



Optical Lithography

Here Is Why

Burn J. Lin

Optical Lithography

Here Is Why

Optical Lithography

Here Is Why

Burn J. Lin

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Lin, Burn Jeng, 1942-

Optical lithography : here is why / Burn J. Lin.

p. cm.

Includes index.

ISBN 978-0-8194-7560-2

1. Microlithography. 2. Semiconductors--Etching. 3. Lasers--Industrial applications. I. Title.

TK7872.M3.L56 2009

621.3815'31--dc22

2009049350

Published by

SPIE

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: +1 360.676.3290

Fax: +1 360.647.1445

Email: Books@spie.org

Web: <http://spie.org>

Copyright © 2010 Society of Photo-Optical Instrumentation Engineers (SPIE)

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

The content of this book reflects the work and thought of the author(s).

Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Printed in the United States of America.



Contents

Preface	xiii
---------------	------

Chapter 1 Introducing Optical Lithography / 1

1.1 The Role of Lithography in Integrated Circuit Fabrication	2
1.2 The Goal of Lithography	4
1.3 The Metrics of Lithography	4
1.4 The Contents of This Book.....	6

Chapter 2 Exposure Systems / 7

2.1 Proximity Printing	7
2.2 Projection Printing and a Comparison to Proximity Printing	9
2.3 Full-Wafer Field.....	11
2.4 Step and Repeat	14
2.5 Step and Scan.....	17
2.6 Reduction and 1X Systems	21
2.7 1X Mask Fabricated with a Reduction System.....	24
2.8 Summary	25
References	25

Chapter 3 Image Formation / 29

3.1 The Aerial Image	29
3.1.1 Effects of a spherical wavefront and deviations from it.....	30
3.1.2 Spherical wavefront	30
3.1.3 The effect of a finite numerical aperture on the spherical wavefront	31
3.1.4 Deviation from a spherical wavefront.....	36
3.1.4.1 The Seidel aberration coefficients	36
3.1.4.2 The Zernike aberration coefficients	38
3.1.5 Imaging from a mask pattern	40
3.1.5.1 Coherent imaging from a mask pattern	41
3.1.5.2 Incoherent imaging from a mask pattern.....	44
3.1.5.3 Partial coherent imaging from a mask pattern.....	45
3.1.6 Spatial frequencies	47
3.1.6.1 Spatial frequencies of an isolated line opening	48
3.1.6.2 Spatial frequencies of line-space pairs.....	49
3.1.6.3 Angular spectrum	56
3.1.7 Imaging results.....	56
3.2 Reflected and Refracted Images.....	60

3.2.1 Methods to evaluate the reflected and refracted image from a mask	62
3.2.2 Impact of multiple reflections on DOF	63
3.3 The Latent Image	64
3.4 The Resist Image	65
3.4.1 The A , B , C coefficients	68
3.4.2 The lumped parameters	71
3.4.3 β and η	83
3.5 From Aerial Image to Resist Image	85
3.6 The Transferred Image	87
3.6.1 Isotropic etching	87
3.6.2 Anisotropic etching	88
3.6.3 Lift off	89
3.6.4 Ion implantation	90
3.6.5 Electroplating	91
References	92

Chapter 4 The Metrics of Lithography / 95

4.1 The Resolution and DOF Scaling Equations	95
4.2 Determination of k_1 and k_3 Based on Microscopy	97
4.3 Determination of k_1 , k_2 , and k_3 Based on Lithography	99
4.3.1 E-D branches, trees, and regions	99
4.3.2 E-D window, DOF, and exposure latitude	101
4.3.3 Determination of k_1 , k_2 , and k_3 using E-D windows	103
4.4 k_1 , k_2 , and k_3 as Normalized Lateral and Longitudinal Units of Dimension	106
4.5 The E-D Tools	106
4.5.1 Construction of E-D trees	106
4.5.1.1 E-D tree construction from E-D matrix linewidth data	106
4.5.1.2 E-D tree construction from E-D matrix edge data	107
4.5.1.3 E-D tree construction from intensity distribution	109
4.5.2 Importance of log scale in the exposure axis	109
4.5.3 Elliptical E-D window	110
4.5.4 EL-versus-DOF tradeoff	111
4.5.5 Incorrect elliptical E-D window	112
4.5.6 CD-centered versus full-CD-range E-D windows	113
4.5.7 E-D window and CD control	115
4.5.8 Application of E-D tools	115
4.5.8.1 Combination of feature types	116
4.5.8.2 Combination of feature sizes	116
4.5.8.3 Combination of cuts for 2D features	119
4.5.8.4 Combination of CD tolerances	120
4.5.8.5 Combination of resist processing tolerances	122
4.5.8.6 Combination of image field positions	124
4.5.8.7 Setting the mask-making tolerance	127

4.5.8.8 Effects of phase-shifting mask errors	130
4.5.8.9 Comparison of experiment and theory	130
References	132

Chapter 5 Components in Optical Lithography / 133

5.1 Light Source	133
5.1.1 Mercury arc lamp	133
5.1.2 Excimer laser	136
5.1.2.1 Operation principle	136
5.1.2.2 Bandwidth narrowing	137
5.1.2.3 Spatial coherence	139
5.1.2.4 Maintenance, safety, and lifetime of excimer lasers	140
5.2 Illuminator	141
5.2.1 Köhler illumination system	142
5.2.2 Off-axis illumination	142
5.3 Masks	142
5.3.1 Mask substrate and absorber	144
5.3.2 Pellicles	145
5.3.3 Critical parameters for masks	146
5.3.3.1 CD control	146
5.3.3.2 Placement accuracy	146
5.3.3.3 Mask transmission and thermal expansion	147
5.3.3.4 Mask reflectivity	147
5.3.3.5 Mask flatness	149
5.3.3.6 Physical size	149
5.3.3.7 Defect level	149
5.3.4 Phase-shifting masks	150
5.3.4.1 Operating principle	150
5.3.4.2 Un-flat BIM is not a PSM	151
5.3.4.3 PSM types and mechanisms of imaging improvement	151
5.3.4.4 PSM configurations	156
5.4 Imaging Lens	159
5.4.1 Typical lens parameters	159
5.4.1.1 Numerical aperture	159
5.4.1.2 Field size	160
5.4.1.3 Reduction ratio	161
5.4.1.4 Working distance	161
5.4.1.5 Telecentricity	161
5.4.2 Lens configurations	161
5.4.2.1 Dioptric systems	162
5.4.2.2 Reflective systems	162
5.4.2.3 Catadioptric systems	163
5.4.3 Lens aberrations	165
5.4.4 Lens fabrication	165
5.5 Lens Maintenance	166

5.6 Photoresists.....	167
5.6.1 Classifications	168
5.6.1.1 Polarity.....	168
5.6.1.2 Working principle	170
5.6.1.3 Imaging configurations	175
5.6.2 Light interactions with a photoresist.....	181
5.6.2.1 Wavelength compression	181
5.6.2.2 Light absorption.....	181
5.6.2.3 Resist bleaching or dyeing	181
5.6.2.4 Resist outgassing	182
5.6.2.5 Multiple reflections.....	182
5.6.3 Profile of resist images.....	187
5.7 Antireflection Coatings	189
5.8 Wafer.....	202
5.9 Wafer Stage	202
5.10 Alignment System	204
5.10.1 Off-axis alignment and through-the-lens alignment	205
5.10.2 Field-by-field, global, and enhanced global alignment.....	206
5.10.3 Bright-field and dark-field alignments.....	207
References.....	208

Chapter 6 Processing and Optimization / 213

6.1 Optimization of the Exposure Tool	213
6.1.1 Optimization of NA	213
6.1.2 Optimization of illumination	217
6.1.3 Exposure and focus	220
6.1.4 DOF budget	222
6.1.4.1 Components in DOF _{required}	222
6.1.4.2 Focus monitoring	225
6.1.5 Exposure tool throughput management.....	231
6.2 Resist Processing.....	238
6.2.1 Resist coating	238
6.2.1.1 Defects	239
6.2.1.2 Resist adhesion.....	239
6.2.1.3 Resist thickness.....	239
6.2.1.4 Resist uniformity	240
6.2.1.5 Saving resist material	241
6.2.2 Resist baking	243
6.2.2.1 Precoating bake.....	244
6.2.2.2 Postapply bake (pre-exposure bake).....	244
6.2.2.3 Postexposure bake.....	245
6.2.2.4 Hard bake	247
6.2.3 Resist developing.....	248
6.2.4 Aspect ratio of resist image.....	251
6.2.5 Environmental contamination.....	252

6.3 k_1 Reduction	252
6.3.1 Phase-shifting masks	253
6.3.1.1 Alternating phase-shifting mask (AltPSM)	253
6.3.1.2 Attenuated phase-shifting mask (AttPSM).....	262
6.3.2 Off-axis illumination.....	268
6.3.2.1 Working principle of OAI.....	268
6.3.2.1.1 Conceptual illustration	268
6.3.2.1.2 3D illumination on 2D geometry	269
6.3.2.1.3 Analytic forms	271
6.3.2.2 Simulation results	275
6.3.2.3 OAI and AltPSM	284
6.3.2.3.1 Comparison of imaging performance	285
6.3.2.3.2 Other considerations	285
6.3.2.3.3 Combination of OAI and AltPSM	285
6.3.3 Scattering bars	288
6.3.3.1 Imaging improvement from scattering bars	288
6.3.3.2 Complications	289
6.3.3.2.1 Restricted pitch	289
6.3.3.2.2 2D features	291
6.3.3.2.3 Mask-making concerns.....	291
6.3.3.2.4 Full-size scattering bar	292
6.3.3.3 Hollow subresolution scattering bars and subresolution assist PSM.....	294
6.3.4 Optical proximity correction.....	294
6.3.4.1 The proximity effect	295
6.3.4.2 Edge corrections.....	296
6.3.4.2.1 Rule-based OPC	298
6.3.4.2.2 Model-based OPC	299
6.3.4.3 Local-dose OPC	304
6.3.4.4 Full-depth OPC.....	306
6.3.4.5 Correction to AEI	308
6.3.4.6 Hot-spot checking.....	309
6.4 CD Uniformity	309
6.4.1 CDNU analysis.....	310
6.4.1.1 Linear model for CDU contributions	310
6.4.1.2 Geometrical decomposition	311
6.4.1.3 Physical decomposition	314
6.4.1.4 CDU summation	314
6.4.2 CDU improvement	316
6.4.2.1 Active compensation with exposure tools.....	317
6.4.2.2 Active compensation on tracks.....	317
References	320

Chapter 7 Immersion Lithography / 323

7.1 Introduction.....	323
7.2 Resolution and DOF.....	326
7.2.1 Wavelength reduction and spatial frequencies	326
7.2.2 Resolution and DOF scaling equations.....	327
7.2.3 Improving resolution and DOF with an immersion system.....	328
7.3 DOF in Multilayered Media.....	329
7.3.1 Transmission and reflection in multilayered media	329
7.3.2 Effects of wafer defocus movements	332
7.3.3 Diffraction DOF	334
7.3.4 Required DOF	335
7.3.5 Available DOF	336
7.3.6 Preferred refractive index in the coupling medium.....	338
7.3.7 Tradeoff between resolution and DOF _{diffrac}	338
7.4 Polarization-Dependent Stray Light.....	340
7.4.1 Imaging at different polarizations	340
7.4.1.1 Recombination of spatial frequency vectors in the resist ..	341
7.4.1.2 Polarized refraction and reflection at the resist surface.....	346
7.4.1.3 Different effects of polarized illumination.....	348
7.4.2 Stray light	351
7.4.2.1 System stray light	351
7.4.2.2 Stray light from recombination of spatial frequency vectors inside the resist.....	352
7.4.2.3 Stray light from reflection off the resist surface	356
7.4.2.4 Incorporating polarization effects to E-D windows.....	356
7.4.2.5 Simulation results with PDS	358
7.5 Immersion Systems and Components	367
7.5.1 Configuration of an immersion system.....	367
7.5.2 The immersion medium	369
7.5.3 The immersion lens.....	372
7.5.4 Bubbles in the immersion medium	372
7.5.5 The mask	376
7.5.6 Subwavelength 3D masks	377
7.5.7 The photoresist	378
7.6 Impact on Technology	379
7.6.1 Simulation for an immersion system	379
7.6.2 Polylayer	381
7.6.3 Contact layer	384
7.6.4 Metal layer	385
7.6.5 Recommendation for the three technology nodes	387
7.7 Practicing Immersion Lithography	388
7.7.1 Printing results	388
7.7.2 Defect reduction.....	390
7.7.3 Monitoring the immersion hood and special routing.....	391

7.7.4 Other defect-reduction schemes	397
7.7.4.1 Wafer and equipment cleanliness	397
7.7.4.2 Wafer seal ring	398
7.7.5 Results	399
7.8 Extension of Immersion Lithography	401
7.8.1 High-index materials	401
7.8.2 Solid-immersion mask.....	402
7.8.3 Polarized illumination	403
7.8.4 Double exposures and pitch splitting	404
7.8.5 Pack-unpack	406
7.8.6 Overcoming the throughput penalty with double imaging	408
7.9 Conclusion on Immersion Lithography	408
References	410

Chapter 8 Outlook and Successors to Optical Lithography /415

8.1 Outlook of Optical Lithography	416
8.1.1 Optical lithography galaxy for logic gates	417
8.1.2 Optical lithography galaxy for contact holes	419
8.1.3 Optical lithography galaxy for equal lines and spaces	421
8.2 EUV Lithography	421
8.2.1 Introduction	421
8.2.2 EUV sources	424
8.2.2.1 Wall-power requirement of EUV systems.....	428
8.2.3 EUV masks	428
8.2.3.1 Configuration of EUV masks	429
8.2.3.2 Random phase shifting.....	429
8.2.3.3 Effects of oblique incidence	430
8.2.3.3.1 Overlay mismatch.....	430
8.2.3.3.2 Pattern shadowing.....	430
8.2.3.3.3 Overlay error from mask flatness	432
8.2.3.4 EUV mask fabrication.....	435
8.2.3.5 Absence of mask pellicle.....	436
8.2.4 EUV projection optics.....	436
8.2.5 Wall-power consumption	438
8.2.6 EUV resist	438
8.2.7 EUV OPC.....	440
8.2.8 Summary of EUVL	441
8.3 Massively Parallel E-beam Maskless Imaging	442
8.3.1 Introduction to e-beam imaging	442
8.3.2 MEB ML2 systems proposed	444
8.3.2.1 MAPPER MEB ML2 system	444
8.3.2.2 LV system with individual sources.....	447
8.3.2.3 HV MEB ML2 system	448
8.3.2.4 MBS MEB ML2 system.....	450
8.3.2.5 Reflected e-beam lithography.....	451

8.3.3 Comparison of the different systems	453
8.3.4 Data volume and the rate of DW systems.....	456
8.3.5 Power consumption of MEB ML2.....	456
8.3.6 Extendibility of MEB ML2 systems	457
8.3.7 Comparison of 4X mask writing to 1X wafer writing.....	457
8.3.8 Applications for MEB ML2.....	458
8.3.9 Summary of MEB ML2	459
8.4 Outlook of lithography	459
8.4.1 Double patterning.....	459
8.4.2 EUV lithography	460
8.4.3 MEB ML2	461
8.4.4 Nanoimprint lithography	461
8.5 Conclusions.....	462
References	465

Index / 469

Preface

Optical lithography is a fascinating field. It requires knowledge of geometrical and wave optics, optical and mechanical systems, diffraction imaging, Fourier optics, resist systems and processing, quantification of imaging performance, and the control of imaging performance. Even the history of its development helps to stimulate new ideas and weed out less promising ones. Practitioners of optical lithography may only have a vague idea of its theory, and likewise, theoreticians may not have the opportunity to practice the technology on the manufacturing floor. Although optical lithography seems to be easy to grasp on the surface, many executives or decision makers often muddle about in this important field without paying due respect to the experts. When it comes to lithography, everybody has an opinion. This book intends to cover the above fields in theory and practice to sufficiently bridge these gaps. Optical lithography need not remain “black magic.” It is a multidisciplinary science and treated as such in this book.

In the late 1980s I began teaching a course called “Here is Why in Optical Lithography” at the SPIE Microlithography Symposium. It was an overwhelming success. The course continued until 2005, when I could no longer afford to spend eight hours of my time at the symposium teaching the course. However, many people told me they missed it, especially those companies that routinely sent their new engineers to the course. Fortunately, Bruce Smith and his students, who audited in my course, eagerly asked me for the course notes, explaining that they could use it in a training course they planned to offer. So, there is a good chance the course material has been put to good use. Quite a few people also asked me whether I had written a book based on the course; however, even though I planned for such a book long ago, my work schedule prevented me from spending the time its development deserved. At one time SPIE even filmed an abbreviated version of the course, but these video tapes only reflect a part of the course, not a complete documentation. In the last two years I expanded the course into a one-semester three-credit course at the National Taiwan University. There, I also felt the need of a book for the course.

Yet, as soon as I started writing, I realized that the book could not be just course notes. It had to stand by itself. The content had to be broadened, so that much more depth could be provided. Since time restrictions (as in a short course) and space constraints (as in a journal article) were no longer limiting conditions, I could cover the entire field of optical lithography and provide as much information as a lithographer ought to know.

This book spans from the early days of the technology up to this very moment, when the lithographic technique for 32-nm half pitch is still a looming question worldwide. This text should be useful for both academics and industry practitioners alike.

Even though the subject has broadened and deepened, I am maintaining the original spirit of my short course, i.e., explaining the reasons and principles behind the techniques we use, so that even novices to the technology can gain insights from it. In addition, students in the field can use this text to delve to great depths. Veterans in the field might enjoy reviewing all phases of the technology that they spent their lives practicing and recalling how innovations have sprung up at the different phases of the technology's development.

I am indebted to my colleagues at TSMC who contributed figures and texts. Among them, Shinn Sheng Yu contributed Section 3.1.4.2, part of Section 3.4.2, Sections 5.7, 6.1.4.2, and 6.4. Ru-Gun Liu contributed Figs. 6.79, 6.87–6.90, 6.94–6.105, and 6.111; Ching-Yu Chang, Figs. 6.39, 6.40, and 7.56; Tsai-Sheng Gau, Figs. 7.45 and 7.69–7.70; Chun-Kuang Chen, Fig. 7.46; Shih-Che Wang, Fig. 7.58; Fu-Jye Liang, Figs. 7.59–7.66; Kuei-Shun Chen, Figs. 7.71 and 7.77; John Chin-Hsiang Lin, Fig. 7.74; and Da-Chung Owe-Yang, Fig. 7.78. Some other SEM pictures are the result of engineering studies in the Nanopatterning Division at TSMC. Figure credit from authors outside of TSMC is acknowledged where the figures appear.

Finally who writes a book without thanking their family for patience and understanding? Here, my wife Sue stands out as a great partner in my life to share my family responsibilities and even take them over in times of need. She is a brilliant professional who managed to have successful careers at IBM and AT&T, and yet has not shortchanged her support to my career and the upbringing of our children, Christiana and John.

*Burn Lin
Hsin-Chu, Taiwan
December 2009*

Chapter 1

Introducing Optical Lithography

Lithography creates a resist image on a wafer. The subsequent etching, lift off, or ion implantation process is masked by the resist image at the areas dictated by the lithography mask. Hence, the thin-film material on the wafer is selectively removed, built up, or its characteristics are selectively altered. Replicating the mask pattern produces the resist image, except when mask making or direct writing on a wafer. Figure 1.1 depicts the mask replication process with an imaging lens. The condenser collects light from the source and illuminates the mask pattern. Light then passes through the imaging lens to form an aerial image to selectively expose the resist. After development, the resist image is produced. Figure 1.2 illustrates various forms of image transfer from the resist to the underlying thin film. The film can be isotropically or anisotropically etched, lifted off, plated, or implanted, using the patterned resist as the mask. Detailed descriptions of these transfer processes are given in Chapter 3.

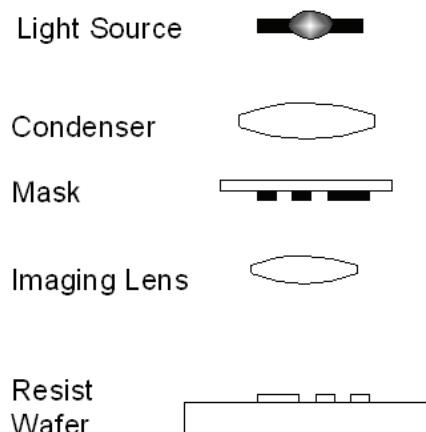


Figure 1.1 Optical lithography replicates the mask pattern through an imaging lens.

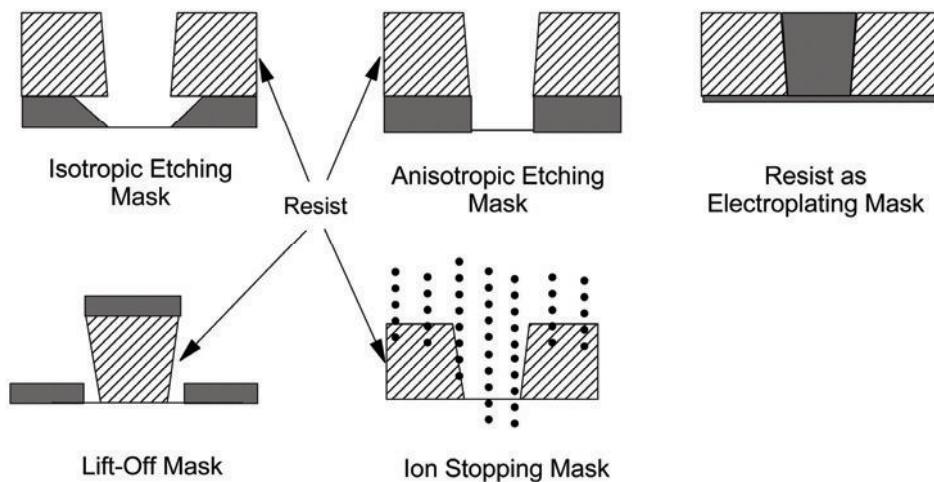


Figure 1.2 Exposed resist image transferred to the underlying thin film by isotropic etching, anisotropic etching, electroplating, lift off, and ion implant.

The image-formation process is facilitated with an information-carrying beam consisting of photons, electrons, or ions. Optical lithography uses photons to carry out this process. Photons with energy ranging from visible to x-ray wavelengths can be used. However, in this book, the wavelengths range between 157 and 436 nm. A majority of these wavelengths are already fully employed in manufacturing semiconductor integrated circuits. The others have been investigated. An outlook is given in Chapter 8 on the prospect of using even shorter wavelengths to produce future circuits.

1.1 The Role of Lithography in Integrated Circuit Fabrication

Lithography is an important part of semiconductor manufacturing technology because it is needed for every masking level. In a typical 0.13- μm CMOS integrated circuit fabrication run with four metal layers, there are more than 30 masking levels using 474 processing steps, in which 212 steps are related to lithographic exposure, and 105 steps are related to pattern transfer using a resist image. Relevant steps are listed as follows.

Lithography exposure-related:

- Wafer cleaning and priming to improve adhesion
- Application of antireflection coating before and/or after resist coating
- Resist coating
- Postapplication bake
- Exposure/alignment
- Postexposure bake
- Resist development
- Resist hardening

- Critical dimension (CD) monitoring
- Alignment monitoring
- Selective removal of antireflection coating
- Stripping of the antireflection coating and the resist

Pattern transfer-related:

- Etching
- Ion implantation
- CD monitoring
- Surface preparation
- Electroplating

Note that the actual sequence and number of processing steps may vary depending on the particular masking level. For example, the bottom antireflection coating at the wafer-resist interface is deposited on the wafer prior to priming and resist coating, if it is inorganic. Otherwise, it is applied between the priming and resist-coating steps. In the latter case, an extra baking step is required for the bottom antireflection coating. A top antireflection coating may be applied after the resist coating and baking, with or without the bottom antireflection coating. Another baking step follows the application of the top antireflection coating. A block diagram showing a possible sequence and relation of these processing steps is shown in Fig. 1.3.

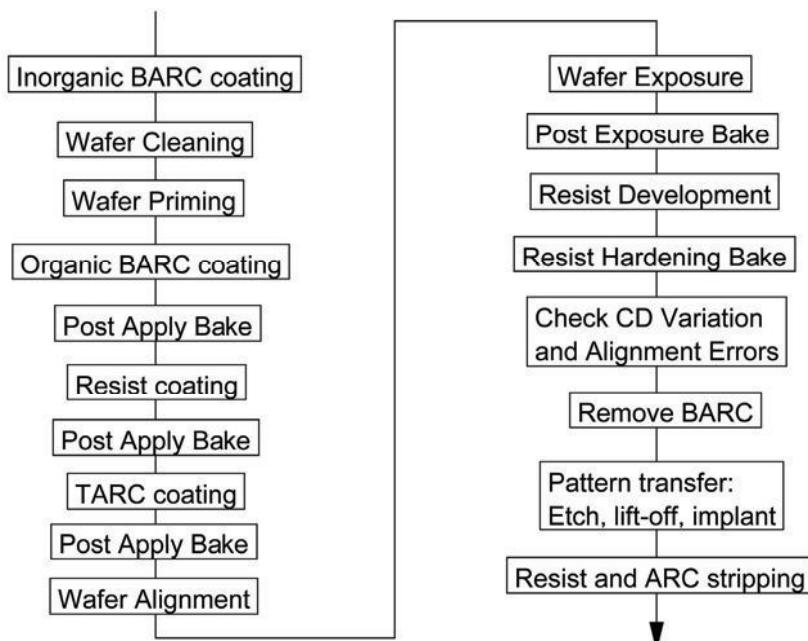


Figure 1.3 Block diagram of lithography processing steps.

Lithography is important not only because it is needed for all masking levels. It is often the limiting factor of entering the next technology node. For every node, the minimum feature sizes and their separations are reduced by a factor of $\sqrt{2}$. Therefore, the succeeding generation to 1- μm lithography is 0.7 μm , followed by 0.5 μm , 0.35 μm , 0.25 μm , 0.18 μm , 0.13 μm , 90 nm, 65 nm, 45 nm, 32 nm, 22 nm, etc., necessitating improvements in resolution and overlay accuracy that call for many of the following improvements: numerical aperture (NA) increase, wavelength reduction, suppression of reflections, better resists, better masks, higher stepping accuracy, higher-precision alignment, less lens distortion, better wafer flatness, and many others. These will be discussed in detail in the subsequent chapters.

1.2 The Goal of Lithography

In order to make the circuit usable, the features fabricated on it must meet certain criteria. The most important one is edge position control. The location of any given edge of a feature on the chip must be within a given tolerance from a nominal position. When the edge meets this requirement, both the linewidth and the overlay controls have been maintained. This is illustrated in Figure 1.4, where each of the six edges of an L-shaped feature must fit within the six windows surrounding the edges. The nominal position, represented by the center line, is where the edge ideally should be. Feature size or feature placement contributes to deviation from the ideal edge positions. Feature-size control and feature placement are often dealt with separately for convenience, even though ultimately the combination of them must meet the edge-placement requirement.

1.3 The Metrics of Lithography

Following the goal above, a measure of success lies in the size of the process window for feature size and placement controls that keep a given feature edge within its tolerance. The feature-size control parameters are depth of focus (DOF) and exposure latitude (EL), whereas the feature placement parameters are

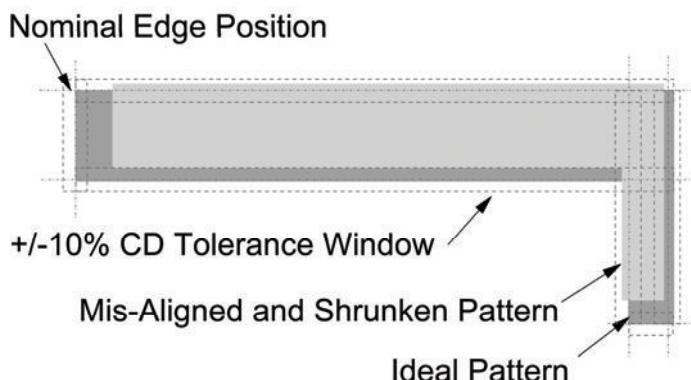


Figure 1.4 Edge placement tolerance.

alignment accuracy, magnification, and rotations. All are mutually dependent and can be set during wafer exposure. Among these five parameters, the exposure-defocus (E-D) window is most frequently used. It is shown in Fig. 1.5, with its mutual tradeoff depicted in Fig. 1.6. Adding alignment accuracy, magnification, and rotations generalizes the metric to include overlay-related parameters. The metrics will be fully described and extensively used in this book.

NA0.6,Wn365,CD.35um,Tol-35/35nm,Elat10.2%,DOF1.47, LW004.lwf

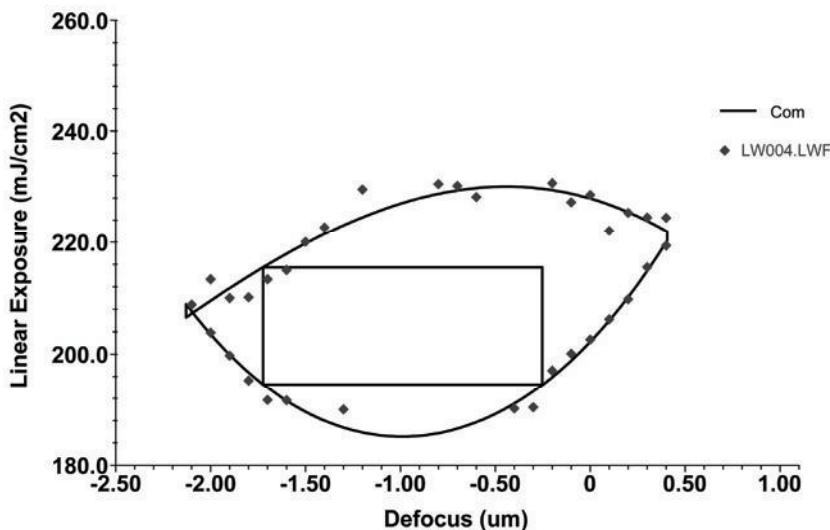


Figure 1.5 A typical E-D window in an E-D tree.

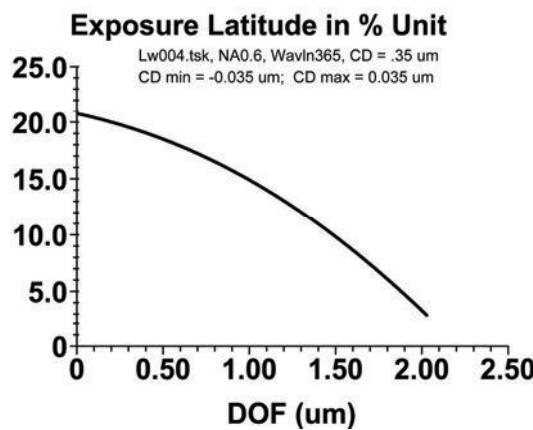


Figure 1.6 Exposure latitude versus DOF tradeoff curve.

1.4 The Contents of This Book

In the following chapters, exposure systems, image formation, metrics of lithography, components of optical lithography, processing, optimization, and the industry outlook will be covered.

Chapter 2 covers exposure systems including proximity and projection systems, step-and-repeat, step-and-scan, 1X, and reduction systems.

Chapter 3 deals with image formation, starting with the formation of aerial images diffracting from the mask, imaging through the lens with its aberrations, followed by creation of the latent image, resist image, and the transferred image.

The metrics of lithography discussed in Chapter 4 contain the basic principle of E-D trees, forests, and windows. Their applications include defining the common and individual exposure latitude and DOF of different feature types and sizes, at different cuts in a feature, a mixture of CD tolerances, resist processing tolerances, mask errors, proximity, and projection printings. E-D tools help to check experiment against theory, characterize lens aberrations, quantify the performance of phase-shifting masks, and guide proximity corrections.

Chapter 5 describes components of optical lithography consisting of light sources, illuminators, masks, the imaging lens, photoresists, wafers, wafer stages, and alignment systems.

Chapter 6 covers processing and optimization. Processing includes exposure, overlay, resist processing, k_1 reduction, optical proximity correction, and improving CD uniformity. Optimization deals with optimizing the NA and illumination coherence, wavelength, bias, off-axis illumination angle, and mask tolerance.

Chapter 7 deals with theory and practice of immersion lithography.

Chapter 8 presents the outlook for optical lithography and its successors, such as EUV lithography using the 13.5-nm wavelength and multiple e-beam maskless lithography systems.

Chapter 2

Exposure Systems

Before getting into imaging theory and practice, I would like to present exposure systems so that readers are better able to comprehend the abstract concepts in Chapters 3 and 4.

An exposure system replicates the mask image to an aerial image in order to expose the photoresist layer on a wafer. After development, this resist image is used as an etch, implant, plating, or lift-off mask for pattern transfer to a thin-film layer on the wafer. There are different systems that can perform the replication function. When the feature size was in the 2- to 5- μm regime, and the budget for semiconductor manufacturing was low, the aerial image in the 10- to 20- μm proximity of the mask sufficiently produced a useable resist image, maintaining feature size and placement control. Below 2 μm , infidelity of the aerial image, defect generation, and alignment viewing difficulties drove the bulk of exposure systems to 1X full-wafer projection printing. The field size of full-wafer systems grew with wafers from 50 to 125 mm in diameter. The ever-increasing wafer size and reducing feature size—together with the associated requirement in feature-size control and overlay accuracy—drove exposure systems to reduction step-and-repeat. Further increases in field size and a reduction in feature size, as well as difficulties in containing the 1X mask making within the total feature size control and overlay accuracy budgets, made it necessary to use reduction step-and-scan systems.

This chapter describes proximity and projection-printing systems, full-wafer, step-and-repeat, and step-and-scan systems. The significance of reduction imaging is also analyzed.

2.1 Proximity Printing

The proximity printing system is the simplest form of lithography printing systems. It consists of a light source, condenser, mask, wafer, and a wafer stage that holds the wafer at a fixed distance from the mask. It also provides the means to move the wafer with respect to the mask for alignment purposes; there are alignment-viewing devices to guide lateral positioning on the wafer. A schematic of a proximity printing system is shown in Fig. 2.1. The mask and wafer are held

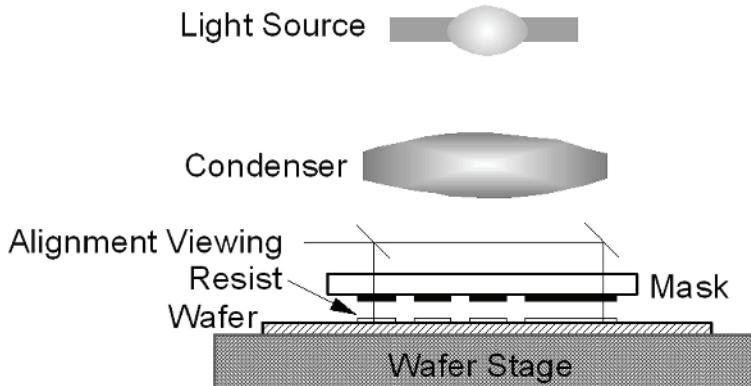


Figure 2.1 A schematic of a proximity printing system.

at a distance of less than 20 μm from each other. A larger distance prevents damage to the mask and the wafer; however, resolution is reduced rapidly as the mask-to-wafer gap increases. In addition, alignment viewing must accommodate two focal positions; thus, it becomes more complicated. Detailed coverage of proximity printing has been given elsewhere.¹ Here, only the bare essentials related to submicrometer lithography will be covered.

The mutual dependence of minimum feature size W , wavelength λ , and mask-to-wafer gap G is estimated by the Fresnel number v :

$$v \equiv \frac{W^2}{\lambda G}. \quad (2.1)$$

This is the parameter that governs near-field diffraction using the Fresnel approximation.² A large v indicates a close optical proximity to the mask. When the Fresnel number is above a given threshold, the square of the feature size is sufficiently large with respect to λG , the image is sufficiently close to the mask so that it acceptably resembles the mask, and the system is considered capable of the resolution derived from it. For example, at $v = 1$, with a wavelength of 250 nm and a gap of 10 μm , the system can support printing 1.6- μm features. In reality, determining the threshold requires the exposure-gap (E-G) methodology similar to that of the E-D methodology to be discussed in Chapter 4. Behaving like the resolution and DOF coefficients k_1 and k_2 in projection printing, the Fresnel number threshold is a function of feature shape, size, proximity environment, their combination, illumination coherence, and many other factors.^{3,4,5} The only missing parameters are those related to the imaging lens, such as numerical aperture and aberration coefficients.

Because of the absence of an imaging lens, it is relatively easy to change the wavelength of the imaging system. Therefore, proximity printing is often the forerunner of a wavelength reduction scheme used to screen potential resist candidates.^{6,7,8,9} It also enables printing at a spectral region where lens material is

absolutely impossible. For example, in the vicinity of a 1-nm wavelength, if a Fresnel number threshold of 1 is adequate, then 0.1- μm features can be delineated with a 10- μm gap. This is the basis of x-ray proximity printing, which claims a very high resolution potential.^{10,11}

The smallest feature produced in manufacturing with proximity printing is 2 μm in ultraviolet (UV) and mid UV, and 1.5 μm in deep UV. Their spectral bands are 450–350 nm, 350–260 nm, and 260–200 nm, respectively. The lower-bound wavelength of 350 nm for UV is chosen so that the popular mercury i-line wavelength of 365 nm is included. Similarly, the upper-bound wavelength of 260 nm for deep UV includes the 254-nm mercury line.

Proximity printing is as old as the IC technology¹² that started in 1958. Up until the 1970s, UV proximity printing was widely used in production. Commercial deep-UV proximity printers were available in the late 1970s until the early 1980s. Commercial x-ray proximity printers are still being used in laboratories today.

2.2 Projection Printing and a Comparison to Proximity Printing

Two problems with proximity printing lead to the use of an imaging lens for lithography pattern transfer. The first problem is related to defects. As the design rule is reduced to 2 μm and below while the wavelength is kept constant in the UV region, the maximum printable mask-to-wafer gap is squeezed to below 20 μm , making it extremely likely that the wafer will accidentally scratch the mask and vice versa. Mask damage issues and edge acuity requirements drove chrome masks to replace emulsion masks. Wafer damage issues drove manufacturers to clean their masks after every 100 exposures. These preventative measures still could not stamp the tide of projection printing, which can print at a much larger working distance. The second problem in proximity printing is related to the absence of reduction capability, which became apparent only after the feature size dropped below 1 μm . Mask making at 1X became quite demanding. With these persistent problems, lithography must rely solely on projection, regardless of wavelength. Even if the wavelength is reduced to 1 nm, as in the case of x-ray proximity printing, the 1X mask requirement is simply too stringent. More discussions on 1X and reduction systems can be found in Sec. 2.6.

A schematic projection-printing system is shown in Fig. 2.2. It looks very similar to the proximity printing system except for the addition of the imaging lens. The alignment system is not shown because there are many different ways to view alignment. The alignment marks may be viewed through the imaging lens or in an entirely separate system. The mask and wafer marks may be viewed together or separately. In a projection system, the mask-to-wafer distance is typically 80 cm or larger. The working distance between the front lens element and the wafer surface is in millimeters. Even more forgiving, the working distance between the mask and the first physical element is in centimeters. This facilitates high-speed movements between the mask and the wafer without any

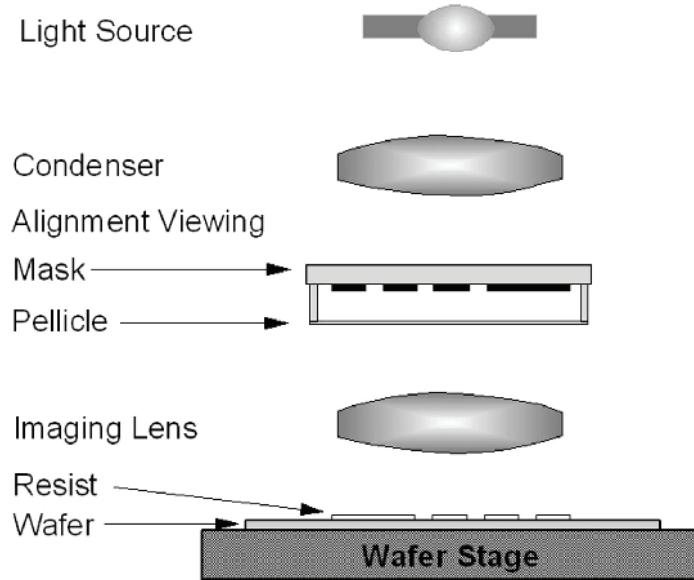


Figure 2.2 A schematic projection-printing system.

danger of damaging either of them. Projection printing also allows placement of mask-protecting pellicles at both sides of the mask. The pellicle is a thin transparent membrane of negligible optical thickness, stretched over a frame attached to the mask. In Fig. 2.2, only the pellicle at the absorber side of the mask is shown. If any foreign particle falls on the pellicle, it is safely kept out of focus from the imaging plane. The system is therefore immune to particulate-induced defects on the mask or due to the wafer touching the mask.

There is a practical difference between proximity printing and projection printing. The proximity image is always a mirror image of the mask, whereas the projected image may or may not be a mirror image, depending on the number of mirrors and whether mirrors are used in the projection lens system. Figure 2.1 shows the mirror-image relationship between the mask image and the resist image. Figure 2.2 shows a projection system without any mirror and thus an unaltered image in the resist. During introduction of full-wafer 1X projection printing using mirrors, an extra mirroring step was introduced to create an image compatible with masks used in proximity printing so that they can be used in both systems. This full-wafer 1X system is discussed in Sec. 2.3 below.

The resolution and DOF of a projection-printing system subscribes to the following relationships:

$$W \propto \frac{\lambda}{NA} \quad (2.2)$$

and

$$DOF \propto \frac{\lambda}{NA^2}, \quad (2.3)$$

where W is the minimum printable feature size, λ is the wavelength, and NA is the numerical aperture of the lens. The proportional constants for Eq. (2.2) and Eq. (2.3) are k_1 and k_2 , respectively. They will be fully discussed in Chapter 4. There, Eq. (2.3) is shown to be only true for dry and paraxial systems, whose NA is relatively small. Here, Eqs. (2.2) and (2.3) are simply used to point out the difference between projection and proximity printing. Note that, similar to proximity printing, resolution increases when wavelength is reduced. However, the improvement is much more efficient. Improvement in resolution is linearly related to wavelength, whereas the improvement is only related to the square root of wavelength in proximity printing. In projection printing, there is a penalty in DOF when the wavelength is reduced, whereas in proximity printing, the imaging impact of a shorter wavelength is always positive. In projection printing, the DOF is separated from the working distance. The former is usually on the order of a few micrometers; the latter is in millimeters. In proximity printing, the DOF and the working distance are identical. It is usually larger than that of projection printing in terms of DOF, but insufficient in terms of working distance.

One important aspect of projection printing is the exposure field size that dictates the size of chips that can be made and how defects are reduced. If many identical chips can be printed in the same field, then a defect on one of the chips has a small impact on yield, even though the same field is repeatedly stepped to cover the entire wafer. On the other hand, if the field is so small that it can only accommodate a single chip, the defect is repeated on every chip on the wafer, and the yield drops to zero. Multiple identical chips make possible the use of chip-to-chip inspection, rather than needing to resort to die-to-database inspection. Chip-to-chip inspection is usually known as die-to-die inspection. In proximity printing, the field size is limited only by the size of the mask substrate and the illumination system. There is no fundamental limit otherwise. Exposing a 100-mm diameter wafer in one shot is not difficult. In projection printing, the field size is dictated by the imaging lens. The field diameter is only in tens of a millimeter, leading to the following discussion of methods for covering the wafer.

2.3 Full-Wafer Field

A full-wafer field exposure system covers the entire wafer in one exposure step. The mask covers a field as large as the wafer. In terms of accommodating as many chips as possible in one exposure, full-wafer field exposure is desirable. It also helps to improve throughput because there is no need to wait for the wafer stage to move through all of the exposure positions, as opposed to when a small field must be stepped through the entire wafer. In the early days, the full-wafer requirement kept projection printing out of manufacturing until a clever scheme enabled it to cover the full-wafer field, as proximity printing is inherently capable. This is the popular all-reflective 1X projection-printing system^{13,14} shown in Fig. 2.3.

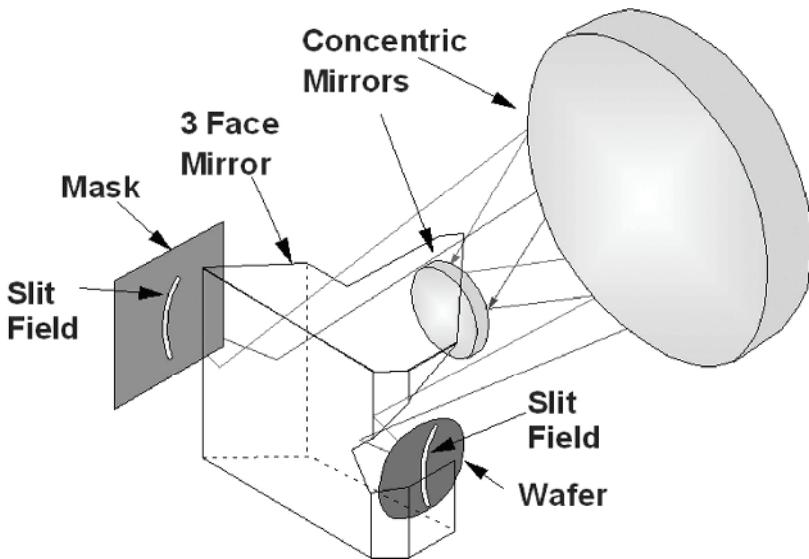


Figure 2.3 A 1X all-reflective, full-wafer field-projection-printing system. (Courtesy of A. Offner.)

The inventors took advantage of the low-aberration ring-field zone of two concentric mirrors and scanned a segment of the ring field to cover the entire wafer. A key component of this system is the three-face mirror assembly. Without it, the mask object ring section and the wafer image ring section are coplanar and must be scanned in opposite directions to cover the entire mask, as shown in Fig. 2.4.

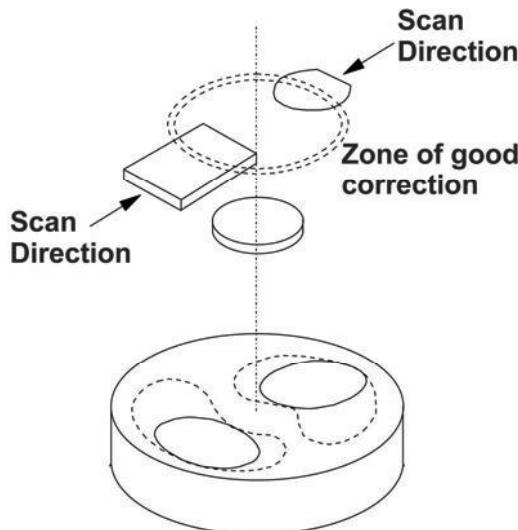


Figure 2.4 Ring zone of good correction, and mask and wafer scanning directions. (Courtesy of D. Markle.)

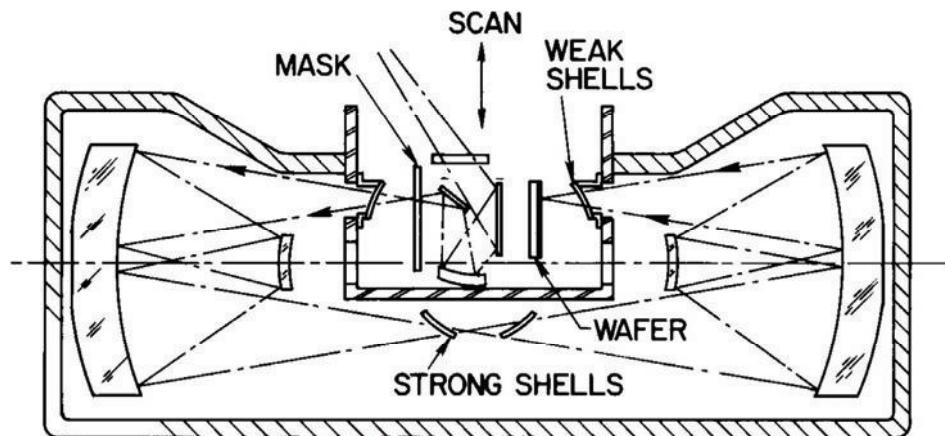


Figure 2.5 The 1X all-reflective, full-wafer field-projection-printing system with the mask-wafer carriage and flexure bearing. (Courtesy of D. Markle.)

Scanning with precision much less than 1 μm between the mask and wafer should obviously be avoided due to difficulty and cost. One mirror at the mask side and another one at the wafer side change the direction of the light beams so that the mask and wafer can now be scanned together in one direction on a single carriage. Actually, there are two mirrors at the wafer side. The additional mirror is inserted to produce a mirror image at the wafer to match that from the proximity printers. The mask and the wafer are mounted on the mask-wafer carriage to pivot against the flexure bearing during scanning, as shown in Fig. 2.5.

Because the law of reflection is independent of wavelength, this all-reflective system is broadband and can be used for new, shorter wavelengths just as conveniently as a proximity printing system. It has all of the advantages of a proximity system without its drawbacks. In fact, this system inherently replicates a correct image of the mask, whereas the proximity printing system produces a mirror image.

Starting in 1974,¹⁴ the 1X projection printer quickly replaced proximity printing to become the industry's workhorse. As wafer size continued to increase, the original 1X all-reflective projection-printing system was expanded in 1984 to include two sets of concentric mirrors¹⁵ and some refractive elements, making it a catadioptric system. A schematic drawing of the system is shown in Fig. 2.6.

This system is capable of covering a 125-mm wafer in full field. Due to scanning, the field is not limited by the lens in the scan direction. For 150-nm wafers, a $125 \times 150\text{-mm}^2$ field is used. The loss of exposed wafer area reduces profit because many more chips would be produced otherwise. So, further growth of the wafer size requires growing the mask and the lens field. Neither is easy. As feature sizes approach 1 μm , making a 1X mask to specification becomes difficult. In addition, the concentric mirror scheme is usable up to about 0.18 NA. Above that, it is difficult to keep out aberrations to achieve the resolution

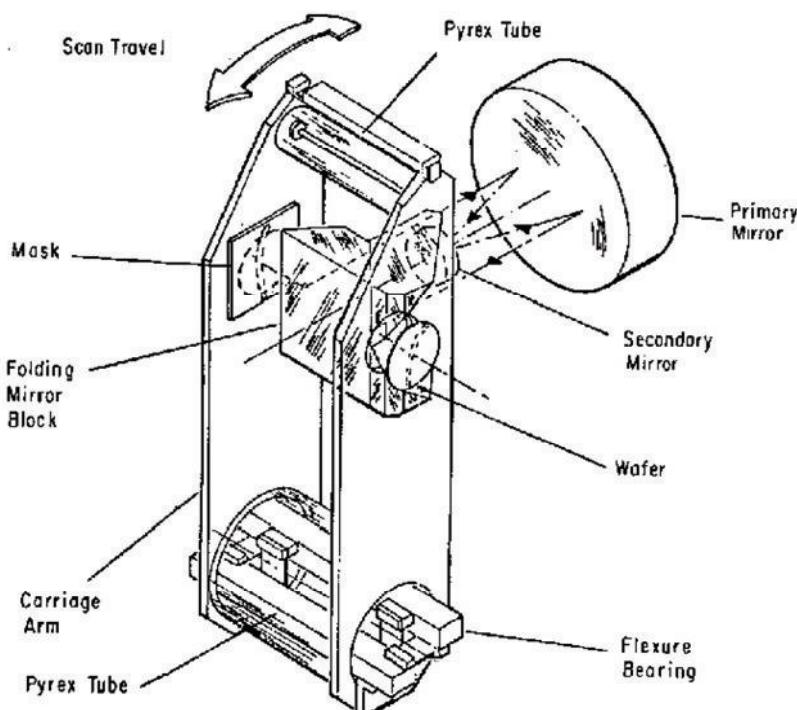


Figure 2.6 Expanded 1X full-wafer projection system using two sets of concentric mirrors and some refractive elements. (Courtesy of J. D. Buckley.)

expected of a higher NA. It became inevitable that the full-wafer field concept could be sustained. Step-and-repeat systems succeeded the full-wafer field systems. The smallest feature piloted for manufacturing is 1 μm , using deep-UV light.ⁱ

An all-reflective system is regaining popularity. With four to six extremely high-precision aspherical mirror surfaces in the imaging lens, designs with an NA in the 0.3 regime for 13.5-nm light have been shown. More of this topic is covered later in Chapter 8.

2.4 Step and Repeat

When the imaging lens dictating field size is smaller than the wafer size, the only way to cover the entire wafer is to step identical exposures through the wafer. Even though, in principle, one can change a mask between exposures, this is rarely done due to throughput and overlay considerations. A step-and-repeat system steps a mask repeatedly to cover the wafer. Figure 2.7 depicts the circular lens field and an exposure field with the unused part of the lens field blocked out by mechanical blades. As many chips as possible are fitted into the exposure field for defect, throughput, and mask inspection considerations.

ⁱ M. Haggi led a team at IBM to pilot such a product in 1982.

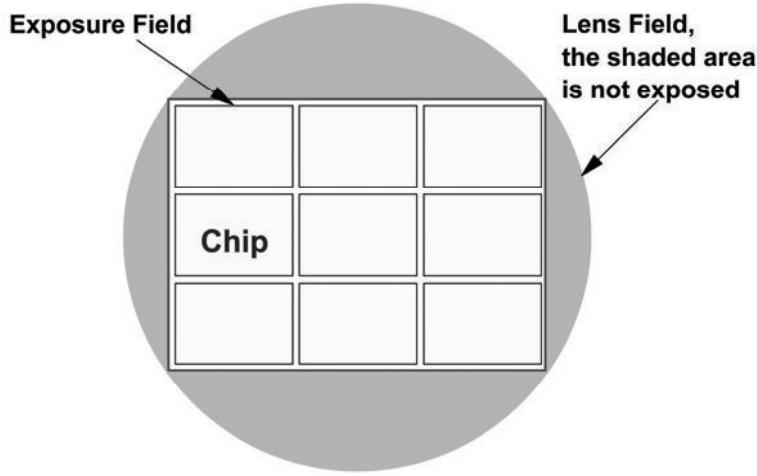


Figure 2.7 Chips, exposure field, and lens field.

The exposure field is stepped repeatedly to cover the entire wafer, as shown in Fig. 2.8. Note that the out-of-bound fields are exposed. The wafer area is very valuable; even though the chips partially or completely out of bounds are lost, the chips exposed in full can still be used as product chips.

A typical step-and-repeat projection-printing system using a refractive lens is shown schematically in Fig. 2.9. This is the most popular configuration for production of integrated circuits with features 1.5 μm and below. Not required by projection printing, but essential to improve mask accuracy, the majority of step-and-repeat projection-printing systems^{16, 17, 18} use a reduction ratio of five. Initially, there were 10X reduction systems.^{19, 20} Recently, 4X systems started appearing.²¹ The tradeoff in reduction ratio is discussed in Sec. 2.6 below.

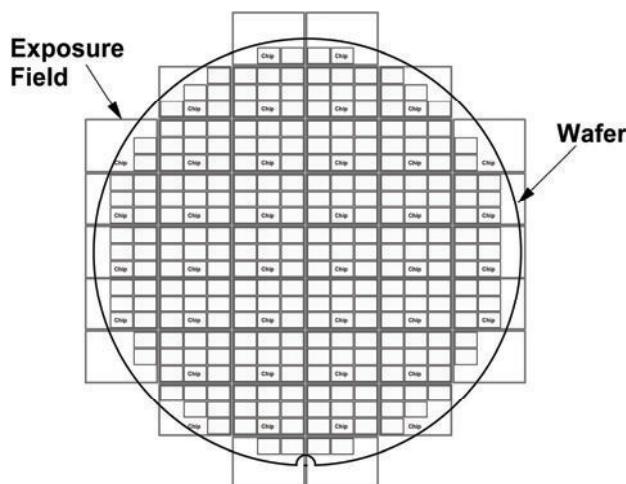


Figure 2.8 Exposure field stepped repeatedly to cover the entire wafer. The spilt-over chips at the wafer edge are not plotted.

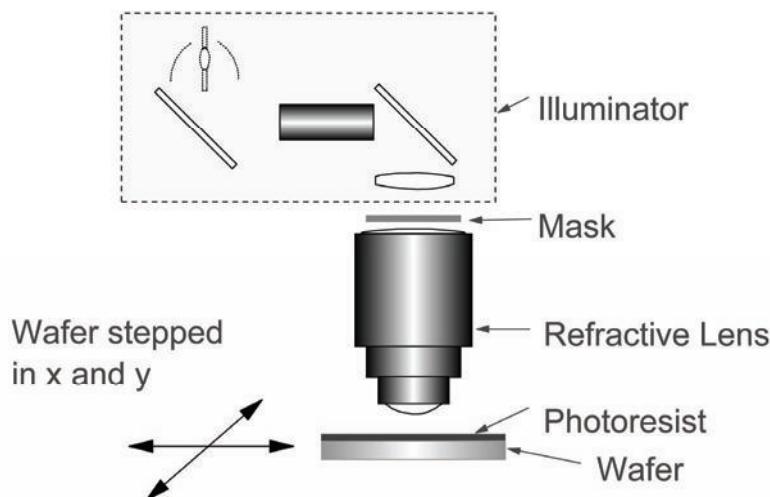


Figure 2.9 A step-and-repeat projection-printing system using a dioptric reduction imaging lens.

The step-and-repeat projection-printing system is not limited to the dioptric (refractive) reduction systems. Figure 2.10 depicts a 1X catadioptric lens²² that is used in a step-and-repeat system, which has proven to be cost effective and remains widely used for technology nodes in the longer submicrometers, as well as for noncritical levels in the shorter submicrometer technology nodes. Short submicrometer is the regime below 0.3 μm ; mid-micrometer is 0.3 to 0.7 μm ; and long-micrometer is 0.7 to 1 μm . Short submicrometer is sometimes referred to as “deep submicrometer.”

Initially, masks used in a step-and-repeat system are called reticles. Presently, they are also referred to as masks. The terms “mask” and “reticle” are almost interchangeable.

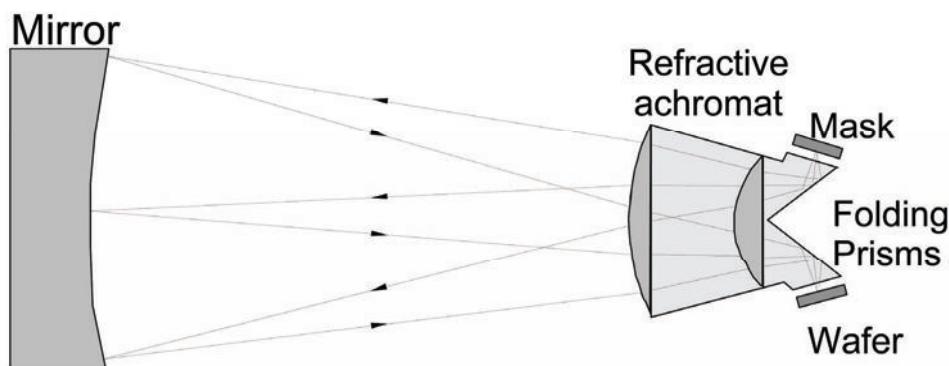


Figure 2.10 1X catadioptric projection lens used for step-and-repeat. (Courtesy of R. Hershel.)

2.5 Step and Scan

The increase in chip size drives the increase in field size, without which wafer throughput, yield, and productivity cannot be maintained. On the other hand, high packing density demands higher resolution that most often calls for increasing the lens NA. High packing density also demands higher overlay accuracy that tightens the lens distortion specifications in addition to the mask fabrication specifications. These competing requirements compound the complexity and cost of the imaging lens. The limit of these three lens parameters has become difficult to exceed. The step-and-scan system was developed to meet these requirements.

The step-and-scan principle²³ is shown in Fig. 2.11. An illuminated slot is scanned over the mask to expose the wafer. This requires that the lens remains stationary while the mask and the wafer move in a controlled fashion. With a 1X system described in Sec. 2.3 above, the mask and the wafer move at the same speed and in the same direction, making it mechanically simpler to scan together in sync. However, 1X systems require unattainable specifications for the mask in short submicrometer applications; a reduction system must be used, requiring the mask to move at a speed higher than that of the wafer by the reduction ratio. For example, with a 4X system, the mask must move exactly four times faster than the wafer. At the end of the travel, the wafer is moved to a new location for the next scanning. This procedure repeats until the entire wafer is exposed. Note that the scanning direction must be reversed from one step to the other, so that the scanning slot can start from where it stopped in the previous step.

