareh, *Fundamentals*, p. 326.
24. Ibid., p. 327.
25. P. Singer, "Many Challenges of Oxide Etching," p. 110.
26. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era—Vol. 1, Process Technology*, p. 556.
27. Y. Wang et al., "High-Selectivity Silicon Nitride Etch Process" *Semiconductor International* (July 1998): p. 238.
28. Y. Lii, "Etching," p. 357.
29. S. Tandon, "Challenges for 300 mm," p. 80.
30. Y. Lii, "Etching," p. 355.
31. Adapted from S. Tandon, "Challenges for 300 mm," p. 80.
32. B. El-Kareh, *Fundamentals*, p. 330.
33. R. DeJule, "Managing Etch and Implant Residue," p. 58.
34. Y. Lii, "Etching," p. 360.
35. D. Taylor, "Wet-Etch Process Improvements Through SPC," *Solid State Technology* (July 1998): p. 119.
36. J. Rembetski, W. Rust, and R. Shepherd, "The Removal of Hard Masks in Semiconductor Processing," *Solid State Technology* (March 1995): p. 68.
37. P. Singer, "Plasma Ashing Moves into the Mainstream," *Semiconductor International* (August 1996): p. 84.
38. Ibid.
39. C. Cheng and J. Oncay, "A Downstream Plasma Process for Post-Etch Residue Cleaning," *Semiconductor International* (July 1995): p. 185.
40. G. Herdt, P. Gillespie, and Y. Wasserman, "Characterization of Damage from Dry Ashing and Residual Removal," *Semiconductor International* (November 1997): p. 80.
41. R. DeJule, "Managing Etch," p. 58.

ION IMPLANT

Intrinsic silicon crystal structure is formed by silicon covalent bonds. As explained in Chapter 2, intrinsic silicon is a poor conductor. Silicon only becomes useful as a semiconductor when its structure and conductivity are altered by adding small amounts of impurities called dopants. This process is referred to as doping. The doping of silicon is fundamental to pn junction fabrication in semiconductor devices. Ion implantation is the primary method for doping wafers.

Doping is widely used throughout wafer fabrication. Dopants are precisely and uniformly placed into the device to alter its electrical performance. Silicon chips require doping with Group IIIA and Group VA dopants

for many different reasons, which will be discussed in this chapter.

The continual shrinking of chip critical dimensions and the subsequent increase in chip packing density force the scaling of all device elements. An example of this is the reduction of the polysilicon gate length to obtain a narrower channel region. A shorter channel length requires shallower junction depths for the source and drain doped regions. Junction depths are ultrashallow for chips designed at the $0.18\text{ }\mu\text{m}$ and below technology nodes. The junction depth requirements decrease to $30 \pm 10\text{ nm}$ for devices with a CD of $0.1\text{ }\mu\text{m}$.¹

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the purpose and applications for doping in wafer fabrication.
2. Discuss the principles and process of dopant diffusion.
3. Provide an overview of ion implantation, including its advantages and disadvantages.

4. Discuss the importance of dose and range in ion implant.
5. List and describe the five major subsystems for an ion implanter.
6. Explain annealing and channeling in ion implantation.
7. Describe different applications of ion implantation.

INTRODUCTION

Doping is the introduction of a dopant into the crystal structure of a semiconductor material to modify its electronic properties (e.g., electrical resistivity). As discussed in Chapter 2, certain Group IIIA and VA elements are used as dopants in wafer fabrication. For review, Table 17.1 on page 476 highlights the four principal dopants used in semiconductor manufacturing.

Dopant species are also referred to as impurities but should not be confused with contaminating impurities. Dopants are introduced into the silicon and other fabrication materials for a variety of reasons. For instance, dopants such as boron and phosphorus are used to form the majority carriers in silicon devices, to form conductive layers in the wafer, or to alter material properties (e.g., doping SiO_2 to form borophosphosilicate glass, or BPSG). The polysilicon gate electrode is doped for conductivity.

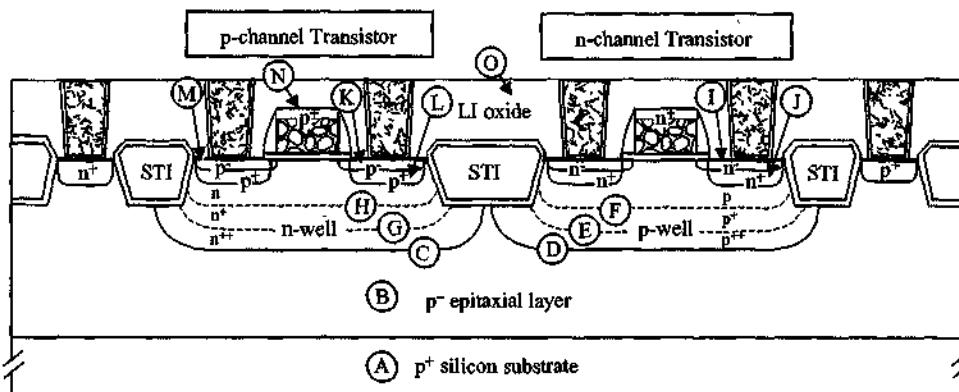
There are two techniques in wafer fabrication for introducing dopant elements to the wafer materials: thermal diffusion and ion implantation. *Thermal diffusion* uses high temperature to move the dopant through the silicon lattice structure. This method is dependent on time and temperature. *Ion implantation* introduces dopants into the substrate through a high-voltage ion bombardment. The dopants are implanted into the wafer through high-energy collisions at the atomic level. In the beginning of semiconductor manufacturing, thermal diffusion was the primary method for doping the wafer. However, due to shrinking critical dimensions and the corresponding scaling requirements, nearly all dopant processes in modern wafer fabrication are now done with

TABLE 17.1 Common Dopants Used in Semiconductor Manufacturing

Acceptor Dopant Group IIIA (p-Type)		Semiconductor Group IVA		Donor Dopant Group VA (n-Type)	
Element	Atomic Number	Element	Atomic Number	Element	Atomic Number
Boron (B)		Carbon	6	Nitrogen	7
Aluminum	13	Silicon (Si)	14	Phosphorus (P)	
Gallium	31	Germanium	32	Arsenic (As)	
Indium	49	Tin	50	Antimony (Sb)	

Note: Common dopants are shaded.

ion implantation (see Figure 17.1). Table 17.2 outlines some critical processes commonly doped with ion implantation versus the limited doping done by diffusion. Figure 17.1 shows these doped regions on a cross section drawing of a CMOS inverter. Note the dopant type (p or n) and concentration (− or +) are also shown. The different doped regions labeled A to O are defined in Table 17.2.

**FIGURE 17.1** CMOS Structure with Doped Regions**TABLE 17.2*** Common Dopant Processes in Wafer Fabrication

Process Step	Common Dopant Species	Ion Implant or Diffusion	Remarks
A. p+ Silicon Substrate	B	Diffusion	The monosilicon crystal is doped during the crystal growing process.
B. p- Epitaxial Layer	B	Diffusion	The epitaxial layer is doped by diffusion during epitaxial silicon growth.
C. Retrograde n-well	P	Ion Implant	Retrograde well has dopant profile with peak concentration buried at a certain depth. The dopant concentration decreases as it approaches the surface.
D. Retrograde p-well	B	Ion Implant	Retrograde well has dopant profile with peak concentration buried at a certain depth.
E. p-Channel Punchthrough	P	Ion Implant	Phosphorous implanted more laterally than source/drain implants to keep electric field of drain from punching through p-channel and reaching source region.

Process Step	Common Dopant Species	Ion Implant or Diffusion	Remarks
F. p-Channel Threshold Voltage (V_T) Adjust	P	Ion Implant	Implant phosphorous to adjust MOS V_T threshold voltage.
G. n-Channel Punchthrough	B	Ion Implant	Boron is implanted more laterally than source/drain implants to keep electric field of drain from punching through n-channel and reaching source region.
H. n-Channel V_T Adjust	B	Ion Implant	Implant boron through oxide layer to adjust MOS threshold voltage.
I. n-Channel Lightly Doped Drain (LDD)	As	Ion Implant	Low-dose implant of arsenic in region adjacent to the n-channel to improve electrical performance by reducing the peak electrical field and therefore hot carriers, which minimizes interface charges trapped in the gate oxide.
J. n-Channel Source/Drain (S/D)	As	Ion Implant	Higher implant dose of arsenic to form n-channel source and drain.
K. p-Channel LDD	BF_2	Ion Implant	Low-dose implant of boron in region adjacent to the p-channel to improve electrical performance between drain and channel region.
L. p-Channel S/D	BF_2	Ion Implant	Higher implant dose of boron to form p-channel source and drain.
M. Silicon	Si	Ion Implant	Implant with nondopant atom to amorphize silicon to reduce transient enhanced diffusion (TED) and channeling effects.
N. Doped Polysilicon	P or B	Ion Implant or Diffusion	Doping of polysilicon gate electrode to make conductive.
O. Doped SiO_2	P or B	Ion Implant or Diffusion	Doped oxide to obtain material benefits (e.g., create better flow and serve as getter).

*Adapted from E. Rimini, *Ion Implantation: Basics to Device Fabrication*, (Boston: Kluwer Academic Publishers, 1995).

Ion implantation typically follows photolithography in the general process flow (see Figure 17.2 on page 478); although, there may be some processes that call for the use of an etched-oxide layer to serve as a mask during either diffusion doping or ion implant. In this case, wafers could flow from etch to diffusion or ion implant. Ion implantation plays a crucial role in chip performance improvement because of the need for smaller dimensions. Transistor performance is increasingly dependent on precise doping profiles in the silicon that are achievable only with ion implant.

Doped Regions

Recall from Chapter 4 that a wafer is uniformly doped with dopant atoms during the crystal growth process to create a p-type or n-type wafer. The type of dopant depends on the manufacturer's specification. During wafer fabrication, dopants are selectively introduced to create the devices on the individual wafers. The selective introduction of dopants into masked openings on the wafer forms a *doped region* where dopants reside within the silicon crystal structure (see Figure 17.3 on page 478). The doped region is characterized by its *dopant profile*, which determines the amount of dopant actually added in the doped region as a function of depth.

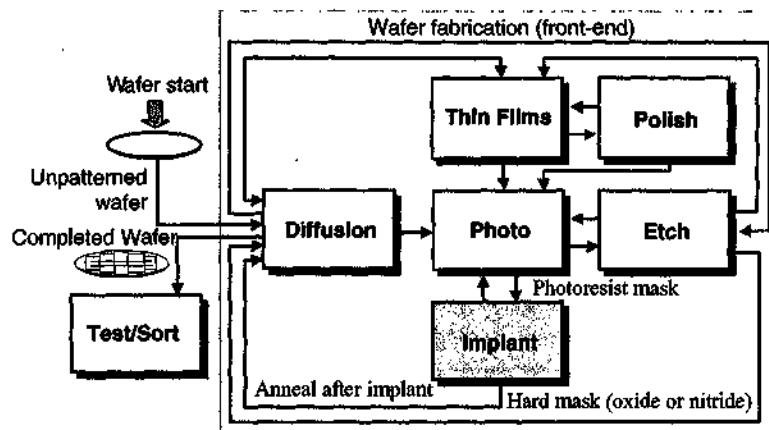


FIGURE 17.2 Ion Implant in Wafer Process Flow
(Used with permission from Advanced Micro Devices)

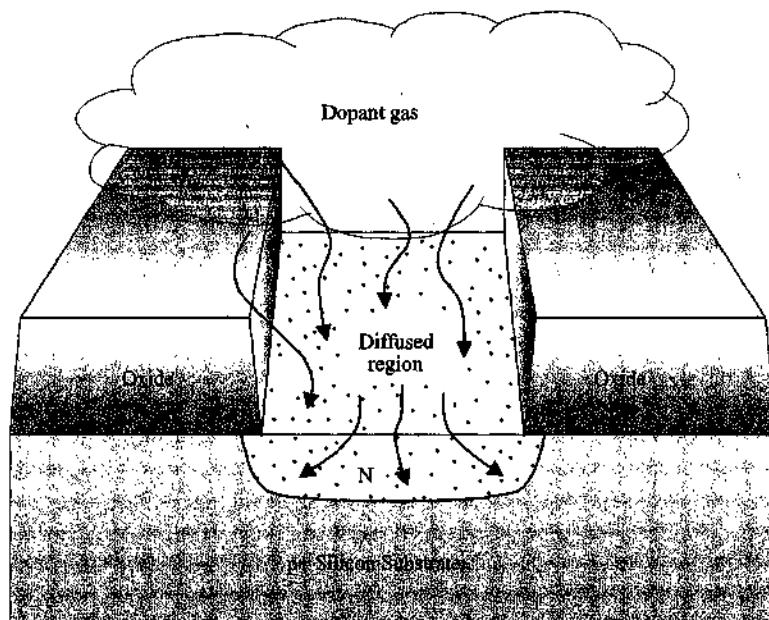


FIGURE 17.3 Doped Region in a Silicon Wafer

The dopant atoms enter the solid silicon through an oxide mask opening. In ion implantation, this mask is often photoresist, but a variety of masks can be used since it is a low-temperature process. Doping by diffusion requires an oxide or nitride mask because this technique is done in a high-temperature furnace.

The doped region can be of the opposite type as the silicon substrate, such as an n-type doped region in a p-type wafer, or it can be the same type with a different dopant concentration (e.g., a heavily doped p⁺ region in a p-type wafer). The location where the conductivity of the doped region makes a change from p-type to n-type, or vice versa, is the pn junction. The exact depth in the wafer where the p-type dopant meets the n-type dopant is the *junction depth*, x_j . At the precise location of the junction depth, the concentration of electrons equals the concentration of holes. Another way of looking at this feature is the net dopant concentration at x_j is zero.

The ultimate interest in wafer doping is related to the individual dopant profiles for the devices after all processing is complete. A wafer undergoes many thermal excursions during processing, such as oxide growth, CVD, and so on, all contributing to undesirable dopant diffusion through the silicon. Dopant diffusion can alter critical parameters (e.g., junction depth and concentration) of the initial doped region and affect device performance. This is an important reason for minimizing the wafer thermal budget.

The main focus of this chapter will be doping by ion implantation. We will first review diffusion to analyze basic concepts and explain how intentional and unintentional diffusion occurs in wafer fabrication.

DIFFUSION

Diffusion is a basic property of matter that describes the movement of one material through another. Diffusion occurs from a region of relatively higher concentration into a region of lower concentration, resulting from the motion of atoms, molecules, or ions. In semiconductor fabrication, high-temperature diffusion is the mechanism used to move a dopant through the silicon crystal lattice. Diffusion can occur as a gas-state, liquid-state or solid-state. An example of liquid-state diffusion is a drop of food colorant in a glass of water. The highly concentrated drop of colorant immediately starts to diffuse into the glass of water and will continue until the entire volume of water is the same color as the drop of colorant. The same type of diffusion occurs in the solid-state when a high concentration of dopant material is introduced to a wafer. The dopant material diffuses through the silicon atoms of the solid crystal lattice.

Diffusion Principles

Thermal diffusion of a solid-state dopant in silicon requires three steps: predeposition, drive-in, and activation. During *predeposition*, or *predep*, wafers are loaded into a high-temperature diffusion furnace and a quantity of dopant atoms is transferred from the source cabinet to the diffusion furnace. The furnace temperature setting is typically 800 to 1100°C for 10 to 30 minutes. The dopants are introduced into a very thin layer of the wafer while maintaining a constant concentration of dopants at the surface. A thin oxide layer, referred to as a *cap oxide*, is grown on the surface to prevent dopant atoms from diffusing out of the silicon. The total number of dopant atoms deposited is referred to as Q . Note that for ion implantation (discussed later in this chapter), Q is referred to as the dose and also represents the number of dopant atoms implanted.

The predeposition step establishes the concentration gradient for the entire diffusion process. The concentration of dopants is highest at the surface and reduces when moving away from the surface, forming a gradient (or slope).¹ This slope establishes the dopant profile, which can be mapped using four point probes (see Chapter 7). Fick's laws, first postulated in 1855, mathematically describe the diffusion process. These laws state in part that the number of particles moving through a cross-sectional area is proportional to the concentration gradient.² Fick's laws are used to predict the dopant concentration at some distance, x , from the surface.

The second part of thermal diffusion is the *drive-in* step. This is a high-temperature process (1000 to 1250°C) used to move the deposited dopants through the silicon crystal to the desired junction depth in the wafer. No additional dopants are added to the wafer. However, the wafer surface oxidizes in the high-temperature environment, which affects the diffusion of the dopant atoms during drive-in. Some dopants, such as boron, tend to move into the growing oxide layer while others, such as phosphorous, are pushed away from the SiO_2 . This modification of dopant concentration due to the silicon surface oxidation is termed *redistribution*.

The third part of thermal diffusion is the *activation* step. Here the temperature of the furnace is raised slightly higher to enable the dopant atoms to bond with the silicon atoms in the lattice structure. This action activates the dopant atoms, which changes the conductivity of silicon.

Dopant Movement ■ Each dopant has a particular *diffusivity* in silicon, which represents the rate (or speed) that the dopant moves in the wafer. The higher the diffusivity, the faster the dopant moves. Diffusivity increases with temperature, which is reflected in a term known as the *diffusion coefficient*, D , of the dopant being diffused. Elements have different diffusion coefficients within the temperature range of the drive-in step that reflect whether they are fast or slow diffusers in silicon. The diffusion coefficient is an equation variable in Fick's laws used to predict the final dopant concentration in silicon.

Within the wafer, dopant atoms move by two different mechanisms: interstitial and substitutional (see Figure 17.4 on page 480). Dopants with high diffusivity, such as gold (Au), copper (Cu), and nickel (Ni), use primarily *interstitial movement* to move between the interstitial space

between regular crystal sites of the silicon lattice. Slower moving dopants, such as arsenic and phosphorus that are common in semiconductor doping, typically use *substitutional movement* where atoms fill empty crystal positions in the lattice.

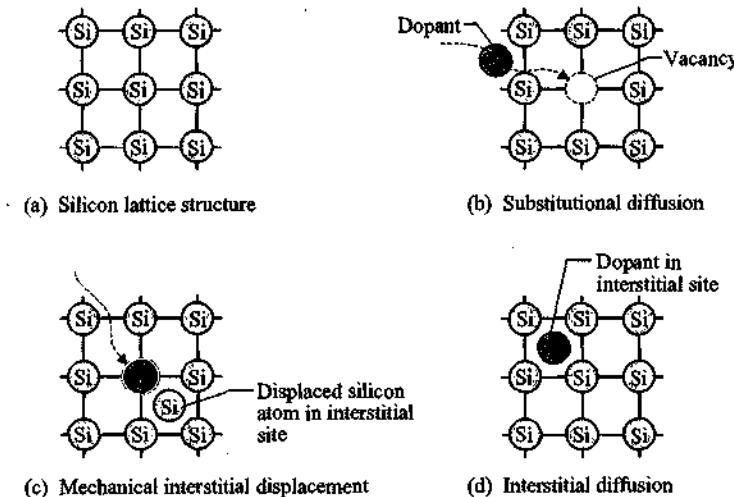


FIGURE 17.4 Dopant Diffusion in Silicon

Dopants are useful to form semiconductor silicon only if they are *activated dopants*, meaning they are part of the silicon lattice structure. An activated dopant can act as a donor or acceptor of electrons and is an n-type or p-type dopant with respect to silicon. If the dopant occupies interstitial space, then it is not activated and is ineffective as a dopant. Thermal energy moves dopants into the regular crystal sites, a process known as *crystal activation*. Activation is a regular part of diffusion because it occurs at a high temperature. For ion implant, crystal activation is done with an anneal step.

Solid Solubility ■ At a given temperature, there is a limit to how much dopant can be absorbed by the silicon. This is referred to as the *solid solubility limit*. This is true for most materials, such as pouring salt into a container of water and watching it dissolve. If salt is continually added, at some point the solubility of the salt in the water reaches a limit where it no longer dissolves. Each particular dopant has a solid solubility limit (see Table 17.3). Note that only a fraction of the dopants in silicon are actually activated and contribute electrons or holes for conduction (about 3 to 5%).³ Most dopants remain in interstitial sites and are electrically inactive.

TABLE 17.3* Solid Solubility Limits in Silicon at 1100°C

Dopant	Solubility Limit (atoms/cm ³)
Arsenic (As)	1.7×10^{21}
Phosphorus (P)	1.1×10^{21}
Boron (B)	2.2×10^{20}
Antimony (Sb)	5.0×10^{19}
Aluminum (Al)	1.8×10^{19}

*SEMATECH "Diffusion Processes," *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 15.

Lateral Diffusion ■ The diffusion mask is either silicon dioxide (SiO_2) or silicon nitride (Si_3N_4) since a photoresist mask could not withstand the high-temperature process. When atoms diffuse into the wafer, they move in all directions: down into the silicon, laterally, and back out of the wafer. *Lateral diffusion* occurs when dopant atoms travel in a direction along the surface of the wafer. Typically, lateral diffusion during a thermal diffusion process is 75 to 85% of the vertical junction

depth.⁴ Lateral diffusion is undesirable in advanced MOS circuits because it can cause a reduction in the channel length, affecting device density and performance.

Diffusion Process

The objective of the diffusion process is to bring the diffusing impurity in contact with a wafer and maintain the specified time and temperature for diffusion to occur. Diffusion takes place in high-temperature diffusion furnaces, discussed in Chapter 10. Typically, one furnace is used for the pre-deposition, drive-in, and activation steps. Quartz and other furnace parts used in diffusion processes should be segregated from other furnace processes to avoid cross-contamination.

Diffusion should produce reproducible results from run-to-run and wafer-to-wafer. The eight steps required to properly perform diffusion in wafer fabrication are:

1. Run qualification test to ensure the tool meets production quality criteria.
2. Verify the wafer properties using a lot control system.
3. Download the process recipe with the desired diffusion parameters.
4. Set up the furnace, including a temperature profile.
5. Clean the wafers and dip them in hydrofluoric (HF) to remove native oxide.
6. Perform predeposition: load wafers into the deposition furnace and diffuse the dopant.
7. Perform drive-in: increase the furnace temperature to drive-in and activate the dopant bonds, then unload wafers.
8. Measure, evaluate, and record the junction depth and sheet resistivity.