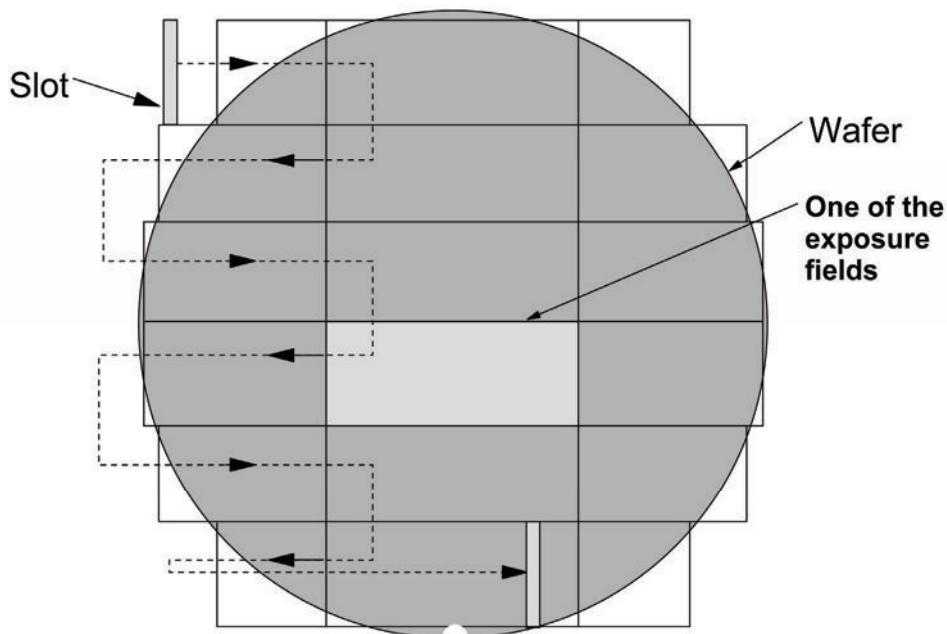


Figure 2.11 Step-and-scan principle. (Courtesy of D. Markle.)

A schematic representation of a step-and-scan projection system is shown in Fig. 2.12. The system shares many basic components of a step-and-repeat system, with the exception of the image-limiting slot and the scanning mechanism for the mask and the wafer. The slot remains stationary with the imaging lens, while the mask is scanned four times faster than the wafer. The majority of step-and-scan systems^{24,25,26} employ a dioptric (refractive) reduction lens. However, an early step-and-scan system uses a catadioptric lens with a beamsplitter,^{27,28,29} as shown in Fig. 2.13. This particular imaging system produces a mirror image so that the mask and the wafer can be scanned in the same direction. It is also arranged to make the mask and the wafer vertical. The other step-and-scan systems need to scan the mask and the wafer in opposite directions, making it slightly better in terms of mechanical balancing, but it is slightly more difficult to control their relative movements. The mask and wafer in these systems are placed horizontally. In either case, there can be mask sagging, but the sagging in the latter system tends to be more symmetrical. The dioptric, catadioptric, and reflective lens systems will be discussed in Sec. 5.5.2.

In the 1X full-wafer field-exposure system, a scanning slit is used, as was shown in Fig. 2.3. The slit has a curvature to follow the contour of the imaging field. Good imaging cannot be maintained outside of the curved narrow imaging field. On the other hand, the slot used to limit the field of a lens used for a full circular field is a straight opening.³⁰ The distance traveled with a slit is longer than that of a slot because of the curvature. As a result, slightly more scanning time is wasted with the slit field. The scanning field of the extreme ultraviolet (EUV) lithography system uses a slit field due to the all-reflective system required of EUV light.

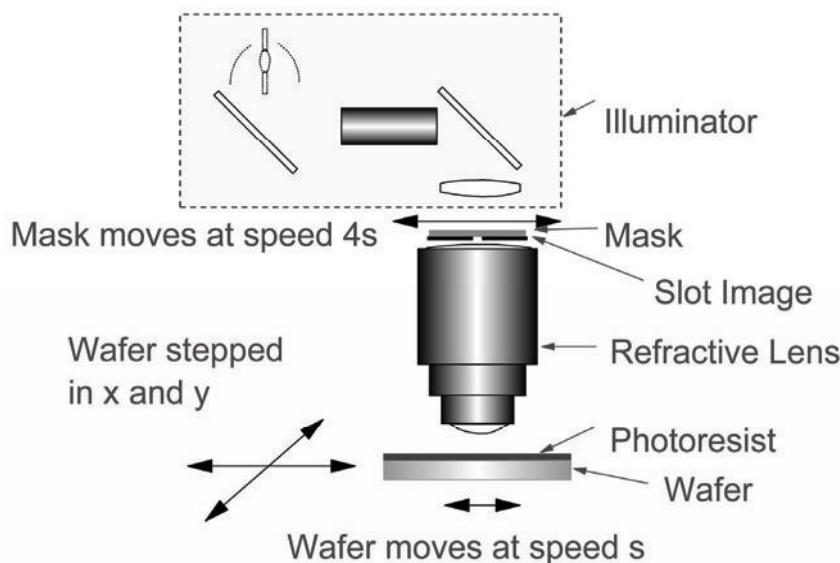


Figure 2.12 A step-and-scan projection system using a dioptric reduction lens.

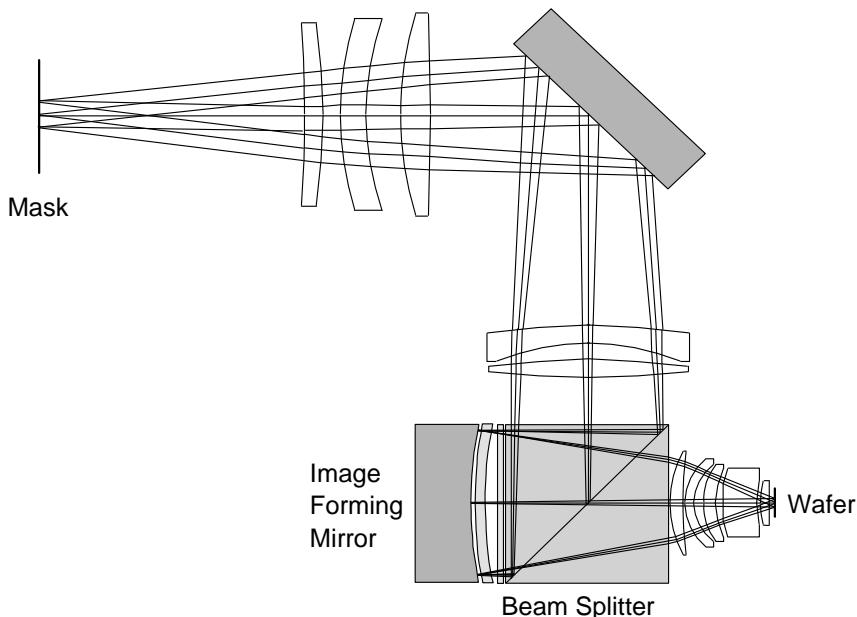


Figure 2.13 A catadioptric lens system with a beamsplitter used in a step-and-scan system. (Courtesy of D. Williamson.)

A major advantage of step-and-scan is the capability to extend field size indefinitely in one direction without adding any burden on the imaging lens. Figure 2.14 shows a typical square stepper field size of $22 \times 22 \text{ mm}^2$ that shares the same circle with a 1:2 stepper field size of $14 \times 28 \text{ mm}^2$. When the same lens is used for scanning, an $8 \times 30 \text{ mm}^2$ slot supports a $30 \times 33 \text{ mm}^2$ field for 150-mm reticles; the field can be extended to 50 mm in the scan direction when 225-mm reticles become available.

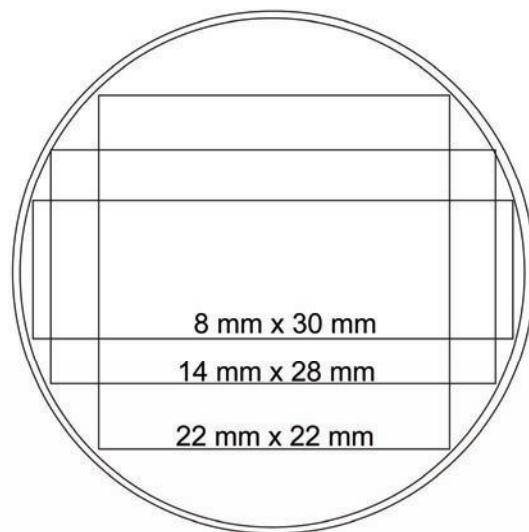


Figure 2.14 Stepper and scanner fields from the same circular field area.

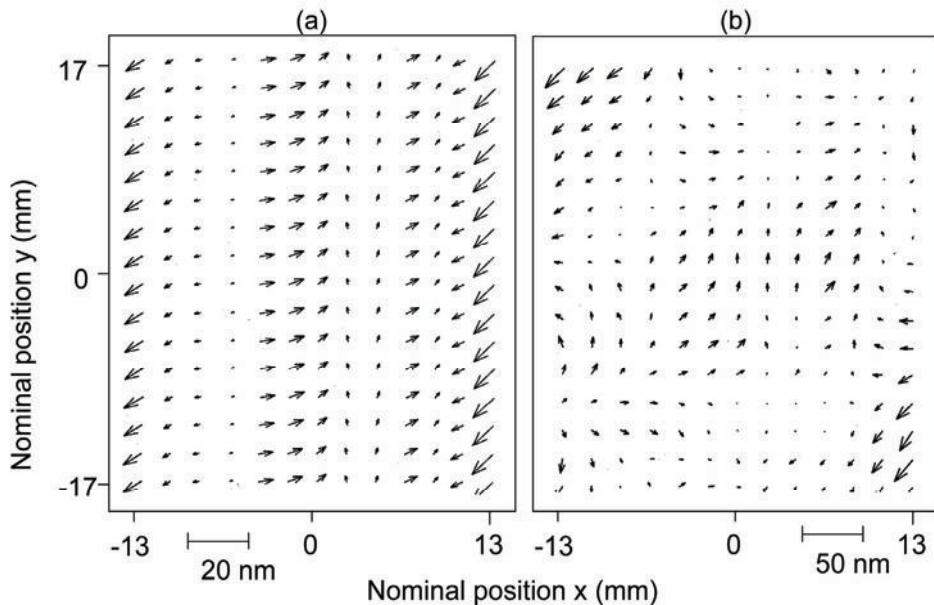


Figure 2.15 (a) Simulated dynamic distortion from measured static distortion. (b) Actual dynamic distortion from the same lens. (Courtesy of de Zwart.)

Another advantage of scanners is the unification of lens distortion in the scan direction. Figure 2.15(a) shows the simulated dynamic distortion map³¹ from the static distortion in a slot area of a stepper lens. When the scanning movement is perfect, this is what is expected. The actual distortion map is shown in Fig. 2.15(b). The difference between them is induced by scanning imperfection and measurement inconsistency. Note that the error vectors in (a) and (b) are different by a factor of 2.5X.

Further advantages of a scanner include in-situ focusing, leveling, and alignment during scan, which improves the lithographic performance. There is no reason the width of the slot should be uniform. Perturbation of the slot width may be used to compensate for illumination nonuniformity. Also, the scanning speed can be modified to produce an exposure gradient.

A step-and-scan system gains its advantages by trading off optical complexity with mechanical precision and complexity. In order to maintain throughput, the wafer stage must move at 250 mm/sec or higher. Modern scanners have achieved speeds in excess of 500 mm/sec. Needless to say, the mask must move at 1 and 2 m/sec, respectively, in a 4X system. Keeping the positioning accuracy to a few nanometers at such speeds cannot be taken for granted by any means.

Two actual step-and-scan systems are shown in Figs. 2.16 and 2.17, respectively.

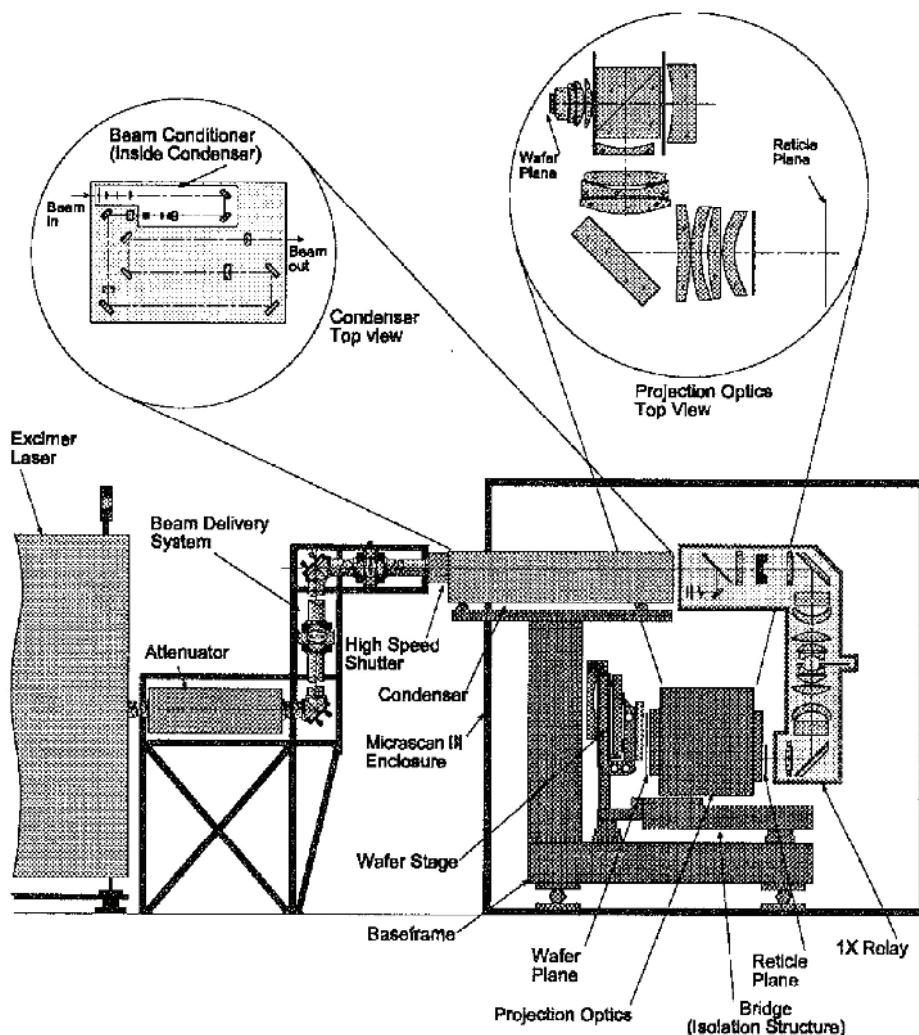


Figure 2.16 SVGL Micrascan™ III block diagram. (Courtesy of ASML.)

2.6 Reduction and 1X Systems

Lithography started with 1X replication because of its simplicity and versatility. As discussed above, 1X systems were popular until 1X-mask specifications could no longer be maintained due to the ever-reducing feature size and the overlay accuracy that followed. The switchover point was 1.5 μm for most manufacturers. From this point on, mask specification was of key importance in selecting the reduction ratio of the replication system.

Table 2.1 is an update of an earlier analysis³⁰ of the CD tolerance contributions from mask fabrication and wafer processing. We assume that the wafer resist feature size is controlled to 10% of the critical dimension. Wafer etch

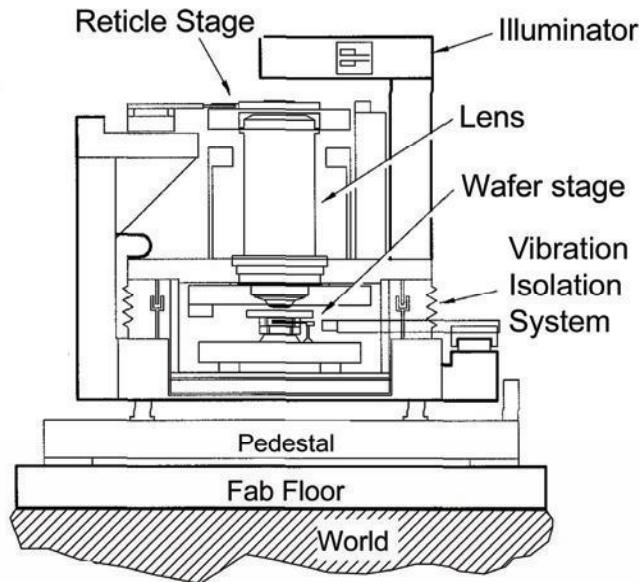


Figure 2.17 Schematic cross section of the ASML PAS500/500 step-and-scan system. (Courtesy of ASML.)

control is given at 6%, and hot-process-induced CD error is 5%. The resist and etch control of mask CD are 8% and 4.8%, respectively—better than those of the wafer. When the same control is applied to a 4X system, the mask control is improved by a factor of 4 to the first order, while the wafer tolerances are identical. The root-sum-square (RSS) of the CD error is 15.75% for 1X systems, compared to 12.9% for 4X systems—a 22% disadvantage. In order to make the CD control performance identical, the wafer resist image error of the 1X system must be reduced to 4.3%, which is quite impossible. Improving wafer and mask resist errors to 5.8% and 7%, respectively, makes the RSS identical for 1X and 4X systems. In either case, there are few reasons that the improvement cannot be applied to 4X systems as well, making it self-defeating for 1X systems to catch up.

Table 2.1 CD tolerance contributions from mask fabrication and wafer processing with MEEF = 1.

CD tolerance	4X	1X	1X, Improve wafer	1X, Improve mask & wafer
Mask resist	8%/4	8%	8%	7%
Mask etch	4.8%/4	4.8%	4.8%	4.8%
Wafer resist	10%	10%	4.3%	5.8%
Wafer etch	6%	6%	6%	6%
Hot process	5%	5%	5%	5%
RSS sum	12.9%	15.75%	12.9%	12.9%

Table 2.2 Same as Table 2.1, except $MEEF = 4$.

CD tolerance	4X	1X	1X, Improve wafer	1X, Improve mask & wafer
Mask resist	32%/4	32%	32%	0%
Mask etch	19.2%/4	19.2%	19.2%	19.2%
Wafer resist	10%	10%	0%	0%
Wafer etch	6%	6%	0%	0%
Hot process	5%	5%	0%	0%
RSS sum	15.7%	39.4%	37.3%	19.2%

Normally, any dimension change is transferred to the wafer linearly; i.e., a 10-nm change on a 1X mask becomes a 10-nm change on the wafer. Similarly, a 40-nm change on a 4X mask translates to a 10-nm change on the wafer. However, when the wafer image is pushed to well below λ/NA , the mask error enhancement factor $MEEF \equiv \Delta CD_{wafer}/\Delta CD_{mask}$ is no longer unity. It can exceed 4 in extreme cases. The contributions from the mask must be multiplied with MEEF in both the 1X and 4X cases. The situation is shown in Table 2.2. The increase in CD tolerance in the 4X case is undesirable, whereas the situation of the 1X case is disastrous. The entire CD tolerance budget is easily taken by the mask tolerance. Making the wafer contribution zero is still insufficient.

Table 2.3 is a similar update of an earlier analysis of the overlay budget. The mask placement tolerance is taken from an e-beam placement tolerance of 30 nm. When two masking levels are aligned against each other, the overlay tolerance is the RSS sum of the two placement tolerances; thus, $\sqrt{2} \times 30$ nm. The overlay error of a 1X system can be 76% more than a similar 4X system, as seen in columns 2 and 3. In fact, even when an e-beam directwrite system is used, the overlay accuracy is still on the order of 70% worse than a 4X reduction system. The original is worse than the copy! This shows the leverage of a reduction system.

Hence, from the point of view of CD and overlay control, a high reduction ratio is always desirable. However, a larger reduction ratio requires a larger mask substrate as well as longer and faster scanning to maintain the same field size on the wafer. These are expensive to implement in commercial systems. Conversely, one can use the same substrate size and reduce the wafer field size. As will be discussed in Sec. 6.15, longer stepper time is required to support smaller field

Table 2.3 Overlay budget contributions from mask fabrication and wafer processing.

	4X	1X	E-beam directwrite
Mask placement tolerance	42/4 nm	42 nm	42 nm
Alignment tolerance	15 nm	15 nm	12 nm
Magnification tolerance	15 nm	12 nm	12 nm
Lens distortion	10 nm	10 nm	0
Laser table tolerance	10 nm	10 nm	10 nm
RSS sum	27.6 nm	48.7 nm	46.8 nm

Table 2.4 CD tolerance analysis of the reduction-fabricated mask situation.

CD tolerance	
4X mask resist	2%
4X mask etch	1.2%
1X mask resist	8%
1X mask etch	4.8%
Wafer resist	10%
Wafer etch	6%
Hot process	5%
RSS sum	16%

size. The productivity loss is also prohibitive. When the mask contribution is no longer dominating, increasing the reduction ratio further has a diminishing return.

2.7 1X Mask Fabricated with a Reduction System

To leverage the advantage of reduction, optically reducing a 4X mask made with e-beam directwrite to make 1X masks has been considered. The tolerance analysis must include the extra step of mask making. For CD control, the tolerance components are shown in Table 2.4. The CD tolerance on the 4X mask indeed enjoys the 4X-reduction leverage. However, the CD control to fabricate the 1X mask is similar to that when fabricating the wafer; nevertheless, because the mask substrate is planar, there is a chance to control the resist and etch images on the 1X mask to 8% and 4.8% instead of 10% and 6%. The result is still worse than 1X lithography using conventional mask making.

The overlay budget using a reduction-fabricated mask is shown in Table 2.5. The contributions from magnification error, lens distortion, and laser table error in reduction mask making are added to the components shown in Table 2.3. The lens distortion component is smaller because an extremely low-distortion lens can be selected for reduction mask making. The resultant overlay tolerance of 32 nm is much better than that of conventional 1X lithography, but it is still 14% worse than conventional 4X-reduction lithography.

Table 2.5 Overlay budget analysis of the reduction-fabricated mask situation.

Overlay error component	Tolerance
4X mask placement	42/4 nm
Reduction-mask-making magnification	15 nm
Reduction-mask-making lens distortion	6 nm
Reduction-mask-making laser table	10 nm
Wafer-printing alignment	15 nm
Wafer-printing magnification	10 nm
Wafer-printing lens distortion	10 nm
Wafer-printing laser table	10 nm
RSS sum	32 nm

2.8 Summary

The progression of optical exposure tools from proximity to projection printing, 1X full-wafer to reduction step-and-repeat, and then to step-and-scan, is given. The interplay of resolution, working distance, field size, CD/overlay control, and productivity changed the configuration of the optical exposure system to the ultimate reduction step-and-scan system in the span of half of a century. Equipped with the global perspective of the optical exposure system, we are poised to move on the next chapter on optical imaging.

References

- 1 B. J. Lin, "Optical methods for fine line lithography," Chapter 2 in *Fine Line Lithography*, R. Newman, Ed., North-Holland, Amsterdam; American Elsevier Publishing Co., New York (1980).
- 2 Equation 2.22 of Ref. 1.
- 3 B. J. Lin, "A comparison of projection and proximity printings—from UV to x-ray," *Microelectron. Eng.* **11**, pp. 137 (1990).
- 4 B. J. Lin, "A new perspective on proximity printing—from UV to x-ray," *J. of Vac. Sci. Technol. B* **8**, p. 1539 (1990).
- 5 J. Z. Y. Guo, G. Chen, V. White, P. Anderson, and F. Cerrina, "Aerial image formation in synchrotron-radiation-based x-ray lithography: the whole picture," *J. of Vac. Sci. Technol. B* **8**, p. 1551 (1990).
- 6 B. J. Lin, "Deep UV lithography," *J. Vac. Sci. Technol. B* **12**, p. 1317 (1975).
- 7 M. Feldman, D. L. White, E. A. Chandross, M. J. Bowden, and J. Appelbaum, "A demonstration of photolithography at 1850 Å," paper given to B. J. Lin in 1975 without specific citation.
- 8 K. Jain, C. G. Wilson, and B. J. Lin, "Ultra high resolution contact lithography using excimer lasers," *Proc. SPIE* **334**, pp. 259–262 (1982).
- 9 J. C. White, H. G. Craighead, R. E. Howard, L. D. Jackel, R. E. Behringer, R. W. Epworth, D. Henderson, and J. E. Sweeney, "Submicron, vacuum ultraviolet contact lithography with an F2 excimer laser," *Appl. Phys. Lett.* **44**(1), p. 22 (1984).
- 10 D. L. Spears and H. I. Smith, "Prospects for x-ray fabrication of Si IC devices," *Electron. Lett.* **8**, p. 102 (1972).
- 11 M. Hasegawa, Y. Nakayama, K. Yamaguchi, T. Terasawa and Y. Matsui, "Printing characteristics of proximity x-ray lithography and comparison with optical lithography for 100-nm node and below," *Proc. SPIE* **3997**, p. 95 (2000).
- 12 J. S. Kilby, "Miniaturized Electronic Circuits," U.S. Patent 3,138,7431 (1964).

- 13 H. S. Hemstreet, D. A. Markle, W. H. Newell, and A. Offner, "Optical projection apparatus," U.S. Patent 4,011,011 (1977).
- 14 D. A. Markle "A new projection printer," *Sol. State Tech.* **17**, p. 50 (June 1974).
- 15 J. D. Buckley, "Expanding the horizons of optical projection lithography," *Sol. State Tech.* **25**, p. 77 (May 1982).
- 16 K. Ushid, M. Kameyama, and S. Anzai, "New projection lenses for optical stepper," *Proc. SPIE* **633**, pp. 17 (1986).
- 17 A. Suzuki, "Double telecentric wafer stepper using laser scanning method," *Proc. SPIE* **538**, p. 2 (1985).
- 18 S. Wittekoek, "Step and repeat wafer imaging," *Proc. SPIE* **221**, p. 2 (1980).
- 19 J. M. Roussel, "Step and repeat wafer imaging," *Proc. SPIE* **135**, p. 30 (1978).
- 20 K. Suwa, K. Nakazawa, and S. Yoshida, "10:1 step-and-repeat projection system," Kodak Micro-Electronics Seminar (1981).
- 21 "Cover story: mix-and-match: a necessary choice," *Semicond. Int.* **23**(2), pp. 66–67 (2000).
- 22 R. Hershel, "Optics in the Model 900 projection stepper," *Proc. SPIE* **221**, p. 39 (1980).
- 23 D. Markle, "The future and potential of optical scanning systems," *Sol. State Tech.* **27**, p. 159 (September 1984).
- 24 S. Murakami, "Optical exposure system—today and future," *Proc. of SEMI Technology Symposium* **94**, p. 397 (1994).
- 25 R. Ebinuma, K. Iwamoto, H. Takeishi, H. Itoh, M. Inoue, K. Takahashi, and M. Ohta, "Imaging performance of scanning exposure systems," *Proc. SPIE* **3334**, p. 437 (1998).
- 26 M. van den Brink, H. Jasper, S. Slonaker, P. Wijnhover, and F. Klaassen, "Step-and-scan and step-and repeat, a technology comparison," *Proc. SPIE* **2726**, p. 734 (1996).
- 27 D. M. Williamson, "Optical reduction system," U.S. Patent 4,953,960 (1990).
- 28 D. M. Williamson, "Catadioptric microlithographic reduction lenses," *Proc. Int. Optical Design Conference*, OSA, **22** (1994).
- 29 D. Williamson, J. McClay, K. Andresen, G. Gallatin, M. Himel, J. Ivaldi, C. Mason, A. McCullough, C. Otiss, J. Shamaly, and C. Tomczyk, "Micrascan III, 0.25 μm resolution step and scan system," *Proc. SPIE* **2726**, p. 780 (1996).
- 30 B. J. Lin, "The paths to subhalf-micrometer optical lithography," *Proc. SPIE* **922**, p. 256 (1988).

- 31 G. de Zwart, M. van den Brink, R. George, D. Satriasaputra, J. Baselmans, H. Butler, J. van Schoot, and J. de Klerk, "Performance of a step and scan system for DUV lithography," *Proc. SPIE* **3051**, p. 817 (1997).

Chapter 3

Image Formation

As discussed in Chapter 1, the goal of lithography is to place the edges of a given mask pattern on the wafer within a specified tolerance. Doing so with effective controls requires a good understanding of the image formation process—from the mask image to the final transferred image. The aerial image is formed from the light that illuminates the mask and carries the information of the mask pattern through the lens to focus on the image plane. The aerial image becomes reflected and refracted as it propagates into the resist and is reflected and refracted numerous times by the resist, as well as by the multiple film layers on the wafer substrate. The superposition of reflected and refracted images in the resist exposes this medium, producing chemical and physical changes. The distribution of these changes is the latent image. The exposed resist has a distribution of dissolution rate according to the latent image. The developer removes the resist according to the dissolution rate distribution and the topography of the resist surface. The developed resist is now the resist image. The transferred image is dictated by the resist image and the characteristics of the transfer process. These images are discussed below.

3.1 The Aerial Image

We define the aerial image as light distribution at the vicinity of the image plane without any resist or multiple-reflecting surface. The image must be treated as light waves in physical optics as opposed to light rays in geometric optics. When the light wave has a perfect wavefront to form the image, there is no aberration to deteriorate the image. The only limit on resolution comes from the finite extent of the wavefront dictated by the aperture of the imaging lens. A larger extent produces a better diffracted image. Such an imaging system is called a diffraction-limited system. We start our investigation using a perfect spherical wavefront with a finite lens aperture to form a point image, which leads to the relationship of diffraction-limited resolution to wavelength and aperture size, followed by the relationship of depth of focus to the same set of parameters. Deviation from the spherical wavefront is treated in Sec. 3.1.4 below. The formation of a diffraction-limited aerial image from a mask pattern and three

different types of illumination will be given in Secs. 3.1.5 and 3.1.7. The concept of spatial frequency and angular spectrum is treated further in Sec. 3.1.6 below.

3.1.1 Effects of a spherical wavefront and deviations from it

Lens designers and lens makers take great pains to make the imaging lens diffraction limited. In other words, their goal is to make every image point in the lens field converge from spherical wavefronts of identical aperture and intensity. Therefore, the spherical wavefront plays a key role in understanding diffraction-limited image formation. A finite aperture further makes the diffraction-limited image aperture limited. Deviation from the spherical wavefront deteriorates the aperture-limited image. In reality, even though the deviation is suppressed to tenths of wavelength, aberrations induced from the deviation cannot be ignored when optical lithography is pushed to its ultimate limit. Thus, Sec. 3.1.4 will cover analytic treatments of deviation from the spherical wavefront.

3.1.2 Spherical wavefront

Let's consider the imaging of a point source. Given a point object, a perfect point image is to be reproduced in the image plane. The perfect wavefront in this case is a spherical wavefront

$$E = Ae^{ikr}, \quad (3.1)$$

where E is the electric field of the light wave, A is the magnitude of the electric field, and r is the distance between the image point P and a point at the wavefront as shown in Fig. 3.1.

$$k \equiv 2n\pi/\lambda_0 \quad (3.2)$$

is the wave number, n is the refractive index of the medium between the spherical wavefront and the image plane, and λ_0 the wavelength in vacuum.

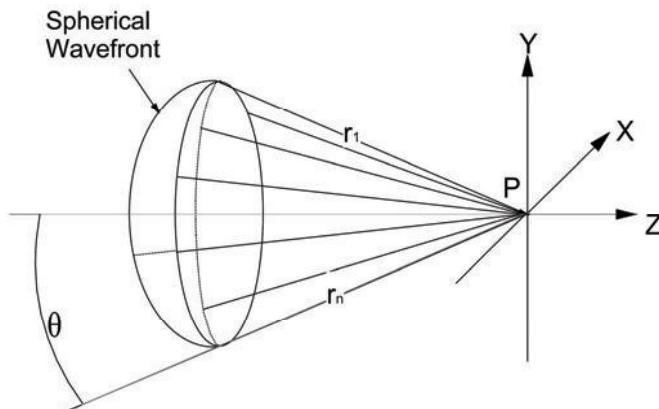


Figure 3.1 A spherical wavefront.

In Fig. 3.1, $P(0,0,0)$ is the origin of the coordinate system at the image plane. All points at the spherical wavefront are equidistant to $P(0,0,0)$, thus, in phase to each other. The electric field (instead of the magnetic field in the electromagnetic wave of the propagating light) is used here because photoresists record light intensity that is proportional to $|E|^2$. This phenomenon is routinely observed in the practice of optical lithography when dealing with standing waves in the resist as will be discussed in Sec. 5.6.2.5. This spherical wavefront produces the strongest E field (thus, the sharpest image) at $P(0,0,0)$, but it is not infinitesimally small because of the finite angle θ sustained by the section of the sphere. In practice, the finite bandwidth of the imaging light also contributes to the noninfinitesimally small image. With an actual lens, the usable section of the spherical wavefront is confined by the size of the exit pupil of the imaging lens. The size of the exit pupil is often conveniently represented by the numerical aperture (NA) of the imaging lens, defined as the product of $\sin\theta$ and the refractive index n of the media between the spherical wavefront and the point P as in

$$NA = n \sin \theta . \quad (3.3)$$

Note that the NA of the lens is different at the mask and the wafer sides in a reduction system. With 4X reduction, the NA at the mask side is 25% of the NA at the wafer side. Even though the mask is four times larger than its image, the solid angle sustained by the imaging lens at the mask side is only 25% as much as that at the wafer side. The NA in optical microlithography usually refers to that at the wafer side because it is used to predict resolution at the wafer.

3.1.3 The effect of a finite numerical aperture on the spherical wavefront

We now show the effect of a finite NA on $E[P(x,0,0)]$. To simplify the calculation, the spherical wavefront is flattened to a cylindrical shape, and the y axis is omitted from integration as shown in Fig. 3.2. Let the NA defining angle be θ_A , and let an arbitrary point R on the wavefront make an angle θ with the optical axis z . On $P(x,0,0)$, the electric field of the image is determined by the accumulation of light on all points of the sphere within the angle θ_A . When x is close to $x = 0$, the electric field vector E_x propagating in the z -direction is

$$E_x [P(x,0,0)] = \frac{1}{2\theta_A} \int_{-\theta_A}^{\theta_A} A e^{iks} \cos \theta d\theta , \quad (3.4)$$

where s is the distance between R and P , and r is the distance between R_0 and P_0 . $RP_0 = R_0P_0 = r$ by definition of a spherical wavefront. $R_0P = RP_0 = R_0P_0$ by small angle approximation. $RP = s = R_0P - x \sin\theta = r - x \sin\theta$. A $\cos\theta$ term is present in Eq. (3.4) because the z -component of light propagation is taken. A

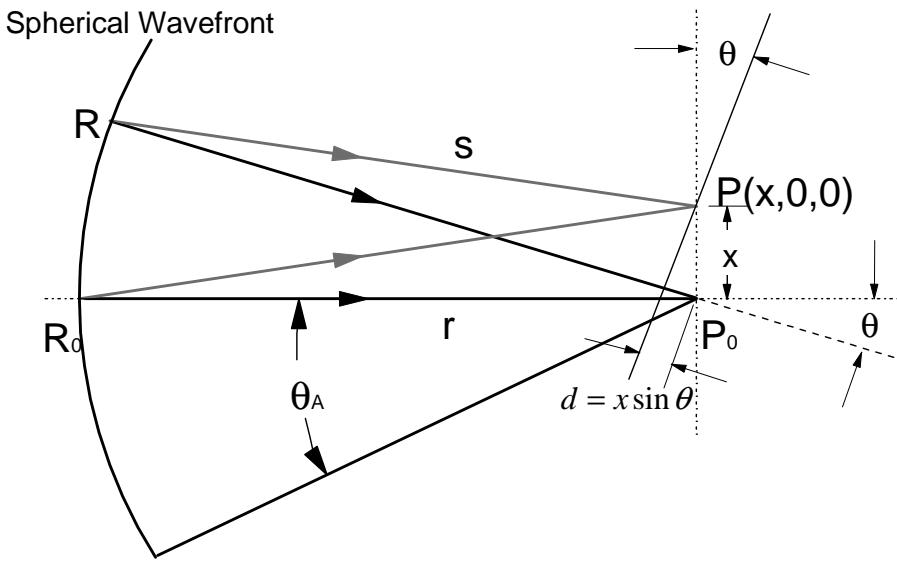


Figure 3.2 Path difference at an off-axis point.

$\sin\theta$ term would be there if the x -component of propagation were considered. Here only the z -component is of interest. The x -component is zero due to symmetry with respect to the z -axis.

$$E[P(x, 0, 0)] = \frac{1}{2\theta_A} \int_{-\theta_A}^{\theta_A} A e^{ik(r-x\sin\theta)} \cos\theta d\theta. \quad (3.5)$$

After integration,

$$E[P(x, 0, 0)] = \frac{A e^{ikr} \sin(kx \sin \theta_A)}{ikx \sin \theta_A}. \quad (3.6)$$

Substituting Eq. (3.2),

$$E[P(x, 0, 0)] = \frac{A e^{ikr} \sin\left(2\pi \frac{NA}{\lambda} x\right)}{2\pi \frac{NA}{\lambda} x}. \quad (3.7)$$

Hence, the intensity at the vicinity of the image point is not zero. It becomes zero only when

$$2\pi \frac{NA}{\lambda} x = \pi, \quad (3.8)$$

i.e.,

$$x = 0.5 \frac{\lambda}{NA}. \quad (3.9)$$

The proportional constant 0.5 may change as a function of the shape of the lens aperture boundary (e.g., square or circular shape). It is designated as k_1 in Eq. (4.1) and is a function of many imaging parameters. Extensive discussions on k_1 will be given in Chapter 4. Here, the main point of Eq. (3.9) is that—even with a perfect spherical wavefront—the image spot has a finite size dictated by λ and NA. When wavelength is reduced, the spot size becomes smaller, whereas a smaller NA produces a larger spot size. The smaller the spot size, the better the resolution. This NA- λ dependency is the basis of Eq. (4.1).

After studying $E[P(x,0,0)]$, we will examine the effects of a finite NA to DOF by studying $E[P(0,0,z)]$ in the vicinity of the focal plane. Equation (4.1) is again used for the derivation. Here, the path difference is between R_0P and RP , which is shown to be $2z \sin^2(\theta/2)$ and depicted in Fig. 3.3.

$$\delta = s - (r + z) = \frac{s^2 - (r + z)^2}{s + (r + z)}, \quad (3.10)$$

where $s = RP$, $r = R_0P = RP_0$, and $z = P_0P$. From the cosine law,

$$s^2 = r^2 + z^2 - 2rz \cos(\pi - \theta). \quad (3.11)$$

Combining Eqs. (3.10) and (3.11) produces

$$\delta = \frac{2rz(\cos \theta - 1)}{s + r + z} = \frac{-4rz \sin^2\left(\frac{\theta}{2}\right)}{s + r + z}. \quad (3.12)$$

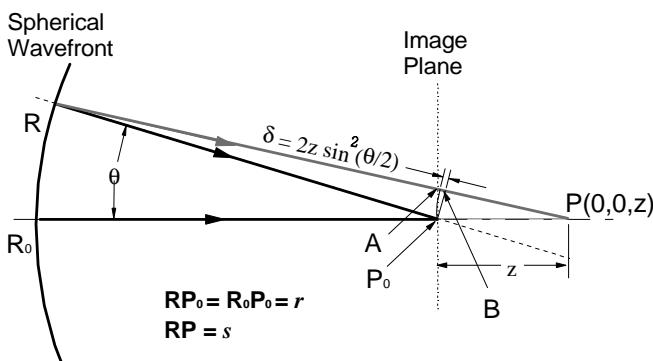


Figure 3.3 Path difference at an axial off-focus point.

Setting $s + r + z \approx 2r$ results in

$$\delta = -2z \sin^2\left(\frac{\theta}{2}\right). \quad (3.13)$$

Substituting into

$$E[P(0,0,z)] = \frac{1}{2\theta_A} \int_{-\theta_A}^{\theta_A} A e^{iks} \cos \theta d\theta, \quad (3.14)$$

$$E[P(0,0,z)] = \frac{1}{2\theta_A} \int_{-\theta_A}^{\theta_A} A e^{ik(r+z-2z \sin^2 \frac{\theta}{2})} \cos \theta d\theta, \quad (3.15)$$

and

$$E_z[P(0,0,z)] = \frac{A e^{ik(r+z)}}{2\theta_A} \int_{-\theta_A}^{\theta_A} e^{-i2kz \sin^2 \frac{\theta}{2}} \cos \theta d\theta. \quad (3.16)$$

A closed-form solution cannot be found for this integration. Instead of relying on a closed-form result, one can derive a similar relationship of the finite size of the image point along the z -axis, just as that in the x -axis, shown in Eq. (3.9). We will sum up the contribution from different parts of the wavefront by $\Delta\theta$, instead of integrating in $d\theta$. We choose $\Delta\theta$ such that there are 10 wavelets from the wavefront to the image point as shown in Fig. 3.4. When $z = 0$, all wavelets from the spherical wavefront are in phase. The intensity is the highest. As soon as z increases, the wavelets are out of phase by the increment $(1/2)z \cdot \sin^2(\theta_A/2)$ and start to have imaginary components, making the total intensity smaller. The first minimum intensity is reached when

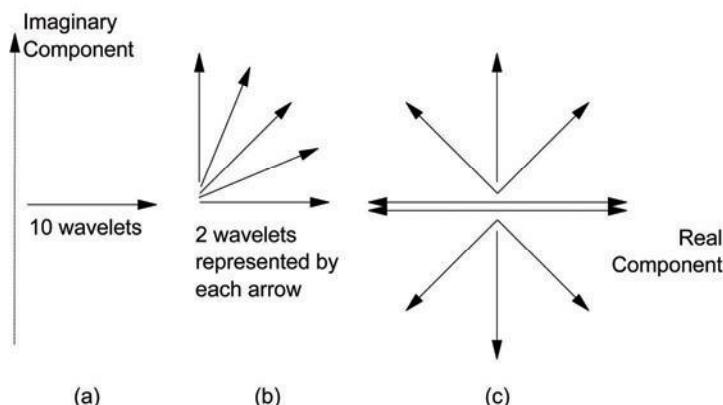


Figure 3.4 Wavelet components toward the image point (a) $z = 0$, (b) $z = 0.125n\lambda NHA^2$, and (c) $z = 0.5n\lambda NHA^2$.

$$2kz \sin^2 \frac{\theta_A}{2} = 2\pi. \quad (3.17)$$

Hence, the intensity reaches the first minimum when

$$z = \frac{0.5\lambda}{n \sin^2 \frac{\theta_A}{2}} = \frac{0.5n\lambda}{NHA^2} \approx \frac{2n\lambda}{NA^2}. \quad (3.18)$$

The numerical half aperture (NHA) is defined as

$$NHA \equiv n \sin \frac{\theta_A}{2}. \quad (3.19)$$

Therefore, the DOF of a point source from a perfect spherical wavefront is not zero. It is dictated by λ and θ_A , just as in the case of lateral displacement. One difference is that it is inversely proportional to $\sin^2(\theta_A/2)$ instead of just $\sin\theta_A$. The factor 0.5 is replaced by k_2 in actual situations, just as k_1 is for lateral displacement. Equation (3.18) is the basis for Eq. (4.2).

We have used simplified analysis to help grasp the physical meaning of image formation. The $\sin^2(\theta_A/2)$ has been derived from Sheppard.¹ For an extensive treatment of imaging a point image from a spherical wavefront, consult Born and Wolf.² They demonstrated that the normalized intensity distribution in the focal plane from a spherical wavefront with circular stop is

$$I(0, v) = \left[\frac{2J_1(v)}{v} \right]^2 \quad (3.20)$$

for small NA, and the normalized intensity distribution along the z -axis is

$$I(u, 0) = \left(\frac{\sin \frac{u}{4}}{\frac{u}{4}} \right)^2, \quad (3.21)$$

where

$$u = \frac{2\pi}{\lambda} \left(\frac{a}{r} \right)^2 z \quad (3.22)$$

and

$$v = \frac{2\pi}{\lambda} \left(\frac{a}{r} \right) x, \quad (3.23)$$

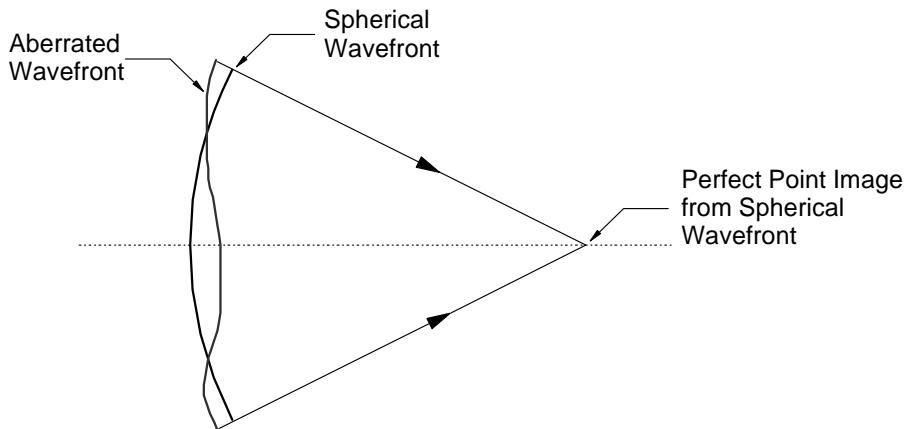


Figure 3.5 Perfect and aberrated wavefronts.

where a is the radius of the circular stop and r is the distance from the wavefront to the image point. The ratio a/r can be considered as a paraxial approximation of $\sin \theta_A$. Hence, Eqs. (3.20) and (3.21) arrive at the same implications of Eqs. (3.9) and (3.18), respectively. Note that the DOF of a diffraction-limited system of arbitrary NA is not inversely proportional to NA^2 but to NHA^2 .

3.1.4 Deviation from a spherical wavefront

The diffraction-limited image is produced by a spherical wavefront. When the wavefront deviates from a perfect sphere to reproduce a point source, the image is no longer diffraction limited. Light distribution near the image is much more complicated to analyze. Figure 3.5 shows an aberrated wavefront superimposed on a reference spherical surface.

With the aberrated wavefront, Eq. (3.1) becomes,

$$E = Ae^{ik(r+\psi)}. \quad (3.24)$$

The aberration is characterized with aberration coefficients. They are simply the coefficient of a polynomial that analytically defines the deviation from a perfect sphere. There are two major types of representations—the Seidel aberration coefficients and the Zernike aberration coefficients.

3.1.4.1 The Seidel aberration coefficients

The Seidel aberration coefficients are based on a power series expansion of the aberration wavefront in x_i , y_i , x_p , and y_p , where x_i and y_i are in the image plane, and x_p and y_p are in the exit pupil plane.

$$\Psi = \sum_l^{\infty} \sum_q^{\infty} \sum_m^{\infty} \sum_n^{\infty} A_{lqmn} x_i^l y_i^q x_p^m y_p^n, \quad (3.25)$$

where m , n , l , and q are positive integers. Many simplifications can take place. If one is only interested in an axially symmetrical optical system, then ψ only contains the combination of $x_i^2 + y_i^2$, $x_p^2 + y_p^2$, and $x_i x_p + y_i y_p$. Furthermore, since only one image point is considered, and axial symmetry is present, any image point with an arbitrary coordinate in x_i and y_i can be rotated to x_i , making y_i zero, thus simplifying the combination to x_i^2 , $x_p^2 + y_p^2$, and $x_i x_p$. The pupil rectangular coordinates x_p and y_p are replaced by polar coordinates ρ and θ .

$$x_p = \rho \cos \theta. \quad (3.26)$$

$$y_p = \rho \sin \theta. \quad (3.27)$$

Equation (3.25) now takes the form of

$$\psi = \psi^{(0)} + \psi^{(2)} + \psi^{(4)} + \psi^{(6)} + \dots \quad (3.28)$$

and

$$\psi^{(2)} = A_{200}\rho^2 + A_{111}x_i\rho \cos \theta + A_{002}x_i^2. \quad (3.29)$$

Note that l represents the order of x_i , m that of ρ , and n that of $\cos \theta$ in A_{lmn} .

$$\begin{aligned} \psi^{(4)} = & A_{400}x_i^4 + A_{040}\rho^4 + A_{131}x_i\rho^3 \cos \theta + A_{222}x_i\rho^2 \cos^2 \theta \\ & + A_{220}x_i^2\rho^2 + A_{311}x_i^3\rho \cos \theta. \end{aligned} \quad (3.30)$$

Given the aberrated wavefront ψ , the intersection of the aberrated imaging ray to the image plane determined by the spherical wavefront is^{3,4}

$$\Delta x = r \frac{\partial \psi}{\partial x_p}, \quad (3.31)$$

$$\Delta y = r \frac{\partial \psi}{\partial y_p}, \quad (3.32)$$

and

$$\Delta z = \frac{r}{NA^3} \frac{\partial \psi}{\partial \rho}. \quad (3.33)$$

In Eq. (3.30), the A_{400} term disappears after the differentiation in Eqs. (3.31) to (3.33). The remaining five coefficients constitute the five primary Seidel aberrations; namely, A_{040} for spherical aberration, A_{131} for coma, A_{222} for

astigmatism, A_{220} for field curvature, and A_{311} for distortion. Detailed descriptions of these aberrations can be found in standard optics textbooks (e.g., Refs. 3 and 4).

3.1.4.2 The Zernike aberration coefficients

The Zernike aberration coefficients are more popular than the Seidel aberration coefficients for optical lithography, because the Zernike polynomials^{5,6,7} are in cylindrical coordinates that fit the geometry of an imaging lens. Imaging lenses are designed and characterized with them. The Zernike polynomials not only form a complete set inside the unit circle, they are also invariant under rotation. In addition, the Zernike polynomials are orthogonal between orders in the radial and the rotational directions.

In the following, we will derive expressions of Zernike polynomials based on simple arguments. With the requirement that Zernike polynomials are invariant in their functional form under rotation, they should be of the following form:

$$Z_j(x, y) = R_n^m(\rho)e^{\pm im\phi}. \quad (3.34)$$

Note that $x = \rho \cos \phi$ and $y = \rho \sin \phi$. Here j is related to the ordering of Zernike polynomials, and the meaning of parameters n and m that describe Zernike polynomials will be evident following the arguments below:

- (a) If $Z_j(x, y)$ is a polynomial of degree n in x and y , then $R_n^m(\rho)$ is a polynomial of degree n in ρ .
- (b) Since $R_n^m(\rho)e^{\pm im\phi}$ is originally a polynomial of x and y , $e^{\pm im\phi}$ must originate from $(\rho \cos \phi \pm i \rho \sin \phi)^m$, and therefore $R_n^m(\rho)$ contains terms in ρ of power of at least m .
- (c) Now, we can rewrite $R_n^m(\rho) = \rho^m Q_{n-m}(\rho)$, where $Q_{n-m}(\rho)$ is a polynomial of degree $n-m$ in ρ . Since the ϕ dependence has been properly taken care of, the remaining ρ dependence, as collected in $Q_{n-m}(\rho)$, should be in powers of the combination $(\rho \cos \phi)^2 + (\rho \sin \phi)^2 = \rho^2$. So, $n-m$ must be even, and $R_n^m(\rho) = t^{m/2} Q_{(n-m)/2}(t)$, where $t = \rho^2$.

Now we are ready to work on the orthogonal property of the Zernike polynomials. In the radial direction,

$$\int_0^1 R_n^m(\rho) R_{n'}^m(\rho) \rho d\rho = \frac{1}{2(n+1)} \delta_{nn'}, \quad (3.35)$$

whereas, in the direction of rotation,

$$\int_0^{2\pi} \cos m\theta \cos m'\theta d\theta = \pi(1 + \delta_{m0}) \delta_{mm'}. \quad (3.36)$$

Equation (3.36) is trivial in the direction of rotation. We will go further in the derivation of Eq. (3.35) below. By requiring

$$\int_0^1 t^m Q_k(t) Q_{k'}(t) dt = a_n^m \delta_{kk'}, \quad (3.37)$$

where $k = (n - m)/2$ and $k' = (n + m)/2$, $Q_{(n-m)/2}(t)$ can be constructed by orthogonalizing the sequence $1, t, t^2, \dots$ with the weighting factor t^m over the range $0 \leq t \leq 1$. The hypergeometric polynomials can satisfy the requirement of Eq. (3.36). Interested readers can consult Born and Wolf⁶ for details.

Conventionally, Zernike polynomials are ordered from small n to large n . For each n , (n,m) starts from (n,n) , then $(n+1, n-1)$, $(n+2, n-2)$, ... until it reaches $(2n,0)$. For each (n,m) , two Zernike polynomials correspond, namely, $R_n^m(\rho)\cos(m\phi)$ followed by $R_n^m(\rho)\sin(m\phi)$ if $m \neq 0$. The situation of $m = 0$ corresponds to only one Zernike polynomial for each n .

Since Zernike has set $R_n^m(1) = 1$, the Zernike polynomials are not normalized. It can be shown that $a_n^m = 1/n + 1$ and

$$\int_0^{2\pi} \int_0^1 Z_j(\rho, \phi) Z_{j'}(\rho, \phi) \rho d\rho d\theta = \frac{\pi}{2(n+1)} (1 + \delta_{m,0}) \delta_{j,j'}. \quad (3.38)$$

In summary, with

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} (-1)^s \frac{(n-s)!}{s! [(n+m)/2-s]! [(n-m)/2-s]!} \rho^{n-2s}, \quad (3.39)$$

Zernike polynomials are represented by

$$Z_j(\rho, \phi) = R_n^m(\rho) \cos(m\phi), \quad (3.40)$$

where $j = [(n+m)/2 + 1]^2 - 2m$ or

$$Z_j(\rho, \phi) = R_n^m(\rho) \sin(m\phi), \quad (3.41)$$

where $j = [(n+m)/2 + 1]^2 - 2m + 1$.

When the system is completely symmetrical, the $\sin(m\phi)$ terms vanish. Some authors count j according to only the situation of Eq. (3.40). That is, $j = 1$ when $n = m = 0$; $j = 2$ when $n = m = 1$; $j = 3$ when $n = 2$ and $m = 0$; etc. In reality, rotational symmetry may be broken due to imperfection when fabricating each optical element in the lens or when assembling them. Therefore, we include the

sine terms in the counting of j . So, $j = 37$ without the sine terms corresponds to $j = 49$ with the sine terms.

Zernike polynomials also provide a way to classify lens aberration. Generally, if m is even, the aberrations are called even aberrations; similarly, odd aberrations refer to odd m . When $m = 0, 1, 2, 3, 4, 5$, respectively, the corresponding aberration is called spherical, coma, astigmatism, 3-foil, 4-foil, 5-foil, respectively. When $n + m = 4$, the aberration is called primary, primary spherical, primary coma, etc.; $n + m = 6$ corresponds to secondary aberrations; $n + m = 8$ corresponds to tertiary aberrations. Note that $j = n^2$ always corresponds to spherical aberration, i.e., $m = 0$, and (n,m) for the next j is (n,n) . This provides a good mnemonic for the classification of Zernike polynomials.

If the deformation of a wavefront is expanded in Zernike polynomials,

$$\Phi = \sum_j c_j Z_j , \quad (3.42)$$

then the mean-square deviation of the wavefront is

$$\overline{(\Delta\Phi)^2} = \frac{\int_0^{2\pi} \int_0^1 (\Delta\Phi)^2 \rho d\rho d\phi}{\int_0^{2\pi} \int_0^1 \rho d\rho d\phi} = \sum_j \frac{1 + \delta_{m,0}}{2(n+1)} c_j^2 . \quad (3.43)$$

This is another advantage of the Zernike polynomials. The magnitude of each Zernike term can be independently summed quadratically for the mean-square deviation of the wavefront. Imaging lenses are often specified this way.

The polynomials can readily be written explicitly for each combination of m and n with Eq. (3.39). For example,

$$R_4^2(\rho) = 4\rho^4 - 3\rho^2 \quad (3.44)$$

and

$$R_6^2(\rho) = 15\rho^6 - 20\rho^4 + 6\rho^2 . \quad (3.45)$$

We now test the orthogonality of these two terms:

$$\int_0^1 R_4^2(\rho) R_6^2(\rho) \rho d\rho = \int_0^1 \left(\frac{60}{12} \rho^{12} - \frac{125}{10} \rho^{10} + \frac{84}{8} \rho^8 - \frac{18}{6} \rho^6 \right) d\rho = 0 .$$

3.1.5 Imaging from a mask pattern

With optical microlithography, one seldom prints point objects, unless they are studying the aberrations of the imaging lens by intentionally creating pinhole objects in the mask. Instead, long lines, rectangles, squares, elbows, T shapes, or a combination of these basic patterns transformed into complicated circuit

patterns are printed. Therefore, the diffraction effects of mask patterns with and without aberrations must be studied. The E-D tools described in Chapter 4 are used to quantitatively study the impact of mask pattern shapes, sizes, and the proximity environment to the processing window. The following sections cover the derivation of the diffracted images from a given mask pattern.

3.1.5.1 Coherent imaging from a mask pattern

The imaging of a point object has been used to gain insights on diffraction-limited imaging and imaging with aberrations. Optical lithographers are concerned with the imaging of actual mask patterns of many shapes, sizes, and relative placements of them. Typical shapes consist of line-space pairs, lines, spaces, contact holes, opaque islands, etc. These objects are subject to diffraction limitations and aberrations just like a point object, but the analysis is more extensive. Let's first consider a coherent imaging situation of an object through a lens, as shown in Fig. 3.6. From geometric optics considerations, all light rays passing through the object opening in the mask are bent by the lens in such a way that they converge to the corresponding geometric image positions in the bright part of the image. From a wave-diffraction point of view, light waves from one object point are modified by the lens in such a way that they travel the same optical distance to the image point. In other words, the same number of peaks and valleys, as well as fractions of them, are experienced when they converge on the image point. The optical distance of all of the waves are made identical by the lens regardless of the angles they make. The thicker part of the lens is traveled by the paraxial waves, while the thinner part of the lens is traveled by the waves further off axis. Because wavelength is reduced by the refractive index of the lens media, the larger distances traveled in the lens media for the paraxial waves make up for the shorter physical distance to the image point. With coherent imaging, all waves from a single object point, as well as from any other object points, have a deterministic phase relationship with each other. For example, point 1 from object point A is 180 deg ahead of point 2 also from object point A, as shown in Fig. 3.6. It remains so at any instance. Point 3 from object point B also remains in phase with point 1 from point A and 180 deg ahead of point 2.

The equation relating the scalar electric field at the mask plane to that of the image plane is^{8,9}

$$E_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_o(x_o, y_o) K(x_i, x_o; y_i, y_o) dx_o dy_o. \quad (3.46)$$

In a well-corrected system, the transmission function K can be taken as

$$K(x_i - x_o, y_i - y_o) = \frac{1}{\lambda^2 r^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\xi, \eta) e^{\frac{-ik}{r} [(x_i - x_o)\xi + (y_i - y_o)\eta]} d\xi d\eta. \quad (3.47)$$

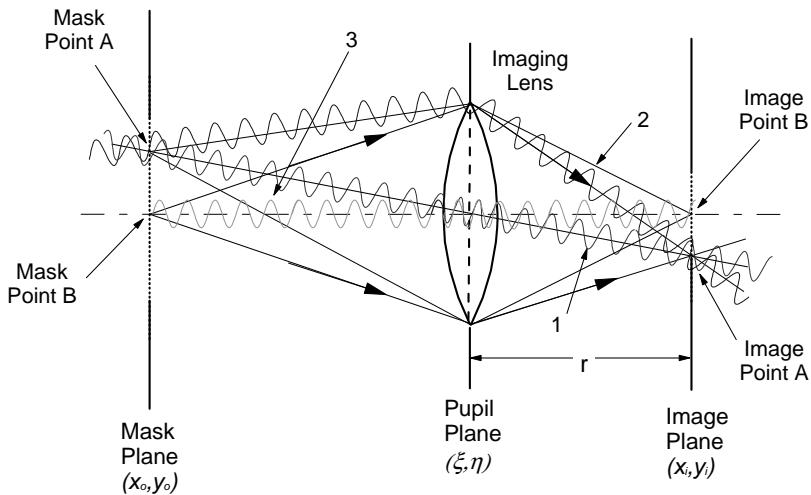


Figure 3.6 Image formation by coherent light.

Equation (3.46) can be dissected into Fourier transform and inverse Fourier transform components with insertion of the pupil function $G(\xi, \eta)$ when performing the inverse transform. That is,

$$E_i(x_i, y_i) = \frac{1}{\lambda^2 r^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x_o, y_o) e^{\frac{ik}{r}(x_o \xi + y_o \eta)} dx_o dy_o \right] G(\xi, \eta) e^{\frac{-ik}{r}(x_i \xi + y_i \eta)} d\xi d\eta. \quad (3.48)$$

Strictly speaking, the electric field in the mask plane is very complicated. It must satisfy Maxwell's equations as well as the boundary conditions.¹⁰ For a mask using chromium as the absorber, the transverse electric field is zero because chromium is a conductor and does not support any transverse electric field on its surface. This supports Eq. (3.49) at the opaque part of the mask for TE incidence. Bedefi¹¹ proved that the transverse magnetic field in the opening is identical to the incident transverse magnetic field. Otherwise, Eq. (3.49) does not hold.