The first six steps are done to prepare wafers for diffusion in the drive-in operation of step 7. A qualification test (or "qual") is performed on the furnace. This test determines the tool's production worthiness in terms of particles and other specified criteria. Once the tool is proven production worthy, the appropriate wafer lots are identified and the proper process recipe is downloaded to the furnace. A thermal profile should always be done to verify furnace temperatures are correct for the process to be performed. In a production environment, a given furnace is dedicated to a specific process step and is maintained at the precise temperature profile.

Wafer Cleaning ■ Wafer cleaning is critical because contaminants can block the diffusion of the dopant atoms into the wafer. Wafers should be cleaned immediately prior to loading the wafers into the furnace to minimize wafer contamination (e.g., native oxide). Cleaning typically includes an immersion in an acid and oxidizer, followed by an etch in an HF solution to remove any remaining oxides (see Chapter 6 for wet cleaning). The use of a dilute HF solution is referred to as a *degaze* step to remove the poor quality native oxide on the wafer surface.

Dopant Sources ■ Although they may have been used in the early days of the semiconductor industry, it is impractical to use pure elements as a dopant source for submicron IC manufacturing. For instance, boron and phosphorus are solids at room temperature with low vapor pressures, making them difficult to melt or vaporize. Dopant concentration is also difficult to control when using a solid dopant source. Dopants are typically supplied as a gaseous or liquid source in the form of a compound. Some of the most common dopants are listed in Table 17.4 on page 482.

A liquid source (bubbler) has a carrier gas (e.g., nitrogen) bubbled through it and is delivered to the furnace tube as a vapor. Oxygen is also supplied for the dopant source to react and form oxides. Using boron as an example, the boric oxide undergoes a second reaction with silicon to form a boron-rich silicon dioxide layer on top of the wafer. This layer serves as the local source of boron for the predeposition step. The following two reaction equations illustrate how the diborane source is converted into a boron dopant:⁵

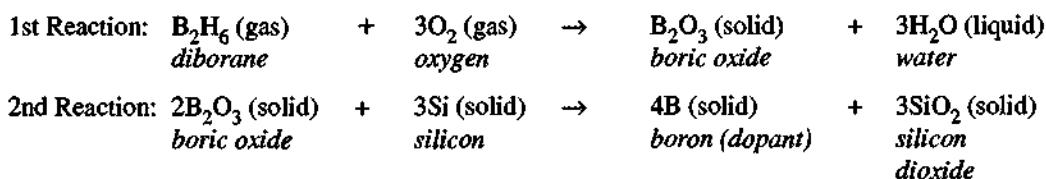


TABLE 17.4* Typical Dopant Sources for Diffusion

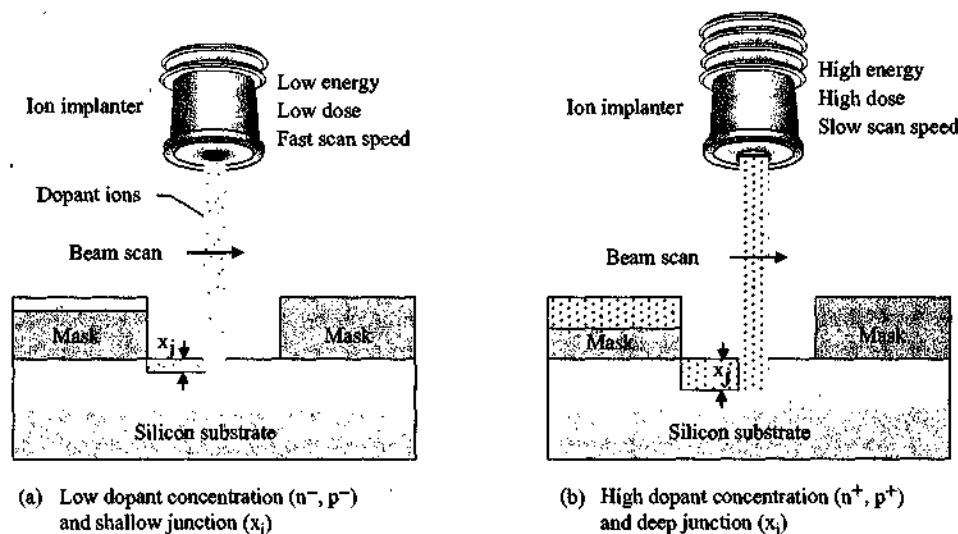
Dopant	Formula of Source	Chemical Name
Arsenic (As)	AsH ₃	Arsine (gas)
Phosphorus (P)	PH ₃	Phosphine (gas)
Phosphorus (P)	POCl ₃	Phosphorus oxychloride (liquid)
Boron (B)	B ₂ H ₆	Diborane (gas)
Boron (B)	BF ₃	Boron tri-fluoride (gas)
Boron (B)	BBR ₃	Boron tri-bromide (liquid)
Antimony (Sb)	SbCl ₅	Antimony pentachloride (solid)

*SEMATECH "Diffusion Processes," *Furnace Processes and Related Topics*, (Austin, TX: SEMATECH, 1994), p. 7.

Gaseous sources are delivered to the furnace directly from the gas cylinder through a mass flow controller. The dopant is diluted in the cylinder with an inert gas to help prevent system corrosion and control the flow of the gas. A point-of-use filter should always be used to prevent fine particulates from getting into the furnace. Some diffusion sources are highly toxic, especially the gaseous sources. The most toxic of the commonly used dopant sources are arsine (AsH₃) and diborane (B₂H₆). Proper storage and usage is required at all times, especially to avoid a gas leak.

ION IMPLANTATION

Ion implantation is a method for introducing controlled amounts of dopants into the silicon substrate to change its electronic properties. It is a physical process—there are no chemical reactions. A number of applications use ion implantation in modern wafer fabrication (see Table 17.2 on page 476). The main application of ion implantation is the doping of semiconductor materials. Each doping application has specific requirements for the concentration of the dopant material and its depth. Ion implantation is preferred over diffusion in nearly all applications because of its ability to repeatedly control dopant concentration and depth. It has become a critical process to meet the challenges of wafer fabrication with sub-quarter micron feature sizes and larger diameter wafers (see Figure 17.5).

**FIGURE 17.5** Controlling Dopant Concentration and Depth

Overview

Ion implant processing is done in one of the most complex semiconductor processing tools, called the *ion implanter* (see Figure 17.6). The implanter has an ion source component that creates positive-charged dopant ions from a source material. The ions are extracted and then separated in a mass analyzer to form a beam of the desired dopant ions. The number of ions in the beam is related to the concentration of dopants introduced into the wafer. The beam of ions is accelerated in a voltage field to attain a high velocity (on the order of 10^7 cm/sec).⁶ Because of the high velocity, the ions have kinetic energy that is used to implant the dopants into the silicon crystal lattice structure of the target wafer. The beam scans the wafer to provide uniform doping across the wafer surface. Implantation is followed by a thermal anneal step to activate the dopant ions in the crystal structure. All implanter processing is done in a high vacuum.

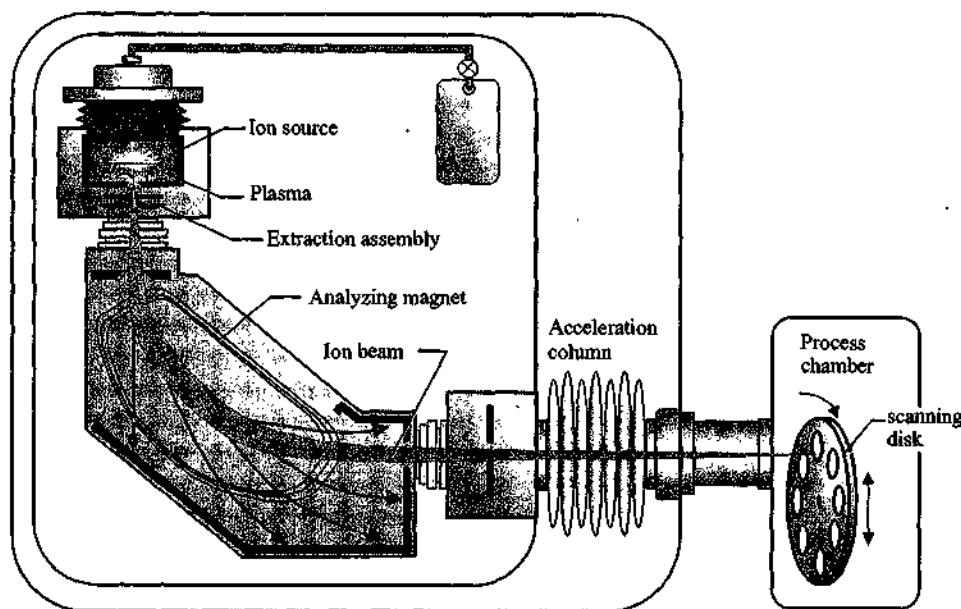
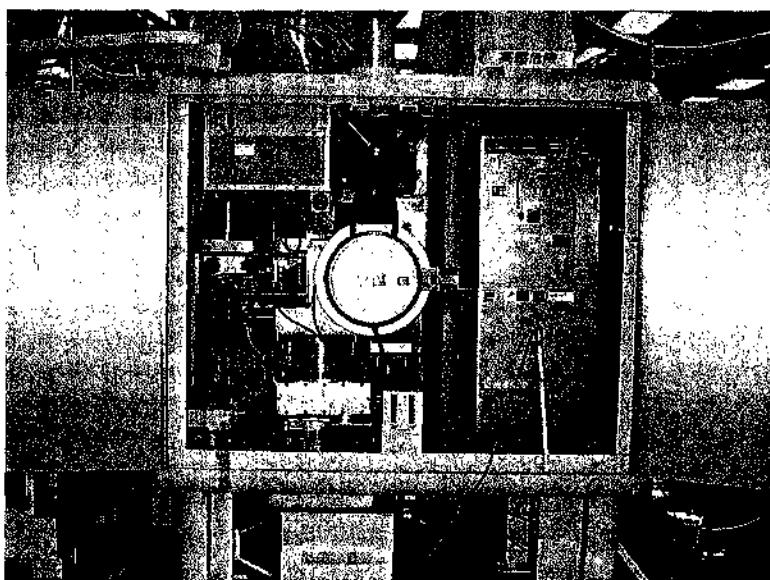


FIGURE 17.6 General Schematic of an Ion Implanter



Ion Implanter
(Photo courtesy of
Varian Semiconductor
Equipment, VISION 80)

There are two major goals for sub-0.25 μm doping requirements:

1. To introduce a uniform, controlled amount of a specific dopant into the wafer
2. To place the dopants at a desired depth.

Advantages of Ion Implant ■ The primary advantages of ion implantation are listed in Table 17.5. These benefits effectively increase the process flexibility for fabricating wafers with sub-quarter micron technology.

TABLE 17.5 Advantages of Ion Implantation

Advantages	Description
1. Precise control of dopant concentration.	Ability to precisely control the number of implanted dopants over a wide range of concentrations, from 10^{10} to 10^{17} ions/cm ² , and controlled to within $\pm 2\%$ over this range. In contrast, diffusion can control dopant concentration to within 5 to 10% at high concentrations and becomes less effective at low concentrations.
2. Good dopant uniformity.	Controls dopant uniformity by scanning the ion beam over a large wafer area, which is critical for smaller feature sizes on larger wafers.
3. Good control of dopant penetration depth.	Controls the ion energy during the implant to achieve good control of dopant penetration depth. This permits design flexibility, such as buried layers where the maximum dopant concentration is at the buried layer and the minimum concentration is at the wafer surface.
4. Produces a pure beam of ions.	Mass separation techniques produce a pure beam of ions free of contamination. Different impurities can be selected for implantation. It is done in a high vacuum that minimizes contaminants.
5. Low temperature processing.	Implantation is performed at moderate temperatures ($<125^\circ\text{C}$) which permits the use of a variety of masks, including photoresist.
6. Ability to implant dopants through films.	Dopants can be implanted through films, such as an oxide or nitride. This condition allows adjustment of the threshold voltage of MOS transistors after the gate oxide is grown. It increases the flexibility to implant at various stages of the process.
7. No solid solubility limit.	The dopant concentration during implantation is not limited by the solid solubility limit of the silicon wafer.

Disadvantages of Ion Implant ■ The major disadvantage of ion implantation is the damage done to the crystal structure when the energetic dopant ions bombard the silicon host atoms. As the high-energy ions enter the crystal and collide with substrate atoms, energy is transferred and some silicon atoms are displaced from the lattice. This action is referred to as *radiation damage*. Most or all of this crystal damage is corrected with a high-temperature anneal step.

Another drawback of implantation is the complexity of the implanter equipment. However, this drawback is offset by the implanter's capability for dose and depth control and overall process flexibility.

Ion Implant Parameters

Ion implant is a flexible process that must meet stringent chip design and productivity requirements for wafer fabrication. The important ion implant parameters are:

- ◆ Dose
- ◆ Range

Dose ■ The *dose* (Q) is the number of implanted ions per unit area of wafer surface, with units of atoms/cm² (also stated as ions/cm²). To count the implanted ions, Q is calculated by the following dose equation:

$$Q = \frac{It}{enA}$$

Where, Q = dose in atom/cm²

I = beam current in coulombs/sec (amperes)

t = implant time in seconds

e = electronic charge equal to 1.6×10^{-19} coulombs

n = charge per ion (for instance, B^+ is equal to 1)

A = area implanted in cm²

One of the main reasons why ion implantation is so important in wafer fabrication is its ability to repeatedly put the same dose into each wafer. The implanter does this by virtue of the positive charge on each ion. When formed into a beam, the flux of positive dopant ions represents a *beam current* measured in millamps (mA). A low-to-medium beam current is from less than 0.1 mA up to about 10 mA, while a high beam current is from about 10 mA up to approximately 25 mA. As seen in the equation above, the magnitude of the beam current is a key variable for defining the dose. If the beam current increases, then more dopant atoms are implanted per unit time. A high beam current is often desirable to increase wafer throughput (more ions are implanted per unit time of production), but can also create uniformity problems.

Range ■ The ion *range* is the total distance an ion travels in the silicon during ion implantation. To characterize range it is necessary to understand energy. When the ions are accelerated due to electric potential difference to a high velocity, they gain energy. The ion energy is kinetic energy (KE) due to its motion, and it typically is expressed in units of joules. However, for ion implantation, the energy is usually stated in terms of the number of the electronic charge times the difference in potential, or electron volts (eV). The equation that describes this energy is:

$$KE = nV$$

Where, KE = energy in electron volts (eV)

n = charge state of the ion (e.g., "+" = 1, "++" = 2)

V = voltage difference in volts

For example, if an ion with a single positive-charge state moves in an electric field with a voltage difference of 100,000 volts (100kV), then its energy is:

$$\begin{aligned} KE &= nV \\ &= (1)(100\text{kV}) \\ &= 100 \text{ keV} \end{aligned}$$

Higher implanter energy means the dopant atom will have an increased range and travel deeper into the wafer. Energy is important in implanters because of the need to control the range and, therefore, junction depth of the dopants. High-energy implanters have an energy greater than 200 keV and may be as high as 2 to 3 MeV. High-energy implant is used in applications such as deep retrograde wells and retrograde triple wells (see Table 17.6 on page 486). A retrograde well has higher doping concentrations deeper in the well than at the surface. Ultralow-energy implanters have energies presently down to about 200 eV to enable doping at very shallow depths for source/drain applications.⁷

There is a *projected range*, R_p , which is how far the implanted ions travel into the wafer, depending on the ion mass and energy, the target mass and the beam direction with respect to the wafer crystal structure (see Figure 17.7 on page 486).⁸ At the same time, not all ions come to stop simply at the projected range; some stop at a shorter distance and others at a greater distance. Ions will also move in a lateral direction. Combining all these ion movements produces a distribution of distances traveled by the dopant atoms implanted in the wafer, referred to as *straggle*, or ΔR_p . The R_p (projected range) indicates how shallow or deep the junction depth can be formed, whereas the ΔR_p (straggle) represents the spread of the implanted species around R_p . As the implant energy of

TABLE 17.6 Classes of Implanters

Class of Implanter System	Description and Applications
Low/Medium Current	<ul style="list-style-type: none"> Highly pure beam currents <10 mA. Beam energy is usually <180 keV. Most often the ion beam is stationary and the wafer is scanned. Specialized applications of punchthrough stops.
High Current	<ul style="list-style-type: none"> Generate beam currents >10 mA and up to 25 mA for high dose implants. Beam energy is usually <120 keV. Most often the wafer is stationary and the ion beam does the scanning. Ultralow energy beams (<4keV down to 200 eV) for implanting ultra-shallow source/drain junctions.
High Energy	<ul style="list-style-type: none"> Beam energy exceeds 200 keV up to several MeV. Place dopants beneath a trench or thick oxide layer. Able to form retrograde wells and buried layers.
Oxygen Ion Implanters	<ul style="list-style-type: none"> Class of high current systems used to implant oxygen in silicon-on-insulator (SOI) applications.

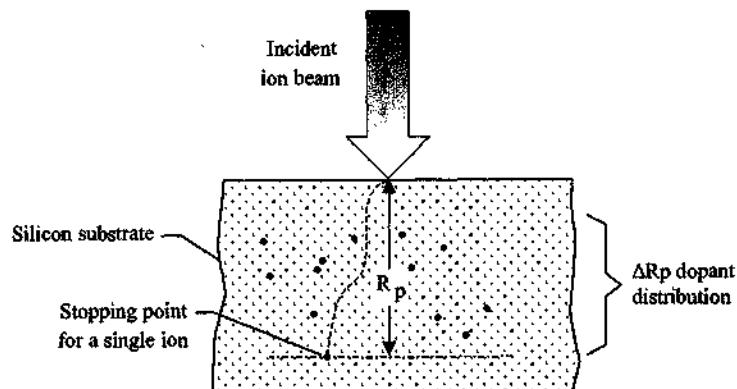


FIGURE 17.7 Range and Projected Range of Dopant Ion

dopant atoms increases, the projected range increases but the peak concentration of dopants decreases because the range straggling increases. Projected range charts are available that predict the projected range for a given implant energy (see Figure 17.8).

The implanted ions will travel some depth into the target wafer before they are brought to rest by energy loss due to collisions with silicon atoms (see Figure 17.9). The two primary energy loss mechanisms are electronic stopping and nuclear stopping.⁹ *Electronic stopping* of dopant atoms is caused by interactions with the target electrons, similar to stopping a projectile in a thick medium, such as a child jumping into a pile of plastic balls. *Nuclear stopping* of implanted ions is caused by collisions between atoms that cause a displacement of silicon atoms. It can be visualized as the collision between two hard spheres, such as billiard balls. Depending on the ion mass and energy, an implanted atom can displace as many as 10^4 silicon atoms by nuclear collisions before coming to rest.¹⁰

The dopant moving through the silicon atoms creates a damage path in the crystal lattice, with the nature of the damage depending on whether the dopant ion is light or heavy (see Figure 17.10). A light dopant atom glances off silicon atoms with little energy transfer and is deflected through a large scattering angle. A heavy ion has a large energy transfer each time it strikes a silicon atom and is deflected through a relatively small scattering angle. Each displaced silicon atom also produces a large number of displacements.

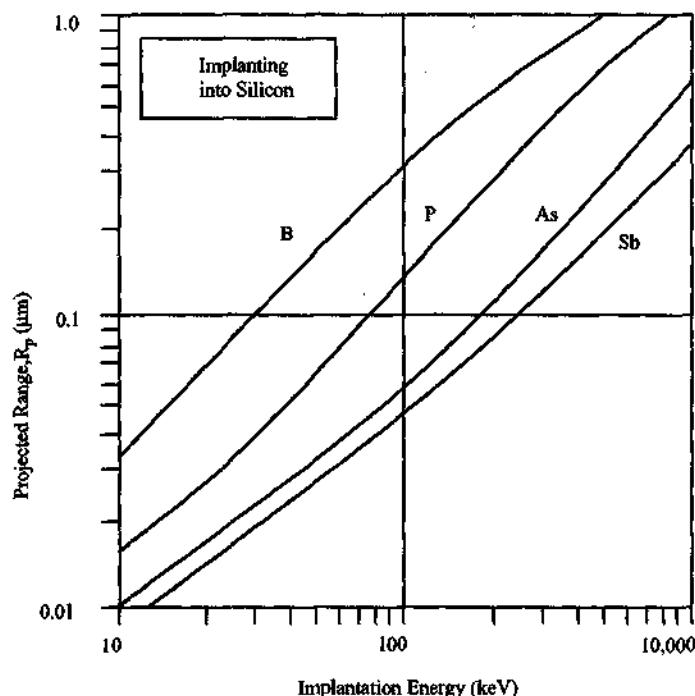


FIGURE 17.8 Projected Range Chart for Implant Energy
Redrawn from B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies*, (Boston: Kluwer, 1995), p. 388.