Figure 3.7 shows the electric field at the mask plane of a 3λ -wide slit opening and its propagation into the image plane. Here TE incidence refers to the electric field perpendicular to the plane of incidence and parallel to the direction of the slit; TM incidence refers to the magnetic field parallel to the direction of the slit. The fine structure disappears at a few wavelengths away from the mask plane. It is more pronounced for TE incidence. With TM incidence, the magnetic field in the opening of the mask follows conventional assumptions [such as in Eq. (3.49)], except that the field in the opaque part of the mask is not zero. Conversely, the electric field in the opaque part is zero with TE incidence, but the field inside the mask opening does not fit conventional assumptions. In addition, there may be singularities at the edge of the mask opening. For illustration of the edge singularities, consult Ref. 12.

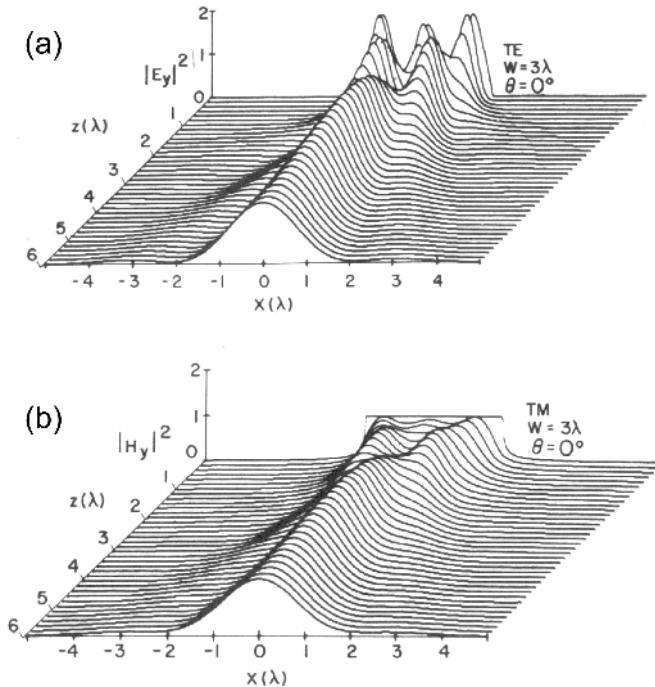


Figure 3.7 Electric field at and from a 3λ -wide slit. (a) TE incidence and (b) TM incidence.

Figure 3.8 shows the fine structure captured in photoresist using exposure of a mask that is coated with resist at the absorber side. The resist is thus exposed at perfect contact to the mask absorber. Unlike resist exposure from conventional proximity printing, the resist development starts at the plane farthest from the mask toward the contact plane. The distance from the contact plane is preserved for measurement after development, and the diffracted image can be recorded at extremely close vicinity to the mask absorber. In this figure, a 6.6λ -wide slit is used, and the distance from the capture plane to the mask plane is 2.1λ . The incident light is unpolarized containing equal TE and TM components.

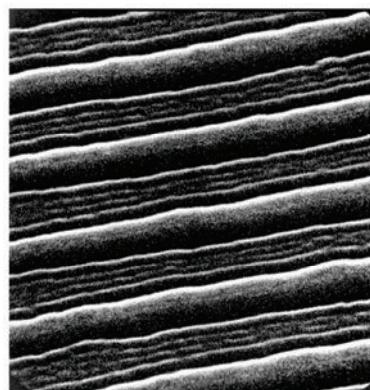


Figure 3.8 Experimental recording of an electric field squared near a 6.6λ -wide slit.

Fortunately, such a highly accurate representation is needed in optical lithography only when the field at close proximity of the mask is of concern, such as in the case of proximity printing;¹³ or when the mask pattern size is of the order of a wavelength or smaller, such as in the case of submicrometer 1X projection lithography or in 193-nm 4X reduction systems for 50-nm features or smaller. Attenuated-phase-shifting masks, with absorbers that transmit light and shift their phase, obviously do not follow Eq. (3.49). For most other purposes, the following approximation for $E(x_o, y_o)$ is usually sufficient.

$$E(x_o, y_o) = \begin{cases} 1, & \text{at openings} \\ 0, & \text{at opaque areas} \end{cases}. \quad (3.49)$$

K , the transmission function of the lens, is closely related to the kernel of Eq. (3.4) or Eq. (3.15). $G(\xi, \eta)$ is the pupil function, which is zero outside the lens aperture and is unity inside when there is no aberration. Otherwise, $G = 1 + \psi/r$, where ψ and r are defined in Eq. (3.24). The coordinates (x_o, y_o) , (ξ, η) , and (x_i, y_i) are depicted in Fig. 3.9. As mentioned earlier, photoresist responds to light intensity, not the electric field. Once the electric field is determined, the intensity distribution is simply

$$I(x_i, y_i) = E^*(x_i, y_i)E(x_i, y_i). \quad (3.50)$$

3.1.5.2 Incoherent imaging from a mask pattern

With incoherent imaging, the phase relationship of different image points is completely random, as depicted by Fig. 3.10. Although the phase relationship within one infinitesimal object point and its conjugate image point remains deterministic, it is impossible to pinpoint the electric field deterministically. However, the intensity distribution is readily evaluated with

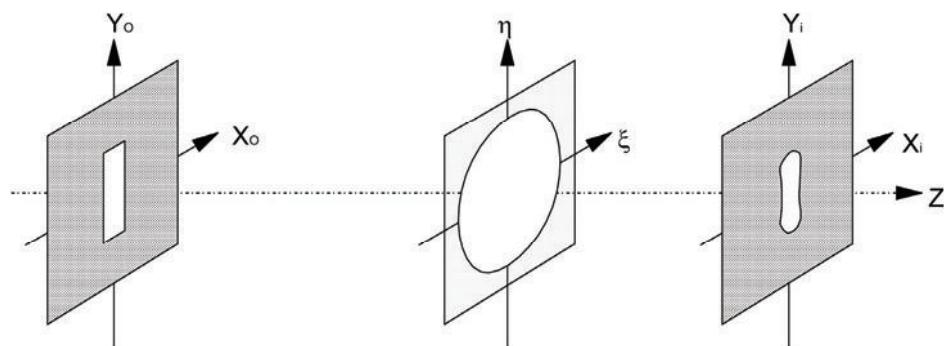


Figure 3.9 Object, pupil, and image coordinates.

$$I_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_o(x_o, y_o) |K(x_i, x_o, y_i, y_o)|^2 dx_o dy_o . \quad (3.51)$$

One can visualize the coherent imaging equation governing the imaging of an infinitesimal object. But with another object point, even if it is only a small finite distance away from the first object point, there is no deterministic phase relationship. The electric field between any two points does not interfere. The only way to account for the simultaneous presence of these points is to sum their contributions by intensity.

3.1.5.3 Partial coherent imaging from a mask pattern

In the actual practice of optical microlithography, the illumination of the mask is neither coherent nor incoherent but rather partially coherent; meaning, the light waves between different image points holds a statistical relationship instead of being completely random. The situation is depicted in Fig. 3.11.

Evaluating the intensity distribution is not as straightforward as Eqs. (3.50) or (3.51). The image intensity distribution from partially coherent illumination on a mask pattern based on Hopkins' analysis is^{14,15}

$$I(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_o(x_o, y_o; x'_o, y'_o) K^*(x_o, y_o; x_i, y_i) K(x'_o, y'_o; x_i, y_i) dx_o dy_o dx'_o dy'_o , \quad (3.52)$$

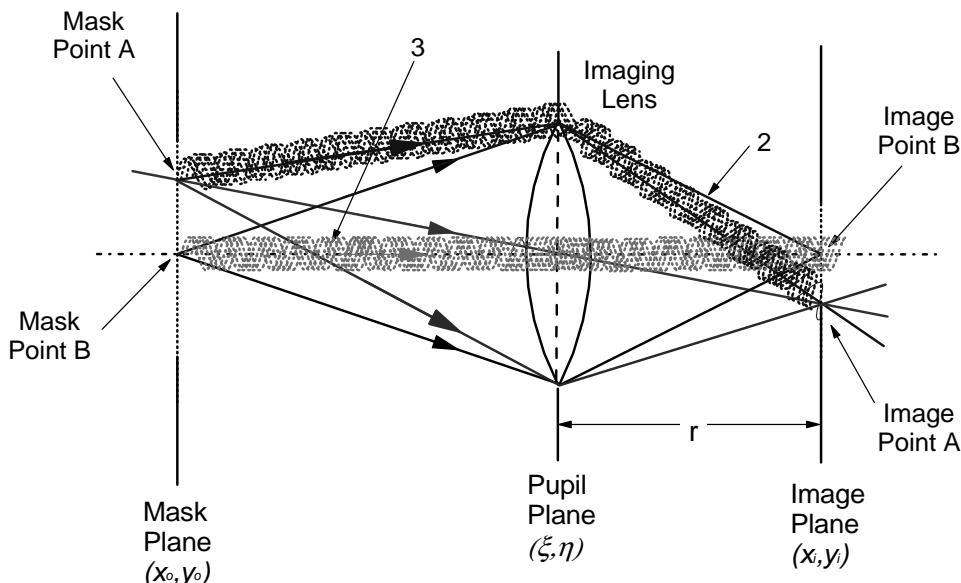


Figure 3.10 Incoherent imaging.

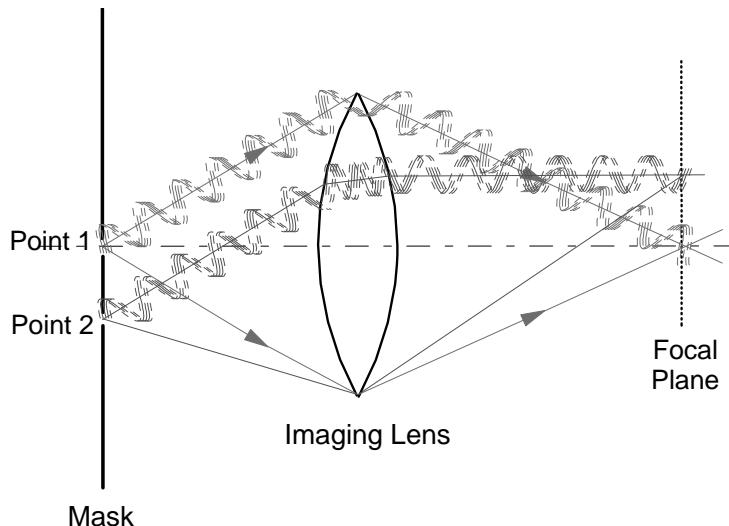


Figure 3.11 Partially coherent imaging.

where $J_o(x_o, y_o, x'_o, y'_o)$ is the mutual intensity at the mask plane, and K is the same transmission function as that in Eqs. (3.46) and (3.51). The mutual intensity at the mask plane can be further decomposed into two components:

$$J_o(x_o, y_o; x'_o, y'_o) = B_o(x_o, y_o; x'_o, y'_o)F(x_o, y_o)F^*(x'_o, y'_o) . \quad (3.53)$$

B_o is the mutual intensity of the illumination and F is the mask transmission function. B_o can be expressed in the Bessel function of the first kind of order 1, $J_1(u_c)$, where u_c is in units of NA_c , the numerical aperture of the condenser. K can also be expressed in $J_1(u_i)$, where u_i is in units of NA_i . It turns out that when the actual calculation is performed, the aperture ratio $\sigma = \text{NA}_c / \text{NA}_i$ is often used. This ratio becomes a parameter to indicate the degree of coherence of the optical imaging system. With a coherent system, $\sigma = 0$, and with an incoherent system, $\sigma = \infty$. When $\sigma = 0$, B_o becomes unity, and Eq. (3.52) becomes Eq. (3.50). When $\sigma = \infty$, B_o is a delta function, and Eq. (3.52) becomes Eq. (3.51).

Though the physical meaning is clear, Eq. (3.52) is time consuming to evaluate numerically because it involves a four-dimensional integral. A good way to make the calculation faster and the mathematics easier to handle is to treat the electric-field propagation as if it were completely coherent with light emerging from an infinitesimally small light source, using Eq. (3.46). Partial coherence is introduced by making the source its actual finite size and integrating the image intensity from Eq. (3.50) over the area of the light source. Yeung¹⁵ pioneered this concept for rigorous electromagnetic diffraction from high-NA imaging optics and through multilayered media as in the resist and thin-film stack found in typical semiconductor substrates.

Using either Eq. (3.50) or Eq. (3.52), spatial frequency space using Fourier transformation can accelerate numerical computation. Not only do the

convolution formulae become simpler, the existing fast Fourier transformation algorithm makes it extremely efficient to perform the Fourier and inverse-Fourier transformations. However, N , the total number of data points for integration, is often required to be an integer power of 2. In addition, a true isolated feature must be approximated with a periodic multiplicity of itself, separated by a large distance between them.

3.1.6 Spatial frequencies

Spatial frequencies play an important role in microlithography imaging for two reasons:

1. They are key in understanding wavelength and subwavelength imaging. The working principle of NA, illumination angle, phase-shifting masks, aperture engineering, etc., are easily understood with arguments given in terms of spatial frequencies.
2. They enable faster computation of the diffraction formulae, as was discussed in Sec. 3.1.5.3.

A given function in 2D space can be decomposed into spatial frequencies using the Fourier integral:

$$\kappa(f, g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(x, y) e^{i2\pi(fx+gy)} dx dy. \quad (3.54)$$

The inverse-Fourier transformation restores $K(x, y)$:

$$K(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(f, g) e^{-i2\pi(fx+gy)} df dg. \quad (3.55)$$

Applying Fourier transformation to both sides of Eq. (3.46) yields

$$E(f, g) = E_o(f, g) \kappa(f, g). \quad (3.56)$$

The physical meaning of Eq. (3.56) is clear. The spatial frequency distribution $E(f, g)$ at the image plane is evaluated by multiplying the spatial frequency distribution at the mask plane $E_o(f, g)$ with the pupil function $\kappa(f, g)$, which is simply unity inside the imaging lens aperture and zero elsewhere in an aberration-free system. The cutoff frequency $f_{cutoff} = g_{cutoff} = \text{NA}/\lambda$ for a lens with a circular pupil. The lens acts as a low-pass filter to relate the spatial frequency components at the mask plane to those at the image plane.

3.1.6.1 Spatial frequencies of an isolated line opening

We now illustrate, with a typical microlithography mask pattern, an isolated infinitely long line opening represented by

$$E(x) = \begin{cases} 1, & -\frac{w}{2} \leq x \leq \frac{w}{2} \\ 0, & \text{elsewhere} \end{cases}. \quad (3.57)$$

Equation (3.54) produces its spatial frequency

$$\mathcal{E}(f)/w = \frac{\sin(\pi wf)}{\pi wf} \quad (3.58)$$

as shown in Fig. 3.12. Hence, a simple line opening consists of a full spectrum of positive and negative spatial frequencies. The amplitude of these frequencies can be positive or negative. Because of the $1/f$ dependency, the significant frequencies in $\mathcal{E}(f)$ are distributed in the low-frequency region. The null points are $f = \pm n/w$, where n is a positive integer. In terms of angular spectrum,¹⁶

$$\sin \alpha = \lambda f, \quad (3.59)$$

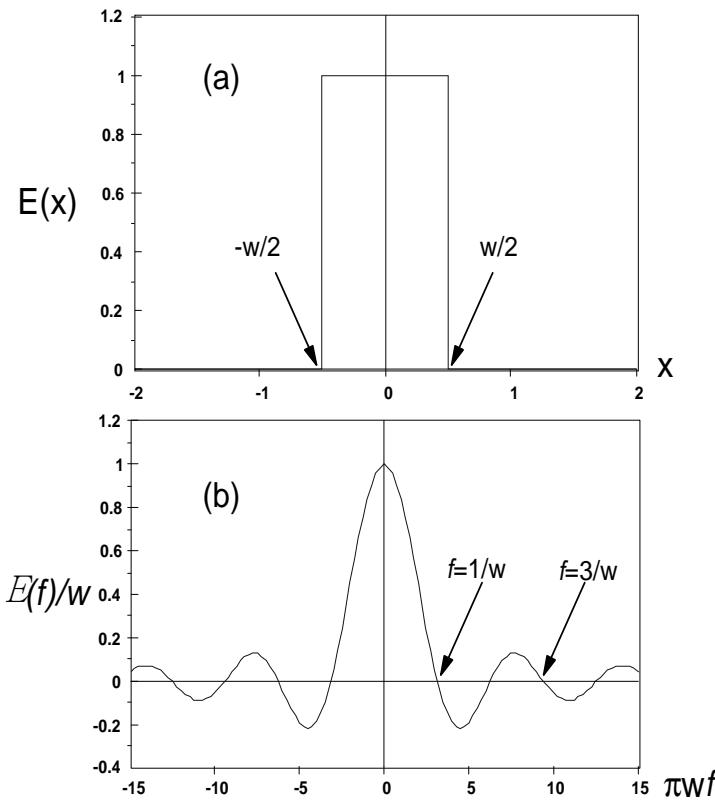
where α is the propagation angle of a plane wave corresponding the spatial frequency f .

$$\sin \alpha = \frac{\lambda}{w}. \quad (3.60)$$

If the numerical aperture of the imaging lens sets the cutoff frequency at the first null positive point, then

$$NA = \frac{\lambda}{w}. \quad (3.61)$$

Modern microlithography aims at resolving features with w values smaller than λ , whereas NA is seldom sufficient to reach a cutoff frequency higher than that at the first null point in the frequency spectrum. Let us consider $w = \lambda$ and $NA = 0.3, 0.6$, and 0.9 , respectively. The pupil functions are superimposed on the frequency distribution shown in Fig. 3.13(b). Equations (3.56), (3.55), and (3.50), respectively, provide the resultant intensity distributions at the image plane in Fig. 3.13(a).

Figure 3.12 (a) $E(x)$, and (b) $\mathcal{E}(f)$.

Note that the resultant images look quite different from those on the mask, even at a very high NA, because w is chosen to be equal to or lower than λ . Fortunately, microlithographers are not concerned with reproducing all of the gray levels in the object. The main purpose of microlithography is to control the image edge position, as stated in Sec. 1.3. The image is kept binary by taking advantage of the resist as a high-contrast recording material and the subsequent pattern transfer processes.

3.1.6.2 Spatial frequencies of line-space pairs

Another frequently used mask pattern in microlithography is the equal line-space pairs as represented by $E(x,y)$ in

$$E(x,y) = \begin{cases} 1, & np - \frac{w}{2} \leq x \leq np + \frac{w}{2} \\ 0, & np - \frac{w}{2} > x, x > np + \frac{w}{2} \end{cases}, \quad (3.62)$$

$n = 0, \pm 1, \pm 2, \dots$

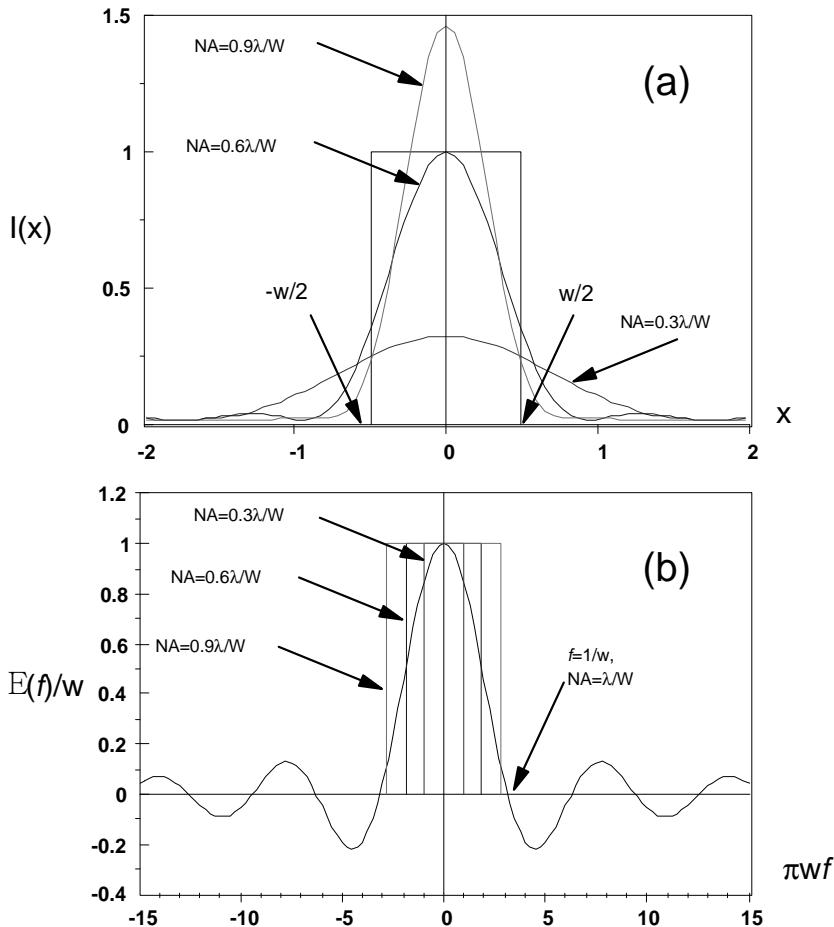


Figure 3.13 Pupil function applied to (a) $I(x)$, and (b) $E(f)/W$.

where w is the width of the openings, p is the periodicity, and n is an integer from negative to positive infinity (in other words, an infinite number of equal line-space pairs). If the number of line-space pairs is truly infinite, then the spatial frequencies are discrete with their amplitudes dictated by Eq. (3.58), as shown in Figs. 3.14 and 3.15. Note that the distance between the frequency spikes is always $1/p$ regardless of w .

The spatial frequency, according to the definition above, is independent of wavelength. It is purely a function of p with amplitude related to w . When the wavelength is put into consideration, the spatial frequency corresponds to a wave traveling in a direction as a function of p and λ , as shown in Fig. 3.16. That is,

$$\sin \alpha = \frac{\lambda}{p} . \quad (3.63)$$

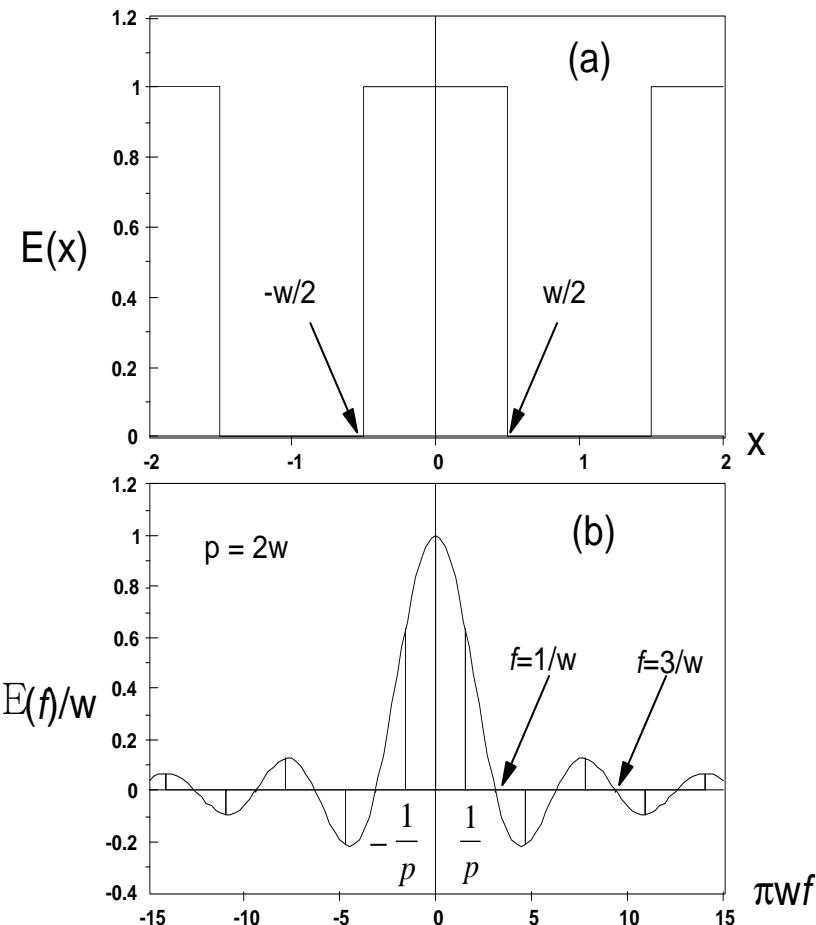


Figure 3.14 (a) Electric field of line-space pairs and (b) corresponding spatial frequency distribution.

When p is large with respect to λ , α is small, and the beam propagates paraxial to the z -axis. The angle increases with smaller p until it equals λ . When p exceeds λ , the beam is no longer a regularly traveling wave, but an evanescent wave instead. The angle θ for the lens NA is independent of p and λ . This angle plays the role of cutting off the high spatial frequencies. When $\sin\theta$ is less than $\sin\alpha$, the lens aperture blocks the spatial frequency λ/p .

Putting all of this together, when light is incident on a mask pattern, it is diffracted into a spatial frequency spectrum at node points $\pm n/p$. The frequencies lower than $1/\lambda$ travel toward the lens at angles $\alpha = \pm \sin^{-1}(\lambda/p)$. Those equal to and larger than $1/\lambda$ will not reach the lens. The lens numerical aperture $n \sin\theta$ further limits the spatial frequencies to $\pm \text{NA}/\lambda$. Note that the vacuum wavelength should be used, otherwise the spatial frequencies are $\pm \sin\theta_{\text{medium}}/\lambda_{\text{medium}}$.

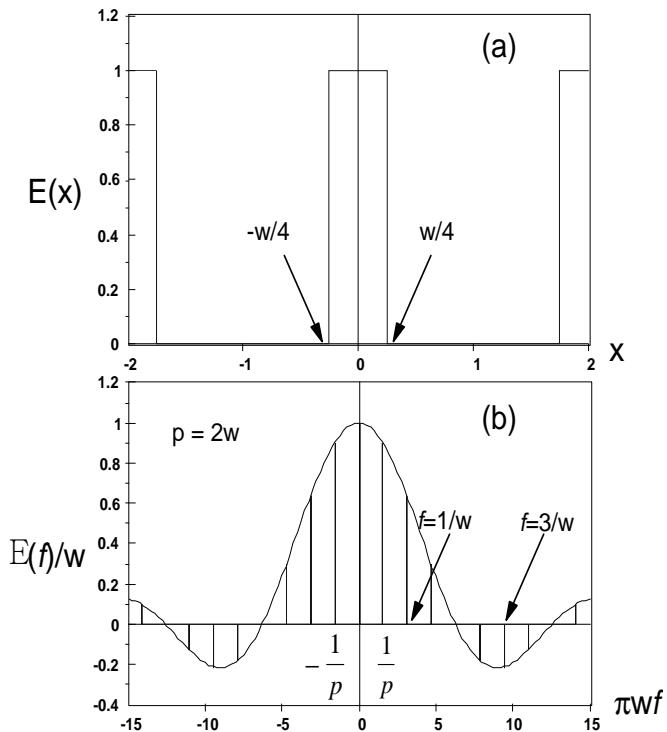


Figure 3.15 Same as Fig. 3.13(a) and (b), except line:space = 1:3.

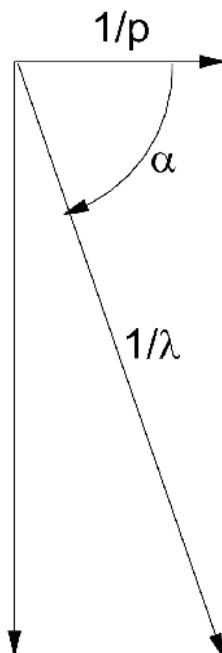


Figure 3.16 Relationship of propagation angle α to λ and p .

Analytically, the spatial distribution of the line-space pairs is

$$\begin{aligned} E(f) &= \sum_{n=-\infty}^{\infty} \int_{np-w/2}^{np+w/2} e^{i2\pi fx} dx \\ &= \frac{\sin(\pi wf)}{\pi wf} \sum_{n=-\infty}^{\infty} e^{i2\pi npf}. \end{aligned} \quad (3.64)$$

The first term is the envelope of the spikes. The summation part¹⁷ is a series of delta functions centering at $f = \pm np$. Using the sum of a geometric series,

$$\begin{aligned} \sum_{n=-\infty}^{\infty} e^{i2\pi npf} &= \lim_{n \rightarrow \infty} \frac{e^{-i2\pi pf} [e^{i(4n+2)\pi pf} - 1]}{e^{i2\pi pf} - 1} \\ &= \lim_{n \rightarrow \infty} \frac{\sin(2n+1)\pi pf}{\sin(\pi pf)} = \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{p}\right). \end{aligned} \quad (3.65)$$

Therefore,

$$E(f) = \frac{\sin(\pi wf)}{\pi wf} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{p}\right). \quad (3.66)$$

Just as in the case of an isolated line opening, one cannot afford the high NA to transmit higher-order spatial frequencies, because higher NA costs more. Most of the time, only the 0th and 1st orders pass through the lens, resulting in a pure sine-wave image plus a constant intensity bias.

Figure 3.17(a) shows the image intensity distribution from the line-space pairs shown in Fig. 3.14. The NA values 0.3, 0.6, and 0.9 are used. With NA = 0.3, there is no structure in the image. It is simply a constant background of intensity 0.25. The reason is obvious from Fig. 3.17(b) because the only spatial frequency accepted by the pupil I is the 0th-order frequency. When NA is increased to 0.6, the 1st-order spatial frequency is admitted, resulting in the sinusoidal image with an electric field that is a single sine wave superimposed with the average electric field amplitude of 0.5. Because the sum of the two 1st-order sine waves is larger than the 0th-order field, the total field at $x = 1$ is negative, resulting in the small positive peaks when Eq. (3.50) is applied to convert electric field to intensity. Further increasing NA to 0.9 does not change the intensity distribution because the cutoff frequency of the low-pass filter still excludes the 2nd-order frequencies.

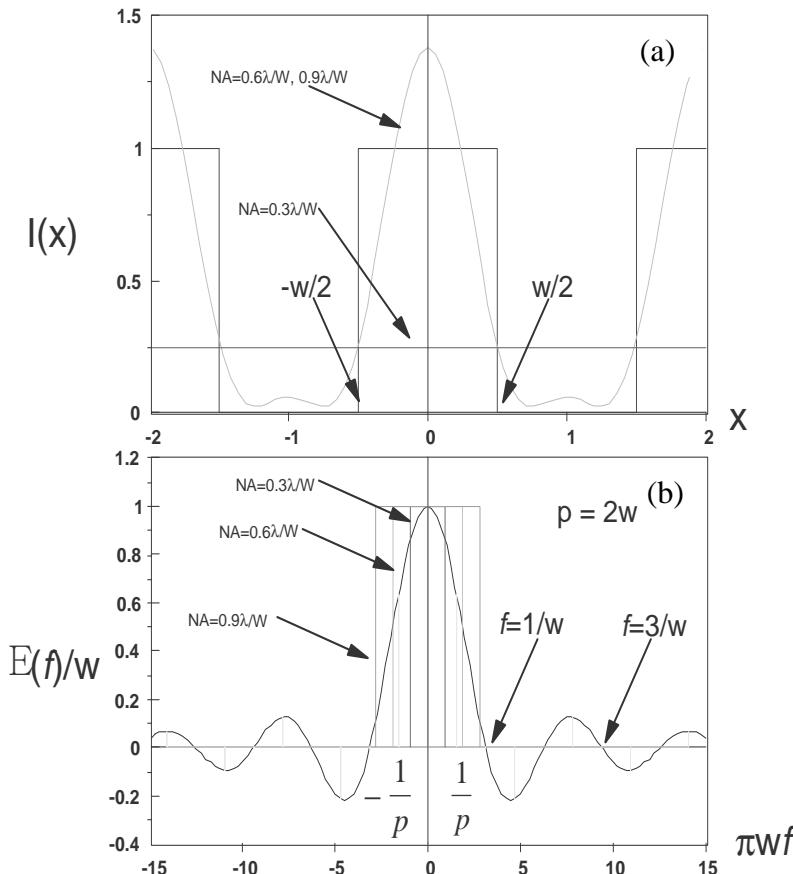


Figure 3.17 (a) Mask and image intensity distribution for 1:1 line-space pairs. (b) The corresponding spatial frequency distribution and the three different pupil functions.

When the transparent-to-opaque ratio is changed to 1:3 instead of 1:1, as shown in Fig. 3.18, the shape of the image intensity distribution is similar to that in Fig. 3.17, except that the magnitude is lower, resulting in an image of smaller openings than the 1:1 pattern at the same exposure level. This can be understood from the frequency spectrum shown in Fig. 3.18(b). Again, only the 0th- and 1st-order frequencies pass through the lens. The stronger 1st-order components bring down the combined intensity because they are negative.

With patterns much larger than the minimum feature size, higher-order spatial frequencies can be present. We now show 1:1 line-space pairs twice as large, equivalent to half the frequency of the line-space pairs shown earlier. As depicted in Fig. 3.19, NA = 0.3 suffices in collecting the 1st-order frequencies, NA = 0.6 collects the 1st- and 2nd-order frequencies. However, there is no 2nd-order component in the original object. Hence, 0.3 and 0.6 NA produce the same image. With NA = 0.9, the 3rd-order frequencies are admitted, resulting in a broader but double-peaked image.

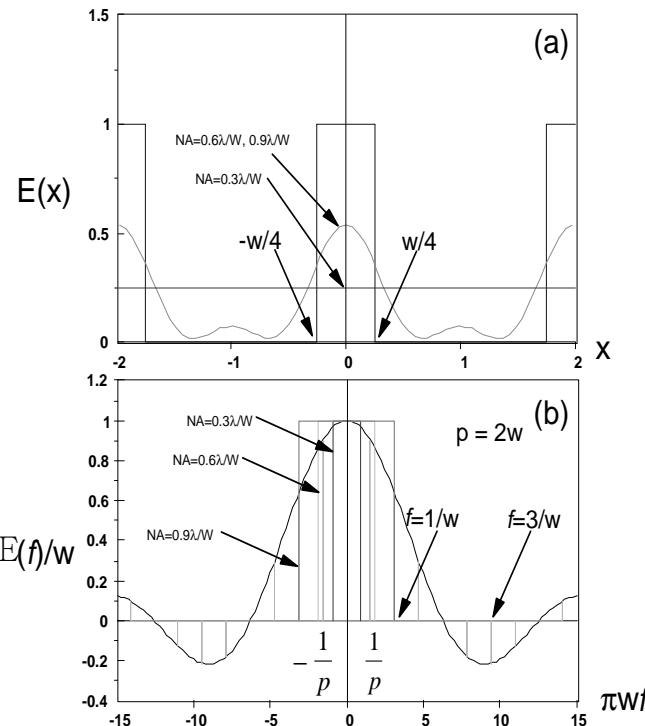


Figure 3.18 Same as Fig. 3.17(a) and (b), except line:space = 1:3.

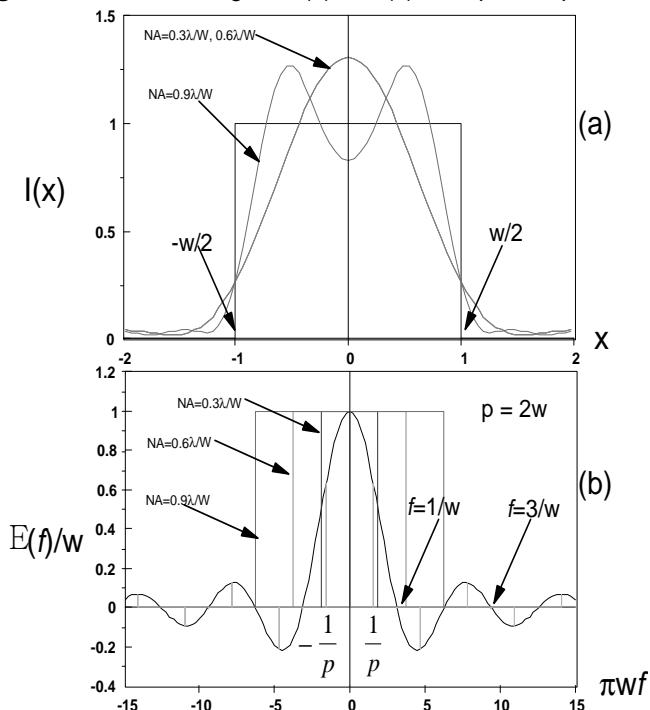


Figure 3.19 Same as Fig. 3.17(a) and (b), except $w = 2$ instead of 1.

3.1.6.3 Angular spectrum

Each spatial frequency f is associated with a pair of angular plane waves. The angles of these plane waves are

$$\theta = \pm \sin^{-1}(\lambda f), \quad (3.67)$$

where $\theta = 0$ is the direction of the optical axis. Using the situations in Secs. 3.1.6.1 and 3.1.6.2 above, when $f = 1/w$ and $w = \lambda$, the angles of the plane waves are $\pm\theta = \pi/2$.

3.1.7 Imaging results

Two examples are given to compare the results from Eqs. (3.46), (3.51), and (3.52). The first example uses a test mask pattern called BigMaC,¹⁸ which was designed for compact placement of a large number of feature combinations. There are minimum and large features with minimum and larger spaces, long and short features, inside and outside corners, and vertical and horizontal features. The minimum feature size is chosen to be $0.4\lambda/\text{NA}$ for this demonstration, as shown in Fig. 3.20.

Figure 3.21(a) shows the diffracted image from the BigMaC mask at the focal plane. The solid line depicts the constant intensity contour at a -0.6 log exposure. The other contour is evaluated at log intensity -0.5 . The distance between the two contours corresponds to a 26% exposure latitude. The dotted line traces a $\pm 10\%$ linewidth control bracket of the original mask pattern. The log intensity level of the two image contours is chosen by centering them to the $\pm 10\%$ brackets. With such small minimum features at $0.4\lambda/\text{NA}$, the diffracted

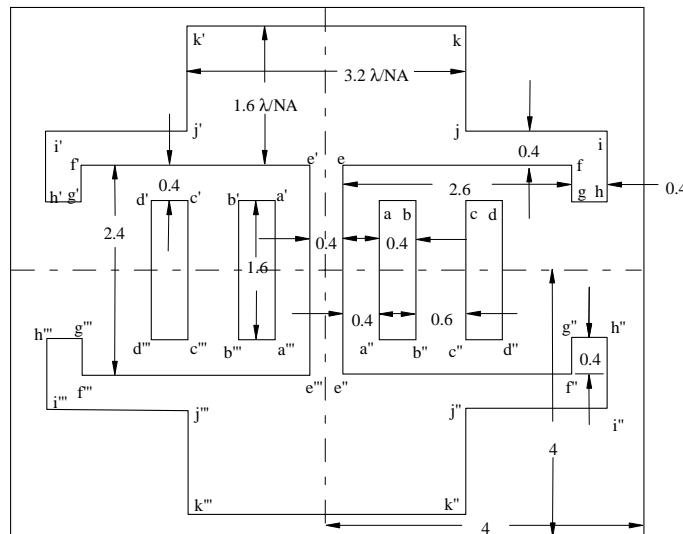


Figure 3.20 Big-small-feature and corner test mask (BigMaC).

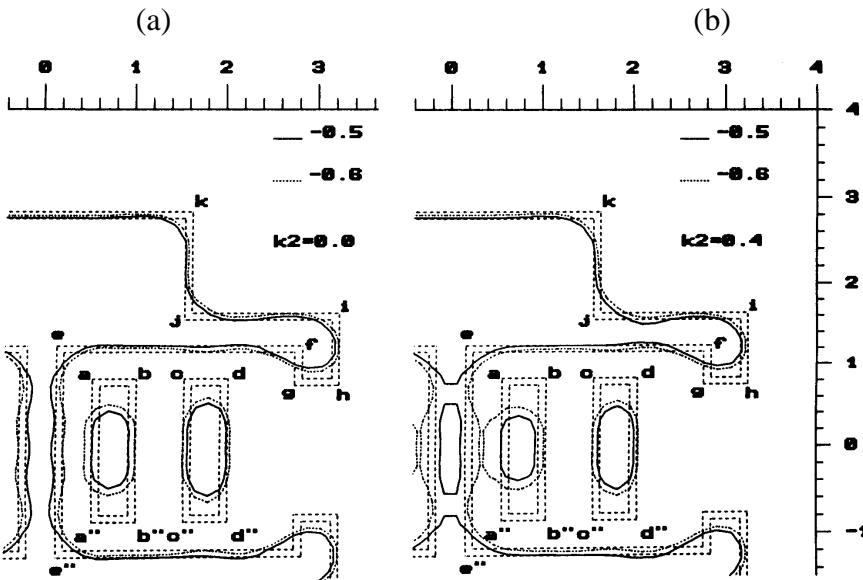


Figure 3.21 Diffracted image from the BigMaC mask at $\sigma = 0.4$: (a) defocus = 0, and (b) defocus = $0.4 \lambda/NA^2$.

image meets the linewidth-control criteria only at limited areas, for example, small parts of edge bb'' , cc'' , ef , hi , ij , jk , and kk' with their symmetrical parts in the other three quadrants. Edge ee'' does not meet the criteria. The lengths of $abb''a''$ and $cdd''c''$ are severely shortened. The latter shortens less because it is farther away from other features. Due to uneven edge misplacements, $abb''a''$ appears to have shifted to the left, whereas $cdd''c''$ shifts to the right. The bend $fghi$ is severely distorted. The centerline segment $ee'e''e''$ suffers from oscillation in edge position. When defocused to $0.4 \lambda/NA^2$, the image is even less useful. The image from a much lower illumination coherence corresponding to $\sigma = 0.8$ is shown in Fig. 3.22. The linewidth oscillation is perceptibly reduced. Optimizing the exposure to produce an acceptable section of ee'' and aa'' results in losing the other sections, especially the arm $efghij$. The situation at defocus = $0.4 \lambda/NA^2$ is similar, though more severe.

Another example uses two adjacent openings of 3:1 height-to-width ratio as depicted in Fig. 3.23.^{19,20} The features' size and distance is $0.987 \lambda/NA$, which is large by modern standards. Even with such large features, corner rounding is easily observed in Figs. 3.24 to 3.26. To reproduce sharp corners require many high-order spatial frequencies.

Incoherent diffraction from the two bars is shown in Fig. 3.24. The intensity at each data point is represented by a number ranging from 0 to 19, with a separation of 1.5 dB between adjacent numbers. Log intensity is used so that the log slope of the optical image is independent of the intensity level. Level 0 contains all intensities higher than or equal to 1, which is the normalized incident intensity. Similarly, level 19 contains all intensity levels equal to and higher than

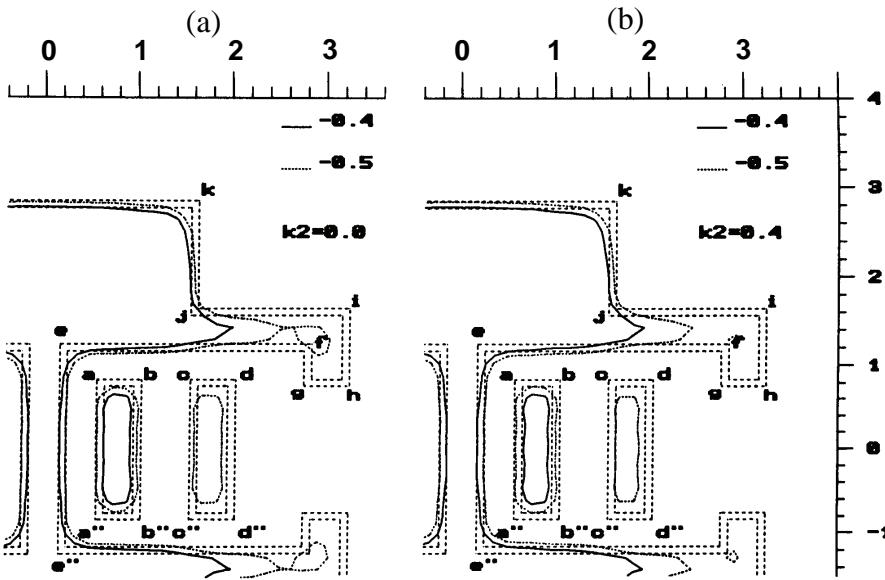


Figure 3.22 Same as Fig. 3.21, except $\sigma = 0.8$.

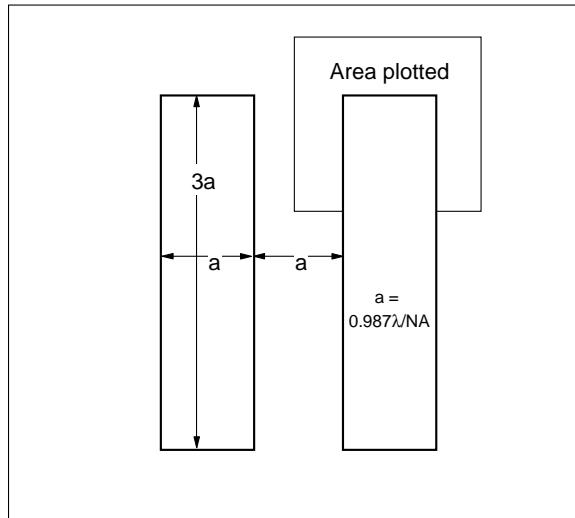


Figure 3.23 Mask pattern used for Figs. 3.24 to 3.26.

that level. The diffracted image in Fig. 3.24(a) exhibits good straight edges, except for the rounding due to loss of higher spatial frequencies. However, the separation of levels is wide, indicating poor linewidth control, because small exposure changes lead to large linewidth variation. The situation is worse at the defocal plane.

Switching over to the coherent image in Fig. 3.25, the image contrast is improved as indicated by the close spacing between exposure levels. However, line edges are wiggled, and there is a bright ghost line at the center. The waviness and influence of the ghost line is even worse with defocus.

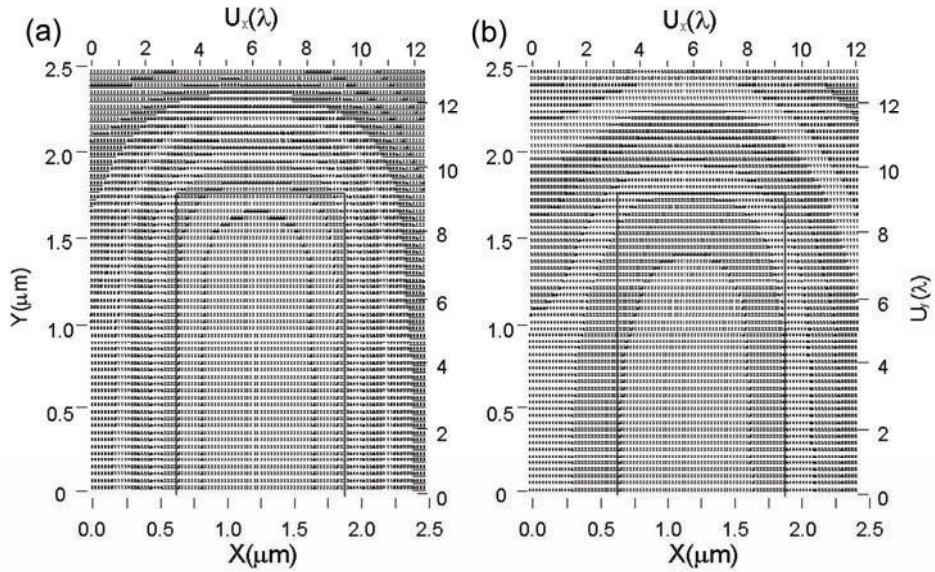


Figure 3.24 Diffracted image from the mask in Fig. 3.23. Illumination is completely incoherent, i.e., $\sigma = \infty$. (a) Defocus = 0, and (b) defocus = λ/NA^2 .

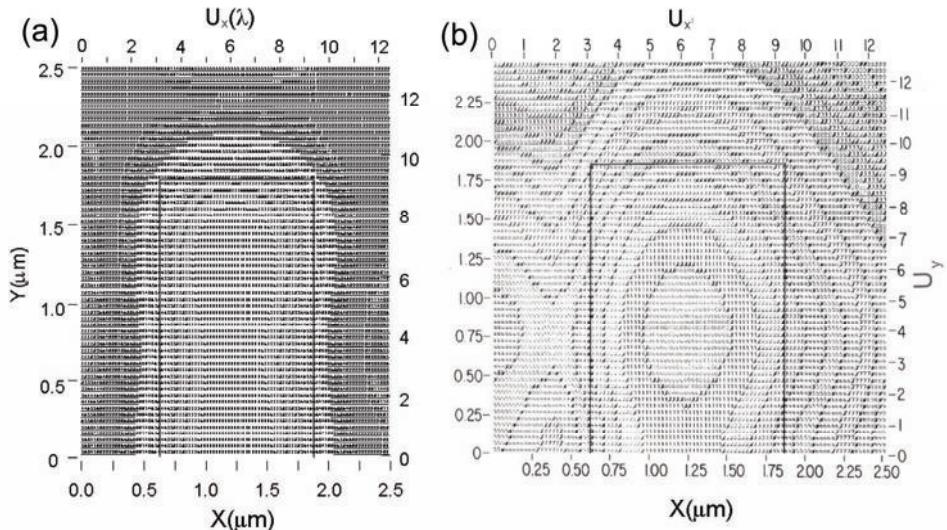


Figure 3.25 Same as Fig. 3.24, except illumination is completely coherent, i.e., $\sigma = 0$. (a) Defocus = 0, and (b) defocus = λ/NA^2 .

The partially coherent image shown in Fig. 3.26 makes a good compromise by eliminating the ghost line while retaining edge straightness and image contrast. This example shows some of the qualitative behavior of 3D diffracted images. Actual optimization of σ will be given in Sec. 7.3.

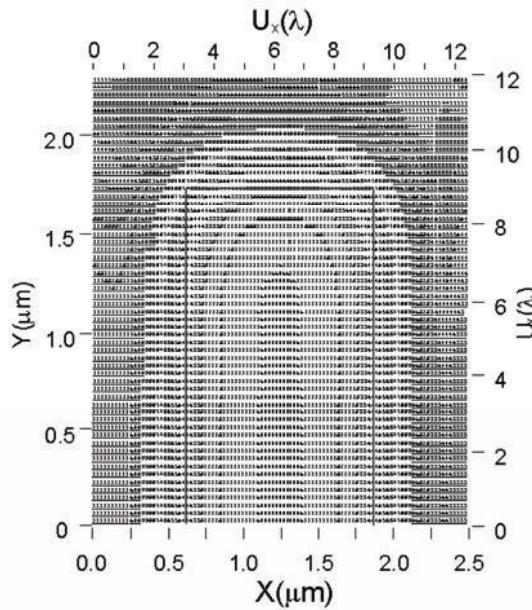


Figure 3.26 Same as Fig. 3.24, except partially coherent illumination $\sigma = 0.78$.

3.2 Reflected and Refracted Images

As soon as the aerial image enters into the resist film, it undergoes a series of changes. As depicted in Fig. 3.27, when the imaging wave reaches the air-resist interface, it refracts into the resist and reflects back to the air medium. To make the drawing easier to see, the image focus is set to the air-resist interface. In actual practice, the image should be focused into the photoresist. In any case, when the waves hit the resist-substrate interface, it refracts into the substrate and reflects into the resist. The refraction angle obeys the law of refraction:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \quad (3.68)$$

whereas the angle of reflection is equal to the angle of incidence:

$$\theta_{\text{incidence}} = \theta_{\text{reflection}}. \quad (3.69)$$

When the photoresist medium absorbs light, the refractive index is a complex number with an imaginary part that is the absorptive component. The angles of refraction and reflection still follow Eqs. (3.68) and (3.69). However, a phase shift at the interface is induced, accompanying the change of the amplitude in the refracted and reflected waves. As soon as the wave leaves the refracting surface, it is absorbed according to the attenuation coefficient α . Let the electric field of the incident wave be

$$E_i(x, z) = e^{i \frac{2n_o \pi}{\lambda} (z \cos \theta_i + x \sin \theta_i)}. \quad (3.70)$$

The reflected field is then

$$E_r(x, z) = C_r e^{i \frac{2n_o \pi}{\lambda} (z \cos \theta_r - x \sin \theta_r)}, \quad (3.71)$$

where the reflection coefficient C_r is a complex number dictating amplitude and phase change at the air-resist interface, and n_o is the refractive index of air. The electric field from points a to b is

$$E_{ab}(x, z) = C_{ab} e^{i \frac{2n_{resist} \pi}{\lambda} (x \cos \theta_{ab} + z \sin \theta_{ab})}, \quad (3.72)$$

where the refraction coefficient C_{ab} is also a complex number dictating amplitude and phase change at the air-resist interface, and n_{resist} is the complex refractive index of the photoresist:

$$n_{resist} = n_{resist, real} + i n_{resist, imaginary}. \quad (3.73)$$

Multiple reflections and refractions at the two interfaces take place until attenuation makes the remaining waves negligible.

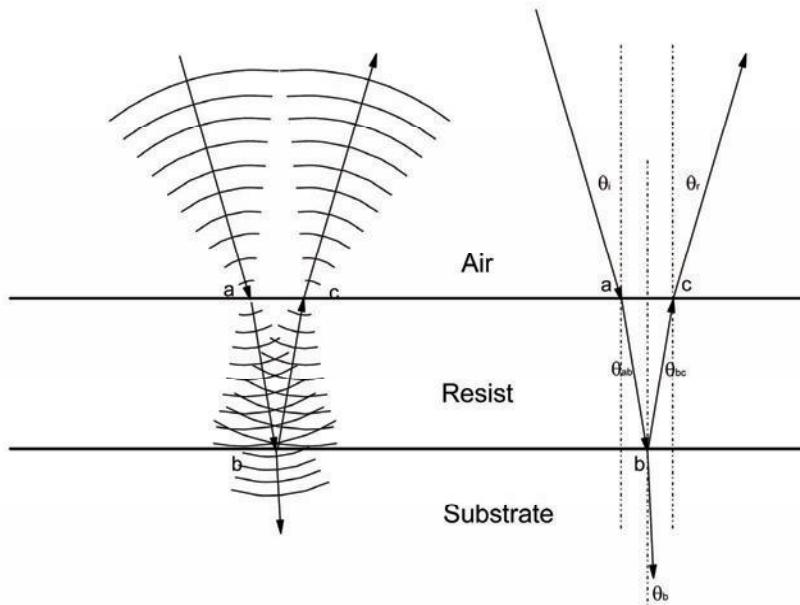


Figure 3.27 Refracted and reflected waves.

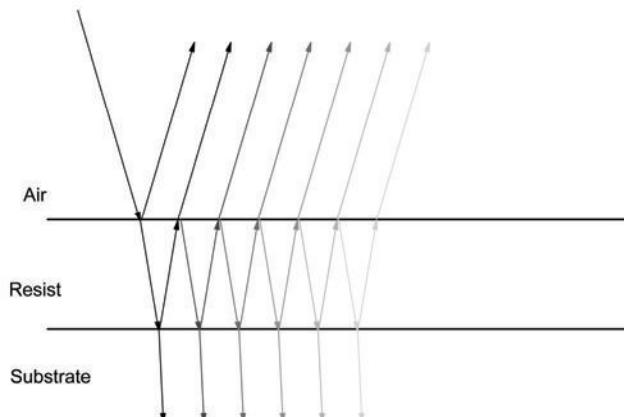


Figure 3.28 Multiple reflected waves in an attenuating media.

3.2.1 Methods to evaluate the reflected and refracted image from a mask

There are two methods to evaluate the reflected and refracted image from a given mask pattern. The first method traces the waves through all of the multiple reflecting and refracting surfaces shown in Fig. 3.28, following Eqs. (3.68) to (3.73). For a given mask pattern, the spatial frequency components are evaluated, then the higher spatial frequencies rejected by the lens's NA are removed. Each of the remaining frequency components is treated as a plane incident wave of angle θ_i to follow through all of the reflections and refractions. Multiple reflections and refractions of plane waves through thin-film stacks are well treated in the literature. One simply applies these equations to each spatial frequency component, then recombines them with partial coherence, as was discussed in Sec. 3.1.5.3. When the resist bleaches or dyes (i.e., its complex refractive index is a function of exposure), the image is evaluated at the initial exposure, then the refractive indices are adjusted for the local exposure, and the reflected and refracted image is evaluated again until the change in refractive index stabilizes. This approach is practiced by the UC Berkeley Electronic Research Laboratory,^{21,22} Mack,^{23,24} and many other researchers in this field. The problem with this approach is that when the thin-film stack consists of topography such as lines running over polysilicon gates, contact holes, metal lines, etc., the reflections and refractions must be traced at each interface, making it extremely tedious.

The most elegant approach to evaluating the reflected and refracted image in the presence of topography and the thin-film stack is to solve Maxwell's equations,²⁵ with boundary conditions given by the topography and the thin films. One can decompose the mask pattern into an angular spectrum of plane waves, then solve Maxwell's equations exactly for each plane-wave component before partial coherent recombination. Going one step further, boundary conditions can be set for the mask opening, thus removing the approximation in Eq. (3.49). Interested readers may consult the publications of Yeung^{15,26,27} and Barouch.^{28,29}

3.2.2 Impact of multiple reflections on DOF

To understand the impact of multiple reflections on DOF, it is best to visualize the image changes caused by multiple reflections. Figure 3.29 shows the electric-field distribution at the interfaces from a $k_1 = 0.35$ line opening illuminated with $\sigma = 0.5$ at a vertical incidence. The beams are vertical but are spread out with a finite angle for the purpose of illustration. Hence, $\cos\theta$ is omitted by assuming a small angle of incidence. Otherwise, the optical distance na would become $n a \cos\theta$. At $z = a$, the air-resist interface where the image is focused, the electric-field distribution is $E(x,0)$ and the reflected image is $R_{ads}E(x,a)$, where R_{ads} is the reflection coefficient of the first interface, a stands for the location of z , d is the downwards direction of the incident light, and s is the state of polarization. Similarly, the transmitted electric-field distribution is $T_{ads}E(x,a)$, where T_{ads} is the transmission coefficient at the air-resist interface. At the resist-wafer interface, the reflected electric-field distribution is $R_{bds}T_{ads}E[x,a + (n + i\alpha)(b - a)]$, where R_{bds} is the reflection coefficient of the resist-substrate interface, α is the absorption coefficient of the multiple-reflecting media, and n is its refractive index. The transmitted electric-field distribution is $T_{aus}R_{bds}T_{ads}E[x,a + 2(n + i\alpha)(b - a)]$. The electric-field distributions at the other points can be expressed similarly by multiplying the electric-field distribution at a reference point with the appropriate reflection and transmission coefficients and adjusting the defocus distance in $E(x,z)$, according to the refractive index and absorption coefficient of the media in which light has traveled.

In Fig. 3.29, 193-nm light is used. The thickness of the resist is 288 nm. The n and α of the resist are taken to be 1.69 and 0.012, respectively, and that of the silicon wafer is 0.863 and 2.747. Hence, $r_{ads} = -0.257 - i0.003$, $T_{ads} = 0.743 - i0.003$, $R_{bds} = -0.385 - i0.656$, $T_{bds} = 0.615 - i0.656$, and $T_{aus} = 1.257 + i0.003$. Note that the outgoing electric-field distribution at the second bounce is larger and wider than that at the first bounce. It is wider because the field distribution

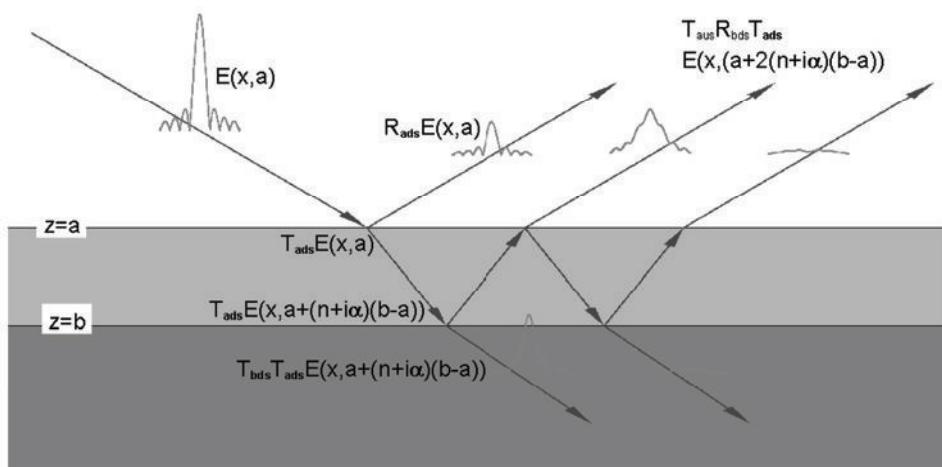


Figure 3.29 Electric-field distribution at interfaces.

spreads out as the defocus distance increases, just as depicted in Figs. 3.33 and 3.34. It is larger because $T_{aus}R_{bds}T_{ads}$ is larger than R_{ads} .