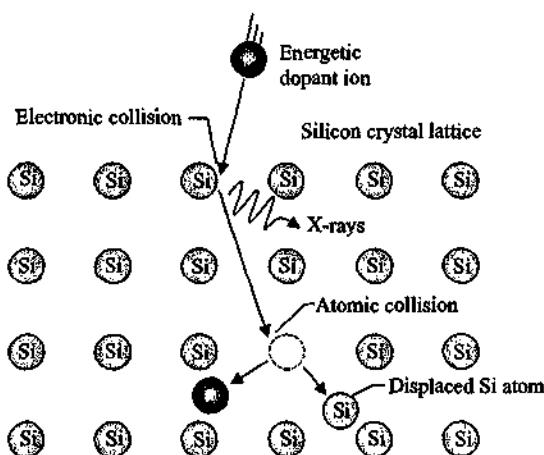


FIGURE 17.9 Energy Loss of an Implanted Dopant Atom

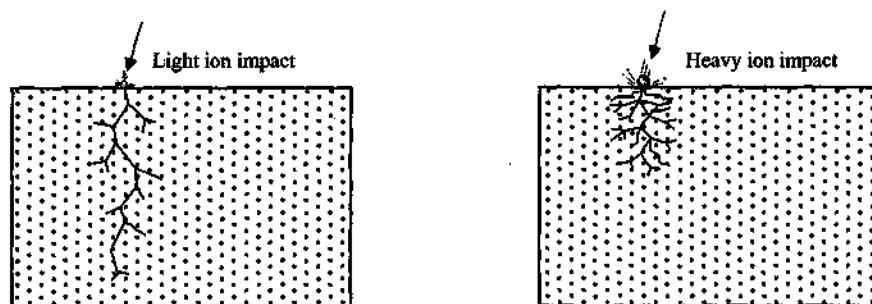


FIGURE 17.10 Crystal Damage Due to Light and Heavy Ions

ION IMPLANTERS

An ion implant tool consists of the following five major subsystems:

- ◆ Ion source
- ◆ Extraction electrode and ion analyzer
- ◆ Acceleration column
- ◆ Scanning system
- ◆ Process chamber

The ion source and extraction assemblies are generally mounted in the same vacuum chamber (see Figure 17.11). The ion source generates positive ions from a specific dopant gas or solid. The circular suppression electrode attached to the extraction assembly uses the electric field of a negative high voltage to draw the positive ions out of the ion source. Electromechanical controls on the extraction assembly aid in positioning the extraction electrode directly in front of the ion source plasma chamber. On older ion implanters, these controls could be manually adjusted from the operator's control panel. Today these controls are set by a process recipe and maintained by the tool's internal computer software.

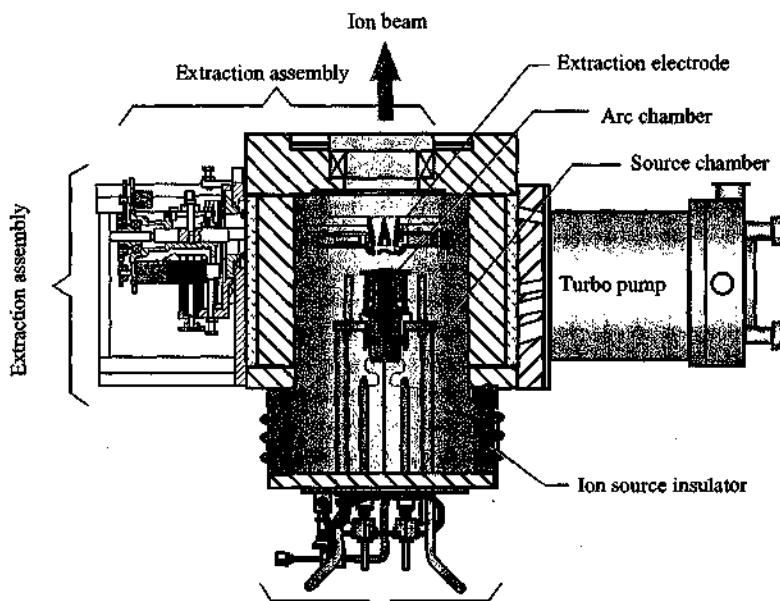


FIGURE 17.11 Ion Source and Extraction Assembly Housing
(Used with permission from Applied Materials, Inc., Precision Implanter 9500)

Ion Source

The species to be implanted must be present in the beam as a charged particle, or ion. Because of their electrical charge, ions can be controlled and accelerated with electric and magnetic fields. Ions for implantation are generated in an *ion source*. A positive ion is formed from a dopant gas source or by vaporizing a solid. Typically B^+ , P^+ , As^+ , and Sb^+ are produced by ionization of atoms or molecules. The most common feed material for a source of dopant atoms is a gas, such as B_2H_6 , BF_3 , PH_3 , and AsH_3 . The gas is packaged in a relatively small cylinder (0.4 to 2.2 liters in size) and diluted with hydrogen to reduce the risks associated with inadvertent gas releases.

Another method of supplying the feed material is by heating and vaporizing a solid material, which is sometimes used to obtain arsenic (As) or phosphorus atoms from solid pellets.¹¹ The solid material is vaporized at about 900°C and the volatile dopant atoms are transported to the ion source chamber. No dilute hydrogen is added, which permits higher maximum beam currents. A disadvantage of a solid source is the long setup time, or tune time, which can be 40 to 180 minutes. Most

of this time is required to heat and stabilize the vaporizer. There is also the need for more frequent maintenance due to the buildup of sputtered materials on the beam-line components. However, many IC manufacturers prefer to use solid dopant sources because it is easier to deal with them from an environmental and safety standpoint.

Ion Creation ■ Ions are generated in the ion source by bombarding the feed gas atoms with electrons. Electrons are often produced by a simple and robust hot tungsten filament source. The *Freeman ion source* is one of the most common electron sources.¹² In this source the rod-shaped cathode filament is held inside an arc discharge chamber with a feed gas inlet. The walls of the arc chamber are the anode. When feed gas is introduced, a plasma is created around the filament in the arc chamber by passing a high current through the filament and applying a voltage of about 100V between the cathode and anode. Positive ions are produced by the collisions between energetic electrons and the feed gas molecules. An external source magnet applies a magnetic field parallel to the filament to increase ionization and to stabilize the plasma. During operation the current between the cathode and anode is adjusted, along with the feed gas, to maintain a stable plasma. A variation of the Freeman ion source is the Bernas ion source that uses a pigtail type of filament and a negative-charged reflector plate to improve electron confinement and efficiency (see Figure 17.12).

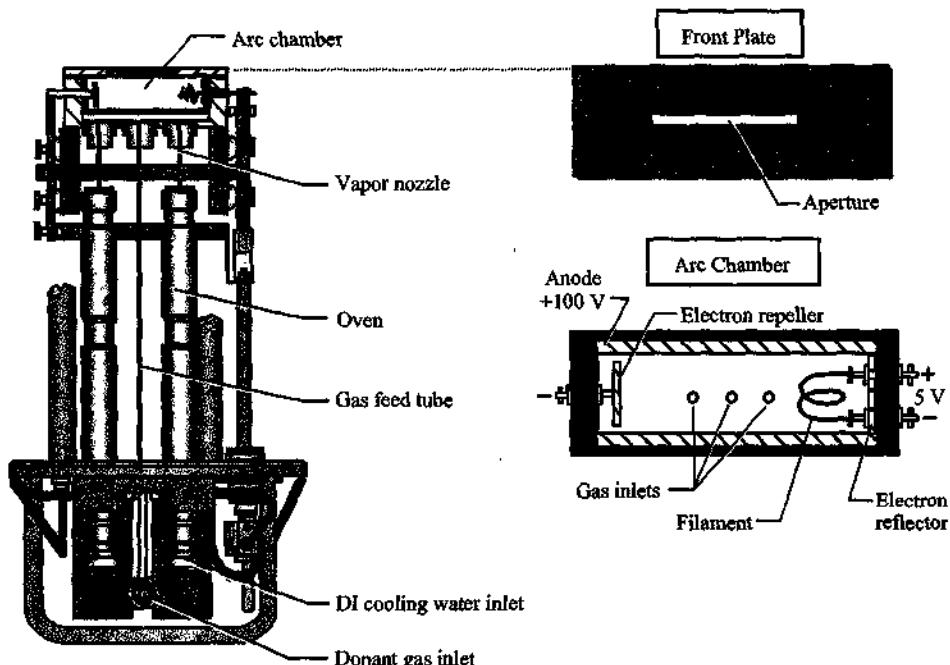


FIGURE 17.12 Schematic of Bernas Ion Source Assembly
(Used with permission from Applied Materials, Inc., Precision Implanter 9500)

The bombardment of electrons breaks up the feed gas molecule into different ion species. Boron trifluoride (BF_3) is often used as a feed gas because boron in its elemental form has a very low vapor pressure, requiring a temperature of 2000°C to vaporize it. BF_3 also has greater mass and lower diffusivity than boron, making it a preferred p-type dopant for forming shallow junctions. In the ion source, BF_3 is broken up into many different species (isotopes), such as B^+ , B_{10}^+ , B_{11}^+ , BF^+ , BF_2^+ , F^+ , and F_2^+ . The B^+ ion is the atom desired to be implanted into the wafer and will be selected when the beam passes through the analyzer magnet.

There are alternative designs under evaluation for ion sources, such as an RF (radio frequency) ion source, cold cathode source, and microwave ion source. An RF source excites the feed gas molecules in a magnetic environment, generating a cooler plasma with the possibility of higher beam currents with a longer ion source lifetime.¹³

Extraction and Ion Analyzer

The traditional implanter extraction system collects all the positive ions created inside the ion source and forms them into a beam. The ions are extracted through a slit in the ion source. They are repelled by the positive bias arc chamber (anode) and are attracted to the negative bias (cathode) of the *extraction assembly* (see Figure 17.13). Since positive ions are like little magnets, each with a positive charge, they are attracted to a negative electric field. The higher the electric field, then the faster the ion will move. At the same time, the faster the ion moves, then the higher is its kinetic energy and the farther it will penetrate into the wafer when it strikes the surface. The negative bias of the extraction electrode also repels any stray electrons in the source plasma and forms the positive ions into a beam. A negatively biased *suppression electrode* is used to focus the ion beam into a parallel beam for efficient beam transport through the implanter. This condition is important for delivering high beam currents to wafers.

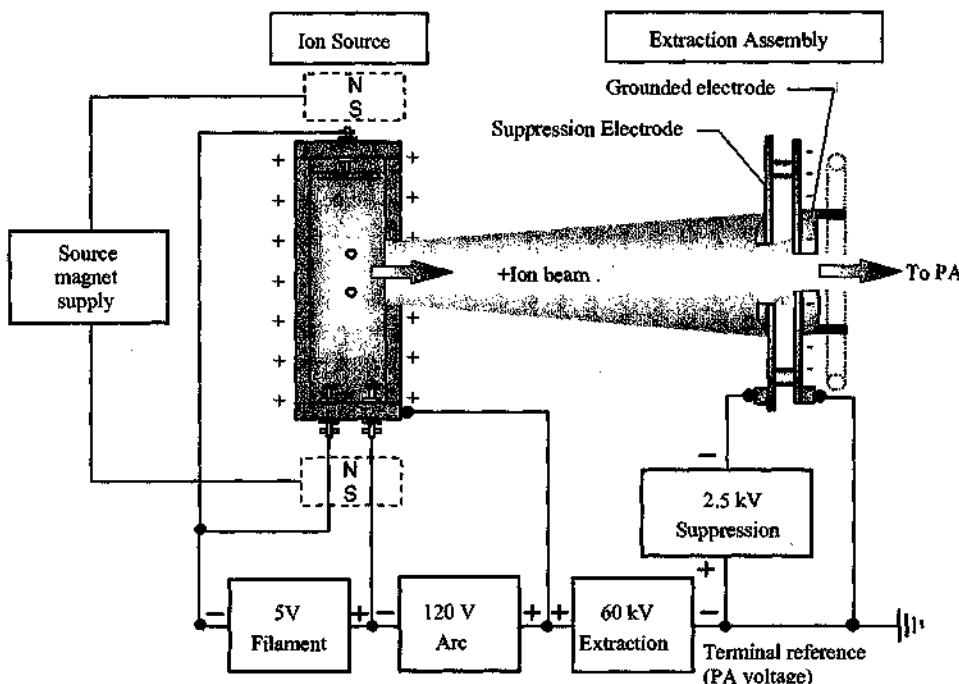


FIGURE 17.13 Interaction of Ion Source and Extraction Assemblies
(Used with permission from Applied Materials, Inc., Precision Implanter 9500)

Mass Analyzer Magnet ■ The ions extracted from the source contain many different ion species and travel at a relatively high speed due to the acceleration provided by the extraction voltage. The species in the ion beam have different atomic mass units (amu). In ion implanters, a magnetic *ion analyzer* separates the desired dopant ion from the main body of ion species. As shown in Figure 17.14, the analyzer magnet is shaped in a 90° angle. The magnetic field of the analyzer magnet causes the ion species to be deflected into an arc. A spectrum of ions is formed based on their amu. For a given field strength, ions with heavy masses are not able to bend at the appropriate angle, while lighter ions bend too easily. There is one ion species, however, that bends just enough at the appropriate field strength to allow it to pass through the center of the analyzer magnet. This is the dopant species that eventually gets implanted into the wafer.

The radius of the arc formed by the ion depends on the mass of the individual species, its speed, the strength of the magnetic field, and the ion's electronic charge. The magnetic field strength is adjusted to match the desired path of the dopant ion. This desirable dopant ion passes through a slit at the end of the analyzer while all other ions strike the walls of the analyzing magnet.

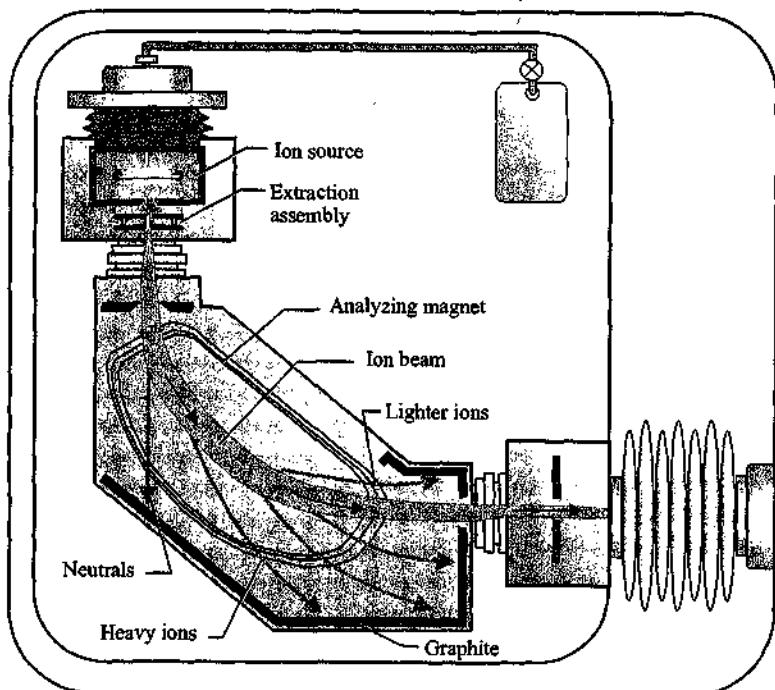
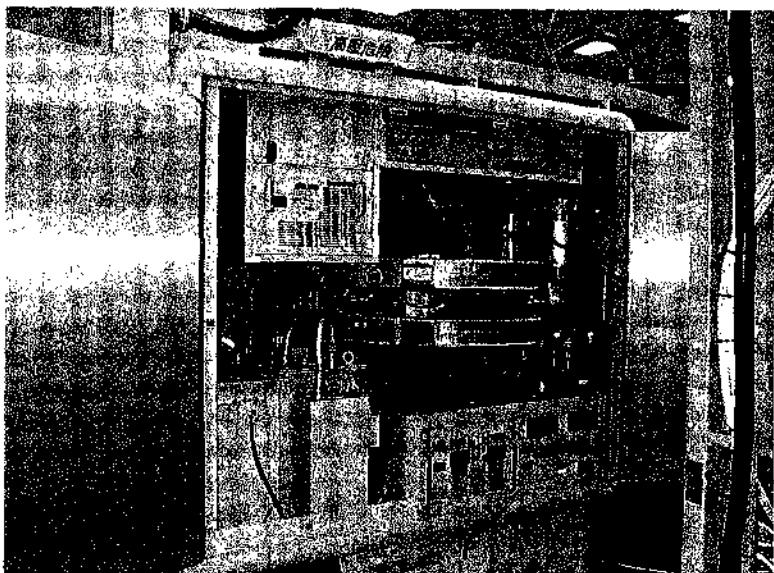


FIGURE 17.14 Analyzing Magnet



Ion Implanter
Analyzing Magnet
(Photo courtesy
of Varian
Semiconductor
Equipment,
ViSiOn 80)

Acceleration Column

To achieve additional ion acceleration (and therefore energy) beyond the analyzer magnet, positive ions are accelerated in an electric field inside an *acceleration column* (see Figure 17.15 on page 492). This acceleration column is a linear design made of a series of electrodes separated by insulators, each with an increasing negative charge. As the positively charged ions enter the column, they start accelerating. In this manner the total voltage difference in the electrode assembly and the acceleration column accelerates the dopant ions. The higher the total voltage, then the higher the velocity and, therefore, energy. Higher beam energy means the dopant ions are implanted deeper into the wafer, while lower beam energy is used for ultrashallow implants. Different beam energies relative to the dose are shown in Figure 17.16 on page 492.

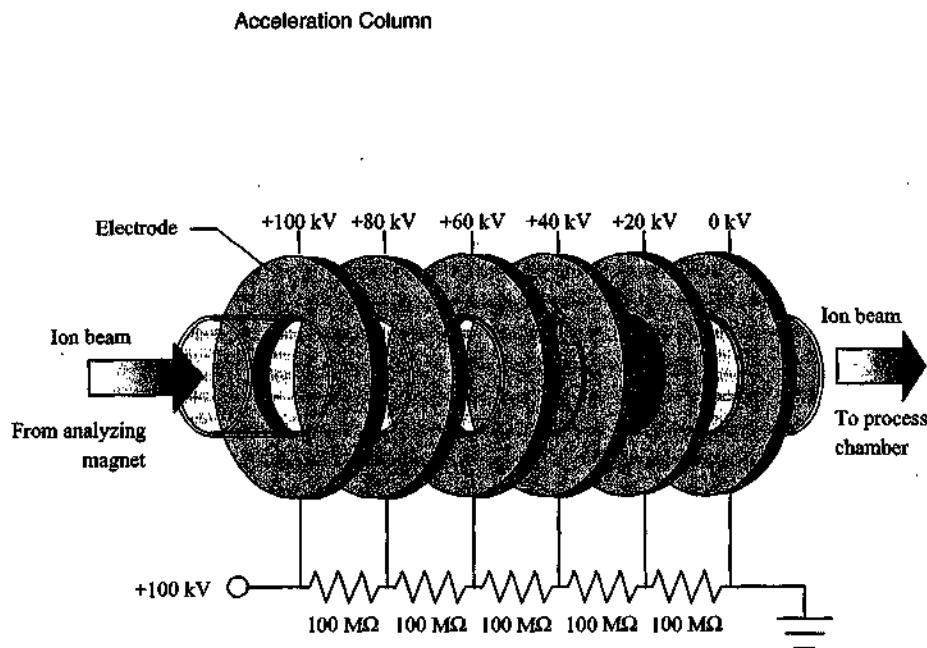
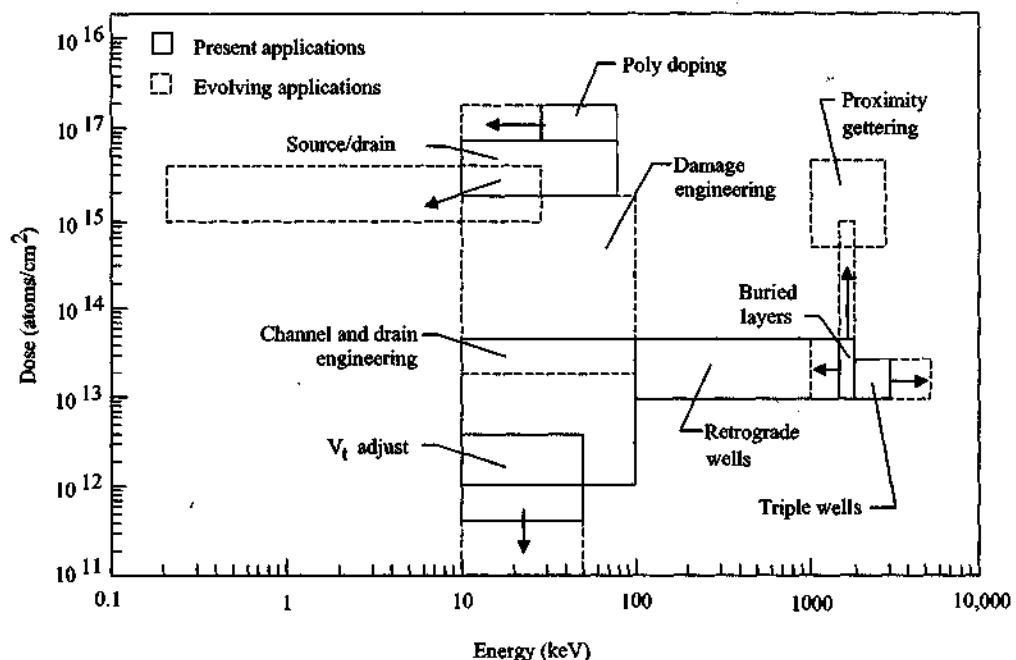


FIGURE 17.15 Acceleration Column

FIGURE 17.16 Dose Versus Energy Map
(Used with permission from Varian Semiconductor Equipment)

The dose versus energy map highlights the importance of energy (range) and dose (concentration) for ion implantation.¹⁴ Beam energy defines the projected range that the ions are implanted in the wafer surface and varies from high energy (>200 keV and into multiple MeV) to low energy (<120 keV and down to 200 eV). Dose is directly related to the concentration of dopants in the wafer and is represented by the beam current, or number of ions in the beam, ranging from low, medium, to high current (>15 mA). Implanters are often classified according to their maximum operating beam current and acceleration energy.

Post-Accelerator ■ There is additional beam focusing that takes place in the *post-accelerator* component of the acceleration column. It typically uses quadrupole lens focusing based on four cylindrical poles in a two-lens system, employing electrostatic or magnetic repulsion to form the ion beam into a focused, circular beam.