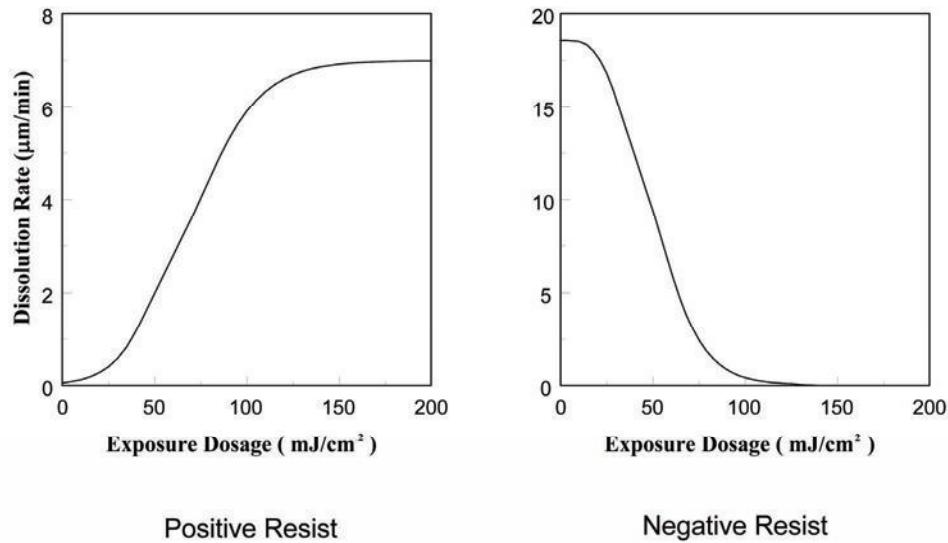
Now, the resultant image is definitely of lower contrast at the best focus position when compared with the image without multiple reflections. Intuition suggests that a higher contrast image may lead to a larger DOF. This is obviously not true when Eq. (3.18) is considered. When the enhancement of contrast is produced by a higher NA, it leads to a smaller DOF. In the situation of multiple reflections, the DOF is also improved, despite a lower image contrast at the best focal point. The situation is similar to the FLEX system,³⁰ in which the superposition of defocused images is achieved by multiple exposures with the wafer purposely offset at different stages of defocus. The averaging effect of these images takes the more-than-sufficient contrast in the focal plane to help the low contrast in the defocal planes. Though they both superimpose images at many defocus planes, multiple reflections are different from FLEX because the former superimposes coherently, whereas the latter superimposes incoherently. Multiple reflections should not be used to extend the DOF. In fact, they should be avoided at all costs because they are the strongest contributor of exposure nonuniformity leading to linewidth-control problems over topography, as will be discussed in Sec. 5.6.2.5.

3.3 The Latent Image

The latent image is the radiation-induced change in the resist before it is developed into the resist image. A simple visualization is to directly relate the complex refractive index distribution $n(x,z)$ in the resist layer to the intensity distribution of the combined reflected and refracted images in the resist layer. The actual combination of these images may be more complicated if the refractive index change is a function of light intensity. One must combine them dynamically in time, as described above in Sec. 3.2.1.

The latent image is not readily visible, since it mostly consists of refractive-index distribution instead of absorption distribution. For some resists with significant complex index change, the latent image can indeed be seen. For example, i- or g-line resists produce a visible latent image when drastically overexposed. Contrast enhancement resists, by definition, emphasize the production of a distinct latent image. Its latent image is readily visible.

However, it is desirable to view the latent image because the image offers an opportunity to see the aerial image captured in the resist media before it is changed into the developed resist image. The latent image is independent of many resist-processing characteristics, and thus can settle doubts on whether the imaging tool or the resist is at fault when things go wrong. Being able to measure it before development facilitates an *a priori* determination of the developing condition and the development end point to control the final developed image. Many attempts have been made to make the latent image visible. It is possible, but the image usually does not offer sufficient contrast to be useful.



Positive Resist

Negative Resist

Figure 3.30 Resist dissolution as a function of exposure.

3.4 The Resist Image

Resist exposure leads to modification of its physical characteristics. Not only is its complex refractive index changed, the dissolution rate of the resist in the developer is also modified according to the exposing image. The dissolution rate distribution $D(x,z)$ is directly related to $n(x,z)$ through an exposure-dissolution relationship, as shown in Fig. 3.30.

With a positive resist, higher exposure leads to a faster dissolution rate, and vice versa for a negative resist. Given $D(x,z)$, the resist image $R(x,z,z_{resist})$ can be evaluated. However, evaluating this image analytically is not trivial because removal of the resist is related to the resist surface opened for dissolution by the developer.

To further visualize the resist-development process, the sequence of developing an idealized dissolution-rate image is depicted in Fig. 3.31. The aerial image $I(x,y,z)$, through the depth of the resist, is assumed to be completely zero in the opaque area of the mask and unity in the transparent area.

$$I(x,z) = 1, \text{ when } x \geq -a \text{ and } x \leq a = 0, \text{ elsewhere.} \quad (3.74)$$

Multiple reflections are neglected by assuming a perfect antireflection coating at the air-resist and resist-wafer surfaces. Resist absorption and diffraction effects are also neglected. Therefore, the latent image is also a simple square function—zero in the opaque area and a constant value in the exposed area. The development rate $R(x,y,z)$ becomes

$$R(x,z) = 10 \text{ nm/sec, when } x \geq -a \text{ and } x \leq a = 1 \text{ nm/sec, elsewhere,} \quad (3.75)$$

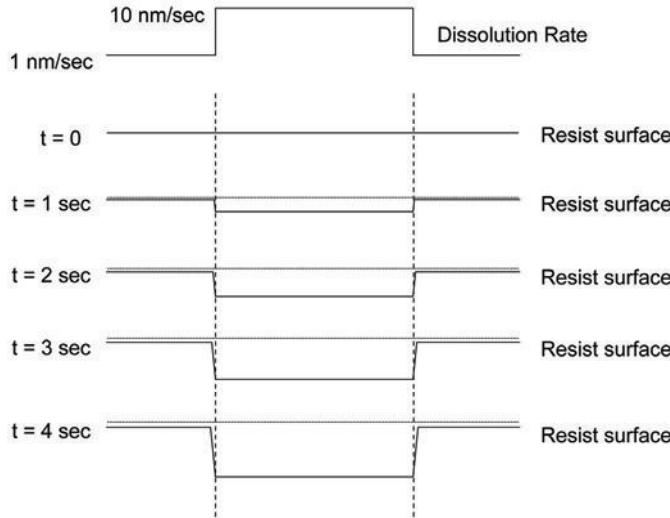


Figure 3.31 Development sequence of an idealized dissolution-rate image.

where there is a finite unexposed development rate, and the resist is kept at a low contrast for illustration purposes. Figure 3.31 shows that because of the finite dissolution rate in the unexposed area, the resist not only loses thickness while being developed, it is also developed sidewise as soon as the vertical developer-to-resist interface is opened up. The developed resist image has a natural overcut profile even if the development rate curve is a perfect square function. This holds for either a positive or negative resist. When optical absorption is taken into account while ignoring the surface effects, the development rate decreases toward the bottom of the positive resist and vice versa for the negative one. The overcut profile is more exaggerated for positive resists, while there is chance for compensation for negative resists, as shown in Fig. 3.32.

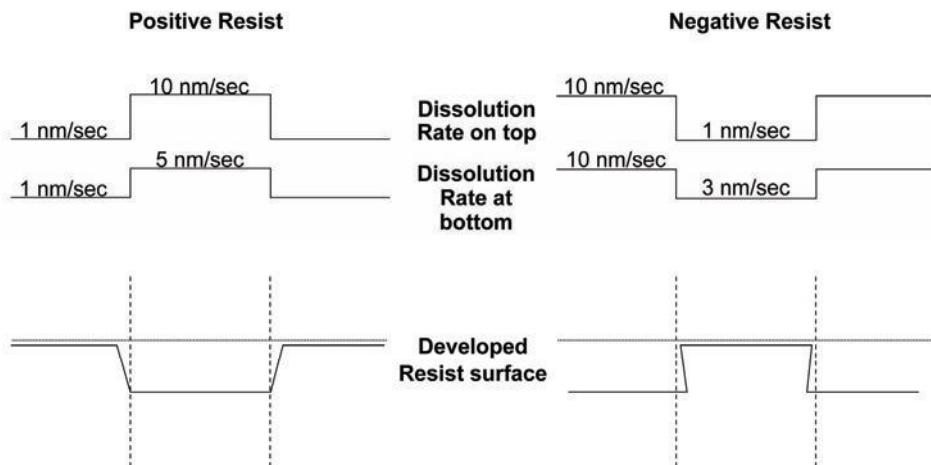


Figure 3.32 Absorptive positive and negative resist images.

Instead of artificial aerial images, we now use the diffracted aerial image from a $k_1 = 0.64$ line opening illuminated with $\sigma = 0.5$ to show the development process depicted in Fig. 3.33. The diffracted image is refracted in the resist media as shown. The diffracted image propagates in the air medium until it hits the air-resist interface at three positions. The top position shows the resist-air interface at $-1.8 k_2$ from the focal point, the middle position is $-0.5 k_2$, and the bottom position is $1 k_2$. As soon as it reaches each air-resist interface, the propagation constant switches to the n and α of the resist; this is what is depicted in Fig. 3.33. The positive-resist dissolution rate curve in Fig. 3.30 is used for all three positions. Each family of the developed image curves represent change of the resist image as a function of development time. The normalized unit of dimension k_1 in the lateral direction and k_2 in the longitudinal dimension are later defined in Sec. 4.4. Figure 3.33 shows that the developed resist image is a strong function of the refracted image. It is predominantly influenced by the refracted image on top, which is the first curve in each family of the refracted image. Also, the defocus resist images are not symmetrical, even though the refracted aerial images are symmetrical with respect to the focal plane marked in the figure. The n and α values in Fig. 3.33(a) are typical of an i-line resist. The developed images look reasonable. The α value in Fig. 3.33(b) is $3 \mu\text{m}^{-1}$. The corresponding developed image is extremely dependent on the refracted image at the resist surface. This is a desirable α value for top-surface imaging systems such as DESIRE.³¹ On the other hand, the absorption is too high for a conventional resist system with resist thickness in the micrometer regime. Note that this is a typical level of absorption for 157-nm resist systems.³² When the resist thickness drops to 100 nm or below, this α value is about right to attain a 30% absorption in the resist layer.

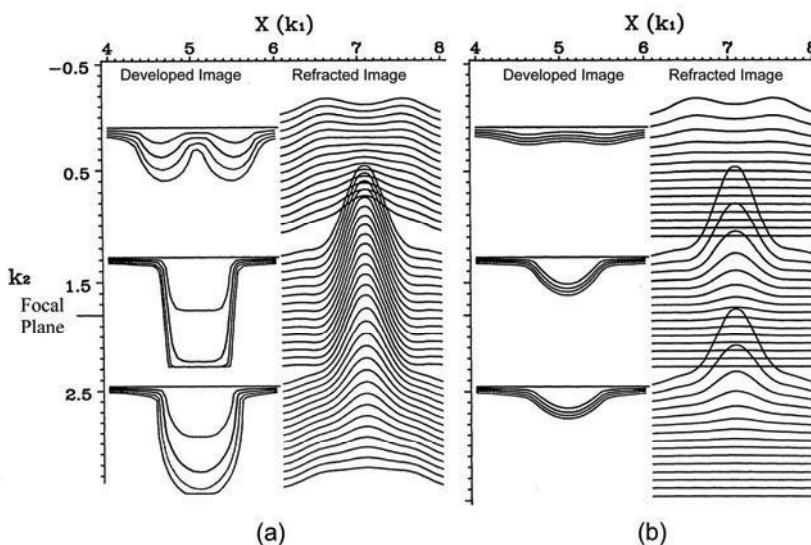


Figure 3.33 Development of resist (a) $n_r = 1.5$, $\alpha = 0.178 \mu\text{m}^{-1}$, and (b) $n_r = 1.5$, $\alpha = 3 \mu\text{m}^{-1}$.

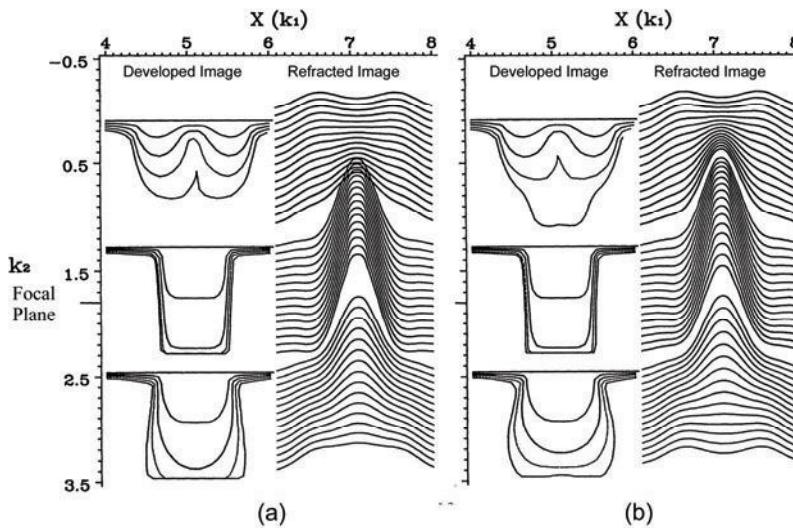


Figure 3.34 Development of resist (a) $n_r = 1.5, \alpha = 0$, and (b) $n_r = 1, \alpha = 0$.

When resist absorption is reduced to 0, as shown in Fig. 3.34, the developed image in focus has a slightly more vertical profile than that in Fig. 3.33(a). However, the developed image at $1.0 k_2$ is too wide because the undamped image is wider at the same threshold. Hence, having a completely absorption-free resist is not necessarily the most desirable, even if we could ignore the energy the resist needs to absorb for chemical reactions. Fig. 3.33(b) shows the case with $n = 0$ and $\alpha = 0$. This resist consumes more DOF; therefore, at defocus = $-1.8 k_2$ (in the refracted image) the refracted image at the resist bottom already becomes single peaked, whereas in Fig. 3.33(a) the refracted image is still double peaked. However, this situation is not bad. The developed resist image is closer to the focused image. The resist image only becomes worse at defocus = $1.0 k_2$, where the refracted image at the bottom of the resist is farther away from focus. Because the developed image is more dependent on the refracted image on top, it is not much more severe than the corresponding image in Fig. 3.33(a). Naturally, when defocus is further out, the difference will be more significant.

We just demonstrated how to simulate the developed image from the refracted image through the rate image. This simulation is possible, but very time consuming. The most accurate model is to divide the resist space into cells and develop the resist cell by cell, according to the development rate of each cell. The development is not only dependent on the amount of exposure the cell received, but it is also a function of the number of facets in contact with the developer.

3.4.1 The A, B, C coefficients

The A , B , C coefficients were developed by Dill et al.³³ As shown later in Eq. (3.123), the complex index of refraction of the resist n_{cr} is not merely a function of space, although this would be true if the resist stayed unchanged when receiving the diffracted light from the mask pattern through the lens. For a real

photoresist, n_{cr} is a function of I_r . Therefore, the resist is bleached or dyed during exposure. On the other hand, I_r is always a function of n_{cr} regardless of bleaching or dyeing. Hence, I_r and n_{cr} are dynamically linked to each other in a loop. Let α be the absorption coefficient of the resist media. Then

$$\alpha = \frac{2\pi}{\lambda} n_i . \quad (3.76)$$

Lambert's law dictates that

$$\frac{dI(z)}{dz} = -\alpha I(z) \quad (3.77)$$

for light propagating in the z -direction. This takes care of the change of intensity as a function of α along the path of the light being absorbed. There must be another equation to govern the change of α as a function of intensity. Let us assume that α is composed of two parts, a constant term and a photoactive term, such that

$$\alpha = Am + B . \quad (3.78)$$

B is independent of the intensity of the incident light, whereas m is dependent on the incident intensity and is the concentration M of the photosensitive component of the resist, normalized to the initial concentration M_0 before exposure. A is the coefficient of m . Then the rate of change of m is

$$\frac{dm}{dt} = -Cm(t)I . \quad (3.79)$$

Now that I and α are dependent on each other, they both must be functions of space and time, as is m . Equations (3.77) and (3.79), respectively, become

$$\nabla I(x, y, z, t) = -I(x, y, z, t)[Am(x, y, z, t) + B] \quad (3.80)$$

and

$$\frac{\partial m(x, y, z, t)}{\partial t} = -Cm(x, y, z, t)I(x, y, z, t) , \quad (3.81)$$

where

$$\nabla I \equiv \frac{\partial I}{\partial x} + \frac{\partial I}{\partial y} + \frac{\partial I}{\partial z} \quad (3.82)$$

is the intensity gradient. Equations (3.80) and (3.81) are expressed in generalized coordinates. The original Dill equations treated I as propagating in the z -direction only. Once generalized, it is not trivial to derive an analytic solution. However, most simulation programs use these equations, either simplified or generalized, to

evaluate I and m numerically. To solve these two simultaneous equations, one needs to know the three coefficients (A , B , C) as well as the initial conditions [$I(x,y,z,t_0)$ and $m(x,y,z,t_0)$]. A and B have the dimension of μm^{-1} and are known as the bleachable and nonbleachable absorption coefficients, respectively. C has the dimension of cm^2/mJ and is known as the exposure rate constant. $I(x,y,z,t_0)$ is usually referred to as the aerial image, that is, the diffracted light from the mask openings through the imaging lens before the light passes through the photoresist. The normalized concentration m of the photosensitive component in the resist is generally homogeneous before patterned exposure.

For a given resist system, the A , B , C coefficients can be evaluated experimentally. The resist is coated on an optically matched substrate that has the same refractive index as the resist. Hence, light propagates without reflections and travels into the resist only once. The reflection at the air-to-resist interface can be calibrated by taking a reading with a blank index-matched substrate. The absorption coefficient is obtained by taking the ratio of the intensity of the incoming light and that of the outgoing light at the actinic wavelength. A spectrophotometer is suitable.

$$\alpha = \frac{1}{d} \ln \left(\frac{I_{in}}{I_{out}} \right), \quad (3.83)$$

where d is the resist thickness. Then

$$A = \alpha_{unexposed} + \alpha_{exposed} \quad (3.84)$$

and

$$B = \alpha_{exposed}. \quad (3.85)$$

Evaluating C is more complicated. Dill et al. provided the following formula:³⁴

$$C = \left\{ \frac{A+B}{AI_0 T(0)[1-T(0)]} \right\} \frac{dT(0)}{dt}, \quad (3.86)$$

where

$$T(0) = \left(\frac{I_{out}}{I_{in}} \right)_{t=0}. \quad (3.87)$$

They also proposed an iteration scheme to determine A , B , and C more accurately by performing a mean square deviation fit to the experimental transmission curve $T(t)$.

The A , B , C coefficients are good parameters to characterize the photoresist and to simulate the resist image. They are readily provided with any commercially available resist system. The concern is in their accuracy. They are definitely a function of processing conditions. For example, a high-temperature, pre-exposure bake may remove some photoactive compound and make A and C smaller. Any baking may change the refractive index of the resist, affecting both A and B . Even if the resist and its processing are maintained very strictly, it is not trivial to perfectly match the indices of the resist and the substrate, especially when the index of the resist changes as a function of exposure. Some work must be done to accommodate the dynamic difference in the indices.

3.4.2 The lumped parameters

The A , B , C coefficients are useful for characterizing a given resist system and simulating the resist image. Once they are determined, one goes through the formulations suggested by Eqs. (3.122) through (3.126) to evaluate the final resist image. The process is physically sound, but tedious and time consuming. In most cases, once the aerial image is obtained, the resist image can be estimated with the lumped parameters, which consist of the resist thickness D , the absorption coefficient of the resist α , and its development contrast γ .

The lumped parameter model (LPM) was initialized by C. A. Mack³⁵ et al. in 1986. Later, he extended the model^{36,38} to the enhanced lumped parameter model to include the effect of resist absorption by assuming a segmented development path in vertical and horizontal directions. In 1996, T. A. Brunner et al.,³⁷ assuming an exponential form of the aerial image near the threshold, obtained analytic expressions for linewidth and sidewall angle of a resist feature for a nonabsorptive resist by analogy to ray tracing in geometrical optics. However, if resist absorption is taken into account, the development path is also approximated by a segmented path, with an additional segment that is diagonal. Here, we show that in such a case analytic expressions can also be obtained by straightforward derivation. Before this, we briefly summarize Mack's theory. The coordinate system is defined such that the z -axis points downward, thus perpendicular to the resist top surface.

Mack's LPM makes the following assumptions:

- The aerial image is separable in its x - and z -dependence [i.e., $I(x,z) = I_x(x)I_z(z)$]. Then, the exposure energy at any point of the resist $E(x,z) = EI_x(x)I_z(z)$, where $I_x(\infty) = 1$ and $I_z(0) = 1$.
- Assuming a constant resist contrast γ , the resist local development rate is

$$r(x,z) = r_0 \left[\frac{E(x,z)}{E_0} \right]^\gamma, \quad (3.88)$$

where E_0 is the energy to clear in the large clear area.

- Assume a segmented development path, which starts at $(x_0, 0)$ and proceeds vertically until $E(x_0, z + \Delta z) < E(x_0 + \Delta x, z)$, then proceeds horizontally to (x, z) . The vertical development time from $(x_0, 0)$ to (x_0, z) is

$$T_z = \int_0^z \frac{dz'}{r(x_0, z')} = r_0^{-1} E_0^\gamma E^{-\gamma} \left[I_x(x_0) \right]^{-\gamma} \int_0^z \left[I_z(z') \right]^{-\gamma} dz'. \quad (3.89)$$

Similarly, the horizontal development time from (x_0, z) to (x, z) is

$$T_x = \int_{x_0}^x \frac{dx'}{r(x', z)} = r_0^{-1} E_0^\gamma E^{-\gamma} \left[I_z(z) \right]^{-\gamma} \int_{x_0}^x \left[I_x(x') \right]^{-\gamma} dx'. \quad (3.90)$$

Therefore, the development time from (x_0, z) to (x, z) is

$$T = T_z + T_x. \quad (3.91)$$

Since $x = 0$ is at the center of the mask feature, the x -coordinate of the final development point is in fact half of the developed resist linewidth (i.e., $x = w/2$). So, Eq. (3.91) can be regarded as an implicit function $E(x)$. After rearranging, we obtain,

$$\begin{aligned} E^\gamma &= r_0^{-1} t^{-1} E_0^\gamma \left[I_x(x_0) \right]^{-\gamma} \int_0^z \left[I_z(z') \right]^{-\gamma} dz' + r_0^{-1} t^{-1} E_0^\gamma \\ &\quad \times \left[I_z(z) \right]^{-\gamma} \int_{x_0}^x \left[I_x(x') \right]^{-\gamma} dx'. \end{aligned} \quad (3.92)$$

To derive the change of linewidth resulting from the change of exposure energy, we differentiate Eq. (3.92) with respect to E , resulting in

$$\frac{dE(x)}{dx} = \frac{E(x)}{\gamma D'} \left[\frac{E(0) I_x(0)}{E(x) I_x(x)} \right]^\gamma, \quad (3.93)$$

where we define $D' = r_0 [I_z(D)]^\gamma T$ and relate energy to clear in the large clear area E_0 to the exposure energy leading to zero linewidth $E(0)$ via $E_0 = E(0) I_x(0)$. Integrating on both sides, we obtain finally

$$\frac{E(x)}{E(0)} = \left\{ 1 + \frac{1}{D'} \int_0^x \left[\frac{I_x(x')}{I_x(0)} \right]^{-\gamma} dx' \right\}^{\frac{1}{\gamma}}. \quad (3.94)$$

- If the intensity of the aerial image in the neighborhood of threshold is further approximated by

$$I(x, z) = I(x_0) e^{s(x-x_0)} e^{-\alpha z}, \quad (3.95)$$

where s is the image log-slope and α is the absorption coefficient, then the analytic expression for resist linewidth can be derived, as shown below. One defines

$$n(x, z) \equiv \frac{1}{r(x, z)} = r_0^{-1} e^{-\gamma s(x-x_0)} e^{-\gamma \alpha z}, \quad (3.96)$$

where we have used $EI(x_0) = E_0$. Then, $\nabla n = n\gamma(-s\mathbf{x} + \alpha\mathbf{z})$. We now define a new coordinate system expanded by x_1 and z_1 . The new orthonormal basis vectors are related to the old ones by

$$\begin{aligned} x - x_0 &= x_1 \cos \phi + z_1 \sin \phi \\ z &= -x_1 \sin \phi + z_1 \cos \phi. \end{aligned} \quad (3.97)$$

In the new coordinate system,

$$n(x_1, z_1) = r_0^{-1} e^{-\gamma \sqrt{s^2 + \alpha^2} x_1}, \quad (3.98)$$

which only depends on x_1 . It can be shown that $n \cos \theta_1$ is a constant, where θ_1 is the angle that the tangent vector of the development path makes with z_1 . When $t = 0$, $x_1 = 0$, $z_1 = 0$, and

$$\cos \theta_1 = \cos \phi = \frac{s}{\sqrt{s^2 + \alpha^2}},$$

we have

$$n \cos \theta_1 = r_0^{-1} \left(\frac{s}{\sqrt{s^2 + \alpha^2}} \right).$$

That is,

$$\cos \theta_1 = \frac{s}{\sqrt{s^2 + \alpha^2}} e^{\gamma \sqrt{s^2 + \alpha^2} x_1}. \quad (3.99)$$

The differential of the development path in the x_1 direction is

$$\Delta x_1 = -r \sin \theta_1 \Delta t = -r_0 e^{-\gamma \sqrt{s^2 + \alpha^2} x_1} \sin \theta_1 \Delta t. \quad (3.100)$$

Integrating both sides of the above equation, we obtain

$$x_1(t) = \frac{-1}{2\gamma\sqrt{s^2 + \alpha^2}} \ln \left[\frac{s^2}{s^2 + \alpha^2} + \left(\gamma\sqrt{s^2 + \alpha^2} r_0 t \right)^2 \right]. \quad (3.101)$$

The differential of the development path in the \mathbf{z}_1 direction is

$$\Delta z_1 = r \cos \theta_1 \Delta t = r_0 e^{\gamma \sqrt{s^2 + \alpha^2} x_1} \cos \theta_1 \Delta t. \quad (3.102)$$

Integrating on both sides, we obtain

$$z_1(t) = \frac{1}{\gamma\sqrt{s^2 + \alpha^2}} \tan^{-1} \left[\frac{\gamma(s^2 + \alpha^2)}{s} r_0 t \right]. \quad (3.103)$$

Transforming back to the old coordinate system, we obtain finally

$$\begin{aligned} x(t) &= x_0 - \frac{s}{2\gamma(s^2 + \alpha^2)} \ln \left[\frac{s^2}{s^2 + \alpha^2} + \left(\gamma\sqrt{s^2 + \alpha^2} r_0 t \right)^2 \right] \\ &\quad + \frac{\alpha}{\gamma(s^2 + \alpha^2)} \tan^{-1} \left[\frac{\gamma(s^2 + \alpha^2)}{s} r_0 t \right] \end{aligned} \quad (3.104)$$

and

$$\begin{aligned} z(t) &= \frac{\alpha}{2\gamma(s^2 + \alpha^2)} \ln \left[\frac{s^2}{s^2 + \alpha^2} + \left(\gamma\sqrt{s^2 + \alpha^2} r_0 t \right)^2 \right] \\ &\quad + \frac{s}{\gamma(s^2 + \alpha^2)} \tan^{-1} \left[\frac{\gamma(s^2 + \alpha^2)}{s} r_0 t \right]. \end{aligned} \quad (3.105)$$

Also, we have

$$\cos \theta_1 = \frac{1}{\sqrt{1 + \left[\frac{\gamma(s^2 + \alpha^2)}{s} r_0 t \right]^2}}. \quad (3.106)$$

Or,

$$\tan \theta_1 = \frac{\gamma(s^2 + \alpha^2)}{s} r_0 t . \quad (3.107)$$

So,

$$\theta(t) = \theta_1 - \phi = \tan^{-1} \left[\frac{\gamma(s^2 + \alpha^2)}{s} r_0 t \right] - \sin^{-1} \left[\frac{\alpha}{\sqrt{s^2 + \alpha^2}} \right]. \quad (3.108)$$

Choosing $t = T$ such that $r_0 T = D$, we obtain the final resist linewidth and resist sidewall angle.

In summary, the lumped parameter model^{38,35,36} makes several assumptions:

- The resist development rate is related to exposure by the γ^{th} power.
- The resist absorption is constant throughout its thickness and is independent of exposure time.
- The development process proceeds independently in the lateral and longitudinal directions.

With these assumptions, the exposure required to produce a resist feature can be related to the aerial image analytically, such that

$$E(W/2) = E(0) \left[1 + \frac{1}{\gamma D_{\text{eff}}} \int_0^{W/2} I^{-\gamma}(x) dx \right]^{\frac{1}{\gamma}}, \quad (3.109)$$

where $E(W/2)$ is the exposure dosage required to develop to the resist image edge $x = W/2$, W is the width of the feature of interest, $E(0)$ is the exposure dosage corresponding to a feature width of zero, and $I(x)$ is the normalized intensity distribution of the aerial image. For a detailed derivation, consult Ref. 36.

The resist development contrast γ can be evaluated with

$$\gamma = \frac{dr / r}{dE / E}, \quad (3.110)$$

where r is the dissolution rate and E is the exposure. Hence, the physical meaning of γ is the slope of the dissolution rate versus the exposure curve normalized at the point of evaluation. That is, all values of r and E are taken at the intended exposure. When a log plot of r versus E is used, the slope is taken directly without having to normalize as follows:

$$\gamma = \frac{d(\ln r)}{d(\ln E)}. \quad (3.111)$$

D_{eff} , the effective resist thickness named by Hershel and Mack,³⁶ is defined as

$$D_{eff} = r_0 t_{dev} e^{\gamma i(D)}, \quad (3.112)$$

where r_0 is the dissolution rate corresponding to an exposure E_0 that just clears the photoresist in a large, uniformly illuminated area. Furthermore, t_{dev} is the time it takes to develop the resist through the entire resist thickness D . Rearranging Eq. (3.112) reveals the physical meaning of D_{eff} :

$$D_{eff} = (t_{dev}) [r_0 e^{\gamma i(D)}]. \quad (3.113)$$

The last group is the dissolution rate at a depth D if the resist is assumed to have a constant γ within the exposure range from the top to the bottom of the resist layer. Hence, D_{eff} physically means the equivalent resist thickness that the developer would open up if development takes place at the development rate of $z = D$ for the duration required to clear the resist if the latter is illuminated with E_0 .

So far the second lumped parameter α has not been used explicitly. It is hidden in D_{eff} . Using

$$t_{dev} = \int_0^D \frac{dz}{r} = \frac{1}{r} \int_0^D e^{-\gamma i(z)} dz, \quad (3.114)$$

Hershel and Mack derived

$$D_{eff} = \int_0^D \left[\frac{I(z)}{I(D)} \right]^{-\gamma} dz. \quad (3.115)$$

We have performed the integration to expose the dependency on α :

$$D_{eff} = \frac{1}{\alpha \gamma} (1 - e^{-\alpha \gamma D}). \quad (3.116)$$

Equations (3.109) and (3.116) are used together to predict the developed resist image from a given aerial image.

The accuracy of the lumped parameter model depends on the validity of the assumptions and the accuracy of the three parameters. The resist development rate can deviate from the γ^{th} power assumption in reality. One should evaluate γ

at the vicinity of the exposure dosage used to achieve the desired linewidth in order to minimize errors. The constant absorption assumption is generally acceptable. Unless the resist of interest is a contrast-enhancement resist that emphasizes its bleaching characteristics, typical photoresists have a small bleaching or dyeing component. The independent developments in the lateral and longitudinal directions do not seem to cause severe problems. In terms of the accuracy of the parameters, γ is most important. D requires a resist thickness measurement that is not an issue. α can be taken from the A, B, C coefficients as

$$\alpha = \frac{1}{2} (\alpha_{unexposed} + \alpha_{exposed}) = B + \frac{A}{2}. \quad (3.117)$$

Figure 3.35 shows the impact of γ on the DOF of a 0.35- μm isolated line opening imaged at $\lambda = 365 \text{ nm}$ and $\sigma = 0.3$. The other two lumped parameters D and α are fixed at $0.9 \mu\text{m}$ and $0.2 \mu\text{m}^{-1}$, respectively. The DOF is evaluated using the E-D tool methodology given later in Chapter 4, setting the CD tolerance to $\pm 10\%$ and the exposure latitude to 10%. From Fig. 3.35, a relatively high γ of 5 reduces the DOF significantly from the aerial image prediction. The aerial image DOF potential is restored only when γ is increased to 18. Above 18, a higher γ cannot further improve the image. For i-line resists, γ rarely reaches 5. With chemically amplified resists, γ can be extremely high. Figure 3.36 depicts the

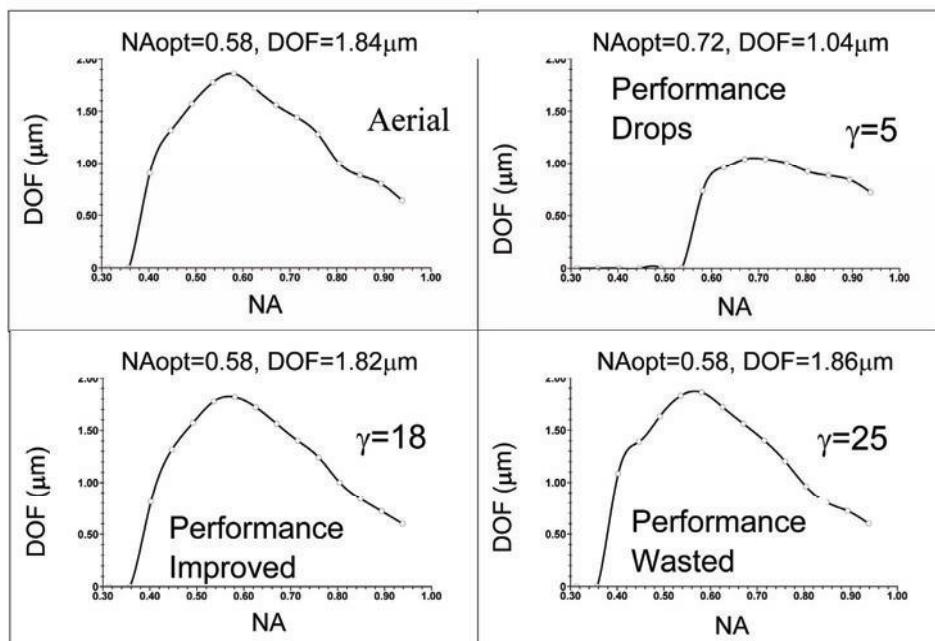


Figure 3.35 DOF versus NA at $\sigma = 0.3$, $\lambda = 365 \text{ nm}$ for a 0.35- μm isolated line opening, $D = 0.9 \mu\text{m}$, $\alpha = 0.2 \mu\text{m}^{-1}$.

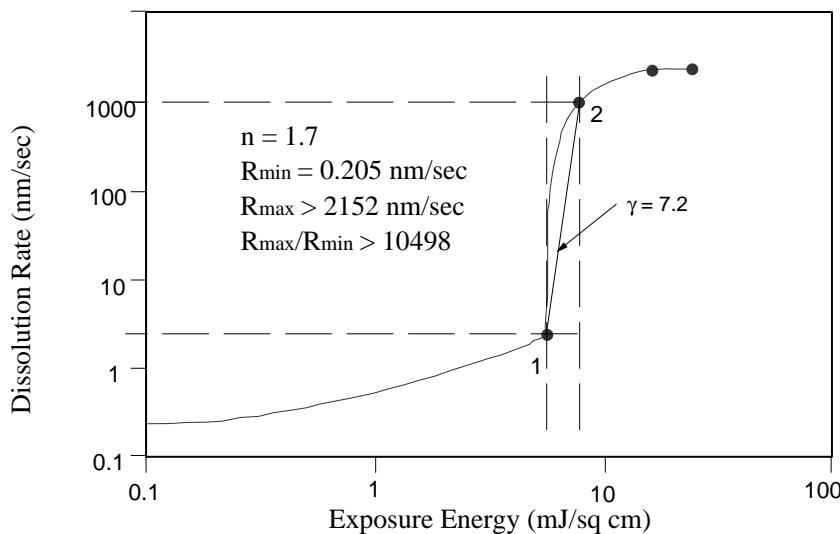


Figure 3.36 UVIII dissolution properties in CD-26. (Image courtesy of W. Conley.)

UVIII dissolution properties in CD-26, as reported by Conley et al.³⁹ With the additional supplemental lines between points 1 and 2, it is seen that the average γ between these two points is 7.2. If evaluated near point 2, the γ value is about 1, whereas if evaluated near point 1, γ is easily above 100. Katanani et al.⁴⁰ plotted γ explicitly for an e-beam chemically amplified resist. They demonstrated that γ is a function of the soft bake temperature and time as well as developer concentration and time, and that the range of γ within their experimental parameters is from 2.39 to 68.6, as shown in Fig. 3.37.

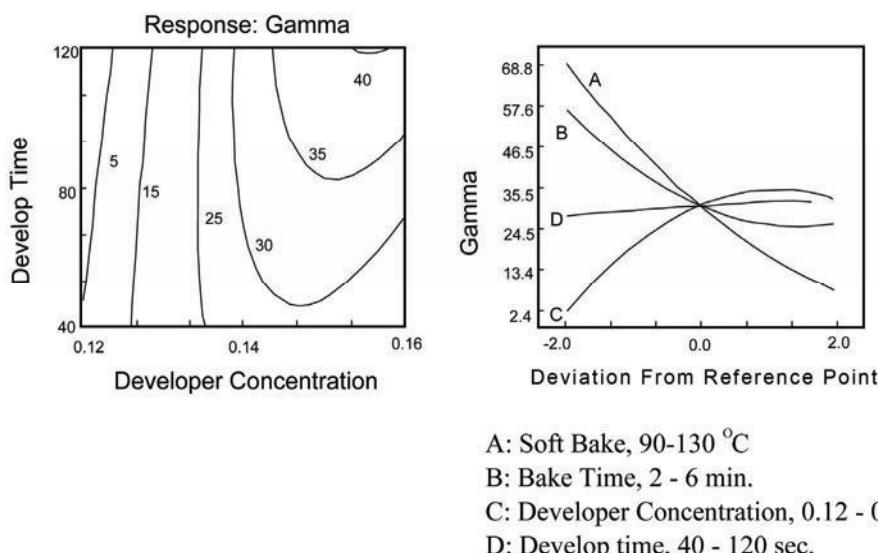


Figure 3.37 γ of an e-beam chemically amplified resist.

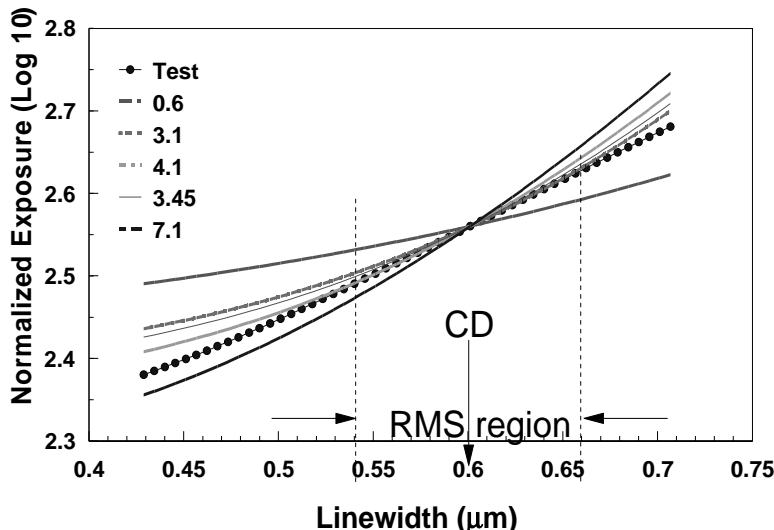


Figure 3.38 Evaluation of γ by matching experimental results.

The most useful way⁴¹ to evaluate γ is by matching the simulated linewidth versus the exposure curve to the corresponding experimental curve. This type of γ is linewidth derived (LWD). The γ value that produces the smallest mean-square error between the two curves is chosen as the γ value of the resist at the given exposure and development environment. Figure 3.38 shows the experimental curve and the simulated curves with a range of γ values. $\gamma = 3.45$ gives the best match. Figure 3.39 shows the E-D tree generated by simulation using LPM and the γ value from Fig. 3.38. The experimental data points are plotted in the E-D plane to check the accuracy of the simulation.

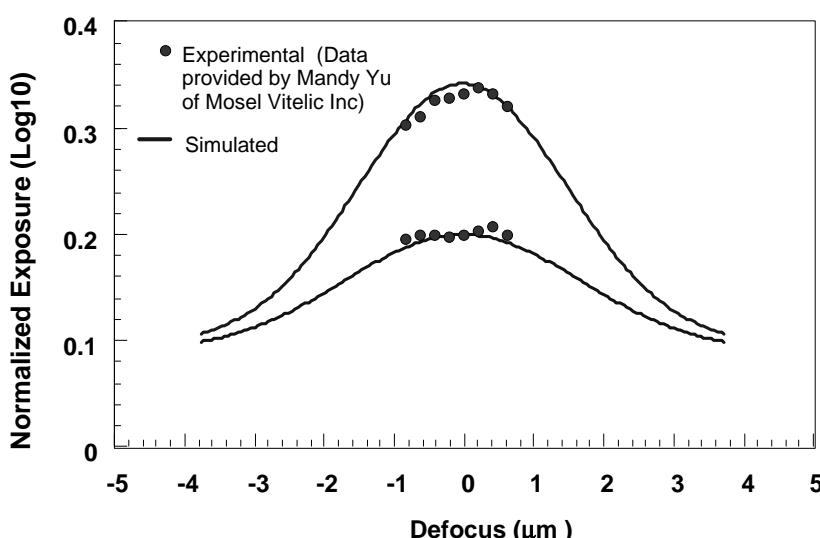


Figure 3.39 E-D tree simulated with γ derived from curve fitting compared with experimental data.

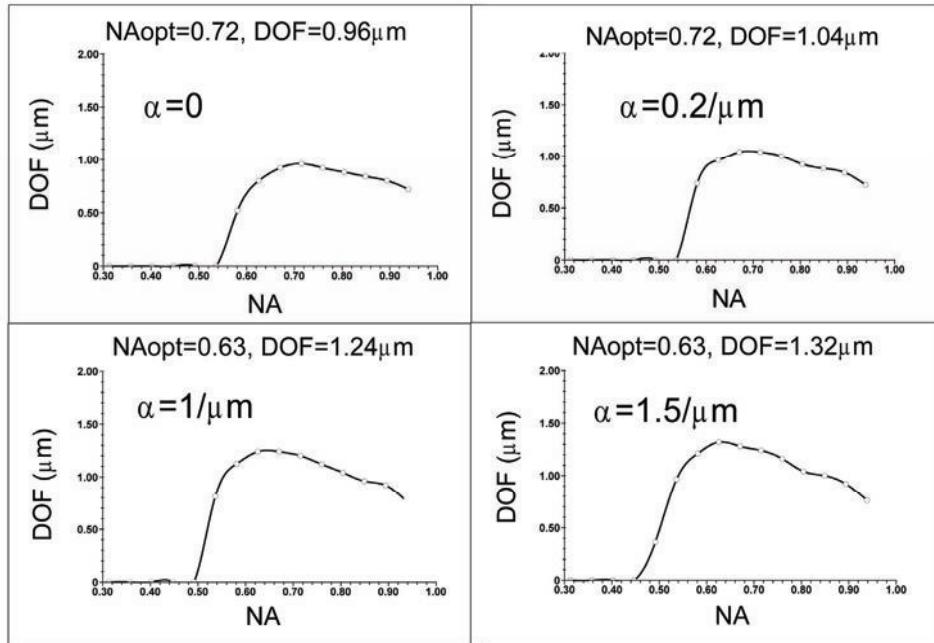


Figure 3.40 DOF versus NA at $\sigma = 0.3$, $\lambda = 365$ nm for a 0.35- μm isolated line opening, $D = 0.9$ μm , $\gamma = 5$.

Figure 3.40 shows the effects of α on the DOF in the same situation as depicted in Fig. 3.35, with γ fixed at 5. The impact of α on the DOF is less significant. However, the DOF seems to be improved by a larger α . This contradicts the knowledge that zero absorption produces a resist profile with less undercut, and that the resist image is closer to the aerial image. Figure 3.41⁴² depicts the difference between the aerial image and the developed resist image. Because the resist is simultaneously developed laterally and longitudinally, by the time the developer reaches the bottom, the developed edge at the bottom already clears the edge predicted by the aerial image. This additional bias penalizes the E-D window if the target CD is set to the mask dimension, which reduces DOF. When resist absorption is high, light intensity attenuates more as it propagates toward the bottom. The light at the bottom is now lower than before, making the resist profile more overcut while the developed edge position at the top of the resist is unchanged because the exposure up there is unaffected by the increased resist absorption. Even though the profile is less appealing, if one strictly judges the resist image by the position of its edge at the bottom, then the resist edge position is closer to that predicted by the aerial image than when the resist absorption is low and the DOF is larger. Additional advantages of high resist absorption are the reduction of multiple interference and standing waves. However, besides the undesirable overcut profile, the exposure time can increase manyfold, lowering wafer throughput significantly.

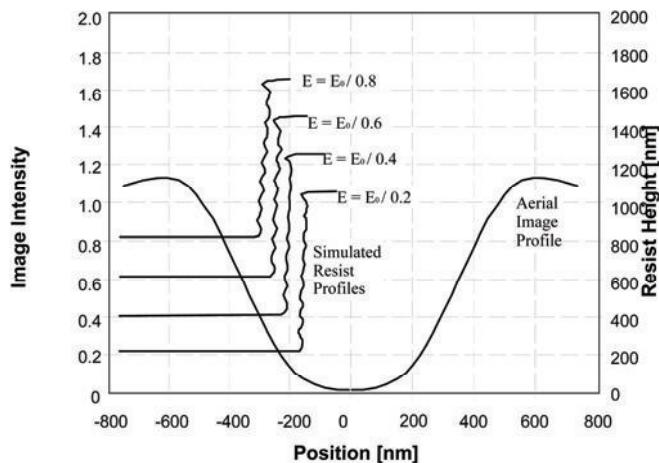


Figure 3.41 Aerial image and resist profiles. (Image courtesy of T. A. Brunner and R. A. Ferguson.)

Figure 3.42 shows the effects of D on the DOF in the same situation as depicted in Fig. 3.35, where α and γ are fixed at $0.2 \mu\text{m}^{-1}$ and 5, respectively. The resist thickness has a low significance in the context of LPM. The DOF is improved slightly only with unusable resist thicknesses in the vicinity of $0.01 \mu\text{m}$. This does not mean that D is not important in imaging. The impact of D is in changing the effective exposure in the resist induced by thickness changes. This translates to a change in the requirement of the exposure latitude. A larger exposure swing for two resist thicknesses to be delineated simultaneously

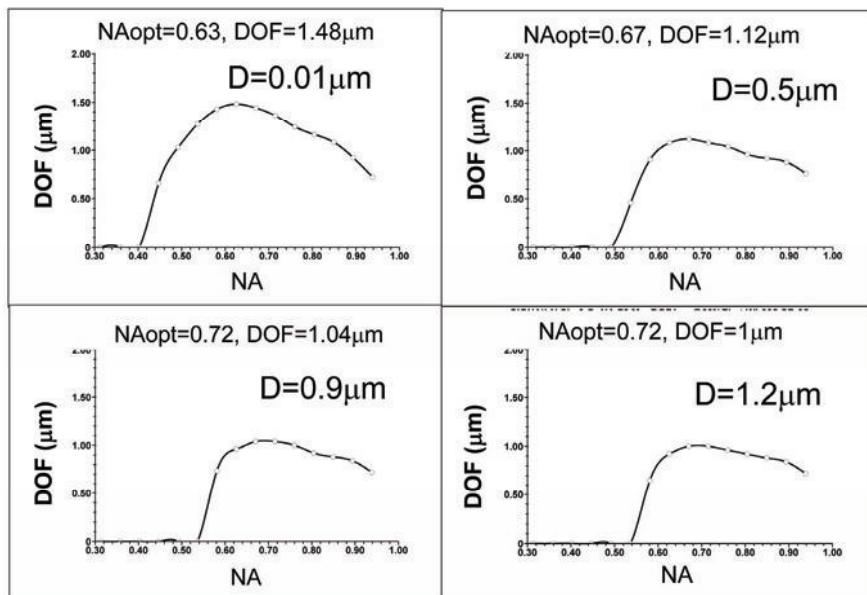


Figure 3.42 DOF versus NA at $\sigma = 0.3$, $\lambda = 365 \text{ nm}$ for a $0.35\text{-}\mu\text{m}$ isolated line opening, $\gamma = 5$, $\alpha = 0.2 \mu\text{m}^{-1}$.

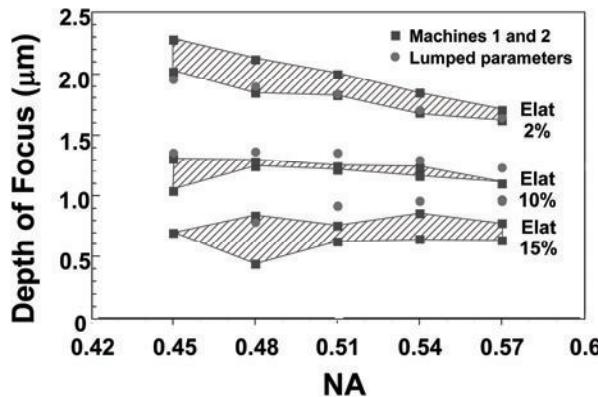


Figure 3.43 Lumped-parameter and experimental results.

requires a larger exposure latitude. Once the exposure swing is taken care of by specifying the correct amount of exposure latitude, the impact of D is similar to that of α .

When $D \rightarrow 0$, $\alpha \rightarrow 0$, and $\gamma > 25$, the resist image predicted with the LPM is practically identical to the aerial image. Figure 3.43 shows that the variation between two exposure tools is larger than the deviation of the results derived from the LPM.

We used a fully simulated resist image from Prolithⁱ to check the effectiveness of LWD- γ .⁴³ Equal line-space pairs with $k_1 = 0.82$ and 0.9 are exposed with 365-nm light at $\sigma = 0.5$ to a resist 800 nm thick, which are typical resist and processing parameters on a planar silicon wafer. The required exposure dosage is plotted as a function of the developed resist-space width. This full simulation result is compared with an LPM-derived result using LWD- γ , as shown in Fig. 3.44. Very good matching is obtained.

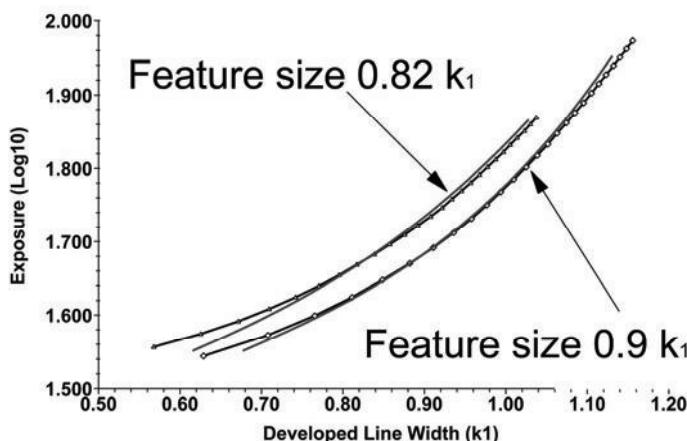


Figure 3.44 Linewidth-derived γ fits a lumped equation for full simulation.

ⁱ Prolith version 5.10. Prolith is a registered trademark of Finle, Inc., a subdivision of KLA-Tencor.

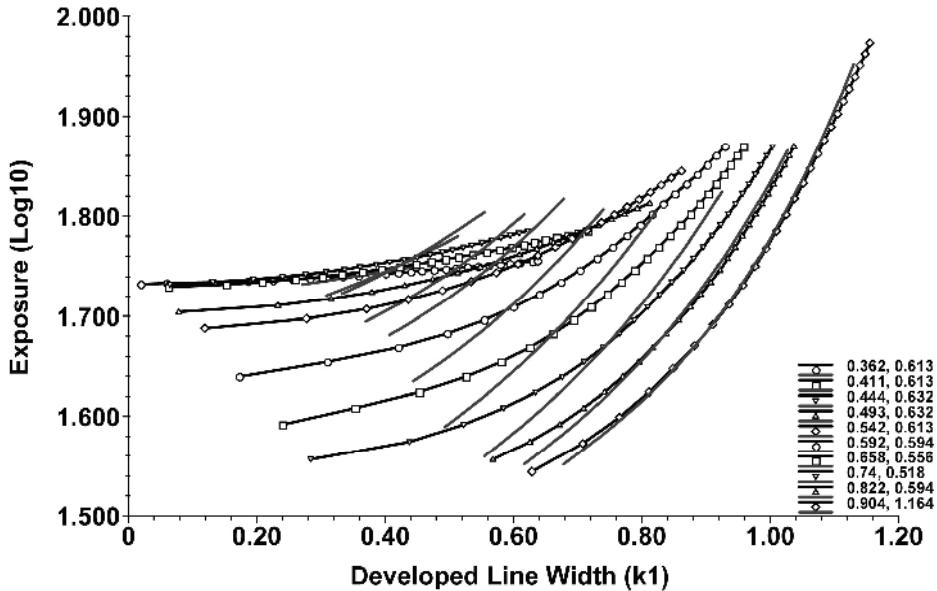


Figure 3.45 Lumped-parameter approximations applied to $k_1 = 0.362$ to 0.904 .

3.4.3 β and η

The LPM works well for $k_1 = 0.82$ or greater. When matching is attempted for smaller k_1 values, the LPM failed. Figure 3.45 shows a comparison similar to that of Fig. 3.44, except that the feature size now ranges from $k_1 = 0.362$ – 0.411 , 0.444 , 0.493 , 0.542 , 0.592 , 0.658 , 0.74 , 0.822 , and 0.904 , with corresponding best-fitted LWD- γ being 0.613 , 0.613 , 0.632 , 0.632 , 0.613 , 0.594 , 0.556 , 0.518 , 0.594 , and 1.164 . The poor-fitting LWD- γ at k_1 less than 0.8 is attributed to the Hershel-Mack lumped parameters ceasing to faithfully derive correct $i\gamma$ values.

Instead of fitting the exposure versus the developed linewidth curve with that derived from the lumped parameters, one simply fits these curves with

$$E(w) = e^{\eta w}, \quad (3.118)$$

where E is the exposure dosage, w is the developed linewidth, and η is the new LPI called LWD- η .⁴³ From Eq. (3.118),

$$\eta = \frac{1}{w} \log_e E(w). \quad (3.119)$$

Hence, η is the log slope of exposure versus the developed linewidth that is different from the familiar log slope of intensity versus displacement in x . The accuracy of η fitting is seen in Fig. 3.46. It is literally the log slope of each $E(w)$

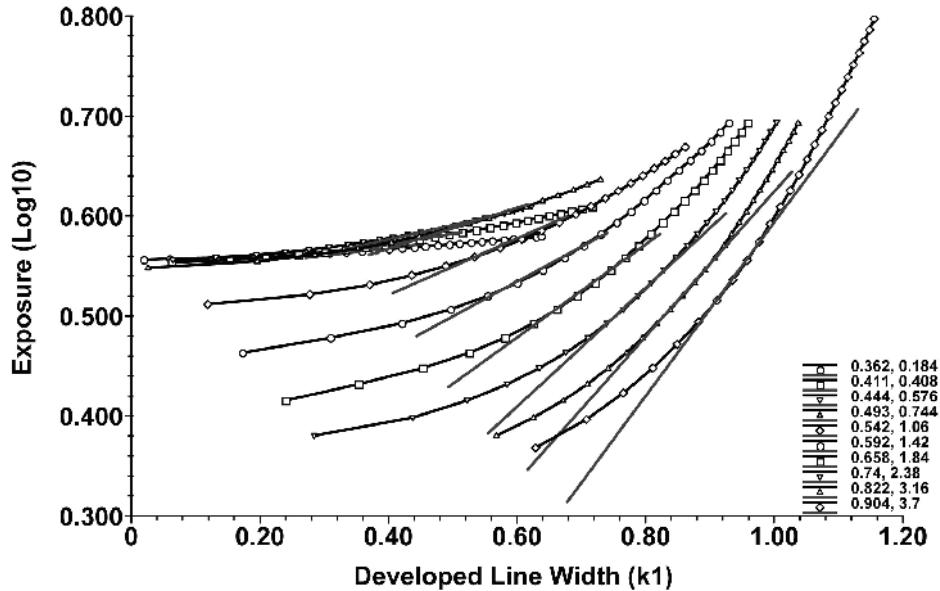


Figure 3.46 Fitting LWD- η from $k_1 = 0.362$ to 0.904.

curve at the nominal k_1 point. For example, the third curve from the right has a nominal k_1 of 0.74. Even though the entire $E(w)$ curve covers a range of $k_1 = 0.28$ to 1.0, the log slope is the tangent at the position $k_1 = 0.74$. The nominal k_1 is the size of the feature to be printed, whereas the k_1 range in the $E(w)$ curve is the range of the developed linewidth.

Motivated by the need of a lumped parameter for all feature sizes and encouraged by the excellent fitting result of Fig. 3.46, a new lumped parameter is found by performing a log-log fit of the $E(w)$ curve. It is

$$\beta \equiv \frac{1}{w} \log_e [10 \log_e E(w)], \quad (3.120)$$

which is taken from the analytic form of

$$E(w) = e^{0.1 \exp(\beta w)}. \quad (3.121)$$

The fitting shown in Fig. 3.47 is much better than that of LWD- γ at small k_1 values. Hence, LWD- β can be used to derive a developed resist image from an aerial image in low- k_1 situations.

While β is a better parameter to replace LPM in low- k_1 situations, η is useful as a lithography performance indicator, as shown in Ref. 43.

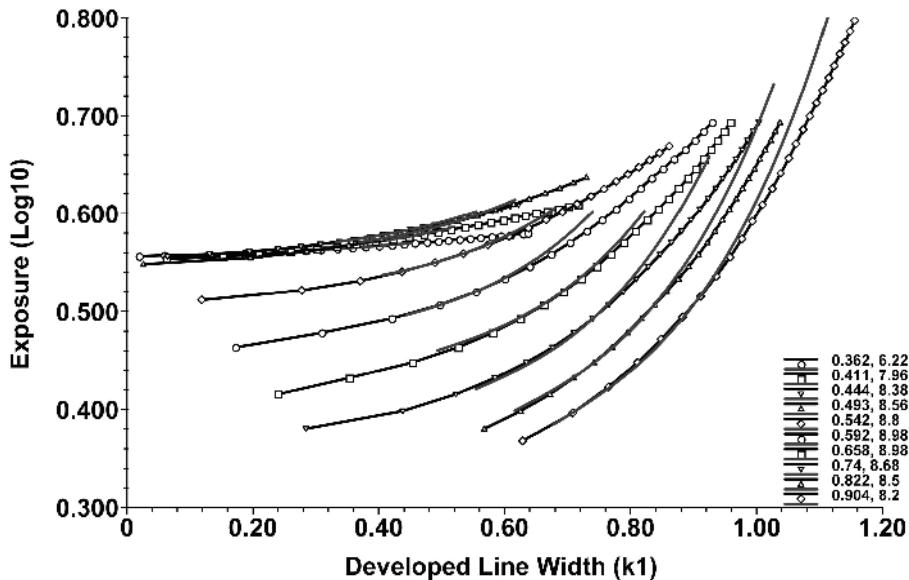


Figure 3.47 Fitting LWD- β from $k_1 = 0.362$ to 0.904 .

3.5 From Aerial Image to Resist Image

We now condense the discussions from Secs. 3.1 to 3.4 to functional forms. We account for light radiating from the light source, going through the illumination system with properly optimized temporal and spatial coherence, then travelling through the mask, propagating through the many refracting and reflecting surfaces of a lens system, and inducing chemical interactions in the resist. Finally, we account for the resist latent image developing into a resist image. What follows are the parameters leading to an aerial image, refracted/reflected image, latent image, rate image, and finally a developed resist image:

- *Aerial image*—The aerial image is the diffracted image from the mask through the lens, but before the image has any interaction with the resist. In Eq. (3.122), the intensity of the aerial image I_a is a function of feature size, shape, and feature combinations; λ is wavelength; NA_i is the numerical aperture of the imaging lens; σ is the aperture ratio NA_c/NA_i between the condenser and the imaging lens; and the lens aberration coefficients are a_1, \dots, a_n .

$$I_a(x, y, z) = f_1(F_{shape}, F_{size}, F_{combi}, \lambda, NA, \sigma, a_1, \dots, a_n). \quad (3.122)$$

This aerial image is assumed to be diffracted by an aberrated lens without consideration of the multiple reflections at the surfaces of the lens elements, or the refraction and reflections at the resist and wafer substrate interfaces.

- *Refracted/reflected image*—The refracted/reflected image I_r in Eq. (3.123) includes these effects:

$$I_r(x, y, z, t) = f_2 [I_a, n_{cr}(x, y, z, I_r), n_{c1}, \dots, n_{cn}, S_1(x, y, z), \dots, S_n(x, y, z)], \quad (3.123)$$

where $n_c \equiv n_r + in_i$ is the complex index of refraction. The real part of the refractive index is n_r , which accounts for the changed wavelength in the media and phase shifting. The complex part of the refractive index is n_i , which accounts for light absorption in the film. The complex index of refraction of the resist layer is n_{cr} . The films above and below the resist layer are represented by n_{c1} through n_{cn} . S_i through S_n are the lens surfaces. Notice that n_{cr} is not only a function of the space coordinates, but also a function of I_r . This is the resist bleaching or dyeing phenomenon. The refractive index changes as a function of the incident light. This change proceeds to perturb the incident light in the resist, creating a feedback loop until bleaching or dyeing stops.

- *Latent image*—The exposure usually terminates before bleaching or dyeing saturates. This results in the expression of the latent image n_{cr} at the end of the exposure, as in

$$n_{cr} \left[x, y, z, I_r(t_{end \ exposure}) \right]. \quad (3.124)$$

- *Dissolution rate image*—The resist dissolution rate can be directly linked to the localized exposure in the resist. This relationship is often plotted as the dissolution rate versus the exposure curve shown in Fig. 3.30. The latent image is also a direct indication of the localized exposure in the resist. Hence, the dissolution rate image is

$$\begin{aligned} \text{Rate}(x, y, z) = \\ f_3(I_r, d, \text{resist variable, thermal history, resist development variables}). \end{aligned} \quad (3.125)$$

- *Resist image*—Given the dissolution rate distribution, $\text{Rate}(x, y, z)$, the resist developer starts at the surface of the resist, carving out new surface shapes as the developer removes the resist differentially according to the rate distribution. The developer continues to adjust its advancement according to the new surface that is available for dissolution and the rate at the very point of dissolving. The resultant image after the developer is removed is the resist image. THK is the thickness of the resist layer.

$$Z(x, y) = f_4 [THK, \text{Rate}(x, y, z), t_{dev}]. \quad (3.126)$$

3.6 The Transferred Image

The resist image can be transferred to the wafer in four ways: etching, lift off, implantation, and electroplating. Etching can be further classified as isotropic or anisotropic. These transfer processes are discussed below.

3.6.1 Isotropic etching

With isotropic etching, the etchant removes the substrate under the resist etch mask at the same rate in all directions. As soon as it penetrates the etch mask, the removal process goes laterally as well as vertically, as shown in Fig. 3.48 by Brandes et al.⁴⁴ After the thin film is completely etched to the bottom, sideways etching continues, further enlarging the etched opening. The surface with a larger curvature etches faster, making material removal faster at the bottom until the profile becomes vertical. The equation governing isotropic etching is

$$Z(x, y) = f_4 (THK, \text{Rate}, t_{dev}), \quad (3.127)$$

where THK is the thickness of the substrate material to be etched. Instead of a function of position, the dissolution rate is constant. Because of the sideways material removal, a small resist spot may be completely etched under and disappear. This is not always undesirable. Small defects can be removed by overetching. Wet etching is generally associated with isotropic etching, except for special etchants having preferential rates along crystal lattices.⁴⁵ Certain types of dry etching, such as plasma etching and ashing, are isotropic. Isotropic etching is not suitable for a high-density or high-aspect-ratio etched image. Anisotropic etching favoring the direction perpendicular to the surface of the substrate must be used.

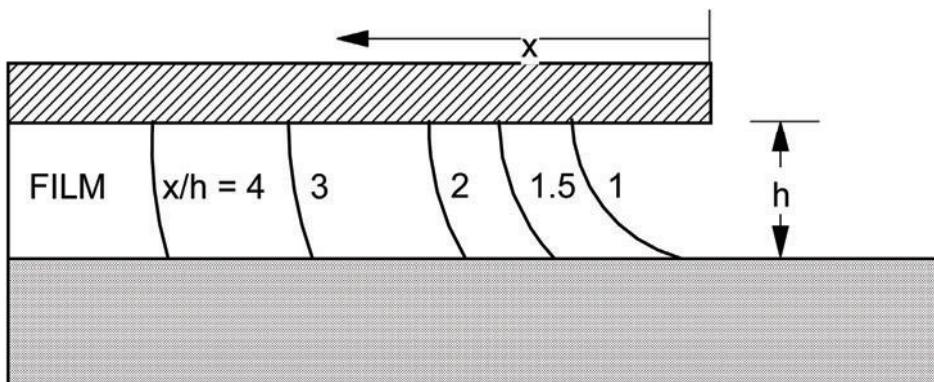


Figure 3.48 Isotropic etch sequence.⁴⁴ (Reproduced by permission of The Electrochemical Society.)

3.6.2 Anisotropic etching

With anisotropic etching, material is removed faster in one direction. Generally, faster removal in the direction perpendicular to the substrate is preferred. Figure 3.49 shows the etch sequence of 10:1 anisotropic etching through a resist etch mask. The etched image is similar, but not identical, to that in Fig. 3.31. In the former case, the resist protects the thin film underneath, while the entire developing resist surface is gradually dissolved by the developer in the latter case.

In a way, surface removal in etching can be treated similarly to resist developing. In the case of isotropic etching, there is an abrupt layer of zero dissolution rate on top, and the dissolution rate in the bulk is uniform. With anisotropic etching, the dissolution rate distribution in the resist is still the same, but there is a directional preference in cell removal, removing more cells in one direction than the other. Similarly, when the etching surface is treated as a propagating wavefront, the wavefront moves faster in one direction than the other. The anisotropic etch equation is

$$Z(x, y) = f_4 (THK, \text{Rate}_{\text{horizontal}}, \text{Rate}_{\text{vertical}}, t_{\text{dev}}). \quad (3.128)$$

Anisotropic etching is used to achieve a high density and high aspect ratio. It is often necessary to produce a completely vertical profile. That cannot be achieved by just increasing the anisotropic etch-rate ratio. Passivation of the etched area helps to prevent sideways etching to produce the anisotropic effect.

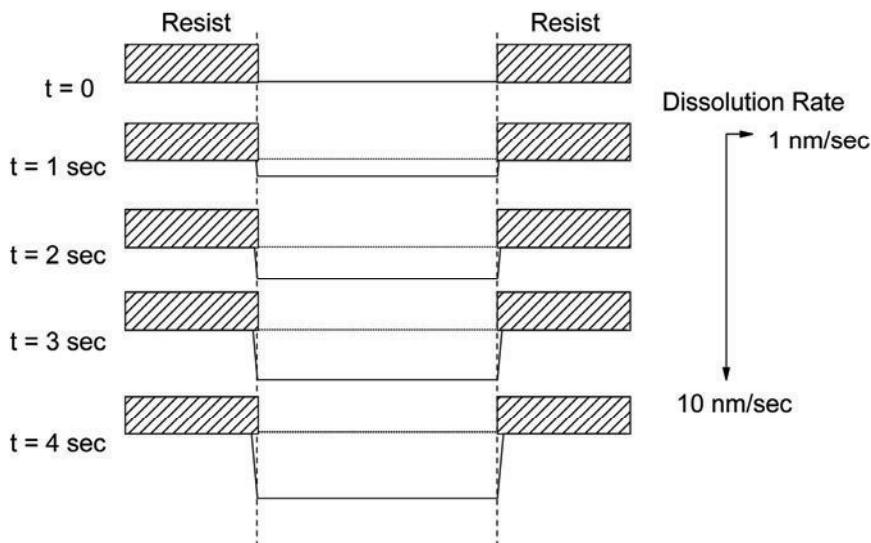


Figure 3.49 Anisotropic etching sequence.

3.6.3 Lift off

Lift off is an additive process that adds material to the resist openings, whereas etching is a subtractive process that removes material through resist openings. As shown in Fig. 3.50, the material to be added to the substrate surface is deposited indiscriminately on the resist image. The unwanted part over the resist-covered areas is removed with the resist by placing the substrate in a resist solvent. The resist image needs to have a vertical or undercut profile. Otherwise, the deposited material would cover the entire resist/substrate surface, preventing any solvent penetration. In addition, the deposition process must be directional (such as in evaporation) to promote the break between the deposited material on the top and bottom. Conformal deposition is not suitable for lift off.

Figure 3.51 shows an SEM micrograph of aluminum deposited over the resist image before lift off. Note that as the deposited material builds up, it expands slightly in the lateral direction. The build up affects the vertical profile because the opening that allows the deposited material to reach the substrate becomes smaller and smaller so that a vertical profile cannot be maintained. Figure 3.50 was drawn with this effect in mind.

The equation for lift off is similar to that for anisotropic etching, except that the profile of the lifted-off image is not dependent on its thickness. Naturally, the thickness of the resist must be larger than the thickness of the lifted-off material, lest the opening close, thus preventing solvent penetration at the pattern edges.

$$Z(x, y) = f_5 (\text{Rate}_{\text{horizontal}}, \text{Rate}_{\text{vertical}}, t_{\text{dev}}). \quad (3.129)$$

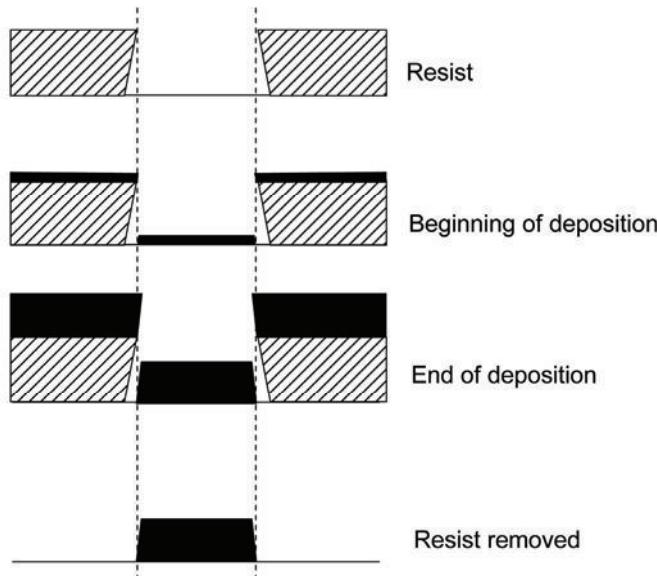


Figure 3.50 Lift off pattern transfer.

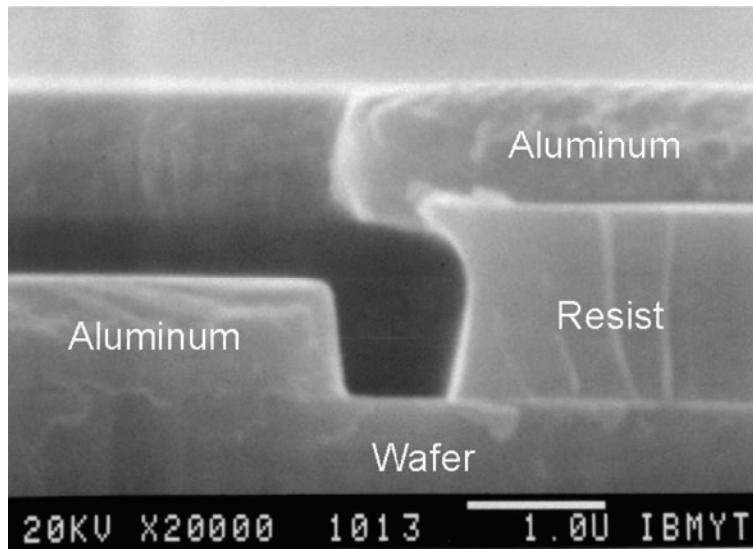


Figure 3.51 SEM micrograph of aluminum deposited over the resist image.

3.6.4 Ion implantation

In an ion implantation image transfer, the spatial distribution function $f(x,y,z)$ of implanted ions is given by⁴⁶

$$f(x, y, z) = \frac{1}{(2\pi)^{3/2} \Delta R_p \Delta X \Delta Y} \exp \left\{ -0.5 \left[\frac{x^2}{\Delta X^2} + \frac{y^2}{\Delta Y^2} + \frac{(z - R_p)^2}{\Delta R_p^2} \right] \right\}, \quad (3.130)$$

where x and y are perpendicular to the incident z -direction of the ions. ΔX and ΔY are the standard deviation of the ion spreads in the x - and y -directions, respectively, and R_p is the so-called straggle in the incident direction, which is simply the standard deviation of ion penetration into the substrate. Equation (3.130) shows that, in addition to traveling in the incident direction, the ions have a transverse spread because of scattering in the resist and in the substrate. Therefore, at a resist edge, even though the ions are stopped in the incident direction by the resist at the resist side, the ions still penetrate under the resist. This happens because the ions that hit the substrate at the uncovered side of the edge are scattered sideways in both the covered and uncovered directions. Note that ion penetration is not perfectly isotropic, the ratio of transverse penetration to the incident penetration is usually not unity. For example, Fig. 3.52 shows that the transverse-to-incident penetration ratio of phosphorus at 200 keV is about 1:2.

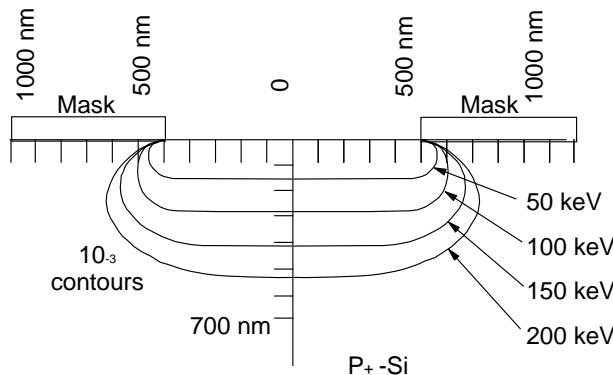


Figure 3.52 Equi-ion concentration curves at 0.1% intervals for phosphorus implantation into silicon at different implant energies.⁴⁷ A 1- μm mask cut is used. (Image courtesy of the Japan Society of Applied Physics.)

In a way, ion implantation is just ion-beam proximity printing. The resist image is a perfectly contacting mask that stops the ion beam. The beam that enters the unblocked substrate is scattered by the substrate material. This can be viewed as an ion-beam proximity effect, similar to the e-beam proximity effect, where the electrons are scattered by the material it penetrates into. From this perspective, ion-beam lithography has become a high-volume production technology that escapes our awareness. Of course, it can also be viewed as a pattern-transfer process, transferring the resist image to an implanted image.

3.6.5 Electroplating

With a plating base under an insulator pattern, metal may be plated up on the plating base at the areas not covered by the insulator pattern. The insulator can be the resist image as well as the other insulator image transferred from the resist image. A schematic drawing of the electroplating process is shown in Fig. 3.53.

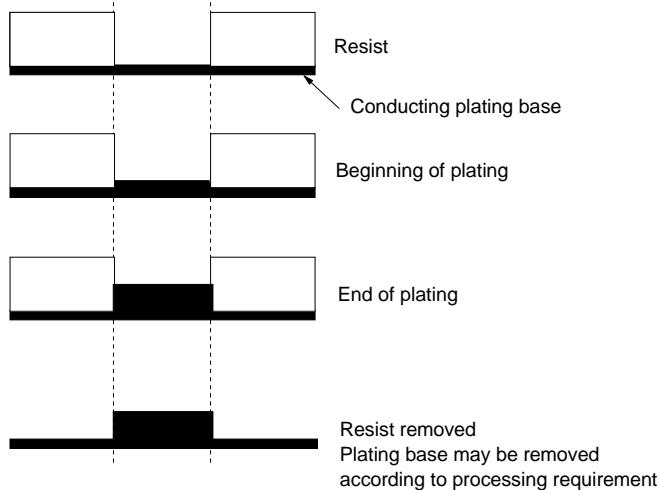


Figure 3.53 Electroplating image transfer.

Similar to lift off, electroplating is an additive process. The metal image conforms to the insulator image faithfully as long as its thickness does not exceed the insulator thickness. Beyond that, because of the absence of constraints, the plated material grows sideways as well as vertically; it may be treated as an isotropic growth under most circumstances. The metal patterns may eventually be connected above the insulator if plating continues. Hence, the plated image is identical to the insulator image below the insulator surface. Above the insulator surface, a spherical wavelet can be integrated from the patterned areas to describe the 3D metallic surface.

References

- 1 C. J. R. Sheppard, "Imaging in a high-aperture optical systems," *JOSA A* **4**, p. 1354 (August 1987).
- 2 M. Born and E. Wolf, *Principles of Optics 6th Edition*, Cambridge Univ. Press, p. 435 (1980).
- 3 Equation 17 of Sec. 5.2 in Ref. 2.
- 4 E. L. O'Neil, *Introduction to Statistical Optics*, Addison-Wesley, Reading, Massachusetts, p. 50 (1963).
- 5 F. Zernike, *Physica I*, 689–704 (1934). English translation, "Diffraction theory of the knife-edge test and its improved form, the phase-contrast method," *J. Microlith. Microfab., Microsyst.* **1**(2), pp. 87–94 (2002).
- 6 Sec. 9.2.2 of Ref. 2.
- 7 V. N. Mahajan, *Optical Imaging and Aberrations*, SPIE Press, Bellingham, Washington, p. 163 (2001).
- 8 Born and Wolf, Sec. 9.5 of Ref. 2.
- 9 J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, p. 108 (1968).
- 10 Secs. 1.1.1 and 1.1.3 of Ref. 2.
- 11 G. Bekefi, "Study of optical diffraction images at microwave frequencies," *J. Appl. Phys.* **24**, p. 1123 (1953).
- 12 B. J. Lin, "Electromagnetic near-field diffraction of a medium slit," *J. Opt. Soc. Am.* **62**, pp. 977–981 (1972).
- 13 R. Newman, *Fine Line Lithography*, North-Holland Publishing Co., Amsterdam, Sec. 1 of Chapter 2 (1980).
- 14 H. H. Hopkins, "On the diffraction theory of optical images," *Proc. Roy. Soc. A* **217**, p. 408 (1953).
- 15 M. S. Yeung, "Modeling high numerical aperture optical lithography," *Proc. SPIE* **922**, p. 149 (1988).
- 16 J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, Sec. 3-7, p. 49 (1968).

- 17 A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, New York, Sec. 3-2, p. 45 (1962).
- 18 B. J. Lin, "Off-axis illumination—working principles and comparison with alternating phase-shifting masks," *Proc. SPIE* **1927**, p. 89 (1993).
- 19 B. J. Lin, "Partially coherent imaging in two dimensions and the theoretical limits of projection printing in microfabrication," *IEEE Trans. Electron Devices* **ED-27**, p. 931 (1980).
- 20 B. J. Lin, Sec. 2.3.2 of Ref. 13.
- 21 W. G. Oldham, S. N. Nandgaonkar, A. R. Neureuther, and M. M. O'Toole, "A general simulator for VLSI lithography and etching processes: Part I - Application to projection lithography," *IEEE Trans. Electron. Dev.* **ED-26**, pp. 717–722 (1979).
- 22 K. K. H. Toh, A. R. Neureuther, and H. Sunami, "Three-dimensional simulation of optical lithography," *Proc. SPIE* **1463**, p. 356 (1991).
- 23 C. A. Mack, "PROLITH: A comprehensive optical lithography model," *Proc. SPIE* **538**, p. 207 (1985).
- 24 C. A. Mack and C. B. Juang, "Comparison of scalar and vector modeling of image formation in photoresist," *Proc. SPIE* **2440**, p. 381 (1995).
- 25 Sec. 1.1.1 in Ref. 2.
- 26 M. S. Yeung and E. Barough, "Three-dimensional nonplanar lithography simulation using a periodic fast multipole method," *Proc. SPIE* **2051**, p. 509 (1997).
- 27 M. S. Yeung and E. Barough, "Three-dimensional mask transmission simulation using a single integral equation method," *Proc. SPIE* **3334**, p. 704 (1998).
- 28 E. Barouch, J. Cahn, U. Hollerbach, and S. Orszag, "Numerical simulations of submicron photolithographic processing," *J. Sci. Comput.* **6**(3), p. 229 (1991).
- 29 E. Barouch, B. Bradie, H. Fowler, and S. Babu, "Three-dimensional modeling of optical lithography for positive photoresists," *KTI Microelectronics Seminar Interface '89*, p. 123 (1989).
- 30 T. Hayashida, H. Fukuda, T. Tanaka, and N. Hasegawa, "A novel method for improving the defocus tolerance in step and repeat photolithography," *Proc. SPIE* **772**, p. 66 (1987).
- 31 F. Coopmans and Bruno Roland, "DESIRE: a novel dry developed resists system," *Proc. SPIE* **631**, p. 34 (1986).
- 32 R. R. Kunz, T. M. Bloomstein, D. E. Hardy, R. B. Goodman, D. K. Downs, and J. E. Curtin, "Outlook for 157nm resist design," *J. Vac. Technol. B*, **17**, pp. 3267–3272, (1999).

- 33 F. H. Dill, A. R. Neureuther, J. A. Tuttle, and E. J. Walker, "Modeling projection printing of positive photoresists," *IEEE Trans. Electron Dev.* **ED-22**, p. 456 (1975).
- 34 F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of positive photoresist," *IEEE Trans. Electron. Dev.* **ED-22**, p. 445 (1975).
- 35 C. A. Mack, A. Stephanakis, and R. Hershel, "Lumped parameter model of the photolithographic process," *Proc. Kodak Microelectronics Seminar 86*, pp. 228–238 (1986).
- 36 C. A. Mack, "Enhanced lumped parameter model for photolithography," *Proc. SPIE* **2197**, p. 501 (1994).
- 37 T. A. Brunner, "Approximate models for resist processing effects," *Proc. SPIE* **2726**, p. 198–207 (1996).
- 38 R. Hershel and C. A. Mack, "Lumped parameter model for optical lithography," Chapter 2 in *Lithography for VLSI, VLSI Electronics - Microstructure Science*, R. K. Watts and N. G. Einspruch, Eds., Academic Press, New York, pp. 19–55 (1987).
- 39 W. Conley, G. Breyta, B. Brunsvold, R. DePietro, D. Hofer, S. Holmes, H. Ito, R. Nunes, G. Fichtl, and J. Thackeray, "The lithographic performance of an environmentally stable chemically amplified photoreist (ESCAP)," *Proc. SPIE* **2724**, p. 34 (1996).
- 40 A. D. Katanani, D. Schepis, R. W. Kwong, W. Huang, Z. C. H. Tan, and C. Sauer, "Process optimization of a positive tone chemically amplified resist for 0.25 μm lithography using a vector scan electron-beam tool," *Proc. SPIE* **2438**, p. 99 (1995).
- 41 B. J. Lin, "Signamization of resist images," *Proc. SPIE* **3051**, p. 620 (1997).
- 42 T. A. Brunner and R. A. Ferguson, "Approximate models of resist development effects," *Proc. SPIE* **2726**, p. 198 (1996).
- 43 B. J. Lin, "A lithography performance indicator (LPI) and a new lumped parameter to derive resist images from aerial images," *Proc. SPIE* **3677**, p. 408 (1999).
- 44 R. G. Brandes and R. H. J. Dudley, "Wall profiles produced during photoresist masked isotropic etching," *J. Electrochem. Soc.* **120**(1), pp. 140–142 (1973).
- 45 M. P. Lepselter, "Beam lead technology," *Bell Sys. Tech. J.* **45**, p. 233 (1966).
- 46 S. K. Gandhi, *VLSI Fabrication Principles—Silicon and Gallium Arsenide*, John Wiley & Sons, New York, p. 317 (1983).
- 47 S. Furukawa, H. Matsumura, and H. Ishiwara, "Theoretical considerations on lateral spread of implanted ions," *Jpn. J. Appl. Phys.* **11**(2), p. 14 (1972).

Chapter 4

The Metrics of Lithography

In microscopy, one is concerned about the closest distance between two objects that can be resolved in the image. The points of concern are quite different in lithography. A low-contrast image with a shallow-intensity slope can be turned into a sharp edge with the photoresist. As mentioned earlier in Sec. 1.3, the location of this edge is of prime interest and must be quantified against operating parameters of the exposure tool during manufacturing, because it determines the feature size that may be related to circuit-switching speed, leakage, resistance, etc. In addition, the edge position also determines whether the image in a given layer can be effectively overlaid with previous or subsequent layers. The five primary parameters that can be adjusted in the field are exposure dosage, focal position, pattern alignment between layers, magnification, and rotation. The last three parameters are mostly related to overlay. Even though magnification can affect feature size in principle, its effect on overlay is much greater than on feature size. The first two parameters—exposure dosage and focal position—determine the process window of the lithographic system in a mutually dependent way. This mutual dependence and the ability to superimpose simultaneous requirements on different features are captured in the exposure-defocus (E-D) tools that are the backbone of lithography metrics.

4.1 The Resolution and DOF Scaling Equations

Section 2.2 introduced resolution and DOF scaling in projection printing using the proportional relationship to wavelength and inverse proportional relationship to the lens NA without introducing the proportional constant. Equations (3.9) and (3.13) used an arbitrary constant of 0.5 as we provided the physical meaning of these relationships. Here, we finally introduce the resolution and DOF scaling equation with the rigorous proportional constants k_1 and k_3 . We wait until now to do so, because the E-D tools discussed in this chapter can unambiguously determine these proportional constants.

The resolution of an optical projection imaging system can be described by the resolution scaling equation:^{1,2}

$$W = k_2 \cdot \frac{\lambda}{NA}, \quad (4.1)$$

where W represents the minimum feature size or half pitch in the image to be delineated, λ is the actinic wavelength, NA is the numerical aperture of the imaging lens, and k_1 is a proportional constant that varies from 0.25 to 1, depending on many factors that will be discussed in the next section. Figure 4.1 depicts the relationship of NA and the collection angle 2θ of the imaging lens. $NA \equiv n \cdot \sin \theta$, where n is the refractive index of the media between the object and the imaging lens.

Similarly, the scaling equation for DOF is

$$DOF = k_2 \cdot \frac{\lambda}{NA^2}, \quad (4.2)$$

where k_2 is a proportional constant for DOF. k_2 also depends on many factors, including k_1 .¹

As shown in Chapter 3, the relationships $W \propto \lambda/NA$ and $DOF \propto \lambda/NA^2$ can be derived from diffraction theory^{2,3} and are paraxial approximations. The large-angle resolution and DOF scaling equations⁴ are

$$W = k_1 \cdot \frac{\lambda}{\sin \theta} \quad (4.3)$$

and

$$DOF = k_3 \cdot \frac{\lambda}{\sin^2 \frac{\theta}{2}}. \quad (4.4)$$

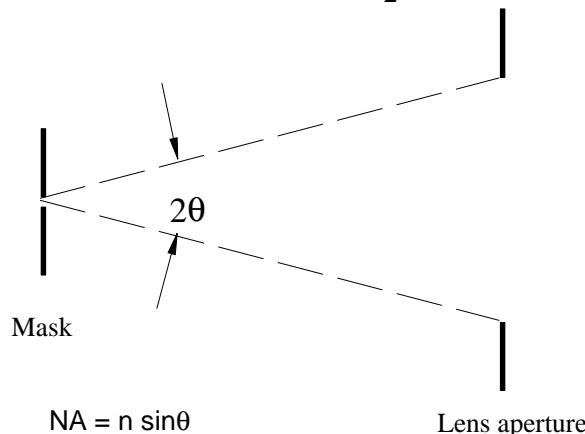


Figure 4.1 Definition of NA: n is the refractive index of the media between the mask and the imaging lens.

The relationships $W \propto \lambda/\sin\theta$ and $\text{DOF} \propto \lambda/\sin^2(\theta/2)$ hold for all illumination conditions. Obviously, k_1 must be adjusted for the degree of coherence in the illumination, and k_3 is a function of k_1 and many other parameters, just as in the paraxial case. Hence, k_1 is the coefficient for resolution, k_2 is the paraxial coefficient for DOF, and k_3 is the coefficient for DOF.

Several observations can be made on these equations:

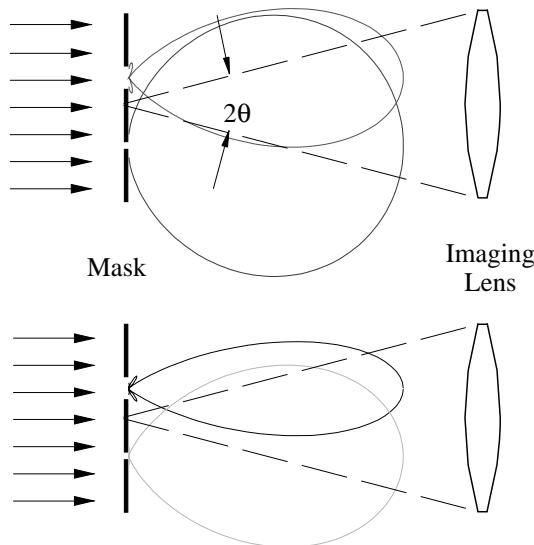
- (1) Resolution is improved by increasing the $\sin\theta$ of the imaging lens. A higher $\sin\theta$ captures a larger range of spatial frequency from the diffracted light from the features on the mask and thus can reproduce the original object with a higher resolution. Note that $\sin\theta$ and NA can be used interchangeably when $n = 1$, i.e., in nonimmersion lithography.
- (2) The tradeoff for higher resolution achieved with a higher $\sin\theta$ is a rapidly decreasing DOF. The different spatial frequency components quickly become out of phase with each other as the image is defocused.
- (3) Resolution is improved by reducing wavelength. A shorter wavelength reduces the spatial frequency in the diffracted light, making it easier for the imaging lens to capture more spatial frequencies.
- (4) With wavelength reduction, DOF is reduced at the same rate as the decrease in feature size. The loss of DOF is slower than that from the $\sin\theta$ increase.

Figure 4.2 depicts the diffracted light from a large object and a small object. The small object contains higher spatial frequencies that must be captured with an imaging lens of a higher NA. When the wavelength is reduced, the spatial bandwidth of both objects becomes narrower, making it possible for an imaging lens of smaller NA to resolve the objects. Figure 4.3 shows DOF versus resolution with λ and NA kept consecutively constant, which illustrates the tradeoff in DOF for higher resolutions, as well as the severity of the tradeoff with wavelength reduction and NA increase.

4.2 Determination of k_1 and k_3 Based on Microscopy

It is desirable to establish the resolution limit and the DOF of a given imaging system. Hence, k_1 and k_3 must be evaluated. In the early stage of optical lithography, the definition of resolution was borrowed from microscopy. After all, a reduction projection imaging system can be viewed as an inverted microscope. Even with 1X projection or proximity-printing systems, diffraction is the dominating phenomenon. In microscopy, two points are considered resolved if their images are separable.

With incoherent imaging, it is agreed that when the center of the Airy disk of one point source falls in the first zero of a second point source, the two points

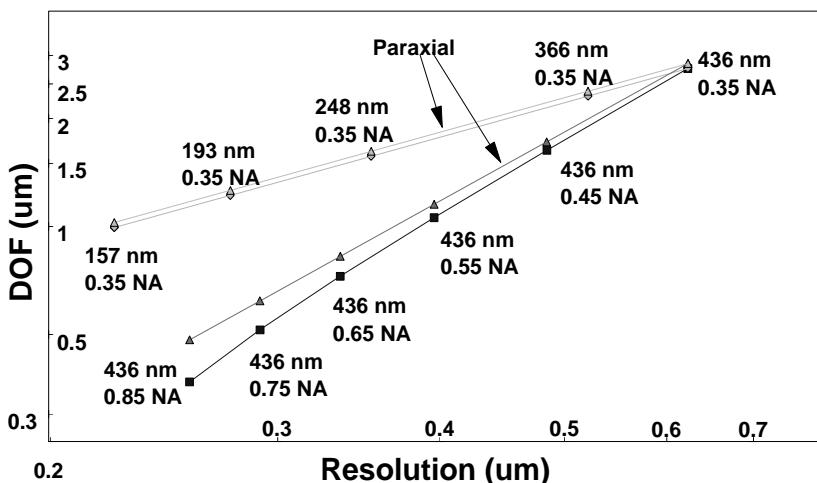


Shorter wavelength

Figure 4.2 Diffracted light from two objects at two wavelengths.

are resolved. Using a line source and a circular aperture, Goodman,⁵ among many other authors, shows that the distance between the center and the first zero is $0.61 \lambda/NA$. This is the classical limit of resolution for incoherently illuminated objects.

For coherent illumination and partially coherent illumination, two points are considered resolved if the peak and valley intensities differ by an agreed upon percentage.⁶ A possible percentage is 26.5%, because it is the same amount as when two points are separated by $0.61 \lambda/NA$ with incoherent illumination. Applying this criterion, the classical limit of resolution ranges from 0.5 to 0.8 λ/NA .

**Figure 4.3** DOF versus resolution at a constant λ and NA.

One can set the same criterion based on intensity difference to determine whether two points are resolved when defocus takes place. Again, besides the certainty of dependence on $\lambda/\sin^2(\theta/2)$, the coefficient for DOF (k_3) is a strong function of the illumination, the normalized feature size k_1 , feature shape, feature proximity, and mask type, as well as the sensitivity and dynamic range of the detector.

The difficulties of this classical derivation are in the arbitrariness and lack of generalization. In terms of arbitrariness, it may be coincidental that an average human eye resolves the incoherent distance because its sensitivity and dynamic range are suitable for detecting the 26.5% difference. The center of one source falling in the first zero with a second source is mathematically clean and physically illuminating, but there is no absolute reason why the resolution must stop there. With modern high-precision instrumentation, those classical limits can easily be exceeded.

In terms of a lack of generalization, it is clear that k_1 and k_3 are dependent on many variables. In a circuit mask, there are many feature shapes, sizes, and proximity conditions that inevitably lead to different k_1 values. The best that one can do to determine the fabrication limit is to identify the situation of the smallest k_1 . However, as will be made clear in Sec. 4.5.2, being able to resolve the most difficult part of the circuit pattern does not guarantee successful delineation of the easier part of the circuit. Each part of the circuit very often calls for different optimized exposure levels. When the exposure levels do not overlap, just resolving the most difficult parts still leaves some easier parts over- or underexposed.

4.3 Determination of k_1 , k_2 , and k_3 Based on Lithography

The lithographer's concern with resolution is not so much to resolve two adjacent lines, but to control the line edges to the specification for the product, as was described in Chapter 1. Namely, the edge placement must be within a window of tolerance, typically taken as $\pm 5\%$ of the critical dimension of the circuit.

4.3.1 E-D branches, trees, and regions

The edge placement criterion necessitates the knowledge of the $\pm 10\%$ CD boundary in the E-D space. Figure 4.4 shows these boundaries for one edge of a 0.26- μm isolated line opening using 248-nm light at NA = 0.59 and $\sigma = 0.5$. These constant edge position contours are E-D branches. The +10% CD and the -10% CD contours form an E-D tree.⁷ Any operating point in the E-D region bound by these two E-D branches will produce an image within the edge-control budget. However, the branches for different object shapes or sizes are different. Hence, if there are three critical edges in the circuit, three E-D trees must be constructed. Only the common region, enclosed by these branches, is acceptable for the combination of features. Figure 4.5 shows the superposition of the E-D

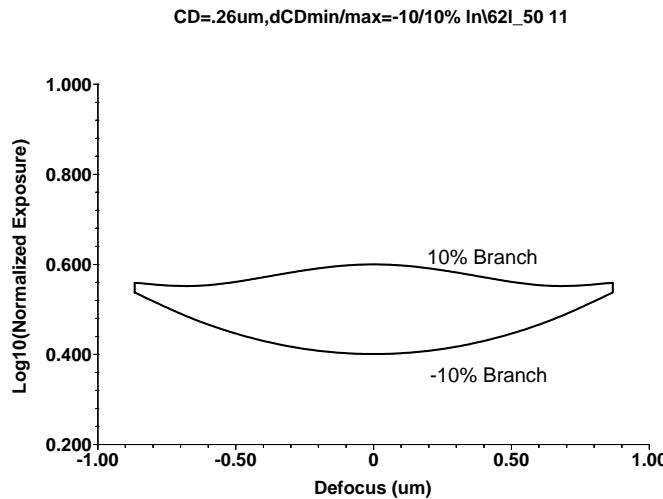


Figure 4.4 Constant edge-position contours of a 0.26- μm line opening, 248-nm exposure, NA = 0.59, σ = 0.5.

tree of one edge of the line opening shown in Fig. 4.4, the E-D tree of an opaque space of the same size, and the E-D tree of equal line-space pairs at a pitch of 0.52 μm . The region overlapped by all three trees is the common E-D region of this feature-shape combination. An operating point in the common region produces the image that meets the edge position requirement of all three edges of concern. The common E-D region is of concern not only for feature-shape combinations. Size combinations, mixtures of CD tolerances, nonuniformity of resist baking temperatures, bake times, development conditions, objects at different lens field positions, different fields, wafers, and wafer batches should be subjected to the common E-D region characterization as well.

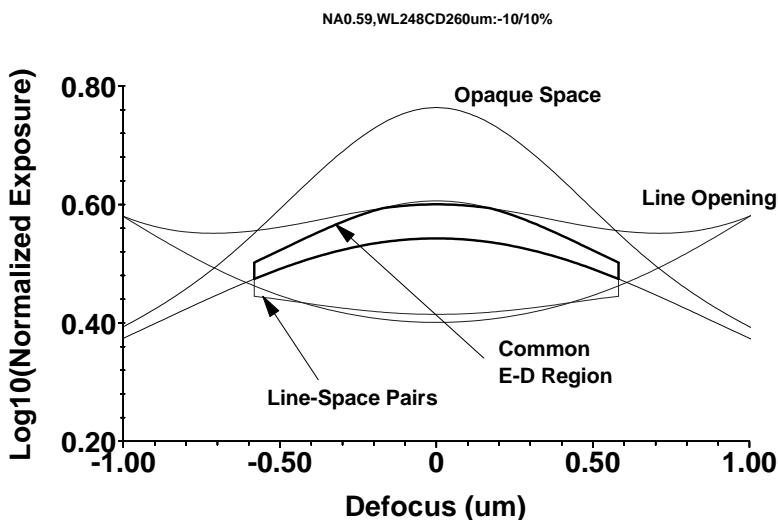


Figure 4.5 E-D trees of a line opening, an opaque space, and line-space pairs.

Strictly speaking, there are two E-D trees per feature because there are two edges in each feature. The common E-D region should be constructed from the superposition of six E-D trees instead of just three, as shown in Fig. 4.4. However, the lens used to image the above features is assumed to be aberration free. Therefore, the two edges of each feature move symmetrically in or out as a function of exposure and defocus, making the two E-D trees of each feature identical. The E-D tree of either edge can be used to represent the tree of the other edge and to represent the E-D tree plotted for constant-line-width contour, in addition to constant-edge-position contour. However, when there is asymmetry in the movement of the two edges, both edge trees must be used, and their common tree must be evaluated to assess the processing window. Figure 4.6 depicts the same object as in Fig. 4.4, but its edges are displaced by $0.01\text{ }\mu\text{m}$ to simulate lens distortion. The E-D trees of the edges are no longer identical to each other in terms of their relative position to the ideal edge positions, resulting in two separate trees whose common region is much less than before. This situation is depicted in Fig. 4.7, where edges 1 and 2 are no longer at the center of the edge placement window. Edge 1 has more room for overexposure, and edge 2 has less room. The consequence is reversed for underexposure. Hence, the E-D tree for edge 1 is higher than the linewidth-based tree, and the E-D tree for edge 2 is lower than the linewidth-based tree.

4.3.2 E-D window, DOF, and exposure latitude

A rectangle can be fitted in the E-D region of Fig. 4.4, as shown in Fig. 4.8. The height of this rectangle is the exposure latitude (EL) of the given E-D tree and the width, its DOF. For this particular rectangle, they are 36.7% and $0.85\text{ }\mu\text{m}$,

$\text{CD}=.26\text{um}, \text{dCDmin/max}=-10/10\%$ 62I50.iDf 16

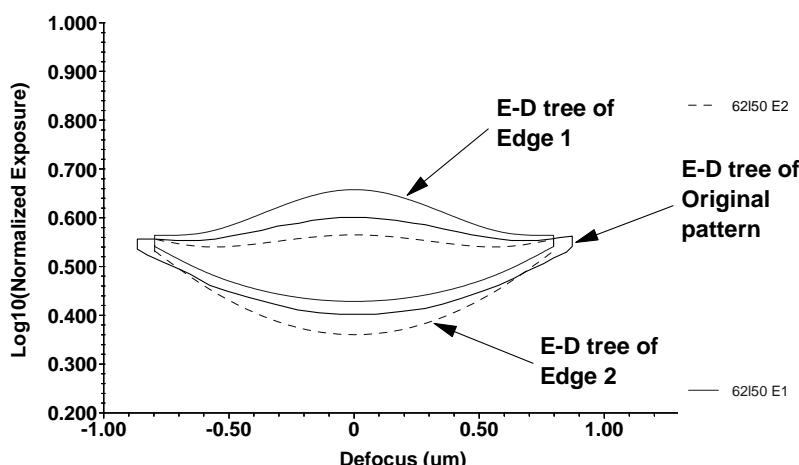


Figure 4.6 E-D tree of edge 1 shows the same line opening as in Fig. 4.4. In E-D tree of edge 2, edges are displaced by $0.01\text{ }\mu\text{m}$.

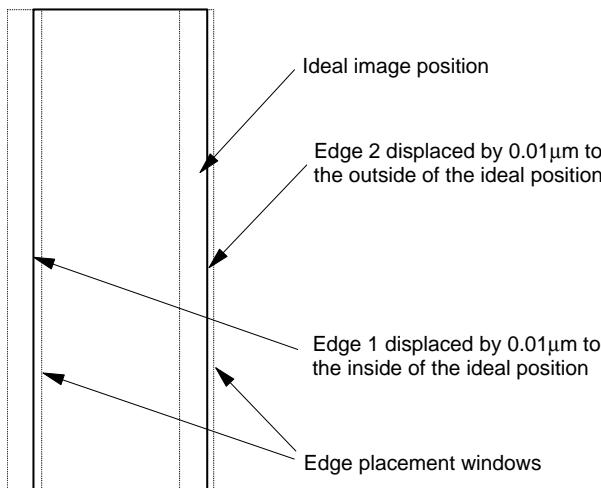


Figure 4.7 Image used in Fig. 4.6.

respectively. If these parameters are maintained, the resultant edge placement meets the desired specification. The parts of the E-D region outside the rectangle are not as useful and cannot be counted toward the total EL or DOF. For example, point A($-0.6, 0.5$), i.e., at defocus = $-0.6 \mu\text{m}$ and \log_{10} exposure = 0.5, cannot support an EL of 36.7%. Point B($-0.6, 0.46$) is within the 36.7% EL but is outside the usable E-D region; therefore, it violates the edge-placement specification.

One can fit another rectangle that has a width (i.e., DOF) of $1.2 \mu\text{m}$. However, the EL is reduced to 20.2%. Consequently, there is a tradeoff between EL and DOF; Fig. 4.9 depicts such a relationship. The EL is not the only parameter that changes with DOF. E-D area, the center of focus, and the center of exposure change according to the size, shape, and location of the new rectangle.

CD=.26um,dCDmin/max=-10/10%,Elat=36.7%,DOF=.85, 62I50.iDf

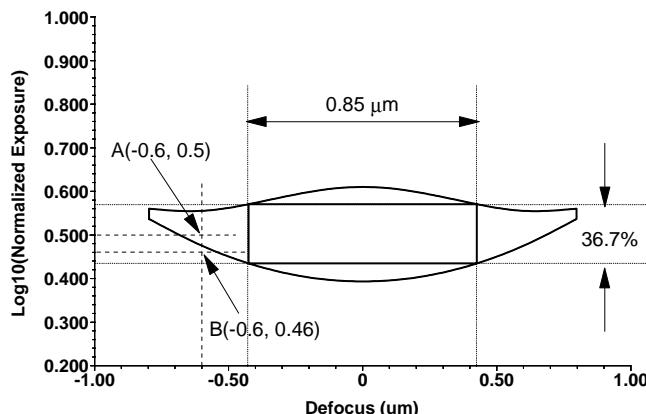


Figure 4.8 An E-D window is constructed in the E-D tree of Fig. 4.4.

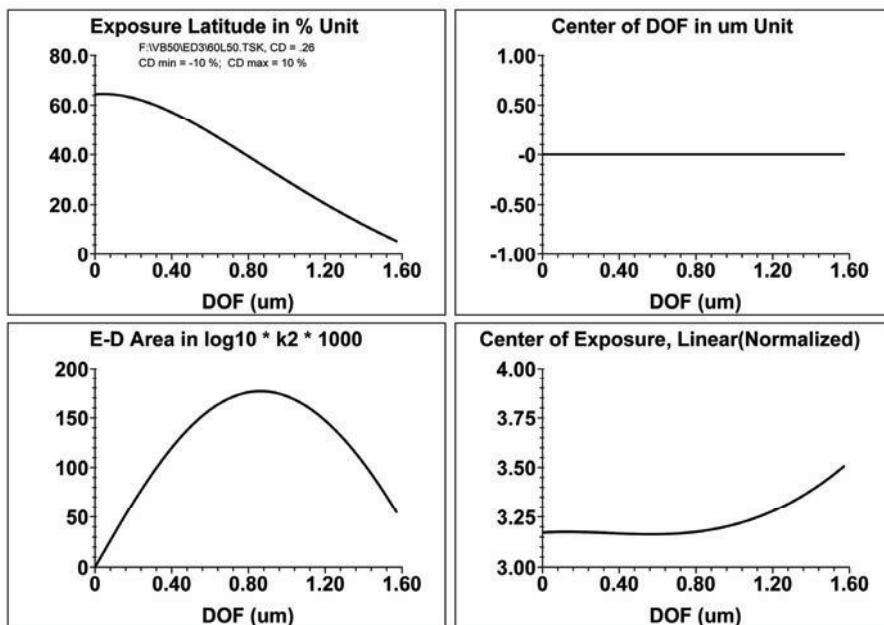


Figure 4.9 Exposure versus DOF, E-D area, center of focus, and center of exposure.

4.3.3 Determination of k_1 , k_2 , and k_3 using E-D windows

Evaluation of k_1 and k_2 using the E-D methodology starts with constructing the E-D tree of each component—namely, each feature size, feature shape, feature orientation, proximity environment, field position, and processing condition. The E-D trees are superimposed to evaluate the common E-D region. A rectangle is fitted to the E-D region. Depending on the lithography environment, the rectangle may be of a fixed height to restrict the exposure tolerance or of a fixed width to restrict the DOF. When neither needs to be restricted, one may choose to use the rectangle of the largest E-D area for the largest possible processing window.

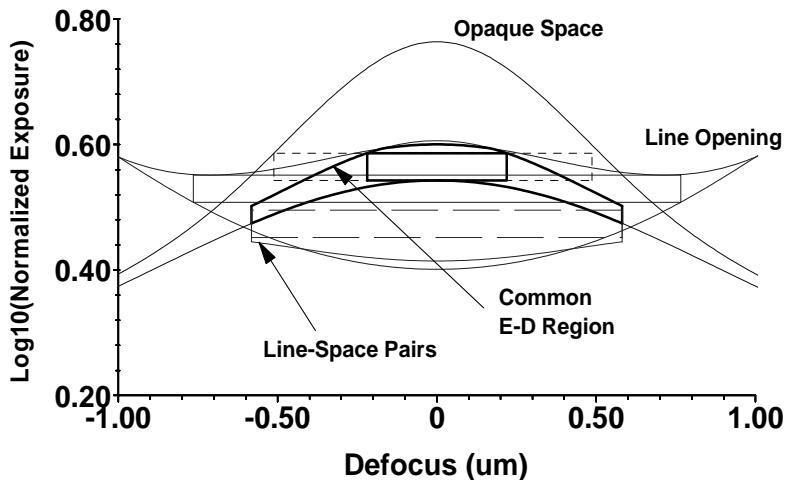
The situation leading to Fig. 4.4 is used as an example here. Rectangles are fitted to each individual tree as well as the common tree, as shown in Fig. 4.10. Here, 10% EL is imposed; therefore, all rectangles are of the same height. Table 4.1 shows the dimension of these rectangles, with the line opening producing the largest DOF, followed by the line-space pairs, then the opaque space. The exposure level required to achieve the largest DOF is different for each feature. The exposure level of the common tree and that of the opaque space are similar because the opaque space is the limiting feature in this combination.

When the E-D windows are optimized for the E-D area, Fig. 4.11 results. The E-D values of the new E-D windows are shown in Table 4.2. Now, the ELs of the three features are no longer identical to each other, and they are much larger than in the previous case, at the expense of DOF. It is only coincidental that the EL of the common E-D window is 10%.

Table 4.1 Dimensions of the common and individual E-D windows in Fig. 4.10.

Feature	EL	DOF (μm)	E-D area $\text{EL}(\log_{10}) \cdot k_2 \cdot 1000$
Common window	10%	0.44	26.8
Line opening	10%	1.53	93.8
Opaque space	10%	1.00	61.2
Line-space pairs	10%	1.16	71.4

NA0.59,WL248,E10%D0.44um,CD260um:-10/10%

**Figure 4.10** Individual and common windows fitted to the E-D tree in Fig. 4.5. All trees have 10% EL.

NA0.59,WL248,E10%D0.45um,CD260um:-10/10%

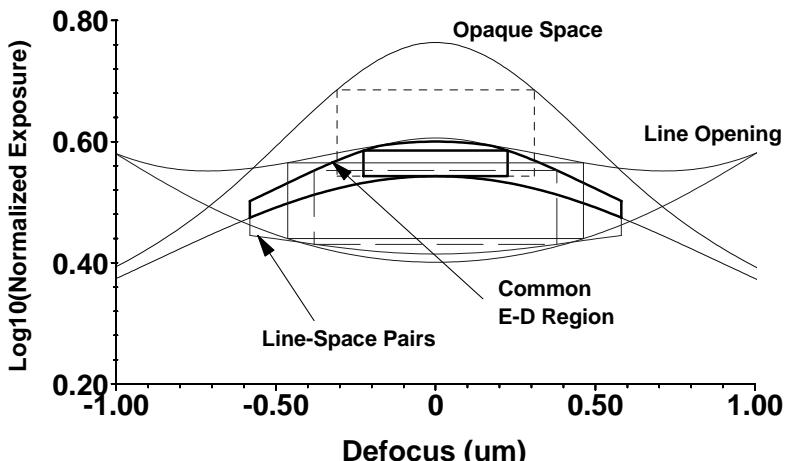
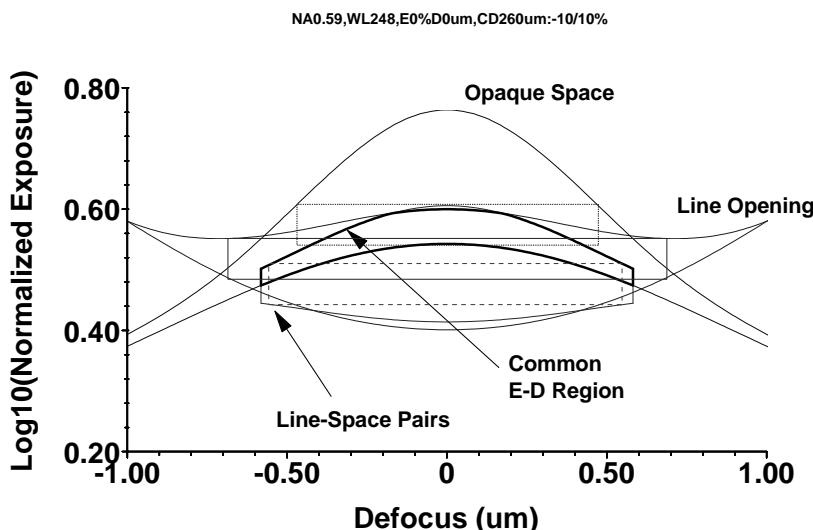
**Figure 4.11** Same as Fig. 4.10, but the optimized E-D area is the criterion for plotting the E-D windows.

Table 4.2 Dimensions of the common and individual E-D windows in Fig. 4.11.

Feature	EL	DOF (μm)	E-D area EL($\log_{10} \cdot k_2 \cdot 1000$)
Common window	9.73%	0.45	26.8
Line opening	28.6%	0.93	162.5
Opaque space	32.5%	0.62	123.6
Line-space pairs	27.9%	0.76	130.0

Using the first situation, the imaging system can resolve $k_1 = 0.26/0.248 \times 0.59 = 0.62$, with $k_2 = 0.45/0.248 \times 0.59^2 = 0.63$, and $k_3 = 0.45/0.248 \times 0.31^2 = 0.17$. However, one must qualify the set of coefficients by identifying feature combination, illumination, and EL. Because of the coincidence in the common E-D window, k_1 , k_2 , and k_3 for the optimized E-D area are also 0.62, 0.63, and 0.17, respectively. In general, the set can be different.

In these two cases, the resolution limit is not exceeded at $k_1 = 0.62$. With the same set of features, the resolution limit may already be exceeded if any of the qualifications are changed. For example, the result for 15% EL is shown in Fig. 4.12. There is no common E-D window because the height of an E-D window with 15% EL is larger than the vertical distance between the two E-D branches of the common E-D tree. The E-D values are shown in Table 4.3.

**Figure 4.12** Same as Fig. 4.10, except EL is increased to 15%.**Table 4.3** Dimensions of the common and individual E-D windows in Fig. 4.12.

Feature	EL	DOF (μm)	E-D area EL($\log_{10} \cdot k_2 \cdot 1000$)
Common window	15%	0	0
Line opening	15%	1.37	125.3
Opaque space	15%	0.94	85.8
Line-space pairs	15%	1.11	101.7

4.4 k_1 , k_2 , and k_3 as Normalized Lateral and Longitudinal Units of Dimension

We have been using k_1 , k_2 , and k_3 as the coefficients for resolution and DOF, as in Eqs. (4.1) to (4.4). They can be defined as normalized lateral and longitudinal dimensions:¹

$$k_1 = \left(\frac{W}{\lambda} \right) NA, \quad (4.5)$$

$$k_2 = \left(\frac{Z}{\lambda} \right) NA^2, \quad (4.6)$$

and

$$k_3 = \left(\frac{Z}{\lambda} \right) NHA^2. \quad (4.7)$$

Examples of the lateral dimension include resolution, linewidth, line length, edge position, distance between features, and any distance or location on the mask or wafer plane. Longitudinal dimensions include DOF, defocus, and film thickness. It is easier to compare systems with normalized dimensions. For example, a resolution of 0.26 μm at 248 nm and 0.59 NA is equal to a resolution of 0.62 λ/NA, or simply 0.62 k_1 units. This is the same level of resolution as 0.38 μm at 365 nm and 0.59 NA, or 0.34 μm at 248 nm and 0.45 NA. All are at $k_1 = 0.62$.

4.5 The E-D Tools

The E-D tools consist of E-D branches, trees, forests, and windows. Everything except E-D forests have been covered in Secs. 4.3.1 and 4.3.2 above. The E-D forest is simply a collection of E-D trees too numerous for the human mind. In this section, the method for constructing E-D trees is explained. Examples of E-D tool applications are also given.

4.5.1 Construction of E-D trees

E-D trees can be constructed from three types of image data—namely, linewidth as a function of exposure and defocus, edge position as a function of exposure and defocus, and image intensity distribution at a series of defocus planes.

4.5.1.1 E-D tree construction from E-D matrix linewidth data

An E-D tree is constructed from linewidth as a function of exposure and defocus by evaluating the constant linewidth contours in the E-D space. Figure 4.13 shows the linewidth using E-D matrix data with the data points marked at the E-D locations. The constant linewidth contours are evaluated using least-mean-square polynomial fitting. Fitting the data is preferable to interpolating it directly,

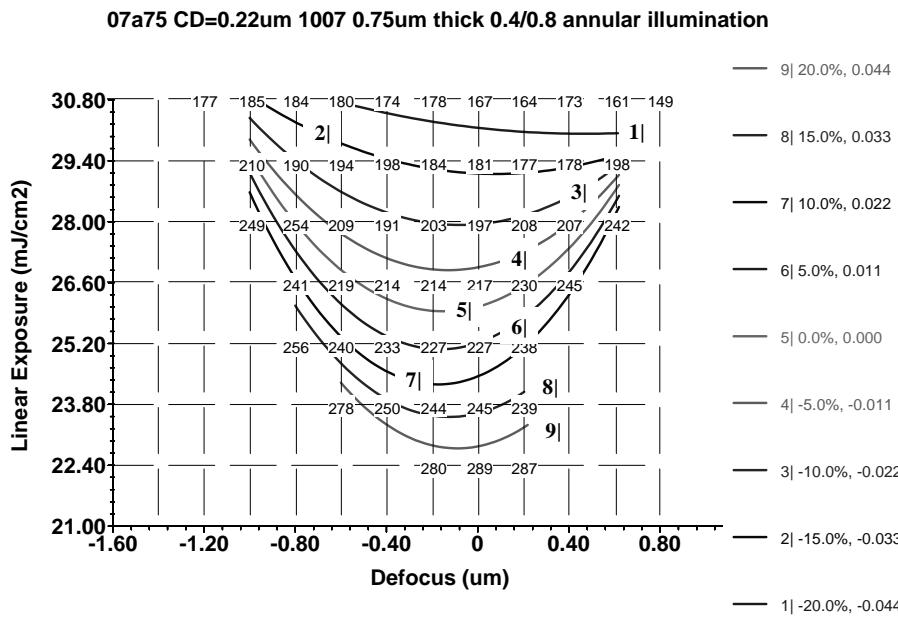


Figure 4.13 Constant linewidth contours from polynomial curve fitting.

unless the data is from a simulation and there is absolutely no possibility of noise in the data. Most experimental data can benefit from least-mean-square fitting. Figure 4.14 shows the same data but with spline-fitted contours. Spline fitting forces the contours to pass through all data points exactly. Experimentally induced noise is preserved.

Let's cover the numerical details in constructing Fig. 4.13. Take the -5% contour, with nominal linewidth $CD = 0.22 - 0.011 = 0.209 \mu\text{m}$, marked as -5% and -0.011 on the chart. The data point is first linearly interpolated for exposure at each defocus point. For example, at $z = -0.4 \mu\text{m}$, $26.6 \text{ mJ}/\text{cm}^2$ produces a 214-nm feature, and $28 \text{ mJ}/\text{cm}^2$ produces a 191-nm feature. From linear interpolation, the required exposure dosage to produce $0.209 \mu\text{m}$ is $26.9 \text{ mJ}/\text{cm}^2$. After the linearly interpolated dosages at all defocus points are evaluated for a 209-nm contour, a polynomial fit is performed. The $26.9 \text{ mJ}/\text{cm}^2$ point becomes $27.16 \text{ mJ}/\text{cm}^2$ after polynomial fitting, as shown in the contour.

4.5.1.2 E-D tree construction from E-D matrix edge data

If edge-position data as a function of the E-D coordinates is available, it is more desirable to construct an edge-based E-D tree from that data. The procedure is similar to that of using linewidth E-D matrix data, except that when curve fitting, one fits constant edge position contours to the data as shown in Fig. 4.15. Fitting can be done by least-mean-square polynomial fit or spline. In Fig. 4.15 only half of the data points are shown because the data is symmetrical in defocus.