High-Current and High-Energy Beam ■ High-energy implanters are used to implant buried dopant layers, such as for retrograde and triple wells. Most commercial high-energy ion implanters employ *linear accelerator* technology to accelerate ions to high velocity and obtain beam energies in excess of 200 keV and into the MeV range.¹⁵ High beam current (up to 25 mA) is desirable to reduce the implant time and increase wafer throughput through the implanter. A linear accelerator is constructed as an alternating series of high-voltage (several 10 kV) electrodes and grounded quadrupole focusing lenses (see Figure 17.17). The timing of the voltage between each electrode and grounded quadrupole is set to match the arrival of the dopant ions at each gap. The ions are accelerated through the gaps, with the final energies up to 20 times higher than the maximum acceleration in a traditional beamline acceleration column. This condition avoids the use of extremely high voltages. The linear accelerator also aids in removing beam contaminants by reducing masses other than the desired species. A traditional bending magnet is located at the end of the linear accelerator that serves as an energy analyzer to ensure that a pure, monoenergetic beam is formed for wafer implant.

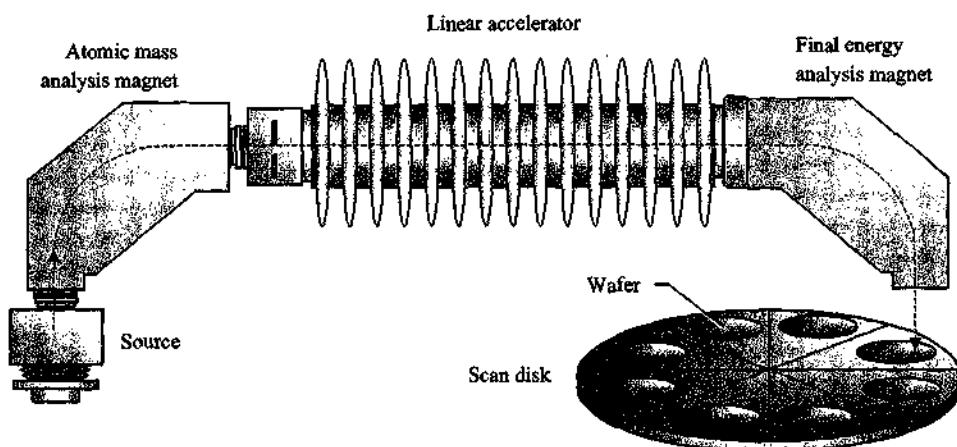


FIGURE 17.17 Linear Accelerator for High-Energy Implanters

High-Current and Low-Energy Beam ■ The need to form ultrashallow implants has promoted the development of high-current, low-energy beams. A low-energy beam accelerates ions to a low velocity so that their projected range is shallow in the target wafer. Maintaining a well-focused beam with low-energy acceleration is difficult because beam size is hard to control and usually increases with high current and low energy. Visualize this action by imagining a bowling ball rolled slowly down a bowling alley lane. It tends to veer to the side and into the gutter much more easily than a bowling ball rolled straight and fast. A method used to achieve a high beam current in a low-energy beam is *beam deceleration*, or *decel*.¹⁶ With this approach, the beam is extracted from the ion source at a high energy and then decelerated down to the desired energy without loss of beam current. An electrode positioned farther down the beamline is used to decelerate the ions from the initial extracted energy. This approach may cause energy contamination of the beam if high-energy ions manage to pass through to the wafer. A variation to decel is the *differential lens*, which combines a short beam path and optimized optics to minimize beam current loss.

Space Charge Neutralization ■ A beam of only positively charged ions is inherently unstable, since like charges repel one another with mutual repulsion. This condition causes *beam blow-up*, meaning the diameter of the beam increases as the ions pass down the beamline, resulting in nonuniform implantation. Beam blow-up is minimized by ensuring that the positive ions in the beam are neutralized with secondary electrons, a condition referred to as *space charge neutralization*.

Secondary electrons are generated when high-energy dopant ions strike a surface (e.g., aperture plates or beam guide assembly) as they pass along the beamline. The secondary electrons become trapped within the beamline due to the negative-biased electrodes along the beam path, including the extraction electrode and the acceleration. The secondary electrons bounce back and forth between the negatively biased plates, unable to leave their specific region in the beamline. The trapped electrons serve to neutralize the beam and prevent or reduce beam blow-up (see Figure 17.18).

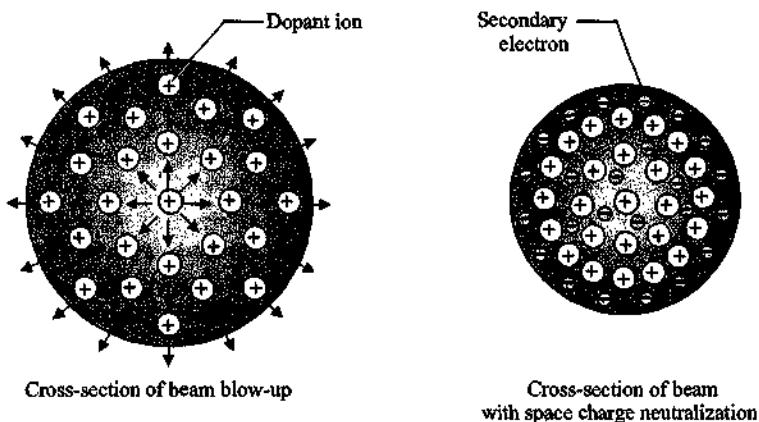


FIGURE 17.18 Space Charge Neutralization

Neutral Beam Trap ■ The positively charged ion beam is formed in a vacuum of less than 10^{-6} torr, but there are always residual gas molecules. Neutral ions are formed in the beam when a dopant ion gains an electron by colliding with a residual gas molecule. They cannot be deflected because of their lack of charge and, if not removed, they will be implanted. A *neutral beam trap* is created by placing a bend in the beamline a short distance before the target chamber using beam-bending electrodes. Since the neutral ions will not be deflected by the electrodes at the bend, they travel straight and collide with a grounded collector plate (see Figure 17.19).

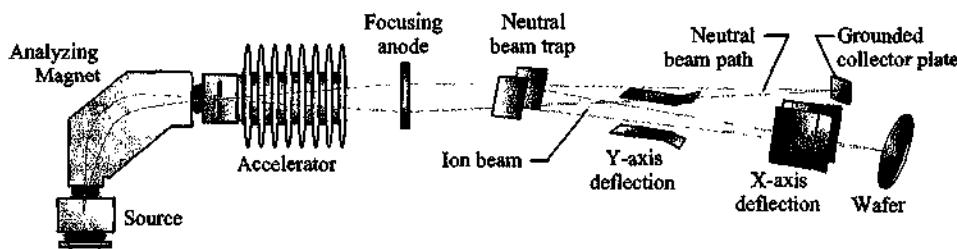


FIGURE 17.19 Neutral Beam Trap
(Used with permission from Varian Semiconductor Equipment)

Scanning System

The goal of the implanter is to form a highly pure ion beam from both a contamination and energy standpoint. This pure beam is implanted into the target wafer. That is, the beam has only the desired dopant ions at a predetermined energy level. Because the focused ion beam is typically small, on the order of 1 cm^2 for medium implanters and 3 cm^2 for high-current implanters, it must be scanned to cover the entire wafer. Wafer scanning is done either by moving the beam over a stationary wafer or moving the wafers through a stationary beam. Scanning plays a critical part in dose uniformity and repeatability.

Traditionally, low- to medium-current implanters keep the wafer stationary, while high-current implanters will keep the ion beam stationary. The different types of scanning systems used in implanters are:

- ◆ Electrostatic scanning
- ◆ Mechanical scanning
- ◆ Hybrid scanning
- ◆ Parallel scanning

Electrostatic Scanning ■ *Electrostatic scanning* deflects the ion beam across a stationary wafer by applying a specific controlling voltage to a set of X-Y electrodes (see Figure 17.20). When the polarity of one side of the electrodes is made negative, the positive ion beam will deflect toward it. By properly positioning the two sets of electrodes and constantly varying the voltage, the beam is made to deflect across a wafer. This type of scanning could be compared to spray painting a surface, where the spray needs to be worked back and forth many times to uniformly coat the surface. An electrostatic scan moves the beam across the wafer (x-axis) 15,000 times in one second while it scans up and down (y-axis) 1,200 times per second. Special care is made to ensure the scan is uniform at the outer edges of the wafer where the scan actually has to stop and reverse directions. In electrostatic scanning systems the wafer can be twisted and tilted relative to the ion beam to achieve desired junction characteristics.

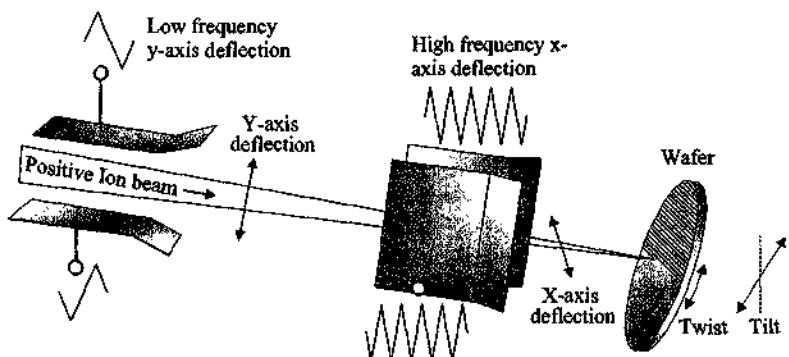


FIGURE 17.20 Electrostatic Ion Beam Scanning of Wafer

Implanting a single wafer at a time is referred to as a serial process and is generally used for low- to medium-current implant applications. In a typical low-current implant operation, a wafer might be scanned for 7 to 10 seconds to obtain uniform dose. Note that this same method of beam deflection can also be accomplished using an electromagnetic field approach instead of an electrostatic approach.

Because the wafer is held fixed during electrostatic scanning with no elaborate mechanical assembly, there is less possibility of particulate contamination. Another advantage to this type of scanning is that electrons and neutrals are not deflected and are eliminated from the beam. The major disadvantage is that the ion beam does not strike perpendicular to the wafer, which leads to undesirable shadowing from the mask material that partially blocks implantation of the ion beam (see Figure 17.21).

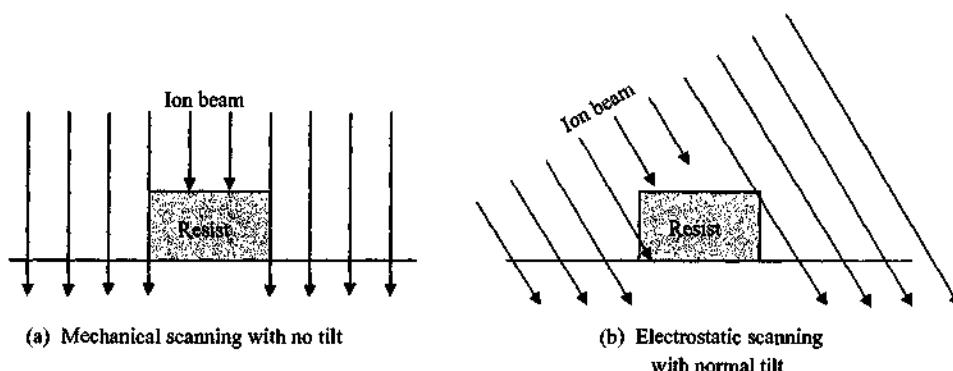


FIGURE 17.21 Implant Shadowing

Mechanical Scanning ■ With *mechanical scanning*, the ion beam is fixed and the wafers are mechanically moved through the beam. This method is generally used for high-current implanters because electrostatic beam deflection is difficult at high currents and energies. The beam size may be 1 cm wide by 3 cm high. Mechanical scanning is done by rotating multiple wafers (e.g., up to 25 200-mm wafers) fixed on the outer circumference of a large wheel assembly disk that is simultaneously moving up and down while rotating at 1000 to 1500 rpm (see Figure 17.22). The disk diameter may be 5 feet or larger. While the disk is spinning, it is also moved up and down so that the beam scans from the outer edge to the inner edge of the wafer. The disk can also be tilted relative to the ion beam to prevent channeling through the interstitial spaces of the silicon lattice (see the following section). Moving the wafers relative to the beam resolves any nonuniformity concerns from shadowing. Mechanical scanning is a batch process that effectively averages the beam power over a very large area. An advantage is the reduced wafer heating from the ion beam energy. Throughput is better than with electrostatic scanning. However, there are potentially more particles generated from the mechanical assembly.