07a75 CD=0.22um 1007 0.75um thick 0.4/0.8 annular illumination

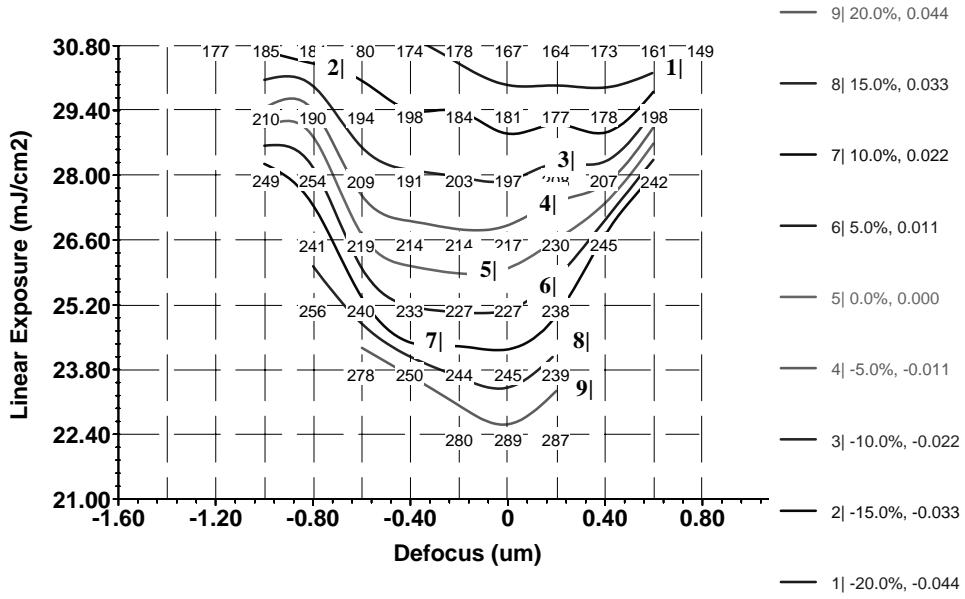


Figure 4.14 Same as Fig. 4.13, except a spline curve fitting is used.

ref CD=500.nm .5k1 .5sig 5x5cell PR914 910R/250P/25X 90C30PrB 100C1PEB 200/.6E 6

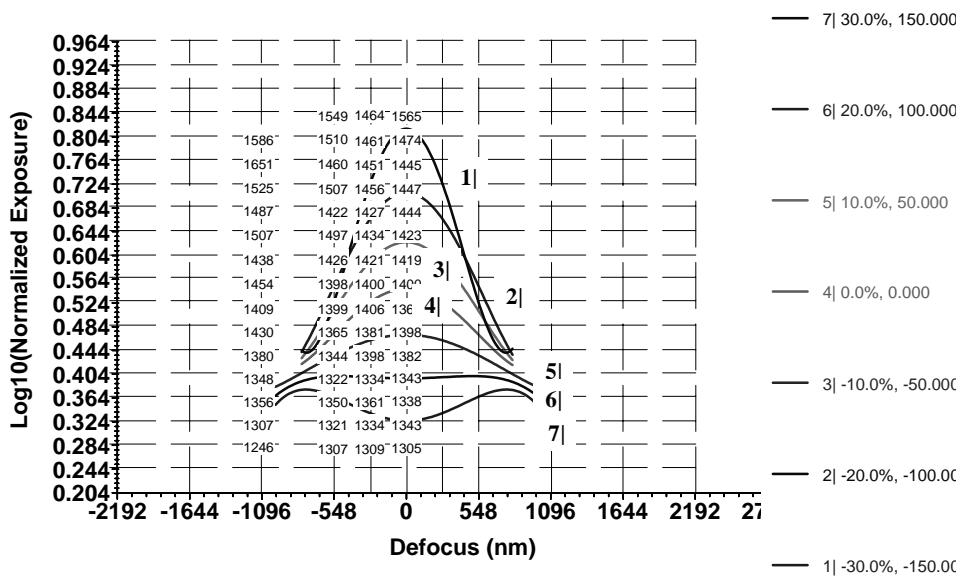


Figure 4.15 Constant edge-position contours from polynomial curve fitting.

4.5.1.3 E-D tree construction from intensity distribution

To construct an E-D branch from the intensity distribution of an image, one first identifies the edge location for the branch of interest. For example, branch 0 is from the location of edge position tolerance 0 on the intensity distribution plot shown in Fig. 4.16. Find the intensity at this edge position at each defocus position. Figure 4.16 uses three defocus positions to produce the points (I_0, D_0, B_0) , (I_1, D_1, B_0) , and (I_2, D_2, B_0) for branch 0 and three other points for branch 1. Here I stands for intensity, D stands for defocus, and B stands for branch. The intensity is converted to exposure by taking the reciprocal of intensity and producing the points (E_0, D_0, B_0) , etc., in the E-D plane, as shown in Fig. 4.17. E-D branch 0 is obtained by joining all E-D points corresponding to branch 0. E-D branch 1 is obtained similarly.

4.5.2 Importance of log scale in the exposure axis

Intuitively, one tends to plot the E-D trees in the linear scale for both exposure and defocus. However, because exposures should be represented by ratios, the log scale is the right way to plot. This way, the horizontal and vertical distances in the E-D region can be compared directly. Let there be three exposures: E_1 , E_2 , and E_3 . If plotted in log, the distance between any two points is $\log E_2 - \log E_1$, $\log E_3 - \log E_2$, and $\log E_3 - \log E_1$, indicating the log of ratios E_2/E_1 , E_3/E_2 , and E_3/E_1 . If $E_2 = 2E_1$ and $E_3 = 2E_2$, then $\log E_2 - \log E_1 = \log E_3 - \log E_2 = \log 2$. But, $E_2 - E_1 = E_1$ and $E_3 - E_2 = 2E_1$. Therefore, the distance between E_3 and E_2 is identical to that between E_2 and E_1 on the log scale but doubled on the linear scale. The EL determined from the distance between E values on the log scale is also identical, as shown by Eq. (4.8).

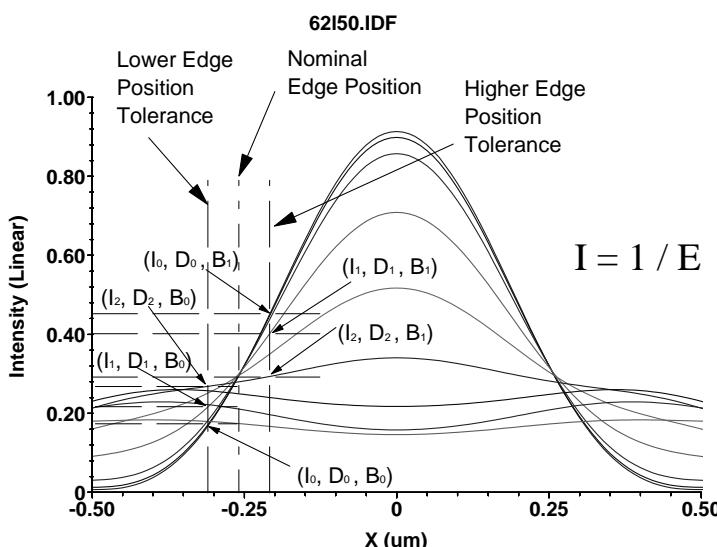


Figure 4.16 The intensity distribution curves at various defocuses and the identification of the (I, D, B) points.

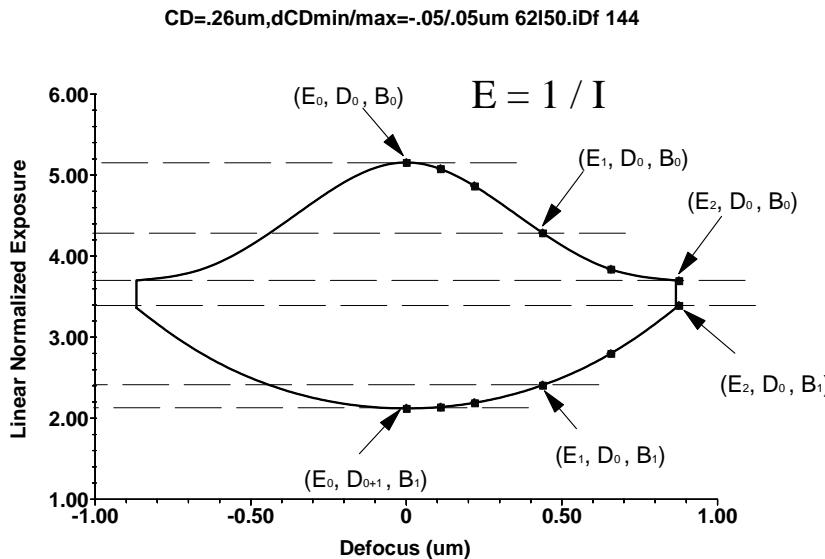


Figure 4.17 The E-D tree and branches composed with the (E, D, B) points from the (I, D, B) points in Fig. 4.16.

$$\text{Exposure latitude} = 2 \frac{E_3 - E_2}{E_3 + E_2} = 2 \frac{E_2 - E_1}{E_2 + E_1} = 2 \frac{e^{(\log E_2 - \log E_1)} - 1}{e^{(\log E_2 - \log E_1)} + 1}. \quad (4.8)$$

Figure 4.18 shows that two E-D trees with an identical EL can appear to be different in size when plotted in linear scale during exposure. Conversely, the trees plotted in log scale correctly reflect the relative size of their EL and thus help the visual comparison of the process window.

4.5.3 Elliptical E-D window

The E-D window defined with rectangles indicates the absolute range of exposure and defocus allowed by the imaging process. It is too strict for manufacturing. The exposure and defocus variation in a high wafer-flow situation often exhibits Gaussian distribution. Violation of the rectangular boundary at the corners is tolerable. This leads to elliptical E-D windows, as shown in Fig. 4.18. The EL and DOF defined by the two diameters of the ellipse are no longer absolute limits, but they are statistical in nature. They are related to a certain number of standard deviation, such as 6σ . The actual number of standard deviation is yet to be defined with a large number of CD versus exposure and defocus data. Figure 4.19 shows the E-D tree of a 220-nm line controlled to ± 25 nm using $NA = 0.55$ and $\lambda = 248$ nm. E-D windows with 8% EL are plotted in elliptical and rectangular forms. The DOF defined by the elliptical E-D window is 1330 nm, whereas the DOF from the rectangular E-D window is 1090 nm. The difference is 22%. When the exposure tolerance is

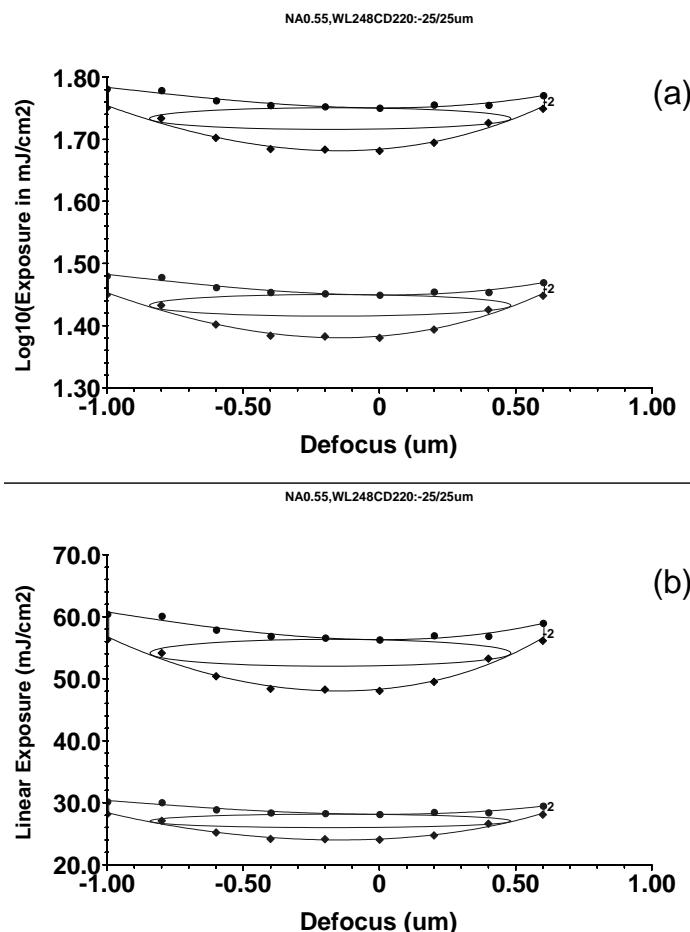


Figure 4.18 Two E-D trees with the same EL but different exposure dosages. (a) In log scale, the trees are identical except for their vertical locations. (b) In linear scale, the tree with more exposure appears to be larger.

changed to 15% as depicted by Fig. 4.20, then the DOF are 960 and 350 nm, respectively, resulting in a change of 267%. At extremely small EL or DOF, the difference in the two types of windows is insignificant; however, near the maximum E-D area, the difference can be substantial.

4.5.4 EL-versus-DOF tradeoff

A series of EL is used to evaluate the corresponding DOF by constructing the respective E-D window. These EL-versus-DOF curves show the tradeoff between EL and DOF, in other words, their mutual dependence. Such a tradeoff curve for the feature used for Fig. 4.19 is shown in Fig. 4.21. The curve based on elliptical E-D windows has a sharper bend than the one based on rectangular windows. This means that the common practice of defining EL at DOF = 0 and DOF at EL = 0 better resembles the situation of the elliptical E-D windows.

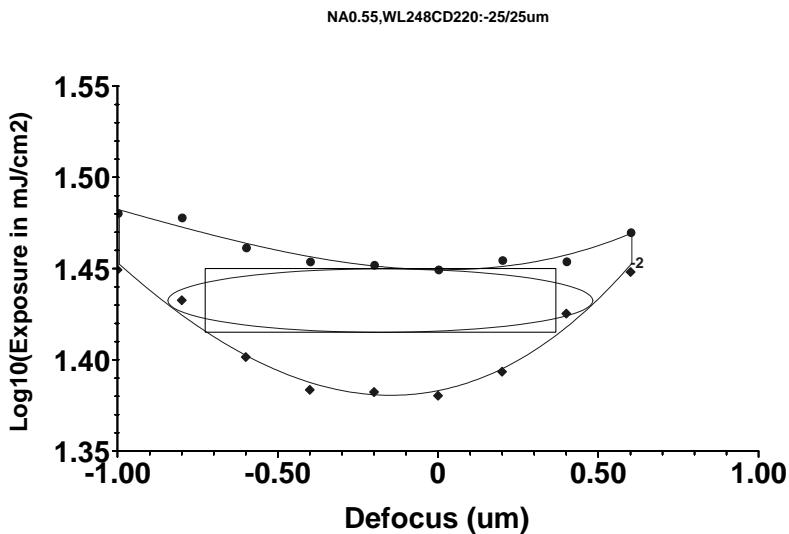


Figure 4.19 E-D tree of a 220-nm line controlled to ± 25 nm. NA = 0.55, $\lambda = 248$ nm. Elliptical and rectangular E-D windows with 8% EL are shown.

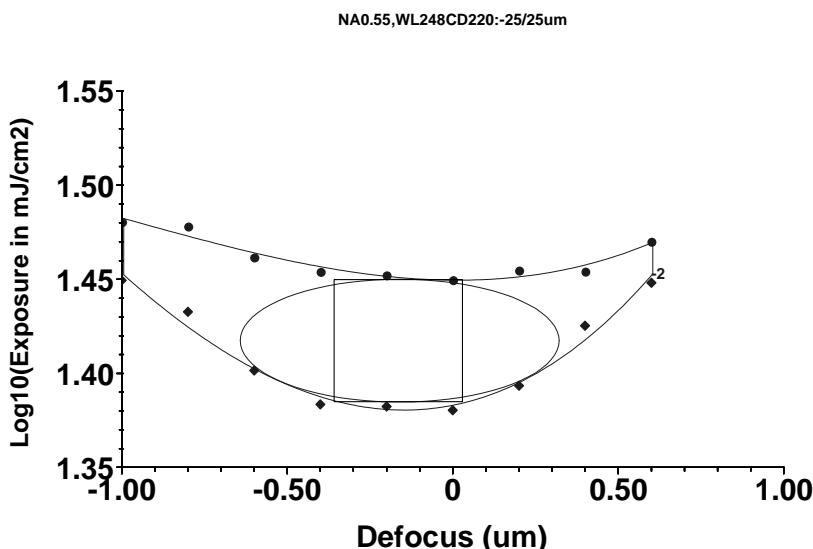


Figure 4.20 E-D tree of the same feature used in Fig. 4.19, except elliptical and rectangular E-D windows are drawn with 15% EL.

4.5.5 Incorrect elliptical E-D window

Many authors have erroneously plotted a perfect ellipse in the linear E-D space. A perfect ellipse only makes sense when plotted in the log space for exposure. The correct method is to work in the log space, construct the ellipse, and then convert it to the linear space. Fortunately, the range of exposure is usually small, making the error not so visible.

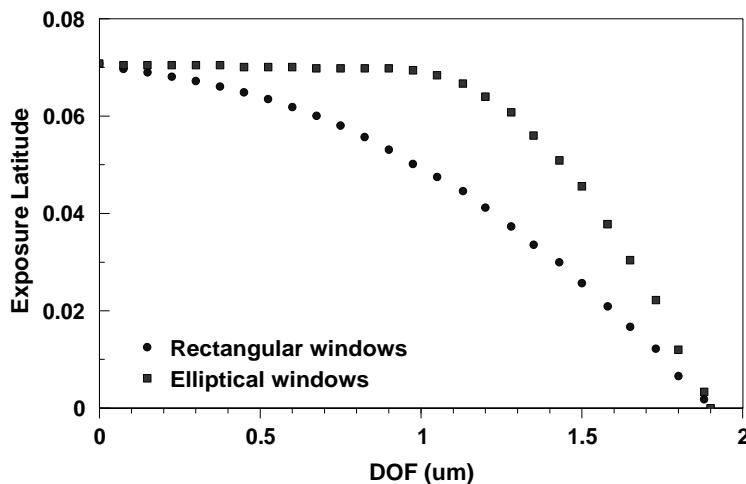


Figure 4.21 EL versus DOF for the 220-nm feature in Fig. 4.19. The curve taken from elliptical windows exhibits a squarer shape.

4.5.6 CD-centered versus full-CD-range E-D windows

Adopting the concept of statistical distribution of exposure and defocus, one assumes that the most popular occurrence of exposure and defocus is at the set point. Hence, setting the center of the ellipse to the $\Delta CD = 0$ contour ensures that the most frequent linewidth is the nominal CD. This ellipse is the CD-centered (CDC) E-D window. What we used to define was the full-CD-range (CDR) E-D window. Both are depicted in Fig. 4.22 with E-D windows of the maximum E-D area. The tradeoff of using the CDC window is a smaller EL and DOF. For example, the EL and DOF of the CDC window are 8% and 1.03 μm , compared with 8% and 1.33 μm of the CDR window.

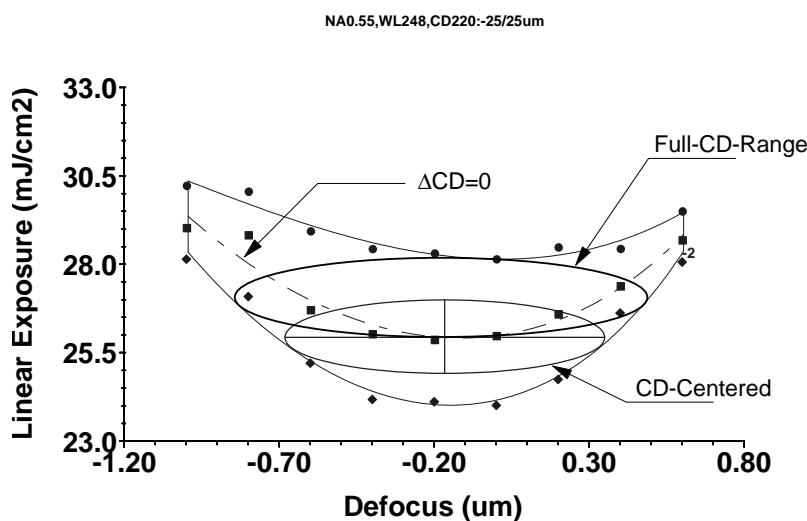


Figure 4.22 CD-centered and full-CD-range E-D windows of the feature in Fig. 4.19.

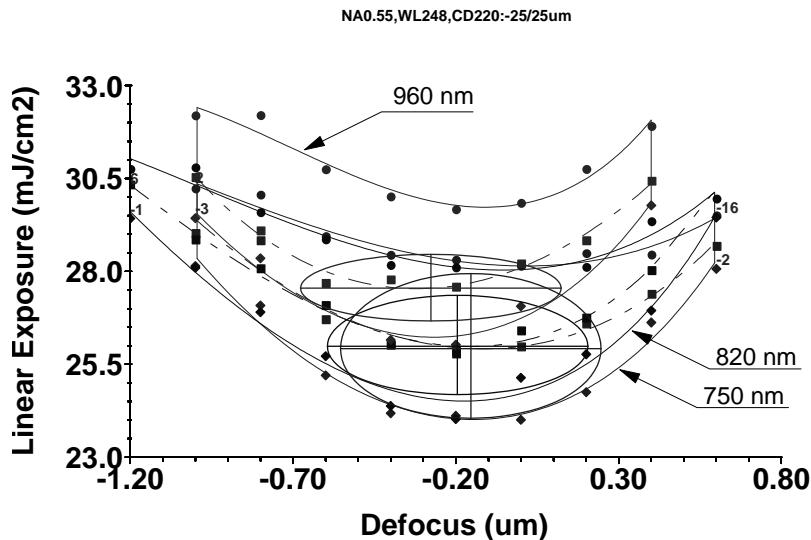


Figure 4.23 CDC E-D windows for three resist thicknesses.

Having a smaller E-D window is not the most severe problem with CDC windows. A common E-D window becomes meaningless because no matter which individual window is centered, there can only be one feature that centers on the nominal exposure. As shown in Fig. 4.23, each E-D window of constant DOF centers at a unique (E, D) coordinate, even though those of 750- and 820-nm resist thickness are much closer than that of 960-nm thickness. In comparison, Fig. 4.24 shows the individual and common CDR E-D windows that share an identical DOF of 800 nm. The E-D parameters are shown in Table 4.4.

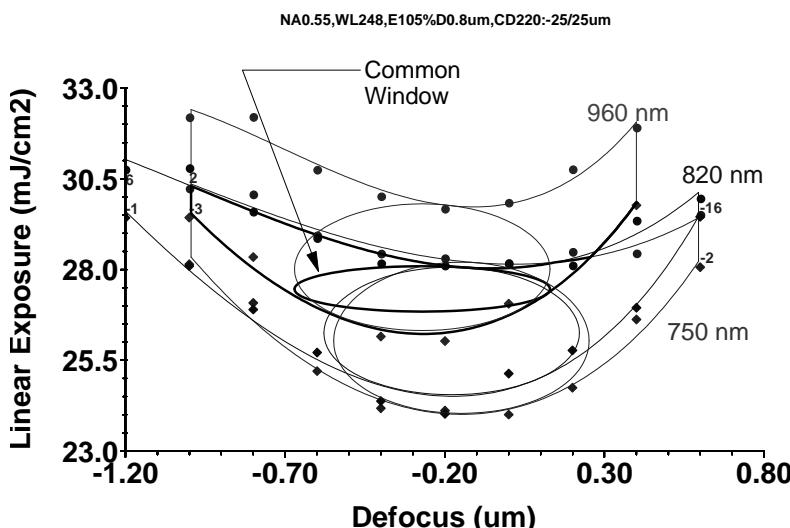


Figure 4.24 CDR E-D windows for three resist thicknesses.

Table 4.4 CDR and CDC E-D window parameters.

Resist thickness (nm)	E_{min} (mJ/cm ²)	EL %	D_{min} (nm)	DOF (nm)	E-D area $EL(\log_{10}k_2 \cdot 1000)$	E_{center} (mJ/cm ²)	D_{center} (nm)
Common	26.84	4.55%	-671	800	19.29	27.46	-271
750	24.03	15.95%	-548	800	67.75	25.95	-148
820	24.55	13.36%	-578	800	56.70	26.19	-178
960	26.32	12.44%	-671	800	52.78	27.95	-271
750 CD _{centered}	24.05	14.98%	-554	800	63.60	25.85	-154
820 CD _{centered}	24.68	10.27%	-595	800	43.54	25.95	-195
960 CD _{centered}	26.66	6.50%	-677	800	27.56	27.53	-277

One way to center all of the windows to the nominal exposure is by mask biasing. Each feature is given a specific bias on the mask to result in the nominal CD at a given identical exposure. For 1D features, this is possible. An example can be seen in Fig. 4.49.

4.5.7 E-D window and CD control

The E-D window defines the EL and DOF of a given feature or a combination of features. When a point in the E-D space is outside of the window, the CD tolerance is exceeded. The point cannot be considered for EL and DOF. However, given two E-D trees enclosing an identical E-D window, the one with more area outside the window has a better chance of meeting the exposure or focus requirement. Thus, this tree indicates better CD control. For example, in the two E-D trees shown in Fig. 4.25, the artificially created E-D tree has a larger area near focus. This artificial E-D tree is defined by the lower branch in the dotted line and the upper branch that shares with the experimental tree. It provides better CD control near focus. Conversely, CD control is better for the experimental tree when focus is not so close to the focal plane.

4.5.8 Application of E-D tools

The power of the E-D tools lies in many aspects. First, they provide a rigorous definition of the EL and DOF of a given optical imaging system as well as their mutual dependence. Second, the components can be superimposed for their common E-D window. Third, systems can be compared with identical criteria. These features lead to many important and interesting applications.

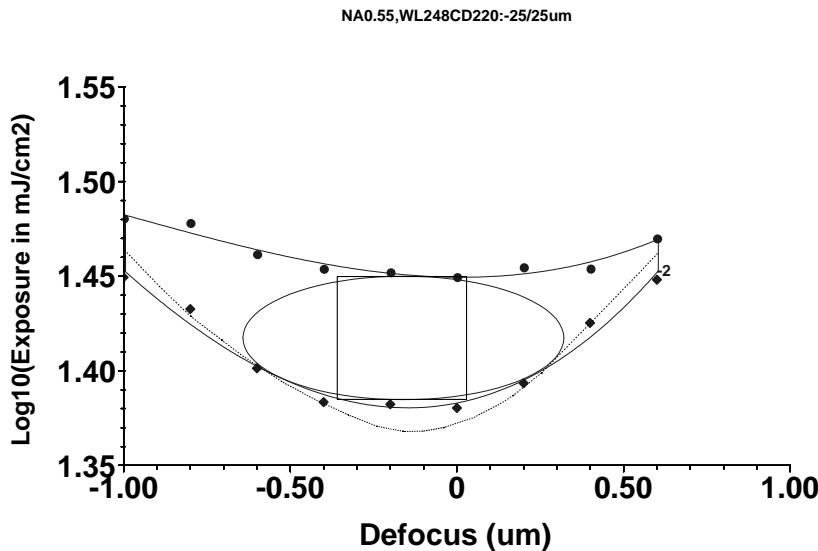


Figure 4.25 Modified from Fig. 4.20. An extra E-D branch using a dotted line is artificially added to illustrate the difference in CD control with different E-D trees sharing an identical E-D window.

4.5.8.1 Combination of feature types

Figures 4.5 and 4.10 are examples of the combination of line opening, opaque line, and line-space pairs. From these figures, one readily observes the distinct shapes associated with different features and their various locations in the E-D space. The individual trees suggest ample processing windows for each feature. However, because they have different exposure centers, the common E-D window is quite small, making it difficult to delineate all three features to the same linewidth criterion at an identical exposure. The E-D trees also show that the line opening and line-space pairs can be biased as a group to produce a larger common window with the opaque line or vice versa.

4.5.8.2 Combination of feature sizes

A common situation in circuit fabrication is to use the same gate length but to allow the gates to be separated differently from each other. This situation is similar to mixing line-space pairs of different line-to-space ratio. With a positive resist, the mask consists of opaque gate lines of identical width. Here, the E-D trees and the common E-D window of a mixture of 1:1, 1:1.5, and 1:2 opaque-to-transparent ratio is shown in Fig. 4.26. The gate width is kept at $0.13 \mu\text{m}$ using $\sigma = 0.5$, $\text{NA} = 0.6$ and $\lambda = 193 \text{ nm}$, i.e., the gate length is at $k_1 = 0.4$. Even though the gate length is identical, because of different separations, there is no common window.

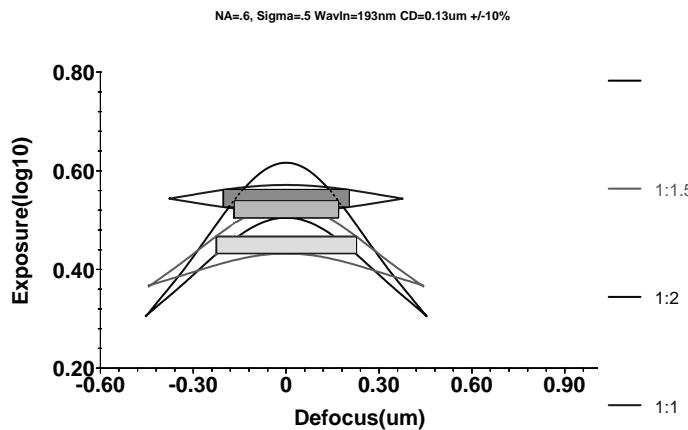


Figure 4.26 Combination of line-space pairs with different line-space ratios of 1:1, 1:1.5, and 1:2.

To find the NA/ σ combinations for a common window, all E-D trees from NA = 0.3 to 0.9 at $\sigma = 0.86$ are plotted in Fig. 4.27. At NA = 0.57 and above, the common E-D window appears. The DOF set at 8% EL is plotted as a function of NA in Fig. 4.28. The DOF of the individual trees is obviously much larger than that of the common tree. Furthermore, the 1:1 feature has a higher peak DOF than that of the 1:1.5 feature, which, in turn, has a higher peak DOF than that of the 1:2 feature. Intuitively, dense features are considered more difficult to print than features with larger separations. E-D tree analysis reveals that even though the features with larger separations have larger EL at the focal plane, the DOF is smaller. This is because the trees have more curvature and can support only a small DOF. Figure 4.29 plots DOF as a function of NA and σ . The DOF is optimized at NA = 0.67 and $\sigma = 0.7$.

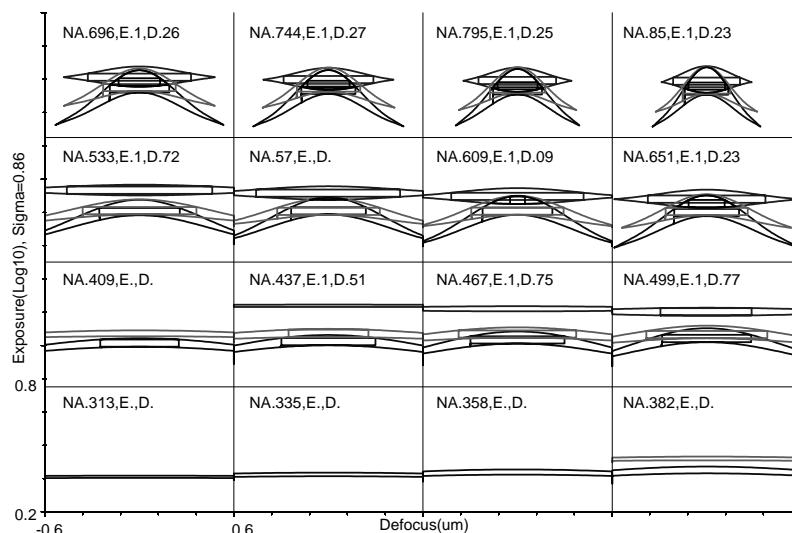
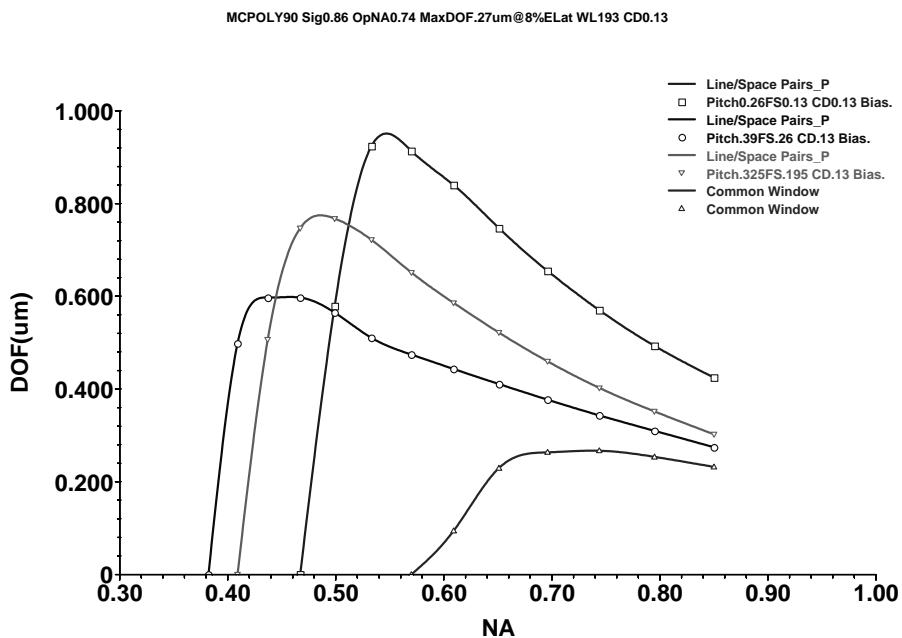
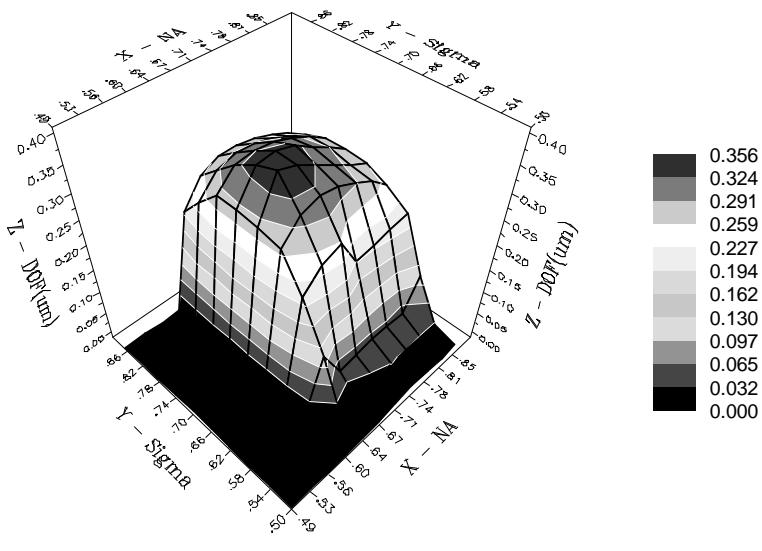


Figure 4.27 Same as Fig. 4.26 but plotted with NA = 0.3 through 0.9, at $\sigma = 0.86$.



Rectangular Window

Figure 4.28 DOF versus NA taken from Fig. 4.27.



Elat = 0.08 Wv ln = 193nm CD = 0.13um

Figure 4.29 DOF versus NA and σ from the features in Fig. 4.27.

Another common situation is to mix contact holes of different pitches. Figure 4.30 shows DOF versus σ from the E-D trees of 0.13- μm contact holes at $\lambda = 193 \text{ nm}$ and $\text{NA} = 0.64$, i.e., $k_1 = 0.43$, with pitches of 0.26, 0.325, and 0.39 μm combined on the same mask. This set of contact holes has hole-to-separation ratios of 1:1, 1:1.5, and 1:2, respectively. An attenuated phase-shifting mask (AttPSM) with 6% transmission is used. The illumination is annular with a 50% inside diameter. We will discuss annular illumination further in Sec. 6.3.2. The trees are constructed at $\sigma = 0.44$ to 1 in 0.04 steps, and the E-D windows have an 8% fixed EL.

The target dimension of the contact holes is 0.13 μm controlled to $\pm 10\%$. However, the actual size is 0.1 μm ; that is, they are biased by 30 nm. Without the bias, the DOF of these holes would be much smaller. These holes must be overexposed to reach the target image size. To further improve DOF, the bias can be individually adjusted. See Sec. 6.3.4 for detailed coverage.

4.5.8.3 Combination of cuts for 2D features

With 1D features, just one cut at the edge of the feature is sufficient to evaluate its E-D characteristics. Combining feature sizes and types still requires one cut per type and per size. If edge E-D trees are needed due to asymmetry of the feature layout or aberrations in the imaging system, the number of cuts may be increased to—at most—two per feature in the layout. With 2D features, the edge of the feature cannot be represented by just one or two cuts. Different locations or edges on one feature often must be maintained at the same edge-control specification.

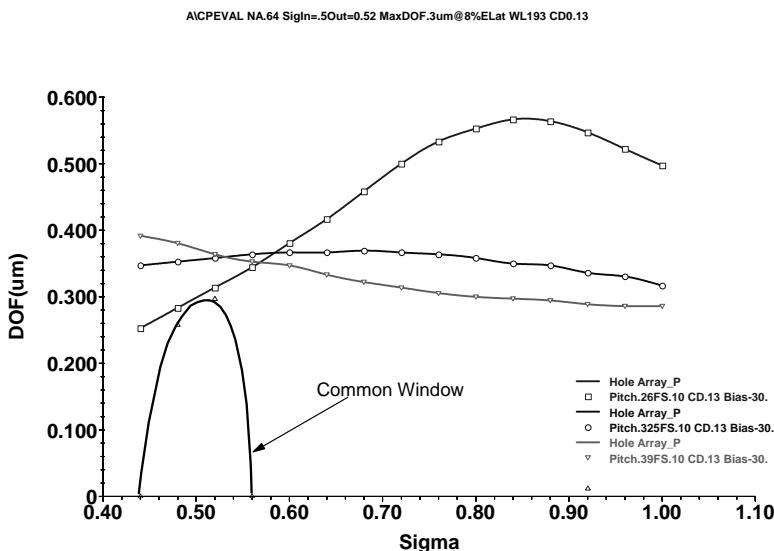


Figure 4.30 DOF versus σ from the E-D trees of 0.13- μm contact holes at $\lambda = 193 \text{ nm}$ and $\text{NA} = 0.64$. This set of contact holes has hole-to-separation ratios of 1:1, 1:1.5, and 1:2, respectively. 6% AttPSM is used. The illumination is annular with a 50% inside diameter.

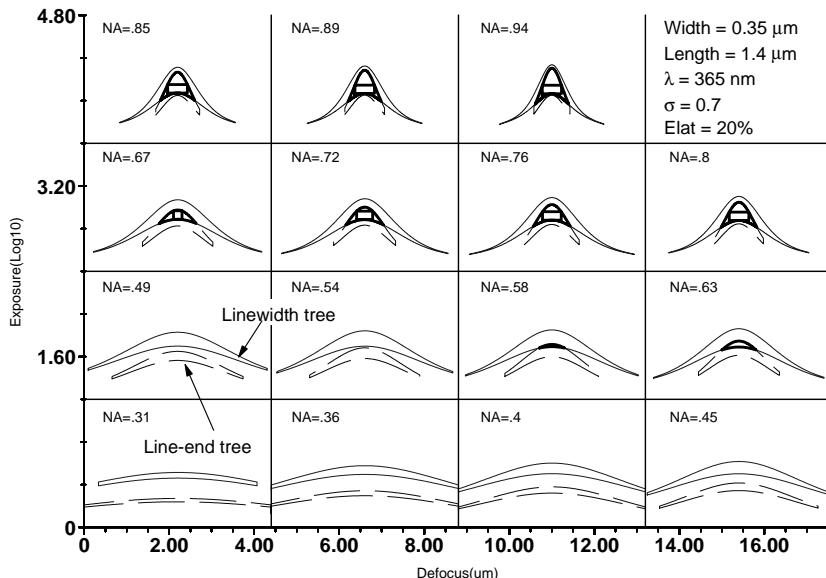


Figure 4.31 E-D trees of an opaque space and its end.

For example, the width and length of a resist line must be kept within the CD tolerance specification. Figure 4.31 shows the E-D trees of a $0.35\text{-}\mu\text{m}$ line $1.4\text{ }\mu\text{m}$ long, illuminated at $\sigma = 0.7$ with a 365-nm wavelength. Both the width and the length are controlled to $\pm 35\text{ nm}$ their nominal edge position. At $\text{NA} = 0.63$ and below, even though each image edge can be controlled within specification at their own exposure dosage as depicted by the existence of their individual trees, there is no unique exposure dosage to satisfy both edges. In practice, because line ends are more susceptible to linewidth variations, a less stringent specification of CD tolerance is given, making the common E-D window larger.

In addition to line ends, the variation of edge position along a line in other 2D situations is also of concern. For example, the CD variation near the junction of a T, an L, or a disconnected T must be represented by many cuts at appropriate places. A typical application will be discussed in Sec. 6.4.2.2 on model-based optical proximity correction, and depicted by Figs. 6.103 and 6.104.

4.5.8.4 Combination of CD tolerances

E-D trees plotted at different CD tolerances can be combined to study the effect of different linewidth-control requirements. Figure 4.32 shows the DOF evaluated with an elliptical E-D window of 10% EL for $0.1\text{-}\mu\text{m}$ positive resist lines in a $0.3\text{-}\mu\text{m}$ pitch, and $0.2\text{-}\mu\text{m}$ lines in a $0.6\text{-}\mu\text{m}$ pitch. The CD tolerance set for the E-D branches is $\pm 10\text{ nm}$ for either line group. It is seen that at a low NA, the $0.1\text{-}\mu\text{m}$ line group gates DOF, whereas when $\text{NA} \geq 0.7$, DOF is gated by the wider line group. When only $\pm 20\text{ nm}$ is required of the $0.2\text{-}\mu\text{m}$ line group, there is a DOF gain at a higher NA because the narrower line group is now gating the entire NA range, as shown in Fig. 4.33.

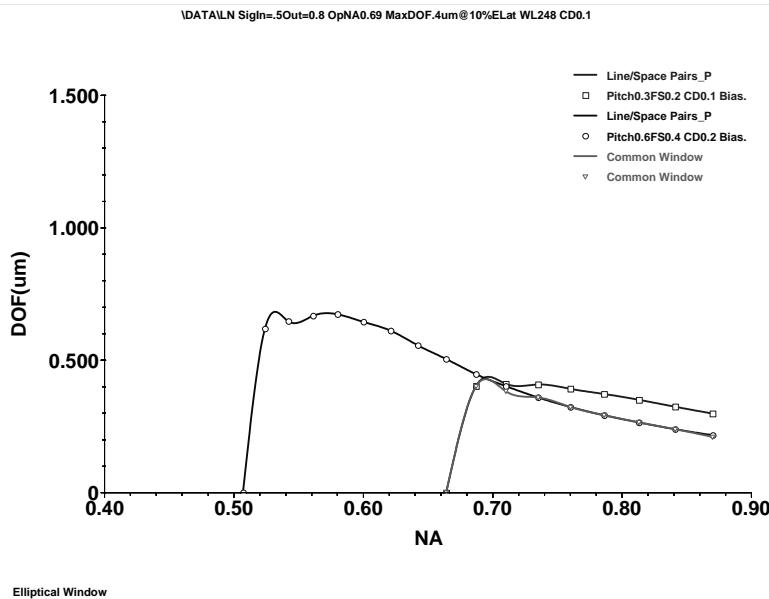


Figure 4.32 DOF versus NA for a 0.1- μm positive resist line in a 0.3- μm pitch, and a 0.2- μm line in a 0.6- μm pitch. CD tolerance is set to ± 10 nm.

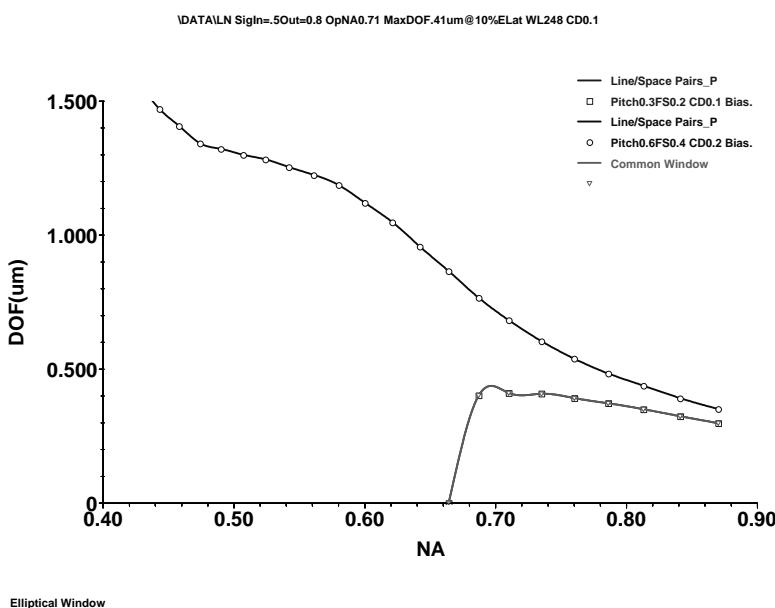


Figure 4.33 Same as Fig. 4.32, except CD tolerance is ± 20 nm for the wider resist line.

4.5.8.5 Combination of resist processing tolerances

When the resist processing conditions change, the E-D trees can take different shapes. As will be discussed in Sec. 5.6.2.5, the resist thickness plays an important role in linewidth control because of multiple reflections. Figure 4.34 shows the E-D trees and windows for a 220-nm line imaged with $NA = 0.55$, $\lambda = 248$ nm, and a CD tolerance of ± 25 nm; the only parameter changed was the resist thickness. Each thickness supports an E-D tree with more than sufficient EL and DOF. However, the center of exposure and DOF change substantially, especially for the 960-nm thickness. If there were no reflection, the adjustment of the focus center would be the difference between 960 and 750 nm divided by the refractive index of the resist, resulting in a focus shift of about 112 nm, assuming a resist refractive index of 1.7. From Table 4.5, the shift of the center of focus is 119.9 nm. Multiple reflections did not significantly shift the focus in this case. However, the exposure center and range are quite different between the 750-nm and 960-nm thicknesses. As a result, the EL and DOF of the common window are extremely reduced. Multiple reflections do affect exposure dosage, as predicted later in Sec. 5.6.2.5.

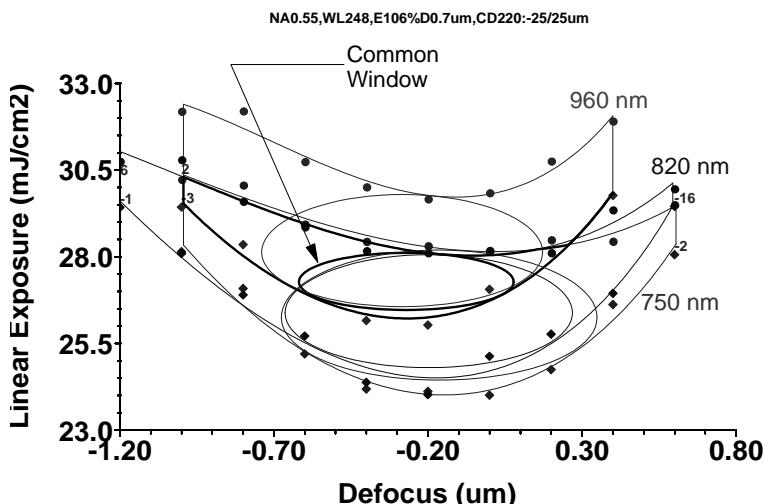


Figure 4.34 E-D trees and windows for resist thicknesses 750, 820, and 960 nm. The feature is a 220-nm line, imaged with $NA = 0.55$, $\lambda = 248$ nm, and a CD tolerance of ± 25 nm.

Table 4.5 E-D parameters for the windows in Fig. 4.34.

Resist thickness (nm)	E_{min} (mJ/cm ²)	EL (%)	D_{min} (nm)	DOF (nm)	$E\text{-D area}$ $EL(\log 10)\cdot k_2 \cdot 1000$	E_{center} (mJ/cm ²)	D_{center} (nm)
Common	26.47	6.04%	-619	697	22.31	27.27	-271
750	24.44	14.25%	-676	1025	77.49	26.19	-164
820	24.80	12.33%	-664	933	60.98	26.33	-198
960	26.56	11.47%	-740	912	55.49	28.08	-284

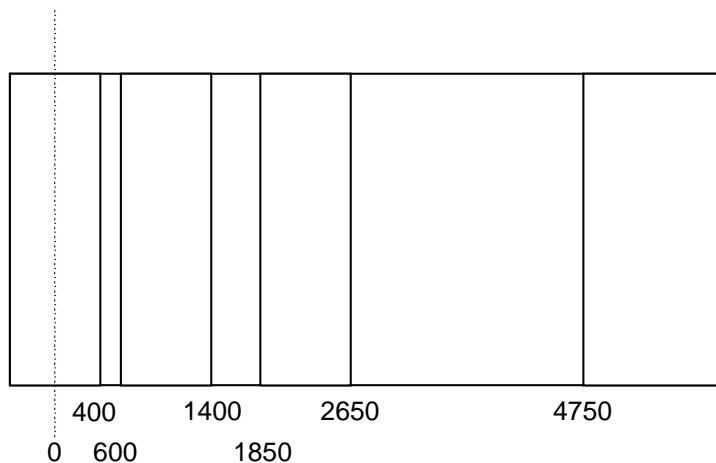


Figure 4.35 A six-edge pattern. The first edge was used for Fig. 4.35 and Fig. 4.37. The first five edges were used for Fig. 4.38.

Figure 4.36 shows the two E-D trees from 100° C and 102° C postexposure bakes simulated with a typical resist. It is based on the 400-nm edge in the mask pattern shown in Fig. 4.35, using $\lambda = 250$ nm, NA = 0.5, and $\sigma = 0.5$. The CD tolerance is ± 30 nm, and EL is 10%. Even though the shapes of the E-D trees are similar, the common window is much smaller because of the large disparity in the exposure centers.

Some processes are more critical than others. The postexposure bake process shown in Fig. 4.36 induces more DOF reduction than the impact of developing time, as shown in Fig. 4.37. Not only are the shapes of the E-D tree at the two developing times similar, they are also very close in position. Hence, a developing time variation of ± 1 sec is seen to be much less significant than a $\pm 1^\circ$ C postexposure bake temperature for this particular resist.

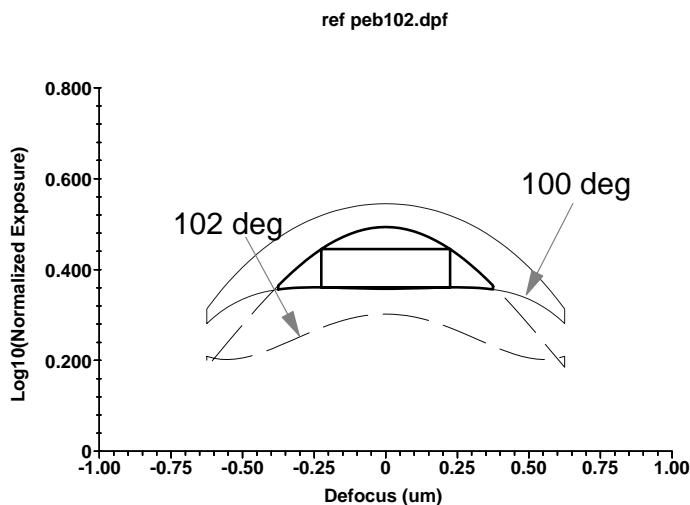


Figure 4.36 E-D trees from $101 \pm 1^\circ$ C postexposure bake.

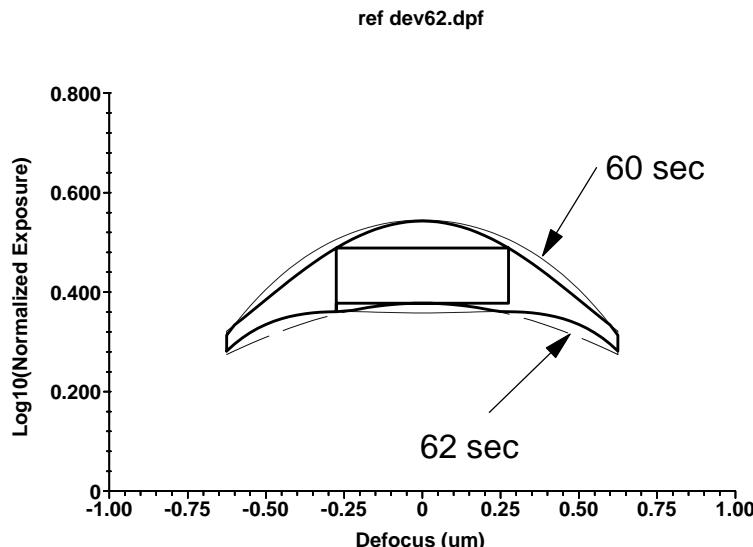


Figure 4.37 E-D trees from 61 ± 1 sec of developing time.

To evaluate the common E-D window realistically, all process tolerances must be considered. In Fig. 4.38, the E-D trees corresponding to the process tolerance of postexposure bake temperature, time, prebake temperature, developing time, and resist thickness are plotted together for the first five edges shown in Fig. 4.35. There are six trees per edge, including the tree for the nominal processing condition. Hence, there are 30 trees in the figure. The common E-D window is greatly reduced. To improve on it, the most restricted tree was identified as the postexposure-bake tree. One must tighten the range of the postexposure bake temperature. Alternately, the nominal baking temperature can be shifted for a larger common E-D window. There is a tradeoff between imaging contrast reflected in the EL or the height of the E-D window and the E-D window matching, reflected in the vertical position of the E-D window. The optimization of this tradeoff can be done unambiguously with the E-D trees. One can work on the next restrictive processing condition until the largest common E-D window is obtained. Alternately, one can use optical proximity correction to shift the position of the E-D trees. This topic will be further covered in Chapter 6.

4.5.8.6 Combination of image field positions

Even the image of an identical feature can vary within a lens field. The imaging field can be tilted. The field curvature of the imaging lens is not absolutely flat. Lens aberrations contribute to field curvature. The illumination is not completely uniform. The mask and proximity correction may have its own CD signature and CD control residue. The wafer may not be totally flat. The topography is often a problem. If the exposure tool is a scanner, scanning error can also contribute to field-dependent CD variation. Figure 4.39 shows the E-D trees at nine field

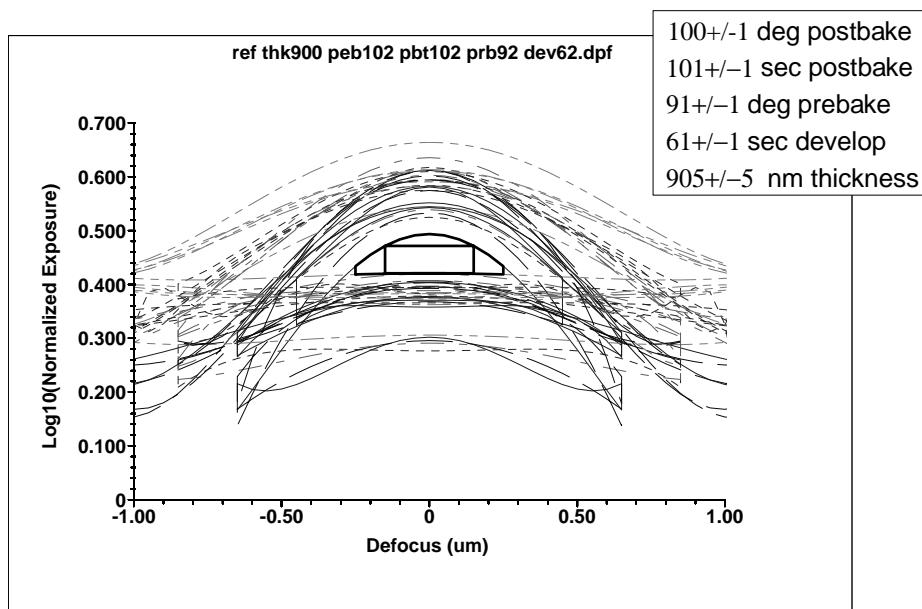


Figure 4.38 Process-tolerance E-D trees and their common window.

positions of a 455-nm y -oriented feature using an $NA = 0.55$, $\lambda = 365$ -nm stepper. Table 4.6 lists the parameters of the E-D tree at these nine field positions. Comparing the E-D area, one sees better imaging performance at the left and upper side of the field. The focus centers indicate that the best imaging field is lower in the middle column.

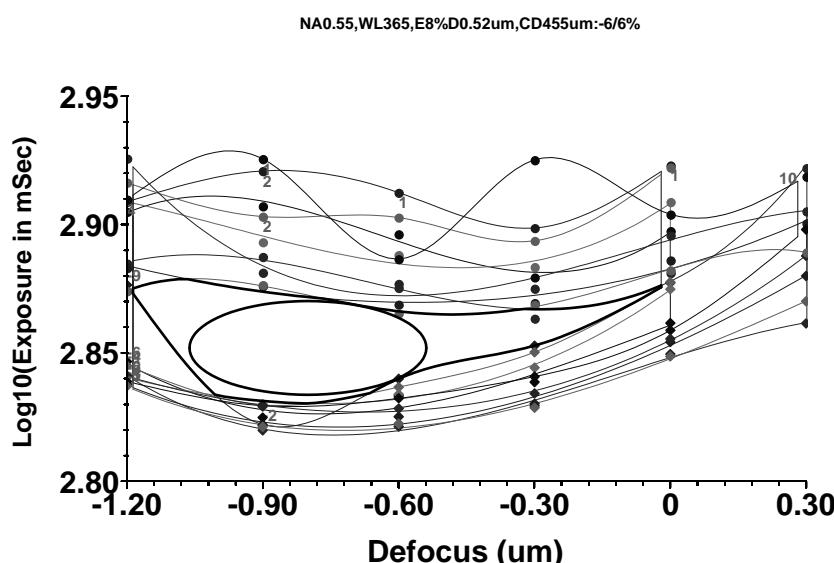


Figure 4.39 E-D trees of a 455-nm feature at nine field locations, with $NA = 0.55$ and $\lambda = 365$ nm.

Table 4.6 E-D window and common window at nine field positions of a 455-nm feature delineated at NA = 0.55 and $\lambda = 365$ nm.

Field location	E_{min} (mJ/cm ²)	EL (%)	D_{min} (nm)	DOF (nm)	E-D area $EL(\log_{10} \cdot k_2) \cdot 1000$	E_{center} (mJ/cm ²)	D_{center} (nm)
Common	22.70	8.39%	-1062	523	15.81	23.65	-800
Left-up	23.30	14.02%	-1170	985	49.80	24.93	-678
Mid-up	22.40	14.75%	-1187	905	48.13	24.05	-734
Right-up	22.10	12.19%	-1182	925	40.64	23.45	-720
Left-mid	23.00	14.00%	-1191	985	49.74	24.61	-698
Mid-mid	22.60	13.09%	-1187	804	37.96	24.08	-784
Right-mid	22.20	10.90%	-1178	945	37.12	23.41	-706
Left-down	22.90	11.40%	-1182	1126	46.25	24.20	-619
Mid-down	22.40	10.22%	-1191	985	36.27	23.54	-698
Right-down	22.10	10.38%	-1178	945	35.36	23.25	-706

Without digging much into what caused such severe field dependence, one simplistic way to fix the problem is to modify the exposure according to the distribution of the exposure center. The resultant improvement can be seen in Fig. 4.40 and Table 4.7. The E-D trees are aligned in exposure, and the resultant E-D window has 9.97% EL and 865-nm DOF, compared with 8.39% and 523 nm before the optimization.

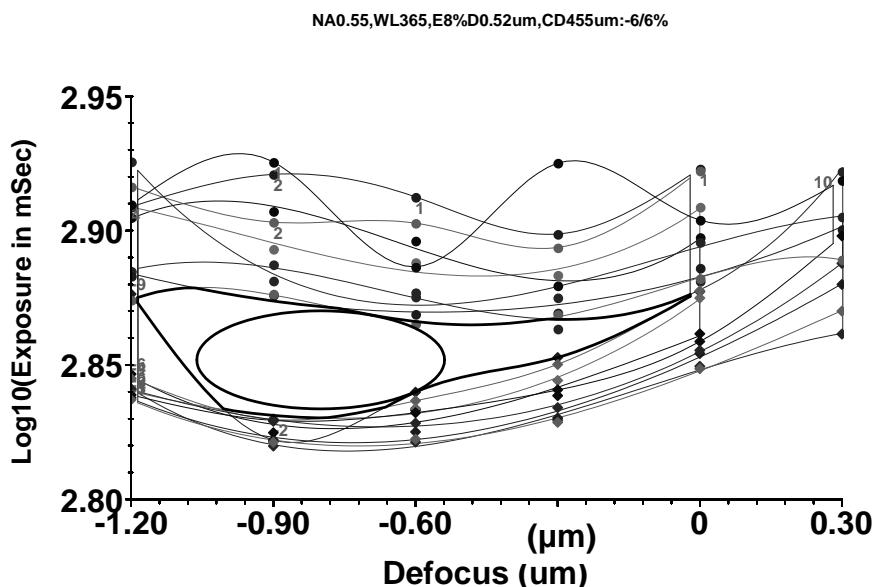


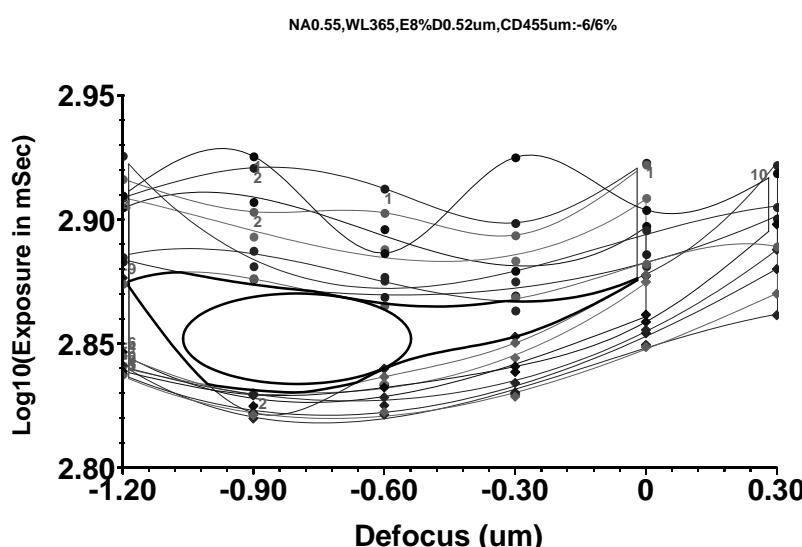
Figure 4.40 E-D tree at the nine positions as in Fig. 4.39, except the exposure at each field is modified according to the exposure center shown in Table 4.7.

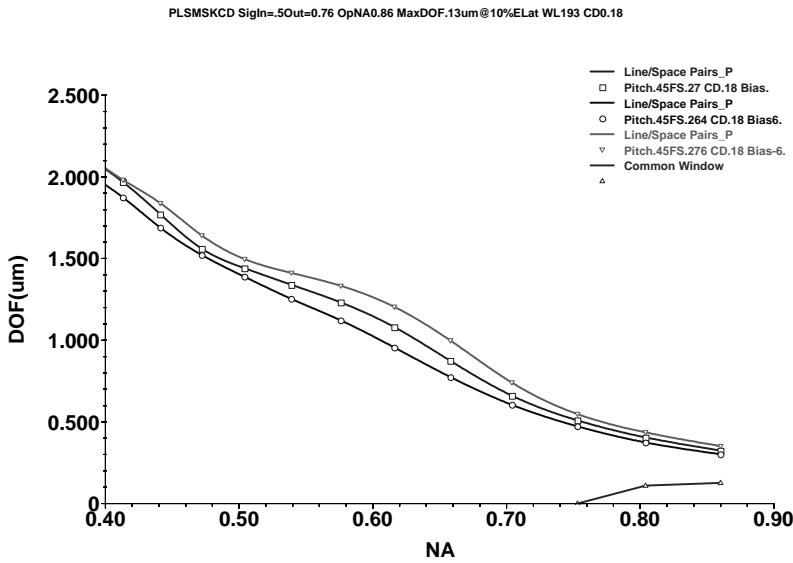
Table 4.7 E-D window of the E-D trees from Fig. 4.40.

Field location	E_{min} (mJ/cm ²)	EL (%)	D_{min} (nm)	DOF (nm)	E-D area EL(log10) k_2 /1000	E_{center} (mJ/cm ²)	D_{center} (nm)
Common	22.86	9.97%	-1154	865	31.05	24.00	-722
Left-up	22.35	14.02%	-1170	985	49.80	23.92	-678
Mid-up	22.41	14.75%	-1187	905	48.13	24.06	-734
Right-up	22.53	12.19%	-1182	925	40.64	23.90	-720
Left-mid	22.54	14.00%	-1191	985	49.74	24.12	-698
Mid-mid	22.55	13.09%	-1187	804	37.96	24.03	-784
Right-mid	22.82	10.90%	-1178	945	37.12	24.06	-706
Left-down	22.66	11.40%	-1182	1126	46.25	23.95	-619
Mid-down	22.88	10.22%	-1191	985	36.27	24.05	-698
Right-down	22.75	10.38%	-1178	945	35.36	23.94	-706

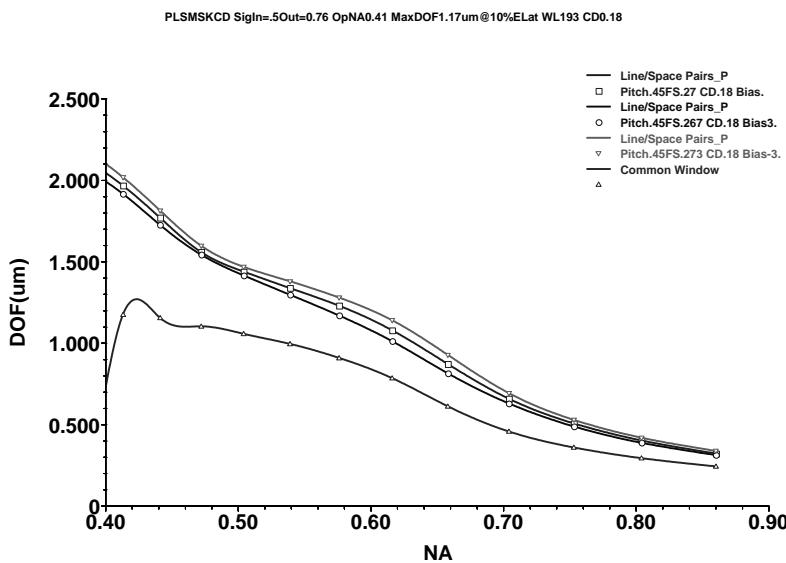
4.5.8.7 Setting the mask-making tolerance

In order to set the tolerance for mask makers,⁸ three E-D trees are constructed corresponding to the maximum and minimum possible CD values on the mask, as well as the nominal CD value. We have chosen 0.18-μm positive resist lines at a 0.45-μm pitch as an example. The exposure wavelength is 193 nm using annular illumination with $\sigma_{out} = 0.76$ and $\sigma_{in} = 0.5\sigma_{out}$. DOF versus NA is evaluated with 10% EL. Figure 4.41 shows the situation with a 10-nm CD control range on the mask. The individual DOF versus NA curves are very close to each other. However, due to spreading of the exposure centers, the common DOF is reduced approximately by half and is completely lost at NA < 0.48.

**Figure 4.41** E-D trees from 0- and ±5-nm mask biases.



Elliptical Window

Figure 4.42 E-D trees from 0- and ± 6 -nm mask biases.

Elliptical Window

Figure 4.43 E-D trees from 0- and ± 3 -nm mask biases.

When the mask CD range is relaxed by just 2 nm, the common DOF is almost completely gone, except for negligible values at NA > 0.75, as seen in Fig. 4.42. Tightening the mask CD range to 6 nm significantly improves the common DOF, as shown in Fig. 4.43.

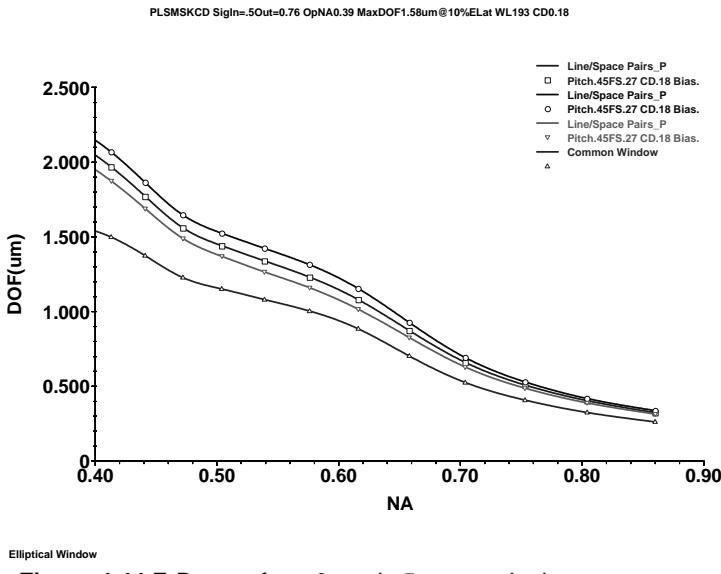


Figure 4.44 E-D trees from 0- and ± 5 -nm mask placement errors.

Mask CD range is not the only mask parameter that needs to be controlled. Placement errors also contribute to a reduction of the common E-D window. The same situation is shown in Fig. 4.41, except that the CD error is replaced with the placement error. Figure 4.44 shows the DOF versus NA curves of the three placement errors on the mask, 0 and ± 5 nm. The common DOF is again seen to be less than any of the individual DOFs. Compared with Fig. 4.41, placement errors in this particular case are more tolerable than CD errors.

Relaxing placement error to ± 8 nm results in the DOF versus NA curves shown in Fig. 4.45. There is clearly more loss in DOF with larger placement errors.

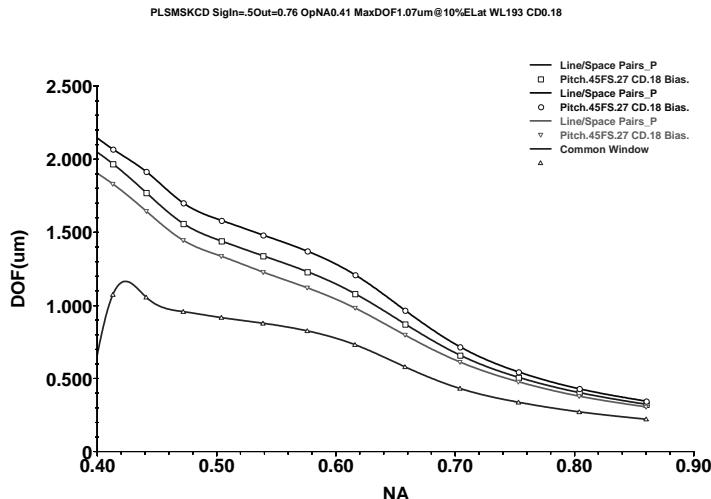


Figure 4.45 E-D trees from 0- and ± 8 -nm mask placement errors.

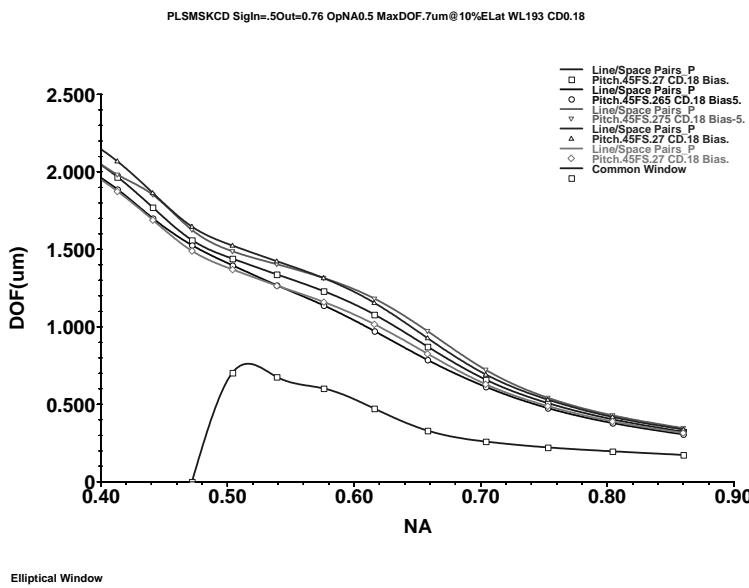


Figure 4.46 E-D trees from 0- and ± 5 -nm CD and mask placement errors.

Realistically, mask CD and placement errors coexist. Figure 4.46 shows the combined effect of ± 5 -nm CD and mask placement errors. In this situation, loss of the E-D window is dominated by mask CD errors.

4.5.8.8 Effects of phase-shifting mask errors

In addition to CD and placement errors, phase-shifting masks can further suffer from phase and transmission errors due to fabrication tolerances. Figure 4.47 contains the E-D tree for five situations—namely, 100% transmission and a 180-deg phase shift, a 170-deg phase shift, a 190-deg phase shift, a 90% transmission at one of the openings, and a 90% transmission at most of the openings. The common E-D window with these mask-making errors has $EL = 7\%$ and $DOF = 0.62 k_2$. This is reduced from 18.4% and $1.33 k_2$ of a perfect alternating phase-shifting mask (AltPSM), shown in Fig. 4.48. The perfect binary intensity mask (BIM) shown in the same figure has $EL = 11.5\%$ and $DOF = 0.75 k_2$. Hence, if one does not specify the AltPSM properly, the imaging performance of AltPSM can be worse than just a simple BIM.

4.5.8.9 Comparison of experiment and theory

The E-D plot is a good platform for comparing experimental and theoretical results, even though experimental results are often obtained in the form of CD as a function of exposure and defocus, whereas theoretical results on aerial images are usually in the form of an intensity distribution. The intensity distribution and CD (as a function of exposure and defocus) can be converted to E-D trees and compared, as exemplified by Fig. 4.49. The DOF from an experiment is 339 nm,

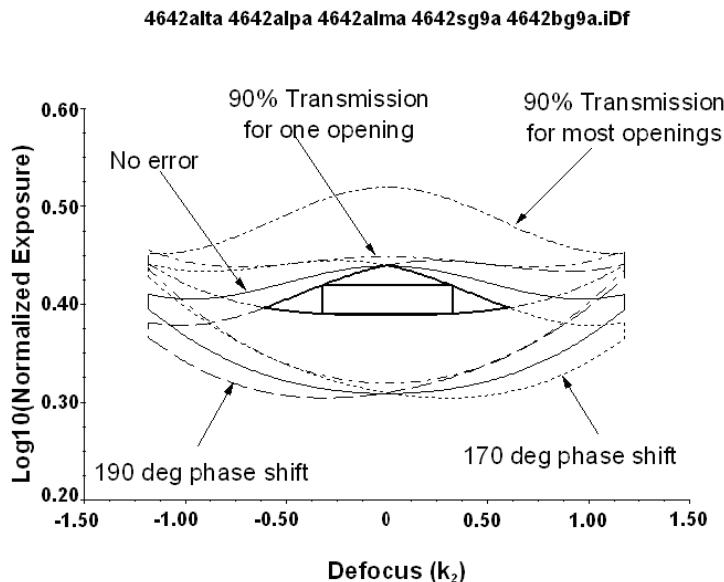


Figure 4.47 AltPSM with phase and transmission errors. $k_1 = 0.46$ and $\sigma = 0.42$.

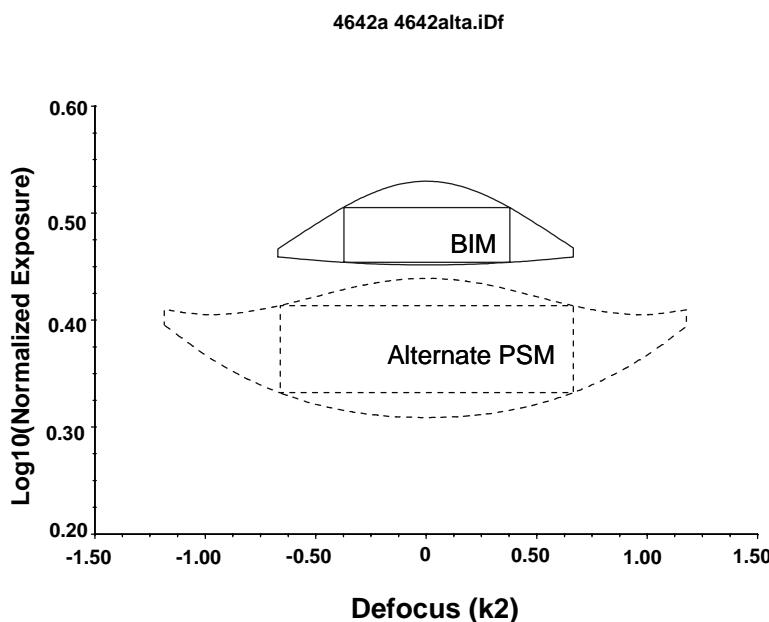


Figure 4.48 Perfect AltPSM taken from Fig. 4.47 and perfect BIM.

and the DOF from a theoretical prediction is 365 nm, both at an 8% EL. The absolute exposure dosage is not calibrated from theory to experiment; thus, there is discrepancy in the exposure dosages. The individual biases used for optical proximity correction are marked in the theoretical data, namely -20, 2, and 20 nm for the 180-, 300-, and 400-nm pitches, respectively. No bias is used for the features at a 600-nm pitch.

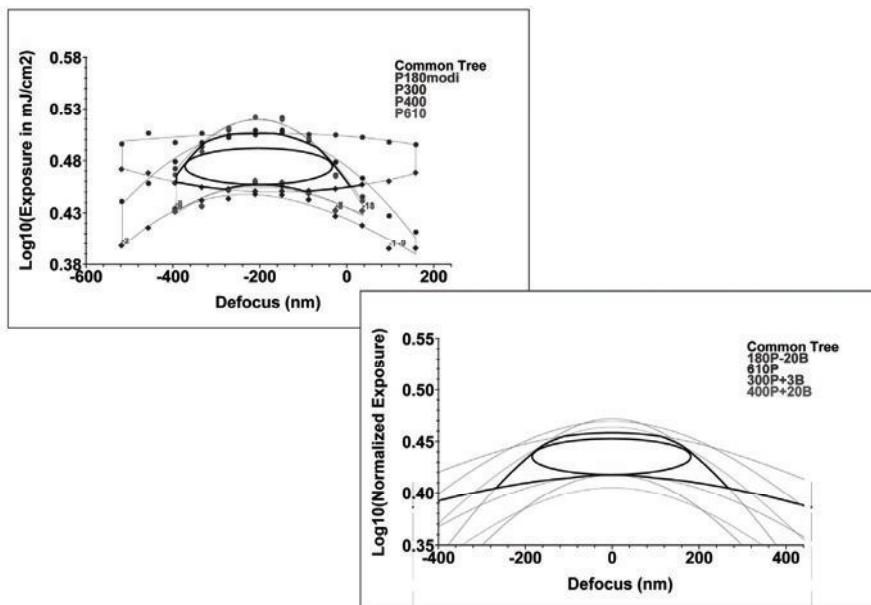


Figure 4.49 Experimental and simulation results from 80-nm polysilicon lines, using a 193-nm water-immersion scanner. The settings are $\text{NA} = 0.8$, $\sigma_{in} = 0.5$, and $\sigma_{out} = 0.8$. The pitches are 180, 300, 400, and 610 nm. Optical proximity correction was used to gain a larger common E-D window.

References

- 1 B. J. Lin, “The optimum numerical aperture for optical projection microlithography,” *Proc. SPIE* **1463**, p. 42 (1991).
- 2 M. Born and E. Wolf, *Principles of Optics*, 6th Edition, Cambridge Univ. Press, Boston, Sec. 8.5 (1980).
- 3 J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, Secs. 6.2 and 6.3 (1968).
- 4 B. J. Lin, “New λ/NA scaling equations for resolution and depth-of-focus,” *Proc. SPIE* **4000**, pp. 759–764 (2000).
- 5 J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, pp. 130 (1968).
- 6 Sec. 8.6 of Ref. 2.
- 7 B. J. Lin, “Exposure-defocus forest,” *Jpn. J. Appl. Phys.* **33**, p. 6756 (1994).
- 8 B. J. Lin, “The relative importance of the building blocks for 193 nm optical lithography,” Lecture notes, first Seminar on 193-nm optical lithography in Taiwan, p. 403 (1998).

Chapter 5

Components in Optical Lithography

An optical-lithography imaging system includes the following key components: a light source that provides the exposure photons with the desired energy spectrum; an illuminator that collects light from the source, adjusts its coherence and incident angles, and delivers light to the mask uniformly; a mask containing the circuit image to be replicated to the photoresist; an imaging lens of the desired NA and field size to reproduce the mask image in the resist by exposure; the resist layer coated on the wafer; the thin-film stack to be delineated by the pattern transfer process or implantation; the wafer held by a chuck on a wafer stage, which can be moved for alignment and field stepping; and devices to guide the alignment movement of the stage. These components are described consecutively in the following sections.

5.1 Light Source

There are two types of light sources for use in optical-lithography exposure tools: the mercury arc lamp and the excimer laser. They are both bright and efficient in their respective wavelength spectrum.

5.1.1 Mercury arc lamp

The mercury arc lamp has been the light source of choice for optical lithography because of its many usable emission lines in the near-UV (350–460 nm), mid-UV (280–350 nm), and deep-UV (200–280 nm) regions. There is even an emission line at 184 nm. Its high brightness, compared to other nonlaser sources, is another reason for its popularity. Figure 5.1 shows the radiation spectrum of a typical mercury arc lamp. Figure 5.2 shows a schematic drawing of a mercury arc lamp. There are pointed and rounded electrodes enclosed in a quartz envelope that contains a noble gas and a small quantity of mercury. During ignition, electric discharge takes place between the two electrodes. As the heat builds up, the mercury is vaporized, and the radiation spectrum becomes dominated by that of the mercury. The pointed electrode becomes a bright radiating spot. The

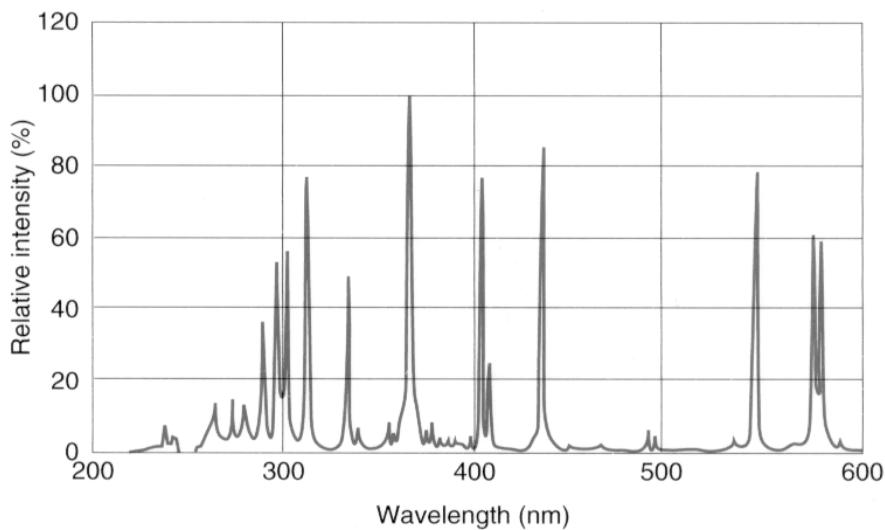


Figure 5.1 Radiation spectrum of a mercury arc lamp.

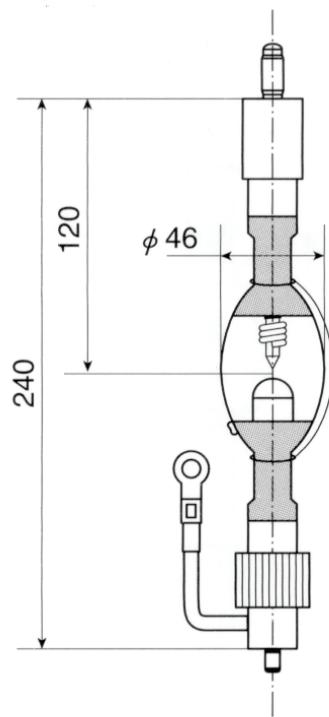


Figure 5.2 Schematic drawing of a mercury arc lamp.

rounded electrode produces a spot with less brightness, which is undesirable for diffraction-limited imaging because the illumination system is often designed for only one point source.

The main problem with mercury arc lamps for monochromatic lithographic imaging is their low optical efficiency. The optical output power is typically 5% of the electrical input power; furthermore, no more than 1 or 2% of the optical energy is used after spectral filtering because refractive projection optics requires a narrow bandwidth on the order of 5 nm. For deep-UV lithography, the mercury arc lamp is even more inefficient because of the low output of the 254-nm line. As the lamp ages, the electrode changes shape due to corrosion. As a result, the shape and size of the bright radiating spot is changed, and the brightness is reduced. The inside wall of the lamp envelope may be contaminated by a light-absorbing deposit that further reduces the light output. Regardless, mercury arc lamps have been the workhorse of optical lithography for 436-nm g-line, 405-nm h-line, and 365-nm i-line systems. They have also been used for deep-UV broadband systems in 4X-reduction, deep-UV scanners,ⁱ 1X full wafer, deep-UV projection mask aligners,ⁱⁱ and 1X full wafer deep-UV proximity mask aligners.ⁱⁱⁱ By optimizing the gas mixture in the lamp and other parameters, the deep-UV spectrum can be enhanced. Figure 5.3 shows the radiation spectrum of an Ushio UX-4000P Hg arc lamp that is optimized for the SVGL Microscan 1 deep-UV scanner. With such applications, the quartz envelope of the lamp needs to have high transmission in deep UV.

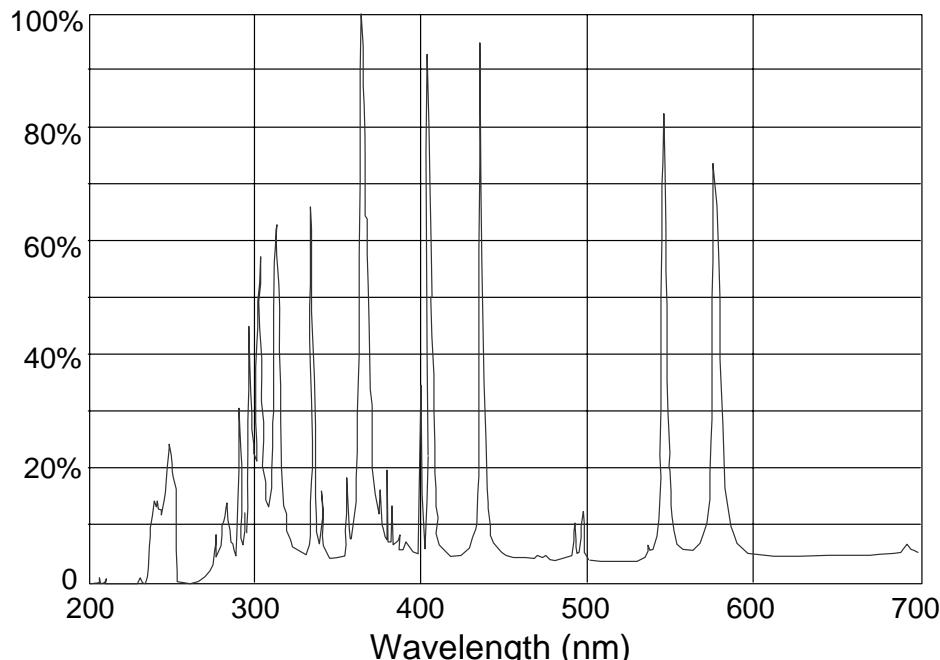


Figure 5.3 Radiation spectrum of a mercury arc lamp optimized for deep UV.

ⁱ Perkin Elmer Microscan 1 system.

ⁱⁱ Perkin Elmer 500 and 600.

ⁱⁱⁱ Canon deep-UV proximity mask aligner.

5.1.2 Excimer laser

When the exposure wavelength is shortened to the deep-UV regime, the low optical efficiency of mercury arc lamps makes it difficult to provide the exposure power required. Furthermore, in the vicinity of 250 nm there are very few optical materials with sufficient transmissivity and mechanical working properties to incorporate into achromatic imaging lenses. For deep-UV dioptric systems, fused silica is the only optical material used for the imaging lens; thus, no correction of chromatic aberration is possible, necessitating an illumination with bandwidth in the picometer regime. Moving the wavelength further down makes the bandwidth requirement even tighter. A laser source is inevitable. However, most laser sources are highly temporally and spatially coherent. The speckles produced by such a source are undesirable for microlithography applications. However, the excimer laser happens to be extremely efficient. Because of its high gain, light needs to travel in the resonator only a few times before it is sufficiently amplified, making it inherently much less spatially coherent than other lasers. The selection of wavelengths—such as 248 nm with KrF, 193 nm with ArF, and 157 nm with F₂—makes the excimer laser a clear choice for deep-UV and shorter wavelengths.

The earliest application of excimer lasers to microlithography is in proximity printing.^{1,2} Subsequently, they are demonstrated on a full-field, all-reflective system³ and a dioptric step-and-repeat system.⁴ A comprehensive treatment of the excimer laser—including its history, operation, design considerations, systems, and maintenance—can be found in Das and Sengupta,⁵ who outline the aspects of excimer lasers most relevant to understanding and practicing microlithography for manufacturing.

5.1.2.1 Operation principle

Microlithography-related excimer lasers are pumped by electric discharge. During the discharge, KrF* excimers are formed. Once they are present, spontaneous emission, stimulated emission, and quenching take place. Spontaneous and stimulated emissions are shown in the following reactions, respectively:



where ν is frequency. The stimulated emission provides the desired optical output. Figure 5.4 shows the energy diagram of a KrF excimer laser. The Kr⁺ + F⁻ band supports stimulated emission, and the Kr + F₂ band supports spontaneous emission.

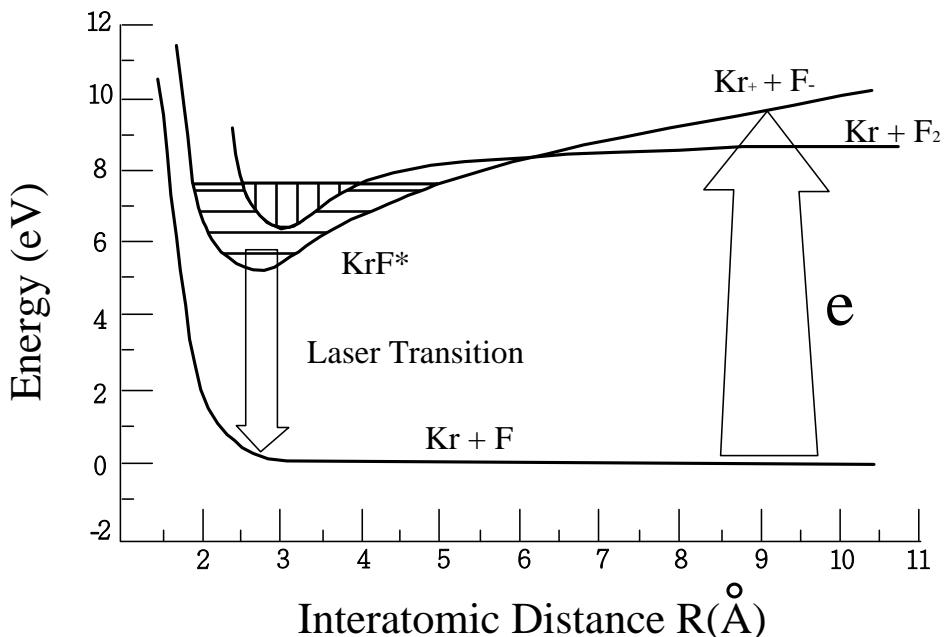


Figure 5.4 Energy diagram of a KrF excimer laser.¹¹ (Image reproduced with permission of Marcel Dekker.)

The discharge voltage is in excess of 10 kV, with the gas in the resonator at several atmospheres in pressure. A pulsed power supply capable of such voltage at a repetition rate between 1 to 2 MHz is used. It consists of a series of π networks of saturable inductors and high-voltage capacitors controlled by a high-speed, high-voltage switch. Typical discharge duration is between 300 and 500 nsec. The optical pulse is about 10 to 20 nsec long.

5.1.2.2 Bandwidth narrowing

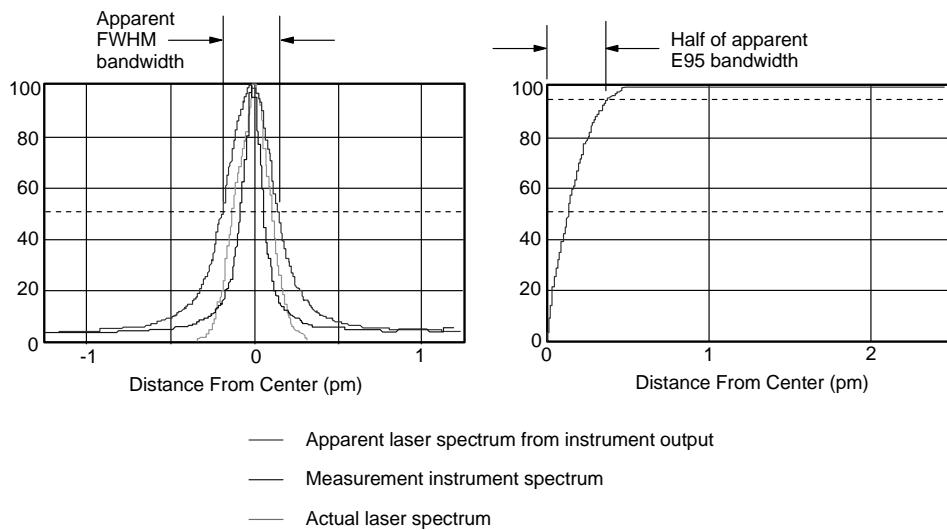
The 248.35-nm pulse from an unregulated excimer laser has a bandwidth of a few hundred picometers. However, a much narrower bandwidth is required to avoid chromatic aberration. Since a single-lens material was initially used in the deep-UV imaging lens, compensation using materials of different dispersion for chromatic aberration was not possible. Only because of recent improvement of CaF₂ materials, the lens in high-end ArF scanners now employs fused quartz and CaF₂ elements. The bandwidth requirement is a function of the dispersion property of the optical material as well as the NA of the imaging lens. Table 5.1 shows the E95 and the full-width half-maximum (FWHM) spectral bandwidth requirement of 248-, 193-, and 157-nm lenses at different NAs. Higher NAs require narrower bandwidths.

Table 5.1 Bandwidth requirements of imaging lenses.

248 nm	< 3 pm E95 at 0.35NA	< 1.4 pm E95 at 0.75NA	< 1 pm E95 at 0.85 NA
193 nm	1.2 pm E95 at 0.75 NA	0.75 pm E95 at 0.85 NA	0.5 pm E95 at 0.93 NA
157 nm	0.5 pm FWHM at 0.8 NA dioptric lens	1 pm FWHM at 0.8 NA catadioptric lens	

The E95 bandwidth is defined as the band where 95% of the entire spectrum energy is located. The FWHM bandwidth is the full width of the spectrum at the half-maximum point. It is used to characterize laser bandwidth until E95 becomes a better quantity to track. Figure 5.5 shows the E95 and the FWHM bandwidths from the apparent laser output that is the raw data from the bandwidth-measuring equipment. This raw data is a convolution of the response spectrum of the measuring instrument. The actual laser output spectrum can be reconstructed by deconvolution.

Table 5.2 shows the dispersion property of fused silica and CaF₂ at 248 and 193 nm.⁶ Higher dispersion calls for a narrower bandwidth in illumination.

**Figure 5.5** Laser output spectra, E95, and FWHM bandwidths.**Table 5.2** Dispersion of fused silica and CaF₂.

	Fused silica (nm ⁻¹)	CaF ₂ (nm ⁻¹)
248.4 nm	-0.00056 ^[7]	-0.00037 ^[8]
193.4 nm	-0.00157 ^[9]	-0.00098 ^[8,9]
157.6 nm	—	-0.0026 ^[8,10]

In order to maintain high efficiency, the wavelength-narrowing feature is incorporated in the laser resonator. Three wavelength-selecting elements have been attempted—namely, etalons, prisms, and gratings—as shown in Fig. 5.6. Etalons have a high wavelength resolution but are highly sensitive to heat. Prisms do not provide sufficient resolution. The combination of a beam expander and a grating (as shown in Fig. 5.7) has been successfully employed.¹¹ Another combination uses a grating for coarse wavelength selection and an etalon for fine selection¹² to overcome the problem of etalon heating, as shown in Fig. 5.8. The etalon in module 3 for wavelength selection is placed at the output end of the amplifying media, needing only to narrow the bandwidth already coarsely narrowed by the grating-prism assembly (1) at the other end. Figure 5.8 also shows other components in a microlithography excimer laser. The bandwidth/wavelength monitor (4) checks the final output bandwidth and compares the output wavelength to an absolute standard. The energy-monitor module (5) regulates the output power for pulse-to-pulse consistency. The windows on the laser tube (2) are oriented at the Brewster angle to discriminate against absorption and scattering induced by the windows. As a result, the output light is polarized in the plane of the drawing.

5.1.2.3 Spatial coherence

Even though great effort has been spent narrowing the bandwidth of the excimer output, its temporal coherence inevitably increases with a narrower bandwidth. More laser speckles are generated. The spatial coherence must be reduced to suppress the generation of the speckles. Inherently, the excimer laser is already of low spatial coherence because only three or four light passes are required through the amplifying media, making it operate in numerous modes.

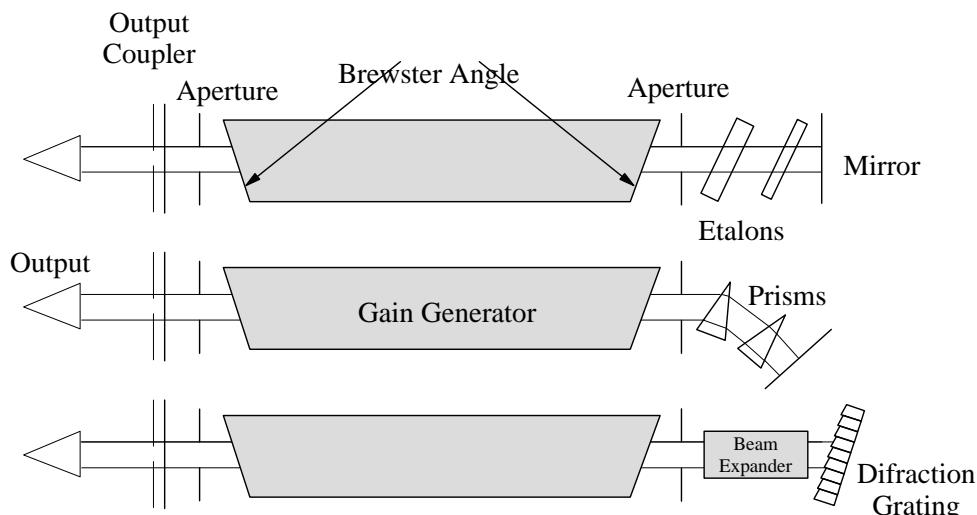


Figure 5.6 Bandwidth-narrowing schemes using basic wavelength-selecting elements.

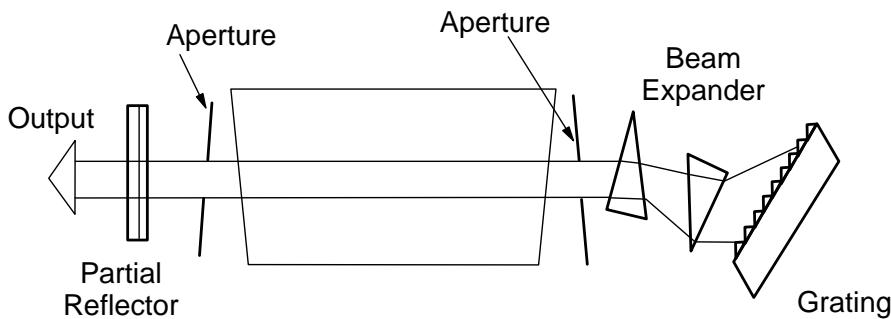


Figure 5.7 Bandwidth narrowing using prisms and a grating.

The typical coherence length of excimer lasers is of the order of a few hundred micrometers (compared to meters for some highly tuned lasers). Nevertheless, further reduction of spatial coherence by scrambling is necessary. Early systems use a rotating mirror, fly's eye lens, or optical fiber to spread out the point source. Modern illuminators incorporate a light rod to internally reflect the light source many times to scramble the light and fill up the entrance pupil. Figure 5.9 shows the earlier schemes using a rotating mirror, an optical fiber, and a modern scheme using a light rod. The light rod scheme is so effective that even conventional point sources, such as mercury arc lamps, can also benefit from it.

5.1.2.4 Maintenance, safety, and lifetime of excimer lasers

The excimer laser is more expensive to purchase and operate than mercury arc lamps. At the very least, it requires maintenance of the laser chamber, the wavelength-stabilizing module, and the bandwidth-narrowing module. The laser also takes up more factory space. One should be extremely careful with radiation-induced damage to the eyes. Unlike the mercury arc lamp, which has a visible output to alert bystanders, the output of KrF and ArF excimer lasers are completely invisible. Damage is realized only after the fact. Therefore, many safety interlocks are incorporated to prevent exposing the laser beam to human beings in any circumstance.

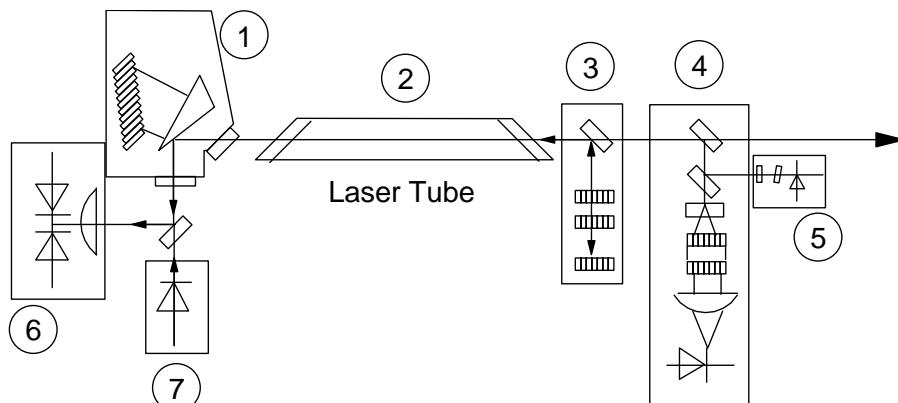


Figure 5.8 Bandwidth-narrowing system using a grating and an etalon.

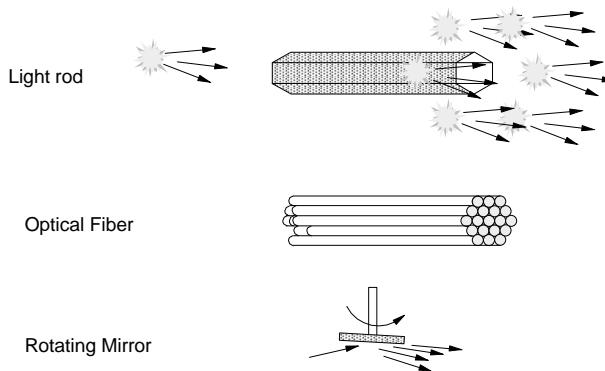


Figure 5.9 Scrambling of a light source.

Operating at many kHz, an excimer laser can deliver tens of billions of pulses per year. Its lifetime and maintenance are measured by the number of pulses it delivers. A modern lithography excimer laser has a lifetime on the order of 15 billion pulses for the chamber and 20 billion pulses for the wavelength-stabilizing and the bandwidth-narrowing modules. Routine maintenance includes cleaning the F₂ trap and the output coupler. The former is required after several hundred new fills and the latter after 7 billion pulses. An excimer laser is usually fine tuned monthly to restore wavelength drift, bandwidth, and energy stability.

5.2 Illuminator

The illuminator of a typical scanner is shown in Fig. 5.10. The laser beam goes through the beam-pointing-and-steering unit and the variable attenuator to adjust its direction, position, and intensity before entering the illumination-shaping optics. There, the beam is shaped for the type of illumination, such as axial, off-axis, Quasar, etc. It passes through the quartz rod to reduce spatial coherence, with its energy being monitored there. The coherence of the illumination is determined at the reticle masking assembly. Together with the imaging lens, they

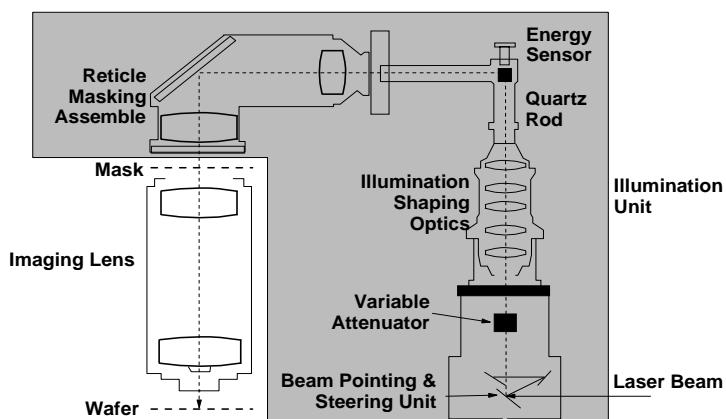


Figure 5.10 Schematic drawing of the illumination unit of an ASML scanner.

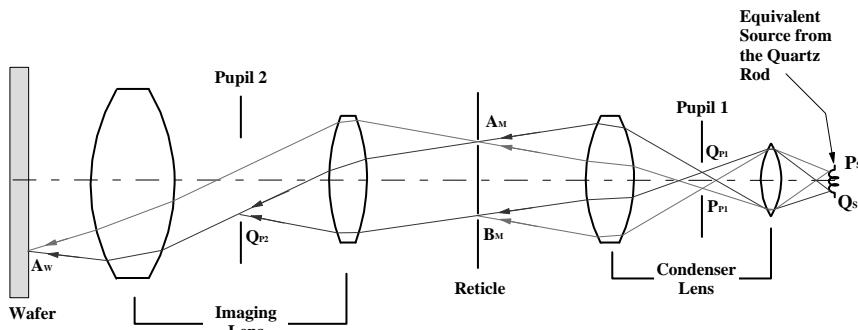


Figure 5.11 Schematic of a Köhler illumination system.

constitute a Köhler illumination system. At the reticle masking assembly, the area outside of the printed area of the reticle can be masked off.

5.2.1 Köhler illumination system

A typical Köhler illumination system¹³ is shown in Fig. 5.11. The key to Köhler illumination is that the source plane is conjugate to the pupil planes, which are, of course, conjugate to each other. Hence, the rays originating from the source point Q_s converge at point Q_{p1} at pupil 1. They converge again at Q_{p2} . The rays from other source points represented by the other ray at the reticle opening A_M go through the imaging lens system including pupil 2 to converge at the wafer at A_w . There are now two conjugate relationships: one links the source plane and the pupil planes, the other one links the reticle plane, wafer plane, and the reticle masking plane in the reticle masking assembly.

5.2.2 Off-axis illumination

The simplest way to produce off-axis illumination is to block off the unwanted central illuminated area with a light-absorbing stop. In fact, this method is very flexible because any arbitrary shape can be made. However, the illumination energy in the stopped area is wasted, which increases the exposure time and reduces the wafer throughput.

Mating wedges are used to produce 1D off-axis light without energy loss, such as in the case of dipole illumination. Depicted in Fig. 5.12(a), the illumination is on axis when the two wedges are in contact with each other. When there is a separation between them, as in Fig. 5.12(b), the illumination is split into two. The distance between the two beams is a function of the separation between the two wedges. For ring illumination, conic mating wedges are presumably the key component.

5.3 Masks

A photomask(mask) consists of patterned absorbers and/or phase shifters delineated on a transparent substrate. According to the presence or absence of the

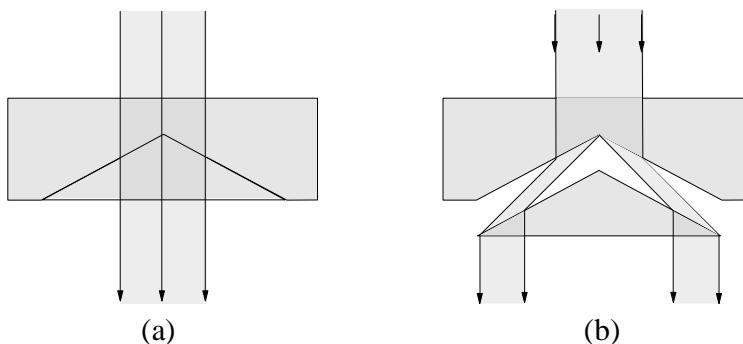


Figure 5.12 Mating wedges to produce off-axis light without energy loss.

absorbers and phase shifters, as well as the number of levels of these patterning materials, the photomask can be classified generically into six types, shown in Fig. 5.13.

Binary intensity mask (BIM)—This is the most commonly used mask type. Absorbers are placed in various areas of the mask, according to the circuit pattern. The absorber completely blocks light as it hits the mask. Alternately, light passes through the mask where no absorbers exist. Note that many people call BIM “binary mask.” This is a misnomer because binary mask can be further subdivided into “binary intensity masks” and “binary phase masks.” Also note that BIM is sometimes confused with chrome on glass (COG); the former denotes the type of mask, and the latter denotes the type of mask blank. COG cannot distinguish itself from a phase-shifting mask (PSM) because many PSMs consist of chrome and quartz, such as AltPSM. The word “glass” is also a misnomer. Soon after projection printing became popular, the mask substrate

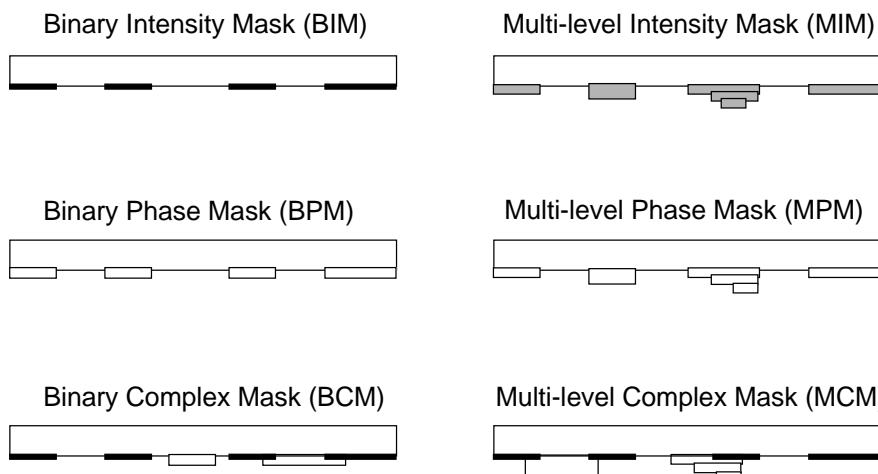


Figure 5.13 Types of photomasks.

became quartz instead of glass to take advantage of quartz's low thermoexpansion coefficient. Deep-UV lithography definitely calls for a quartz substrate due to its transparency in the shorter wavelengths.

Binary phase mask (BPM)—There is no absorber on the mask. The only pattern-dependent variation on the mask is a 180-deg phase shift. This is typified by an unattenuated (chromeless) PSM.

Binary complex mask (BCM)—The absorber passes or blocks light. The phase shifter shifts light by 180 deg. There are phase-shifted and nonphase-shifted areas on the mask. BCM constitutes the bulk of PSMs, including the alternating, the attenuated, the subresolution-assist, and the rim types. The attenuated type is included because the distribution of the absorber/phase shifter is still binary, even though the absorber carries the function of a phase shifter and does not completely absorb. However, when regular absorber patterns are included in these masks, then they become a multilevel complex mask, described below.

Multilevel intensity mask (MIM)—An intensity mask may contain many gray levels for specific applications, such as an exposure calibration mask that contains many calibrated gray levels to facilitate rapid and economical determination of the exposure dosage. MIMs may also be used for threshold leveling,¹⁴ which modifies the background transmission on the mask according to the pattern density to produce large exposure-defocus (E-D) windows.

Multilevel phase mask (MPM)—The phase shifters produce many levels of phase on the mask; presumably, 0 deg, 90 deg, and 180 deg, or 0 deg, 60 deg, 120 deg, and 180 deg. This type of mask is very rarely used.

Multilevel complex mask (MCM)—Either the absorber or the phase shifter is multilevel. The latter is exemplified by the AltPSM in combination with multilevel phase-shifting edges.¹⁵ In order to remove phase-shifting conflicts, extra phase-shifting edges are introduced, and multilevel phase shifts of a small increment (such as 60 deg) are implemented at these edges to reduce their image contrast so that they are not printed. A multilevel absorber in an MCM is exemplified by AttPSMs having completely opaque areas in addition to the transmissive absorber/phase shifter. The need for MCM AttPSMs will be presented in Sec. 5.3.4.3.

PSM is not a generic name. A PSM may be a BPM, BCM, MPM, or MCM. For a detailed description of the various types of PSMs, consult Sec. 5.3.4.3.

5.3.1 Mask substrate and absorber

Fused silica has been used as a substrate material for two reasons:

- (1) Its thermal expansion coefficient at 0.5 ppm/^oC is an order of magnitude smaller than that of other candidates. With the exception of proximity

printing, which requires matched thermal expansion coefficients between the mask and the wafer, it is very desirable to have thermal stability in the mask.

- (2) Fused silica is transparent from the visible region to below 200 nm. It is usable for g-line (435.83 nm), i-line (365.02 nm), KrF (248.35 nm), and ArF (193.39 nm) wavelengths. Even at the F₂ (157.63 nm) wavelength, where only a few fluoride compounds (such as CaF₂) transmit, fluorinated quartz is acceptable as a mask substrate. If CaF₂ was necessary, the cost of mask blanks in 157-nm lithography would be prohibitive.

Similar to the mask substrate material, the mask absorber has also survived many generations of optical lithography. In the proximity printing days, the emulsion mask absorber quickly gave way to chromium because of its hardness and smaller thickness; it has remained ever since. A photoresist material¹⁶ has been proposed for the mask absorber. By using the resist directly, the chromium deposition and delineation steps can be eliminated. Silicides, such as MoSi₂, have also been proposed as mask absorbers because of their desirable processing characteristics and favorable absorption spectra. However, the inertia of low-defect, chrome-on-quartz processing technology is simply too overwhelming. MoSi₂ is primarily used as the phase-shifting absorber for the AttPSM. Its property can be fine tuned to exhibit a 6% transmission and a 180-deg phase shift at the same thickness.

5.3.2 Pellicles

A pellicle is a transparent membrane¹⁷ mounted on a frame that is attached to the mask substrate, as shown in Fig. 5.14. It is usually an organic polymer a few micrometers in thickness. The 6-mm frame thickness is carefully chosen so that if a particle not too large in size is attached to the pellicle, the particle is completely out of focus and can, at most, induce a tolerable exposure change in its locality. It is inconceivable that any mask not be protected by a pellicle because great care has been taken to ensure that the mask induces no defect. If not protected, there will be repeating defects on every wafer at every exposure field. It is not easy to keep the mask defect free, even if it has been fabricated that way. If only one foreign particle falls on the mask during transportation, loading/unloading to the exposure tool, exposure, storage, or any other type of handling, the damage is done. Hence, pellicles are indispensable.

When mounting the pellicle to the mask, care should be taken to prevent trapping or producing particles within the cavity. Pellicle mounting should not induce mask distortion. Remounting should be kept to a minimum. To reduce the number of remountings, mask-inspection tools can often be used without removing the pellicle. However, pellicle removal, and thus remounting, is inevitable when mask cleaning or repair is needed.

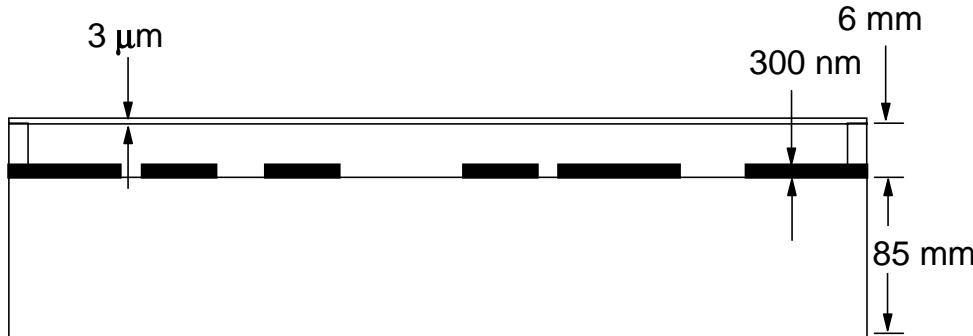


Figure 5.14 A mask and its pellicles.

Glue is used to attach the pellicle frame to the mask. Great care must be taken in selecting the glue to prevent stress production, contamination, or generation of particles during actinic wafer exposures. In addition, the pellicle material must withstand prolonged exposures without reducing its transmissivity or smoothness. Sometimes tiny holes are drilled at the pellicle frame to prevent trapping of harmful vapor generated during exposure. However, this leaves the possibility of contamination from the outside. The usage and prohibitions of pellicles alternate.

5.3.3 Critical parameters for masks

5.3.3.1 CD control

The mean and standard deviation of the critical dimension (CD) designated by the circuit designer are important parameters for the mask. The CD mean must be as close to the desired value as possible. The desired value must be the target value on the wafer multiplied by the reduction factor of the imaging system. A slightly different value may produce a larger E-D window, depending on the feature shape, feature size, feature environment, the optical train, the resist characteristics, and processing conditions. This value is referred to as mask bias. The standard deviation consumes the total wafer CD control budget.

5.3.3.2 Placement accuracy

Each feature on the mask must be located at the desired position with respect to other features on the same mask, as well as on other masks whose images must be overlaid together. Hence, placement error contributes significantly to the overlay accuracy budget. Placement errors are often induced by the mask-making machine, mostly due to difficulties in controlling the position of the writing beam. Additionally, the interferometrically controlled table that moves the mask to each writing field may introduce smaller errors.

5.3.3.3 Mask transmission and thermal expansion

The transmission of the mask blank should be above 90%. From the point of view of mask heating, a higher transmission is desirable. After reaching the material transmission limit, the only way to increase transmission is to reduce the mask thickness. In the early days of microlithography, mask thickness was in the order of 90 mils. It was increased to one quarter inch (0.25 in.) to reduce mask sag in a step-and-repeat exposure tool. The thickness has remained at one quarter inch for decades.

Mask heating causes thermal expansion. With a thermal expansion coefficient of 0.5 ppm/ $^{\circ}$ C, raising the temperature of the mask substrate by 1° C expands a 150-mm square mask by 75 nm at each side. This can easily be corrected by fine tuning the magnification of the exposure tool, especially in a projection tool. In lensless systems, such as proximity printing, mask thermal expansion can be detrimental to overlay because wafers at different stages of processing can exhibit subtle magnification changes. In x-ray proximity exposure tools, the mask must be thermally controlled to manage random magnification errors. Systematic errors were managed by preparing the mask according to the expected magnification change. The solution in e-beam proximity printing¹⁸ is most unique. The tilting angle of the beam is adjusted to compensate for any predictable overlay errors, including magnification changes, with the advantage of covering only a small area in the field.

5.3.3.4 Mask reflectivity

One of the main concerns with chromium as a mask absorber is its high reflectivity at the side facing the imaging lens. This reflection can reduce image contrast, as seen in Fig. 5.15. Points a and b are the edges of an opening in the mask. The absorber can be found in the areas outside of a and b. We shall assume that these two points image perfectly at their conjugate points a' and b' on the resist surface. In the best case, assuming there is no other thin-film layer on the wafer, the wafer itself serves as the reflection plane for points a' and b'. Point b' is now mirror imaged to point b'', which has a conjugate at b'''. The light ray that passes b'' and b''' hits the chromium part of the mask. This illumination is reflected back onto the photoresist at the areas that are supposed to be dark. The stray reflection is induced at the vicinity of the image edge that is most susceptible to stray-light-induced image deterioration. Hence, image contrast is reduced. Figure 5.15(a) depicts ray tracing through a simple lens, and Fig. 5.15(b) through a double telecentric lens system. In both cases the image contrast is affected by mask reflectivity. More discussion on the telecentric lens system can be found in Sec. 5.4.1.5 and Fig. 5.29.

Figure 5.16 shows linewidth versus defocus plots at a range of exposure dosages for mask absorber reflectivity of 66%, 33%, and 10%. The 10% result clearly confirms better linewidth control and exposure tolerance. It is desirable to reduce absorber reflection to 10% or lower.

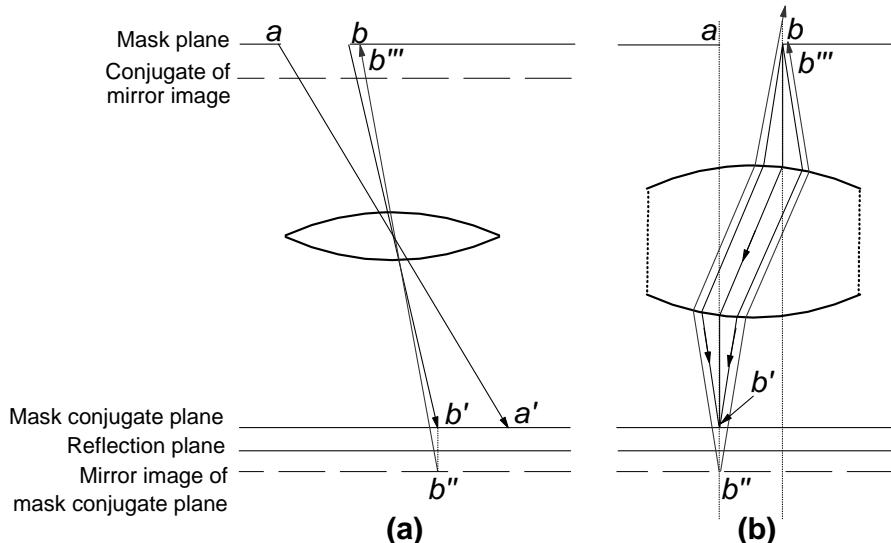


Figure 5.15 Stray light induced by a reflecting mask absorber. (a) Ray tracing through a simple lens; (b) ray tracing through a double telecentric system.