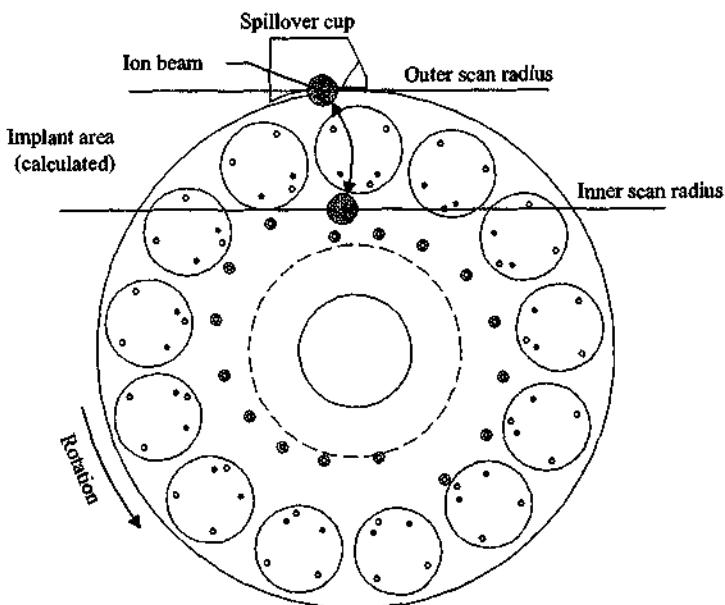


FIGURE 17.22 Mechanical Scanning of Implanted Wafers
(Used with permission from Varian Semiconductor Equipment, VIIision 80 Ion Implanter)

Hybrid Scanning ■ In the *hybrid scanning* system, the wafers are placed on a disk and rotated to scan in the y axis. The beam is then electrostatically (or electromagnetically) scanned in the x-direction. This method is generally used in a serial process for low- to medium-current with a single wafer.

Parallel Scanning ■ In electrostatic scanning, the beam is not scanned perpendicular to the wafer surface, which causes shadowing. *Parallel scanning* is a method used to reduce shadowing and channeling (see the following section) by scanning the ion beam to within $<0.5^\circ$ of the wafer surface.¹⁷ In the parallel scanning method, the beam is first scanned electrostatically. Then the beam passes through another set of magnets which corrects the beam angle and forces it to strike perpendicular to the wafer surface.

Wafer Cooling ■ The energy of the ion beam striking the wafer is converted into heat and will cause the wafer temperature to increase. Wafer cooling is used to control the temperature and prevent problems caused by heating, usually controlling the wafer temperature to $<50^\circ\text{C}$. Photoresist will blister and flake off if exposed to temperatures above 100°C and will become difficult to remove during photoresist stripping.¹⁸ This feature is especially critical during high-dose implantation. If the wafer temperature exceeds 300°C , the electrical properties of the device are affected and

partial annealing could occur that could alter the sheet resistivity of the wafer. Factors that affect how much the wafer temperature rises during implantation include ion beam energy, implant time, scan speed, and wafer size.

The two techniques in wide use for wafer cooling in ion implantation are gas cooling and elastomeric cooling.¹⁹ For *gas cooling* the wafer is sealed against a platen (a cooled plate, often with internal circulating water) and a gas such as helium is introduced behind the wafer. The gas provides a thermal conduction path to carry heat from the wafer to the platen. In *elastomeric cooling*, the metal platen is covered with a thin layer of elastomer material. This elastomer is in contact with the backside of the wafer and conforms to the wafer surface to maximize heat transfer between the wafer and platen.

Wafer Charging ■ During implantation, as the ion beam strikes the wafer, some positive ions accumulate in the masking layer, causing *wafer charging*. This condition can lead to a significant electric charge buildup on the wafer, especially for high-current implanters. Such wafer charge buildup can change the charge balance in the ion beam and lead to beam blow-up, causing substantial dose variations across the wafer. Wafer charging can also damage a surface oxide, including damage to the gate oxide leading to device reliability problems.

Control of wafer charging has traditionally been done by flooding the surface of the wafer with low energy electrons using an electron shower known as *secondary electron flood* (see Figure 17.23). In this method, primary electrons with the energy of a few hundred eV are directed to a target located close to the ion beam path. When these primary electrons strike the target, a cloud of low-energy (<20 eV) secondary electrons is generated.²⁰ These secondary electrons are trapped by the ion beam and are used to neutralize positive charge buildup on the wafer surface. It is important that no high-energy primary electrons reach the wafer or damage may occur to gate oxides.

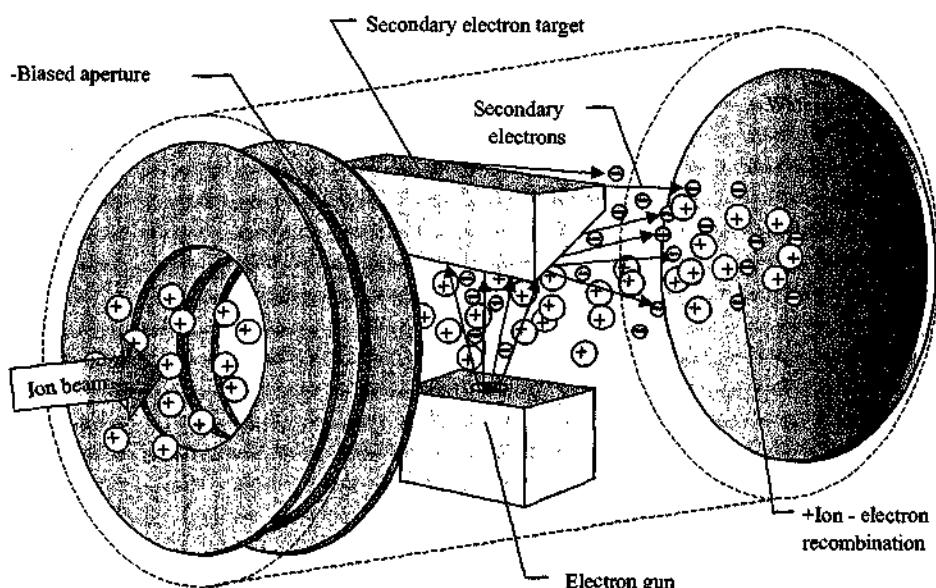


FIGURE 17.23 Electron Shower for Wafer Charging Control
(Used with permission and adapted from Eaton NV10 Ion Implanter, circa 1983)

Control of wafer surface charging is often done now by immersing the ion beam and wafer in a stable, high-density plasma environment, referred to as a *plasma electron flood* system.²¹ It is based on the extraction of electrons from a plasma (usually Ar or Xe) maintained in an arc chamber located close to the beam path and wafer (see Figure 17.24 on page 498). The plasma is filtered so that only secondary electrons reach the wafer surface and neutralize the positive charge buildup. The main advantage of plasma flood over the electron shower is that no high-energy electrons are generated in the plasma, which means only low-energy electron energy is used. Plasma flood effectively reduces wafer charge buildup and damage.

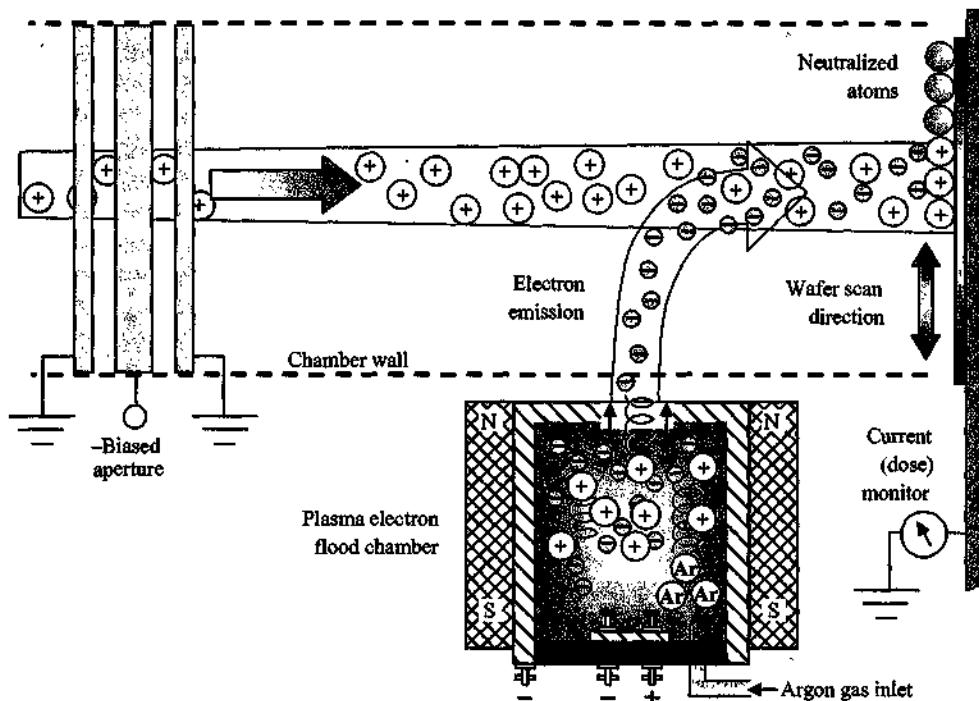


FIGURE 17.24 Plasma Flood to Control Wafer Charging

Process Chamber

The implantation of the ion beam into the wafer takes place in the *process chamber*. The process chamber is a major implanter subassembly that includes the scanning system (discussed above), *end station* with vacuum load locks for loading and unloading wafers, the wafer handler system, and the computer control system. There are also methods for dose monitoring and channeling control (see the following section). The end station can be large if used for mechanical scanning. It is pumped down to the vacuum required for implantation with multiple roughing, turbopumps, and cryopumps to reach base pressure (e.g., typically 10^{-6} torr).



End Station for
Ion Implanter
(Photo courtesy
of International
SEMATECH)

Load/unload systems in the end station use robotic handling to move wafers between an input station and the scanning disk in the target chamber (see Figure 17.25). Cassettes are loaded into the input rack and load locks are used to initially seal the input chamber. A roughing pump lowers the pressure around the cassettes. When pressure is low enough, a turbopump continues the pumpdown