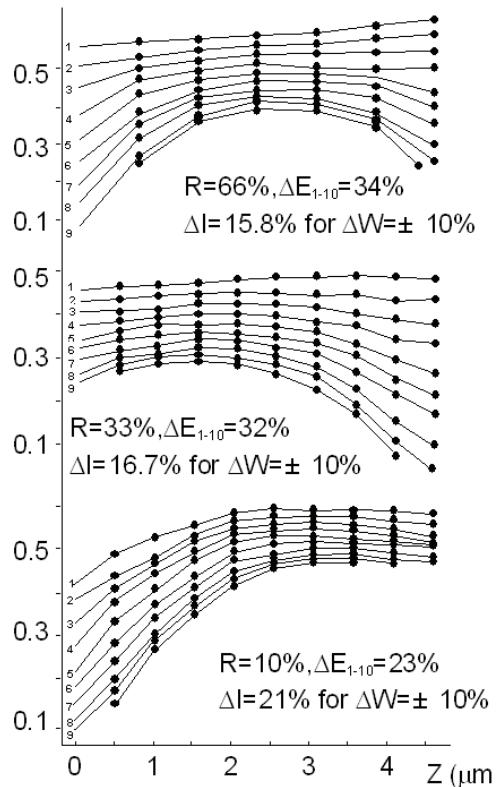


Figure 5.16 Linewidth versus defocus at different exposures for three levels of mask reflectivity.

5.3.3.5 Mask flatness

The flatness of the mask must be a small part of the depth-of-focus (DOF) budget of the entire imaging system. With a reduction system, the flatness requirement is very relaxed because the DOF tolerance on the mask size is M^2 , with M being the reduction ratio. For example, with a mask flatness of 500 nm, the focusing error at the wafer side of a 4X reduction system is 31 nm. In addition to ensuring that the two surfaces of the mask blank are well polished and parallel to each other, the mask blank must have sufficient strength to support itself without sagging, which can induce DOF error.

5.3.3.6 Physical size

A standard photomask is $6 \times 6 \times 0.25$ -inches in size. The thickness is chosen to minimize sag that consumes the DOF budget, as was mentioned in Sec. 5.3.3.5. The lateral dimensions are chosen to accommodate the field size requirement of the IC industry, plus the four sides outside the printable area. These sides are preferably opaque in order to frame the printable areas accurately, even though dynamically adjustable blades at the four sides are provided on the exposure tool for coarse framing purposes. With the advent of step-and-scan systems and increasing pressures to increase the field size, 9×9 -in. substrates have been considered. At one point, a 9-in. scanning range was reserved on ASML scanners in anticipation of future needs. However, moving to 9×9 -in. substrates is not a simple matter. All mask-making equipment, including the writer, track, etcher, and inspection and repair tools, must be changed. Also, because mask makers prefer a square substrate for easier processing, and scanner makers prefer the rectangular substrate to reduce the size and weight of the reticle scanning stage, it is not easy to agree whether a 9×9 -in. or 6×9 -in. substrate would be better.

5.3.3.7 Defect level

No printable defect is allowed on the mask for obvious reasons, as was discussed in Sec. 5.3.2. The defect level is usually specified by the maximum allowable defect size as a function of its vicinity to patterned features on the mask. A programmed defect mask (PDM) consisting of artificial defects varying in size and distance to patterned features is often used to empirically determine defect printability for specification. Simulation is helpful to gain insight for designing the PDM. Because real defects are usually irregular in shape, its printability must be further qualified with a so-called aerial image measurement system (AIMS), which consists of a small-field imaging lens and illuminator in an actinic wavelength to mimic the same $\lambda/NA/\sigma$ condition. Such an optical system is much easier to build when compared to a scanner or stepper. The NA needs only be $1/M$ of the NA of the exposure system. Therefore, 0.18 NA works for 0.72 NA of a 4X reduction system. Ultimate qualification lies in exposing the mask in the production exposure system and inspecting the printed wafer with wafer-inspection tools.

5.3.4 Phase-shifting masks

A mask that contains phase-shifting areas with or without energy-absorbing areas is a PSM. It can be BPM, BCM, or MCM, as defined in Sec. 5.3 above. Here, the types of PSM and their configurations are discussed. The imaging performance and their comparisons were briefly given in Sec. 4.5.2.11. PSM applications and additional imaging performance information will be given in Chapter 6, where their fabrication, inspection, and repair aspects are also included. I wrote an overview article as early as 1990, when many different types of PSMs first started to bloom.¹⁹

5.3.4.1 Operating principle

Phase-shifting masks take advantage of the interference effect in a coherent or partially coherent imaging system to reduce the spatial frequency of a given object, to enhance its edge contrast, or to achieve both, resulting in a combination of higher resolution, larger exposure latitude, and DOF. The shifting of phase is accomplished by adding an extra patterned layer of transmissive material on the mask, as shown in Fig. 5.17. As light propagates through the substrate and the extra layer, its wavelength is reduced from the wavelength in air by the refractive index of the substrate and the extra layer. Comparing the optical paths through the extra material and without it, there is a difference of $(n - 1)a$, where n is the refractive index of the extra layer and a is its thickness. The phase difference θ is

$$\theta = 2\pi \frac{a}{\lambda} (n - 1). \quad (5.3)$$

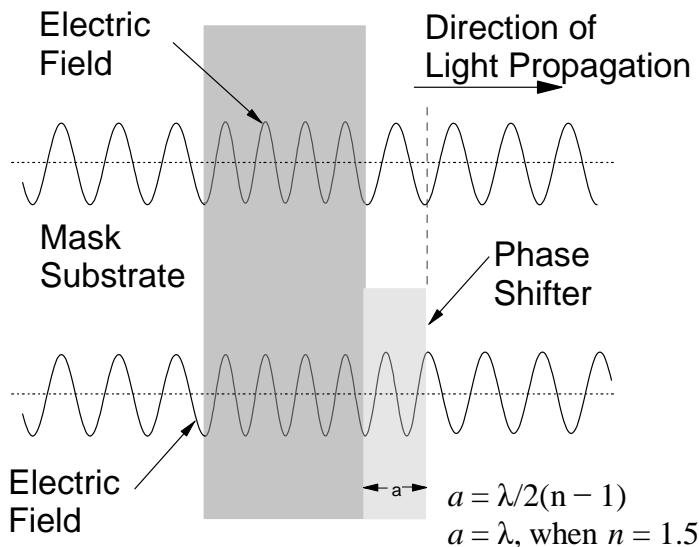


Figure 5.17 Principle of phase shifting.

Usually, a phase shift of π is desirable. One reason for this is that the electric-field amplitude range of ± 1 , according to the π shift, is the maximum amplitude range that can be produced. A second reason is that when any change of phase due to phase-shifter-thickness tolerance takes place, the rate of phase change is minimal with respect to the change of the phase-shifter thickness. In other words, $\partial \cos\theta/\partial\theta$ is a minimum at $\theta = \pi$. Making $\theta = \pi$ in Eq. (5.3) yields

$$a = \frac{\lambda}{2(n-1)}. \quad (5.4)$$

This is the thickness required to induce a π shift. Because the popular phase shifter SiO_2 has a refractive index of approximately 1.5, a is now approximately equal to the actinic wavelength. Note that the phase shift can be any odd number of π , such as $(2m+1)\pi$, where $m = 0, 1, 2, \dots$, and that phase shifting is relative. Phase shifting being relative means that either the high-refractive-index material in a phase-shifting mask or the low-refractive-index air path can be viewed as the phase shifter. To simplify the discussion, the extra layer is often called the phase shifter unless specifically defined otherwise.

5.3.4.2 Un-flat BIM is not a PSM

Newcomers to PSM technology often ask why a BIM with a flatness variation larger than or close to $\lambda/2$ is not a PSM. Phase shifting is effective only when the illumination is substantially coherent. It starts to be ineffective for $\sigma > 0.5$. The region where phase shifting is effective is on the order of the coherence length of the illumination. It is often a few micrometers at most, whereas flatness variation on BIM is gradual. The 0.5- to 2- μm -thickness variation spreads over different areas of the mask on the order of millimeters. How a PSM improves the optical image is explained in Sec. 5.3.4.3 below for each type of PSM.

5.3.4.3 PSM types and mechanisms of imaging improvement

Instead of using generic names, as in the beginning of Sec. 5.3, PSMs are usually named by the location or type of phase shifters. Following are the major types of PSMs.

Alternating phase-shifting mask (AltPSM)—This BCM system is characterized by phase shifting every other transparent element in a closely packed array.²⁰ The AltPSM approach is shown in Fig. 5.18(b) in comparison to BIM in Fig. 5.18(a). The former electric field amplitude is now -1 at the shifted areas on the mask. This negative amplitude effectively reduces the spatial frequency of the electric field so that it is less prohibited by the lens transfer function and forms a higher-contrast amplitude image at the wafer plane. When

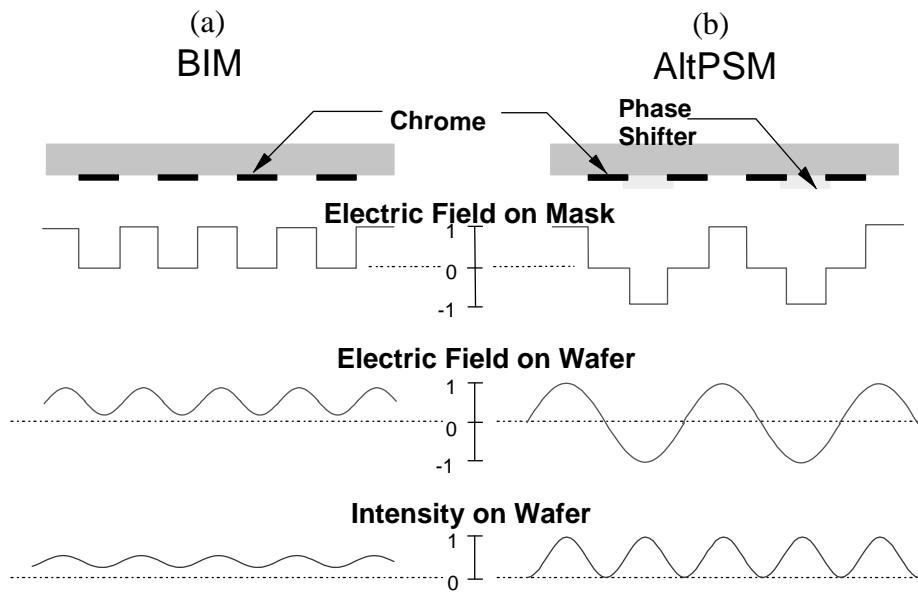


Figure 5.18 BIM and AltPSM.

this electric field is recorded by the photoresist, only the intensity, which is proportional to the square of the electric field amplitude, can be recorded. Hence, the reduced spatial frequency is doubled back to the original frequency, but the image results in a much higher contrast. In addition to the reduction of spatial frequencies, the electric field must pass through 0 to -1 assures zero intensity at the wafer; thus, it helps to improve edge contrast. Therefore, the AltPSM system benefits from the reduction of spatial frequencies as well as the enhancement of edge contrast. The foregoing illustration treats the illumination as completely coherent. In a real exposure system, the σ of the illumination is not 0, but between 0.3 and 0.4.

If coherence is lost, an AltPSM just becomes BIM. Figure 5.19 shows that with incoherent illumination the electric field fluctuates randomly in phase, and the spatial frequency doubling effect shown in Fig. 5.18(b) is not possible. Summing these into intensity simply produces an intensity image identical to that in Fig. 5.18(a).

The change of DOF of 0.35- μm line-space pairs at $\lambda = 365 \text{ nm}$ is shown in Fig. 5.20 as a function of σ . At $\sigma = 0.5$, DOF drops by 60% from the value at $\sigma = 0.3$. At $\sigma = 0.55$ and beyond, the image does not support any DOF at the given CD tolerance of $\pm 10\%$ and exposure latitude of 20%.

Subresolution-assisted phase-shifting mask (SA PSM)—AltPSM opens the door for phase-shifting technology by demonstrating its potential. However, AltPSM requires closely packed patterns to be effective. In actual circuit layouts, there are many situations where critical dimensions are sufficiently far away from any adjacent patterns to provide phase shifting. In order to provide

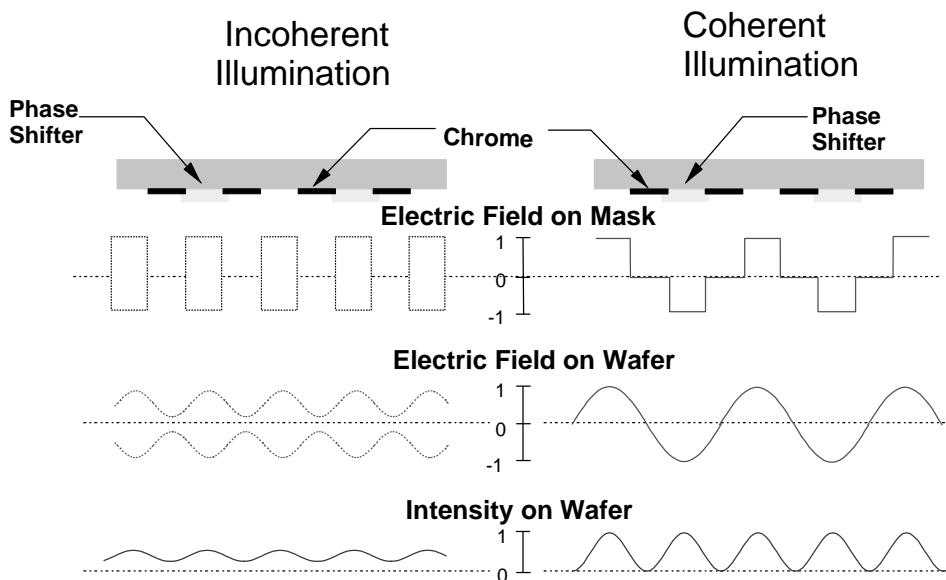


Figure 5.19 Incoherent illumination on AltPSM.

phase shifting for isolated openings, such as contact holes and line openings, another form of BCM uses subresolution phase shifters near isolated openings,²¹ as shown in Fig. 5.21(a). These phase shifters are smaller than the resolution limit of the optical imaging system; thus, they cannot be printed. Their sole function is to enhance the edge contrast of the pattern of interest.

GNA\LSAL NA.32 OpSigma.3 MaxDOF4.99um@20%ELat WL365 CD.35

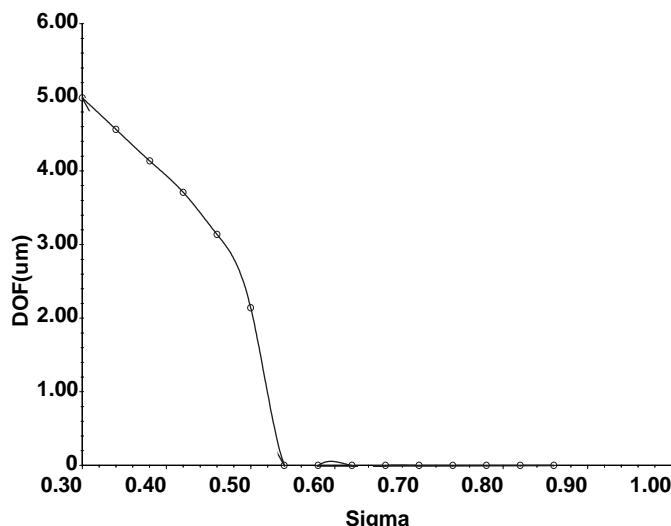
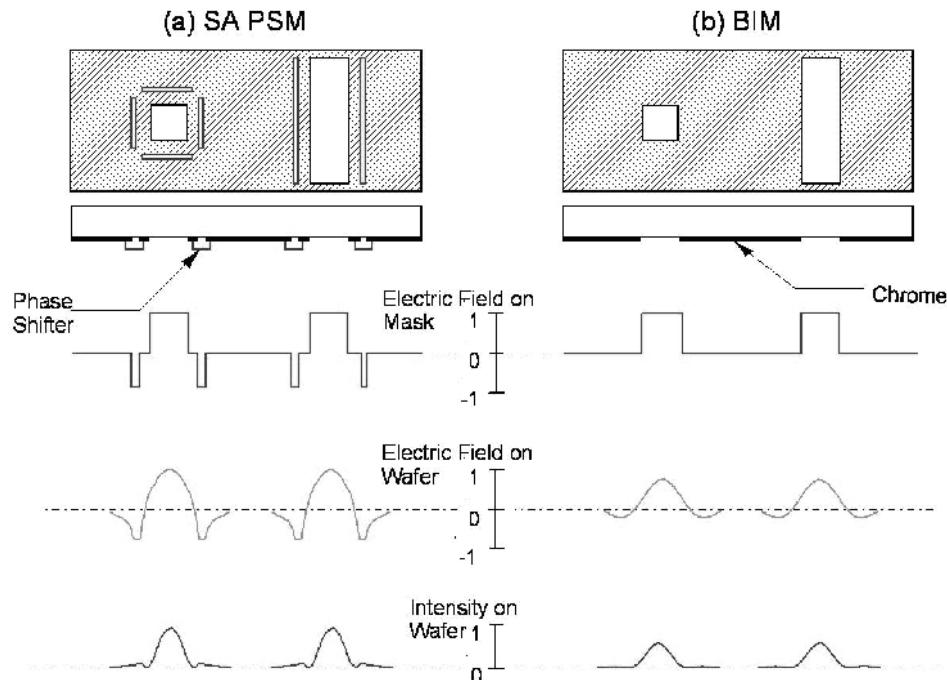


Figure 5.20 DOF of 0.35- μ m line-space at 365 nm for AltPSM as function of σ .



Rim phase-shifting mask (rim PSM)—SA PSM and AltPSM are still limited by their inability to provide phase shift to opaque patterns. Rim PSM,²² yet another form of BCM and shown in Fig. 5.22(a), overcomes such a problem and can be applied to an arbitrary mask layout. Here, phase shifting only takes place at the rim of the mask patterns. The center of the patterns is blocked by the absorber to prevent large areas of negative amplitude from producing bright areas where they are supposed to be dark. Again, bright areas result from negative or positive field amplitudes because the photoresist can only detect intensity that is proportional to the square of the electric field. Note that edge contrast enhancement is now the sole image-improving function of these phase shifters.

Attenuated phase-shifting mask (AttPSM)—The AttPSM^{23,24,25} is a BCM that applies to arbitrary mask layouts, just like rim PSM. It can be implemented on either a transmissive or reflective mask. The dark areas of the mask can be phase shifted to π but with an attenuated amplitude to prevent producing too much light in these areas, as shown in Fig. 5.22(b). The negative amplitude provides the desired improvement in image edge contrast, and the attenuation prevents the negative amplitude from becoming too large and subsequently exposing the resist. Regular absorbers often must be included at the borders of the mask plate to block off light at the boundary of the exposure field and to provide regular reticle alignment marks. Therefore, AttPSM often takes the form of MCM.

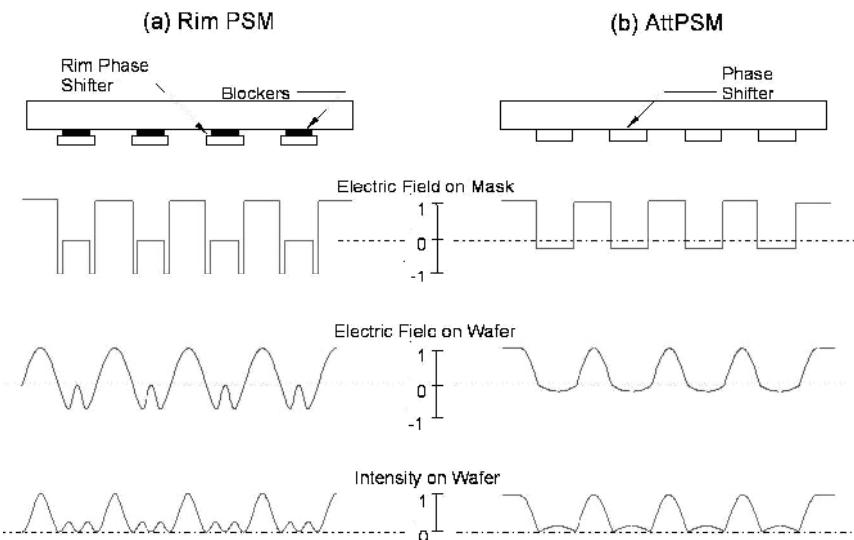


Figure 5.22 Rim and AttPSM.

Phase-shifted edge (PS edge)—A phase-shifted edge²⁶ is simply the boundary between the phase-shifted and the unshifted regions. Generically, it is a BPM. A high-contrast dark-line image following the boundary is produced, as shown in Fig. 5.23(a).

Covered phase-shifting edge (CPS edge)—This is similar to the phase-shifting edge, except that the edge is no longer just the boundary.²⁷ It is covered with an absorber as well, as shown in Fig. 5.23(b). The covered phase-shifting edge can be viewed as one half of each opening in the AltPSM.

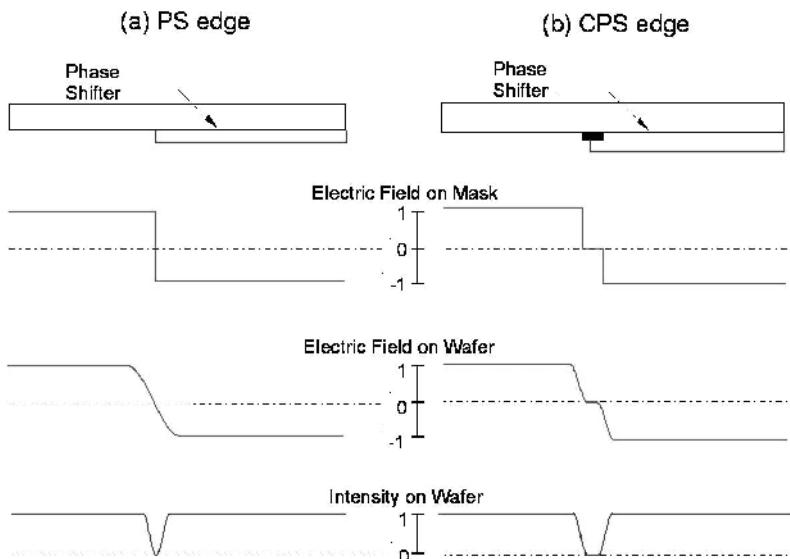


Figure 5.23 Phase-shifting edge and covered phase-shifting edge.

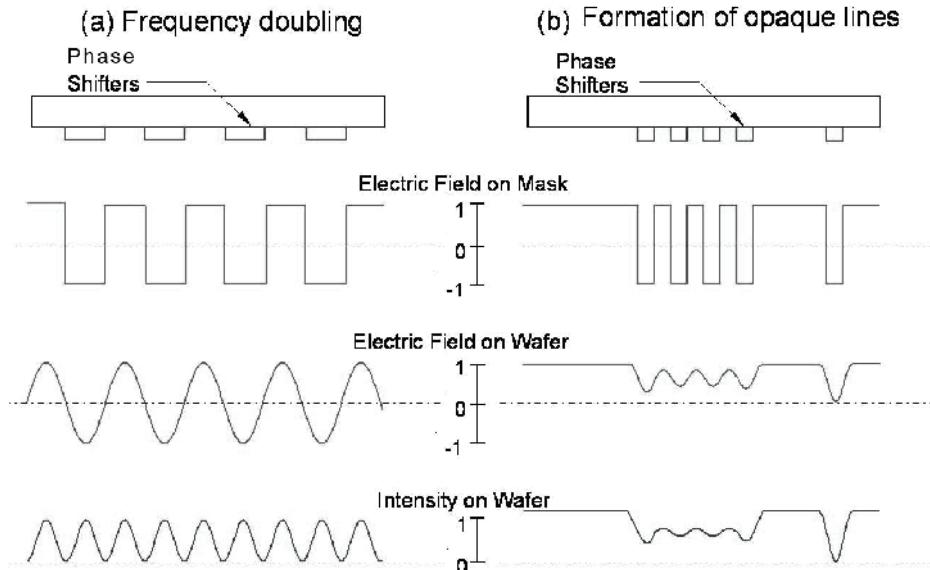


Figure 5.24 Configuration of PS edges.

5.3.4.4 PSM configurations

With the basic PSM types, many different configurations can be built:

Configurations from phase-shifting edge—Two phase-shifting edges of opposite polarities can be combined to form an unattenuated PS line. When the edges are close to each other, opaque images are formed, as shown in Fig. 5.24(b). The unattenuated PS lines with large edge separations can be combined periodically to double the spatial frequency, as shown in Fig. 5.24(a). These lines can be double exposed orthogonally to produce isolated opaque images,²⁶ as shown in Fig. 5.25. With a positive resist, the resultant image is a pedestal; with a negative resist, we get a hole. Note that the image in the resist is rotated by 45 deg. This is due to the isointensity lines in the 45-deg and 135-deg directions resulting from the orthogonal superposition of the intensity distribution of the PS edge shown in Fig. 5.23.

Combination of AltPSM and rim PSM—The AltPSM is a strong phase-shifting scheme that works only on closely packed patterns, whereas the rim PSM is a weak phase-shifting scheme that works better on patterns with larger separations. It is possible to combine them on the same layout to take advantage of their complementary characteristics.^{28,29} Figure 5.26 shows the combined layout and the fabrication steps. A three-level exposure is used to write the mask. The partially exposed areas define the phase-shifter areas. The unexposed areas define the areas to be covered with chrome. The exposed areas become the transparent unshifted areas.

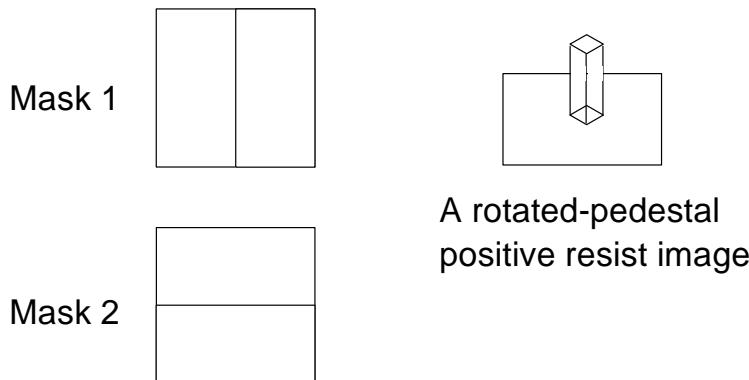


Figure 5.25 Double exposure of orthogonally displaced unattenuated PS lines.

Combination of AltPSM, AttPSM, and unattenuated phase-shift mask (UttPSM)—A combination of AltPSM and AttPSM is also possible. Actually, one can even add UttPSM in the combination.³⁰ The combination process is not trivial as the Alt/rim combination, but the imaging performance of AttPSM is better than that of rim PSM, as will be shown in Sec. 6.3.1.2. Figure 5.27 shows a mask containing AltPSM, AttPSM, and UttPSM simultaneously. To make such a mask, one uses a quartz substrate containing a layer of attenuating phase shifting with a π shifter on top of it. The attenuation is typical for what is required of an AttPSM, e.g., 6%. The amount of phase shifting for the attenuating phase shifter is not critical. We have drawn 30 deg for illustration purposes. The key is to etch out a total of 2π phase shift from the attenuating phase shifter and the phase

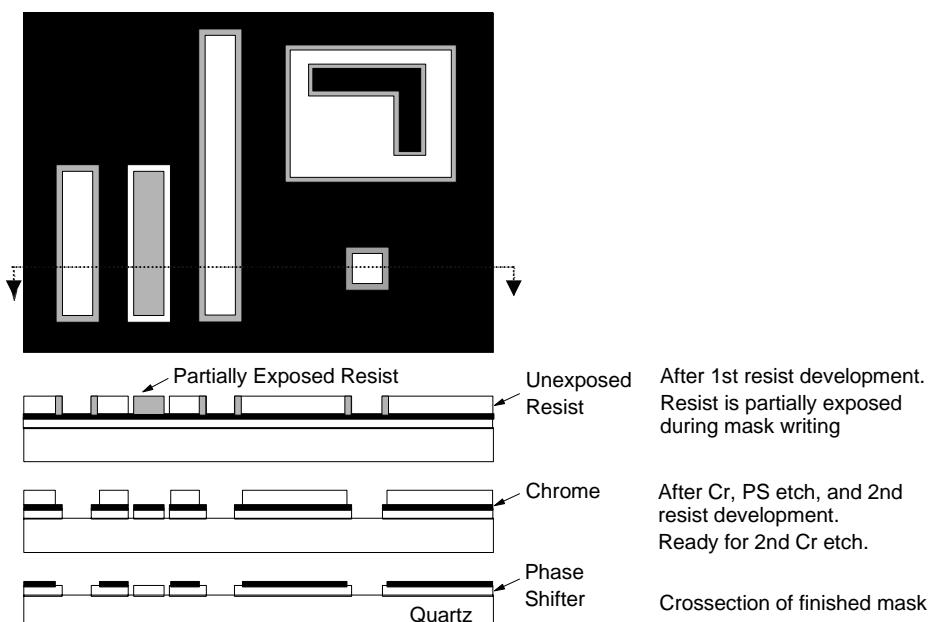


Figure 5.26 Combination of AltPSM and rim PSM.

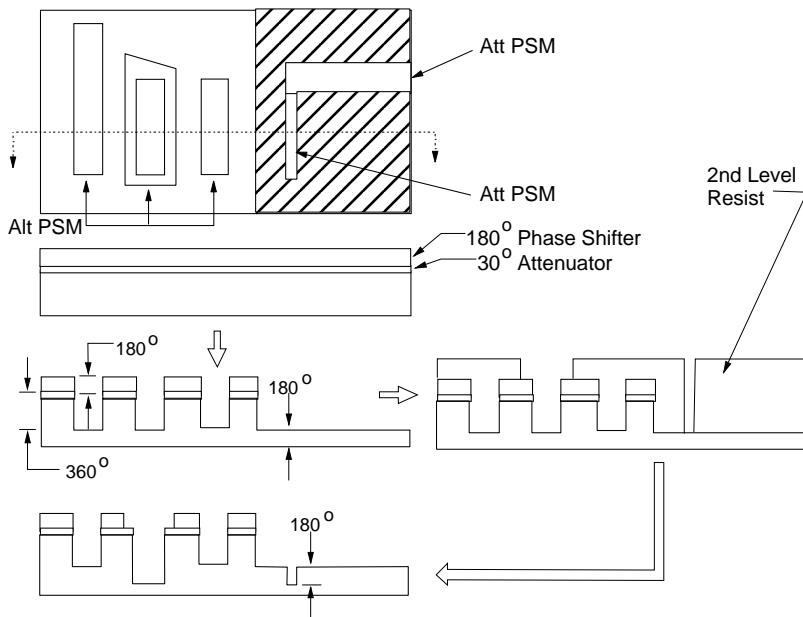


Figure 5.27 Combination of AltPSM, AttPSM, and UttPSM.

shift in the quartz substrate. The first patterning delineates all areas that shift phase by 3π , which is effectively just π . Then the zero phase and the UttPSM areas are opened with another layer of resist and patterning. The total etched depth in these areas is either π or 4π . The former is for UttPSM, and the latter is for zero phase shift.

This AttPSM/AltPSM/UttPSM combination is later replaced by the simpler combination of AttPSM with off-axis illumination, which is a very popular combination for semiconductor manufacturing starting from the 0.18- μm node and is still used for the 65-nm node and below. The performance of this combination will be given in Chapter 6.

Combination of CPS edge and BIM—The gate length of metal-oxide-semiconductor (MOS) devices is very critical, whereas the dimension elsewhere at the gate level is more forgiving. Therefore, the CPS edge can be combined with BIM by double exposure to precisely control the gate length.³¹ Shown in Fig. 5.28, mask 1 is a BIM consisting of the complete layout of the gate level, except that the gate length is intentionally made larger so that the BIM can support the entire layout. After exposing mask 1, mask 2 (which contains the CPS edge at all of the critical gates) is exposed over the enlarged gate areas defined by mask 1. This narrower pattern with higher contrast sets the gate length in the developed resist image and improves the linewidth control. This combination has propelled the company that owns the patent to fame. Besides the obvious handicap of lost productivity with double exposure, a minimum distance

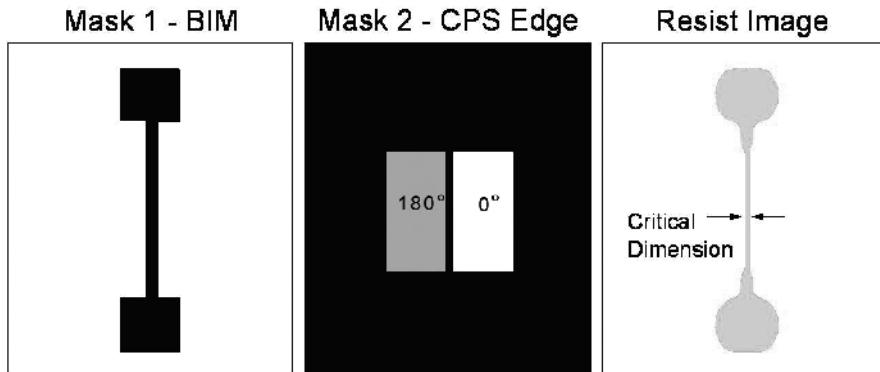


Figure 5.28 Combination of CPS edge and BIM.

between the CPS edges must be set, which limits the reduction of cell size. The popular combination of off-axis illumination with AttPSM can be optimized to provide similar imaging performance as this combination. Eventually, higher cost and longer cycle time of the CPS mask also limits the manufacturing application of the technique. Consequently, this manufacturing company is no longer in existence.

Possible configurations are not exhausted with the above examples. Some combinations improve imaging performance. Some are just the opposite. It is important to evaluate imaging performance quantitatively, as we did in Sec. 4.5.2.11.

5.4 Imaging Lens

The function of the imaging lens is to reproduce the mask pattern at the wafer with the given reduction ratio. It is a crucial part of the entire imaging system. Its power and quality has a direct impact on the resolution of the imaging system.

5.4.1 Typical lens parameters

5.4.1.1 Numerical aperture

The power of the lens is represented by its numerical aperture (NA), which is defined as the sine of the half angle that the lens aperture sustains from the object or from the image. In a 1X system the NA at the image side and at the object side are the same. With a reduction system, the NA is smaller at the object side by the reduction factor. When the medium has a numerical index other than 1, then

$$NA \equiv n \sin\theta, \quad (5.5)$$

where n is the refractive index of the medium between the front surface of the imaging lens and the wafer. This relationship was depicted in Fig. 4.1. The numerical aperture determines the angular extent of the imaging light: a larger NA captures a wider angular spectrum for imaging; therefore, it resolves a

smaller feature. The resolution of a given imaging lens is related to NA by the resolution scaling equation introduced in the earlier chapters:

$$MFS = k_1 \frac{\lambda}{NA}, \quad (5.6)$$

where λ is the wavelength and k_1 is the scaling coefficient for resolution. More discussion on the scaling equation and k_1 were given in Chapter 4.

The NA cannot be increased at will. First, it is progressively difficult to keep the lens free of aberrations. A larger number of lens elements requiring higher precision in the elements and in their assembly makes the imaging lens extremely expensive. It has been estimated that the difficulty in designing and building a lens is proportional to the fifth power of NA.³² Second, a larger NA may lead to unacceptable DOF as predicted by the scaling equation for DOF:

$$DOF = k_2 \frac{\lambda}{NA^2}, \quad (5.7)$$

where k_2 is the scaling coefficient for DOF and was discussed in Chapter 4. The shrinking of DOF at a large NA and the loss of resolution at a low NA point to the existence of an optimum NA, which will be discussed in Chapter 6. Even though Eqs. (5.6) and (5.7) have been widely adapted, there are difficulties applying them to modern imaging optics because of the paraxial nature of k_2 and ambiguity in applying them to immersion lithography.³³ A new set of equations replaces them:

$$W = k_1 \frac{\lambda}{\sin \theta} \quad (5.8)$$

$$DOF = k_3 \frac{\lambda}{\sin^2(\theta/2)}. \quad (5.9)$$

The derivation of Eqs. (5.8) and (5.9) can be found in Sec. 3.1.3, and their application to immersion lithography will be covered in Sec. 7.1.

5.4.1.2 Field size

The field size of a lens has a direct impact on how many chips can be exposed in one exposure step; this is discussed in Chapter 2. From the optical point of view, even though the image field of a lens is circular, a microlithographic lens is usually specified in terms of certain rectangles or the largest square that can be used. For economical reasons, it is more important to populate the wafer with rectangular fields than to fully utilize the circular field. The unused areas are blocked off with movable blades.

Similar to the difficulty in designing and building lenses with a large NA, it is also very difficult to make diffraction-limited lenses with a large field size. The difficulty has been estimated to be proportional to the third power of field size.³² Increasing field size worsens lens distortion. Other aberrations, such as astigmatism and coma, are also field-size dependent.

5.4.1.3 Reduction ratio

The reduction ratio M of the lens has impacts on CD control and overlay accuracy, as seen in Tables 2.1 to 2.3. It increases the DOF by M^2 and reduces NA by M on the mask side. Making a reduction lens is more difficult than a 1X lens because there is no longitudinal symmetry to take advantage of. The reduction ratio has been kept at five for a long time because it is a good compromise between the difficulty of mask making, field size, and mask size. With step-and-scan mask aligners, M is taken to be four to accommodate the increased field size. Further reducing M can have a severe impact on CD control and overlay accuracy. Therefore, the tendency is to use larger mask blanks instead of dropping M further. Of course, there are large-field, low-resolution mask aligners that use an M smaller than four.

5.4.1.4 Working distance

The working distance is the closest distance between the first lens surface and the wafer. It is usually between 5 and 10 mm. The distance must be small to keep the NA large, but a reasonably large distance is required to facilitate defect-free fast stepping. With immersion imaging, the working distance usually drops to below 3 mm so that the thermal control and inhomogeneity of the immersion fluid does not contribute significantly to imaging errors. More details on the working distance in immersion lithography will be given in Sec. 7.5.2.

5.4.1.5 Telecentricity

When the principle ray of the imaging beam at any point in the image field is perpendicular to the image plane, such as the wafer plane, the system is telecentric on the image side. Being telecentric makes the lateral location of the image point insensitive to defocus; thus, it reduces overlay errors. Modern steppers are telecentric on both the mask side and the wafer side to ensure the stability of the reduction ratio.³⁴ Both types of telecentric systems are depicted in Fig. 5.29.

5.4.2 Lens configurations

There are basically three types of imaging lenses based on the type of lens elements. The dioptric system consists exclusively of refractive elements. All-reflective systems contain only reflective elements. Catadioptric systems mix refractive and reflective elements.

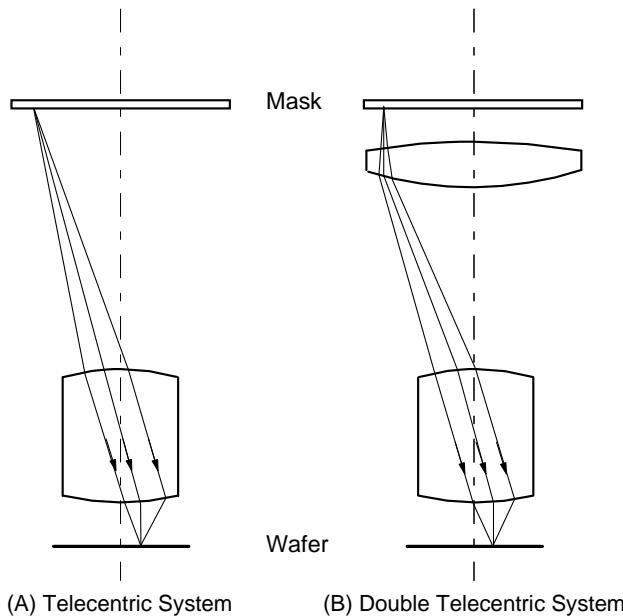


Figure 5.29 Telecentric imaging systems.

5.4.2.1 Dioptric systems

The majority of microlithographic lenses are dioptric. An example is given in Fig. 5.30. Here, a 248-nm 0.8-NA 4X-reduction lens consisting of 20 fused-silica elements is shown.³⁵

The advantage of dioptric systems is in maintaining a single optical axis from the mask to the wafer. This makes for easier assembly and alignment of the elements. However, because refractive elements are inherently dispersive, i.e., the refracting angles are wavelength dependent, many more elements and a mixture of lens material with mutually compensating dispersion characteristics are required to suppress chromatic aberrations. As the difficulties of chromatic aberration increase, the usable bandwidth of the exposure light is narrowed. For i-line steppers, the bandwidth is 6 nm, compared to 10 nm for g-line steppers. At a wavelength of 248 nm and shorter, the practice is strictly to use one single type of material while placing the burden entirely on a very narrow bandwidth. Fortunately, with excimer lasers the bandwidth can be reduced significantly without inducing a substantial loss of energy. The bandwidth requirement in this case is between 0.1 and 10 pm, depending on whether a dioptric or catadioptric system is used, the number of transmissive materials, and the center wavelength.

5.4.2.2 Reflective systems

A well-quoted example of a reflective imaging system was shown in Figs. 2.3 and 2.4. The two concentric mirrors reproduce the mask pattern in a ring region to another part of the same ring. Hence, if the mask is blocked by a plate with a

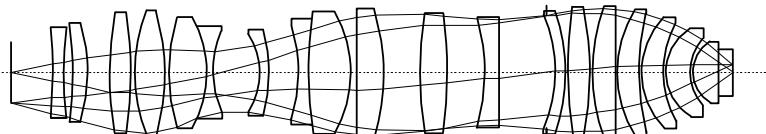


Figure 5.30 A 248-nm 0.8-NA 4X-reduction dioptric lens system.

slit that is a section of the ring, and the same is done with the wafer, the mask and wafer can be scanned at the same speed to reproduce the image of the entire mask. Three plane mirrors mounted on a single block of glass are used to bend the light. This way, the mask and the wafer now face each other and can be moved in the same direction. Previously they were on the same plane and had to be scanned in opposite directions. The configuration of two concentric mirrors is the only all-reflective imaging scheme that has been successfully used in manufacturing. Being completely reflective, the bandwidth of the system is unlimited except for the efficiency of the mirror coatings. This system has been used at near-UV to extreme-UV wavelengths to demonstrate imaging and develop exploratory photoresist systems. Hence, the advantages of a reflective system are a small number of optical elements and freedom from chromatic aberration.

The disadvantage of a reflective system, with the exception of two concentric mirrors, is the difficulty in keeping all of the optical elements on one straight optical axis. Aberration correction is also difficult, because introducing extra elements almost always follows with an increase in the number of optical axes. The surface and assembly tolerances must be at least twice as good as that of a refractive system because any irregularity affects both the incident and the reflected beam on each element. More details on reflection projection optics will be given in Sec. 8.2.4.

5.4.2.3 Catadioptric systems

There are both refractive and reflective elements in a catadioptric system. Usually, there is a spherical or aspherical mirror that performs the major part of imaging. Refractive elements are inserted to remove aberrations. There may be planar reflective elements to bend the rays for better arrangement of the mask, the wafer, or other optical elements. This is exemplified by Fig. 2.7, which shows the 1X catadioptric lens used in Ultratech mask aligners. The two 45-deg mirrors redirect light from the mask to the wafer. The main mirror and the refractive achromat are also shown. Because the incoming and the outgoing beams to and from the mirror cannot share the same optical axis, only a semicircular field can be used, as shown in Fig. 5.31.

Figure 2.10 showed a catadioptric lens system with a beamsplitter. It is similar to the one used in the SVGL step-and-scan mask aligner. Again, there is a mirror that is responsible for reduction imaging. The beamsplitter is induced to keep the incoming and outgoing beams to and from the mirror out of each other's way, so that the entire mirror can be used. There are also refractive elements to correct for aberrations.

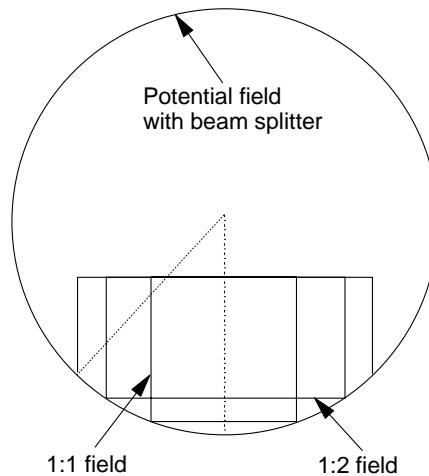


Figure 5.31 A semicircular lens field and the exposure fields from the lens system in Fig. 2.10. The entire circular field can be made available with a beamsplitter.

There are several considerations for introducing a beamsplitter into the lens system. First, a conventional beamsplitter splits light into two paths. Hence, it reduces the intensity of each beam by half. One of the beams is not only wasted; if not treated properly, it can contribute to stray light. The useful beam is split into two again when it travels back to the beamsplitter, before exposing the wafer. Only 25% of the illumination is used, and the system can be prone to stray light. To overcome this problem, a polarizing beamsplitter³⁶ is used. As shown in Fig. 5.32, light in the s-polarization mode is incident on the beamsplitting plane of the cube. The beamsplitting plane is composed of a multilayer stack to form a reflecting and refracting interface. This stack is designed to satisfy the Brewster's

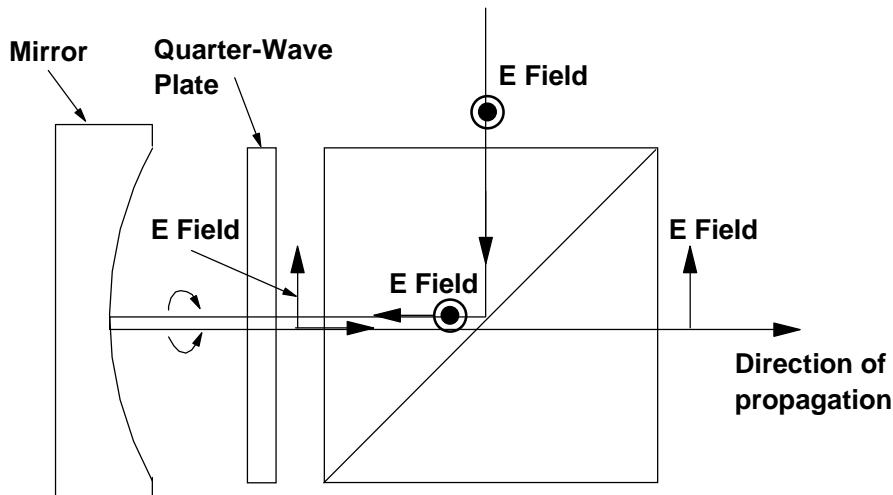


Figure 5.32 The working principle of a polarizing beamsplitter of the catadioptric lens system in Fig. 2.10.

condition for beams at a 45-deg incident angle, such that the reflected beam is in the s-polarization mode and the refracted beam is in the p-polarization mode. Since the incident beam is s-polarized, it is completely reflected toward the quarter-wave plate. The beam becomes circularly polarized after leaving the quarter-wave plate. It is reflected by the mirror and enters the quarter-wave plate again. The beam that reenters the beamsplitter is now in the p-polarization mode whose reflection at the Brewster interface is zero, and refraction into the other side of the cube is 100%.

The second consideration with a beamsplitter is that it must be a very large and thick optical element, in order to have a large field size. The homogeneity and durability requirements of the element are not easy to meet.

5.4.3 Lens aberrations

Lens aberration arises from the deviation of a spherical wavefront that converges to an image point, as was discussed in Sec. 3.1.4 and shown in Fig. 3.5. The deviation of a spherical wavefront is caused by

- (1) finite design residue,
- (2) imperfect surface configuration of the optical elements in the imaging lens,
- (3) placement error of the optical elements due to or after their assembly.

The deviation of a spherical wavefront can be quantified with Zernike polynomials or Seidel coefficients. Because of cylindrical symmetry, lens aberrations in optical microlithography are almost always specified in terms of Zernike polynomials, as was covered in Sec. 3.1.4.2. Note that the $\sin(m\phi)$ -dependent terms should be included because reasons 2 and 3 above induce asymmetry. Reason 3 can take place in the field due to lens heating or a change in temperature or humidity of the lens environment. Mechanical impact to the lens also affects lens aberrations.

Generally speaking, spherical aberrations enlarge image points as if they are defocused. Coma induces displacement in two adjacent images. Astigmatism causes CD variation between features of different orientations. Spherical aberration is uniform across the lens field. Coma and astigmatism are field dependent. Distortion does not affect resolution, but it changes the image placement and is sensitive to field location. These characteristics help to sort out lens-induced and mask-induced distortions.

5.4.4 Lens fabrication

With modern electronic computation capability and the availability of lens-design software, lens design has become less formidable than fabricating the highly sophisticated lens. Each refractive surface must be polished to a $10\text{--}20 m\lambda$ error. For mirror surfaces, the precision required is doubled. When the elements are assembled, they must be centered to better than a fraction of a micrometer.

Surface errors of individual elements are inspected with a wavefront interferometer and corrected before assembling. Subsystems assembled with several elements are also checked with the wavefront interferometer. Assembly errors will be evaluated from the interferometer result and a computer. The subassembly will be reworked if necessary. Because of the uncertainty of this trial-and-error approach, many lens manufacturers have developed high-precision assembling techniques that produce instant feedback to eliminate rework. Even so, manufacturing a high-NA imaging lens is a highly skillful and laborious process. When demand exceeds supply for high-density semiconductor chips, the semiconductor factory capacity is usually limited by the production of mask aligners. They, in turn, are limited by the production capacity of lens suppliers. The lens suppliers may be dependent on the availability of high-quality lens materials. At one time, SiO_2 was the gating material. Later, lens fabrication was gated by CaF_2 , which is required to reduce absorption loss and to achromatize 193-nm imaging lenses. It is also the sole refractive material for 157-nm lenses. The limitation of producing it in a quantity and quality that meets the more stringent requirement of 157-nm imaging is one of the key factors stopping 157-nm lithography from being further developed, despite an enormous amount of development from the equipment and material suppliers.

5.5 Lens Maintenance

The care in assembling the lens does not guarantee that the lens is free from damage or deterioration. There are many things that users should be aware of and pay attention to:

Prevention of mechanical shocks—Any mechanical shock or vibration that can change the assembly precision of the lens must be avoided. The highest possibility of these disturbances comes from transporting the lens from its factory to the semiconductor manufacturing facility. Once there, locating or relocating the lens can also be problematic. Accelerometers are used to monitor the transportation of lenses, steppers, and scanners for quality control of the transportation process.

Prevention of environmental changes—It is well known that slight temperature and barometer changes in the lens environment can change the refractive index of the lens elements.³⁷ They can also affect the assembly precision, if the lens mounting materials are not carefully engineered. Hence, the imaging lens in modern steppers or scanners are enclosed in environmental chambers regulating the temperature to better than 0.1°C . The lens itself is maintained at a designated pressure. It is crucial that the mask aligner's power stays on continuously so that these controls work without interruption. Otherwise, it takes a long time to restabilize or restore the system.

Surface contamination—Any dust particle that falls on any of the lens surfaces becomes a scattering center that produces stray light and flare, thus reducing the image contrast and, hence, linewidth control. Therefore, the lens is assembled in a dust-free environment. In the field, the lens should not be subjected to any particulate contamination. When the photoresist outgasses, there is a possibility of accumulating a layer of polymer material on the front surface of the lens. This also must be avoided.

Radiation damage—The glues for lens elements, antireflection coatings on the lens surfaces, and the lens materials themselves can deteriorate as exposure dosages accumulate. Extremely-high-peak exposing pulses can accelerate the deterioration. Lenses have been known to require refurbishing at the lens factory or even replacement after a year or two. However, there is not much the mask aligner user can do to prevent this kind of damage, except to avoid unnecessary exposures and accidental introduction of high-peak pulses into the optical train.

5.6 Photoresists

The photoresist is a light-sensitive polymer that undergoes light-induced chemical reactions to produce a dissolution rate as a function of the exposure level. The difference in dissolution rate in the exposed and unexposed areas causes the resist to be removed selectively while being developed. The areas still covered with resist are protected when the wafer is subjected to etching, whereas the areas absent of resist are etched.

The resist is used in semiconductor processing in many ways, as shown in Fig. 5.33. The resist can be used as a protection mask for etching. The areas exposed to the etchant are removed either isotropically or anisotropically. It can also be used as a lift-off mask. In this case, an undercut profile in the resist is produced so that a material anisotropically deposited on the wafer is separated at the edges of the resist, resulting in the removal of the material with the resist when it is stripped. The resist can also be used as an ion-stopping mask to selectively implant ions into the semiconductor.

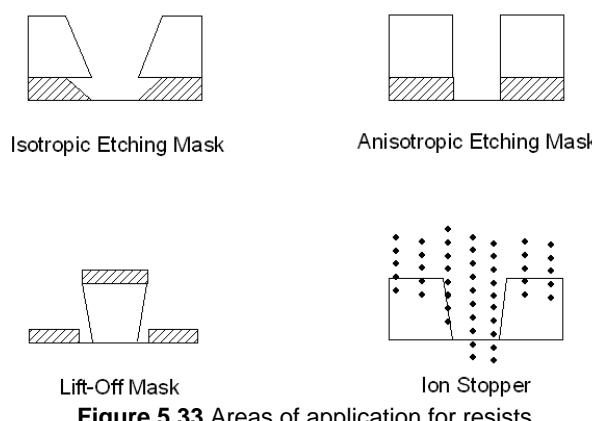


Figure 5.33 Areas of application for resists.

Resist		
Imaging Configuration	Working Principle	Polarity
Bulk Imaging	One-Component	Positive
Surface Imaging	Two-Component	Negative
Multilayer	Chemically Applied	
Contrast Enhanced	Image Reversal	

Figure 5.34 Classification of resist systems.

5.6.1 Classifications

The photoresist can be classified in many ways, namely, in polarity, in working principle, and in configurations shown in Fig. 5.34.

5.6.1.1 Polarity

There are positive and negative resists. The positive resist becomes more soluble in the exposure areas than the unexposed areas and vice versa for the negative resist, as shown in Fig. 5.35. The difference between the resist and the latent image profiles in this figure will be explained in Sec. 5.6.2 below.

A positive resist is desirable in many respects:

Optical imaging—The diffracted light from isolated openings, either lines or holes, always shows better defocus characteristics than light from isolated opaque features, such as isolated opaque lines or islands. The difference in aerial images between a hole and an opaque island of the same size is seen in Fig. 5.36.

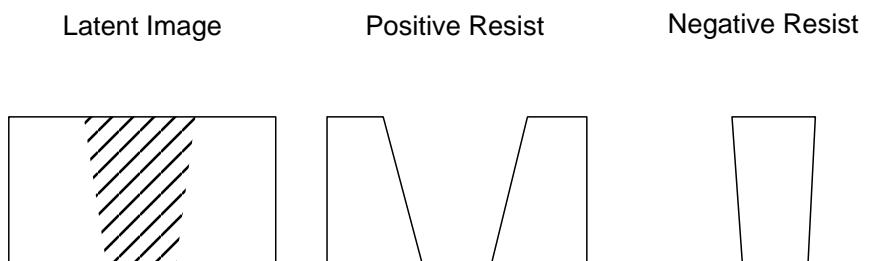


Figure 5.35 Positive and negative resists.

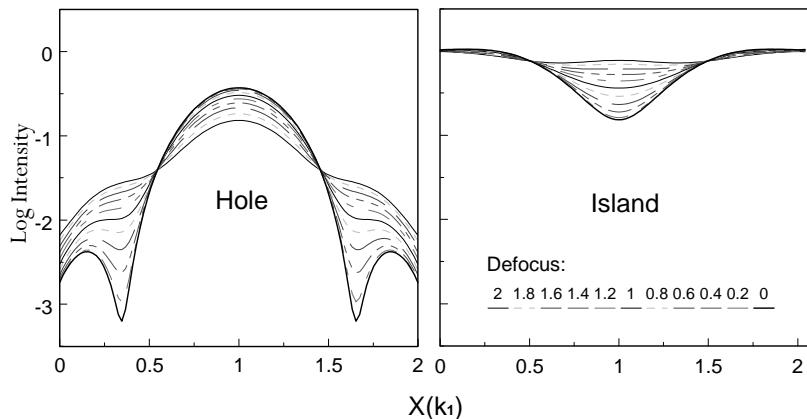


Figure 5.36 Aerial images of a hole and an opaque island.

Protection from ghost lines—As seen in Fig. 5.37, at a sufficient distance from focus, the aerial image degenerates into ghost regions, producing extra lines that are not on the mask. When the top surface of the resist is above the ghost regions, the unexposed positive resist prevents the exposed ghost images from developing. Note that this protection is lost if the top surface of the resist intersects the ghost image, as in the case of incorrect focus.

Freedom from shrinkage—Earlier negative resists exhibit significant shrinking after developing. Not only is the resist thickness reduced and its profile wrinkled, but also the stress can bend the resist line out of its nominal

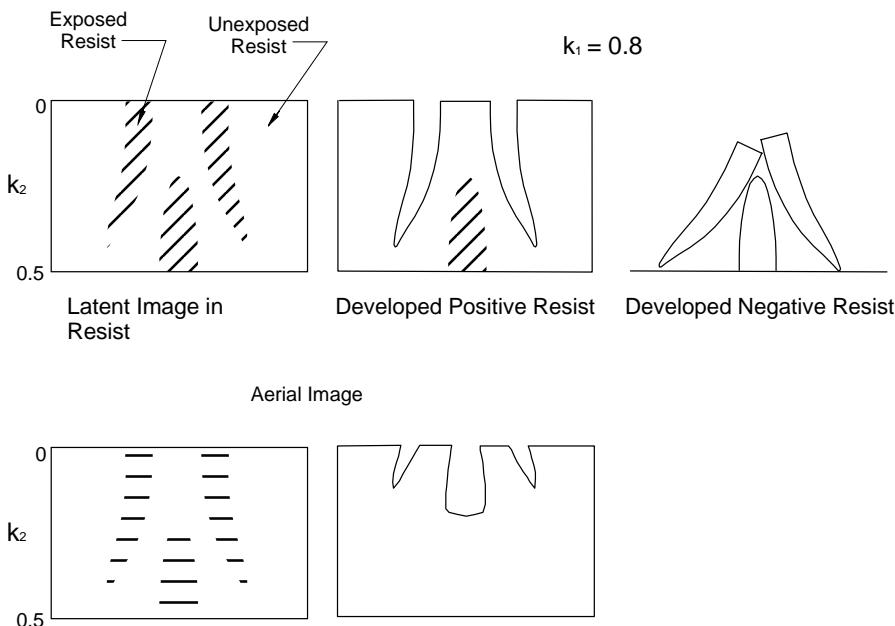


Figure 5.37 Protection of ghost regions with a positive resist.

position. Modern negative resists do not suffer from this shrinkage problem. However, the contrast of negative resists is usually lower than that of positive resists. In addition, negative resists tend to require solvent developers rather than aqueous developers. The latter is much more welcomed in a factory.

Negative resists are more desirable than positive resists in certain circumstances. For example, when high-resolution isolated resist lines must be printed, negative resists support a larger DOF because the aerial image of an isolated line opening is much better than that of an isolated opaque space, as discussed above.

5.6.1.2 Working principle

One-component positive resist systems—The one-component positive resist system works on the difference in dissolution rate induced by changes in molecular weight. High-energy radiation chain scissions a larger polymer molecule into smaller ones, as shown in Fig. 5.38. The change can be of several orders of magnitude, as reported by Moreau,³⁸ who also presented the following relationship between dissolution rate and molecular weight changes:

$$\frac{R}{R_o} = \left(\frac{M_o}{M} \right)^{-a}, \quad (5.10)$$

where R is the dissolution rate of the exposed resist, R_o is the dissolution rate of the unexposed resist, M_o is the initial molecular weight, M is the molecular weight that leads to R , and a is the developer's kinetic solubility parameter.

Polymethylmethacrylate (PMMA) is a well-known one-component positive resist that is sensitive to deep UV, EUV, e-beam, and ion-beam radiations.

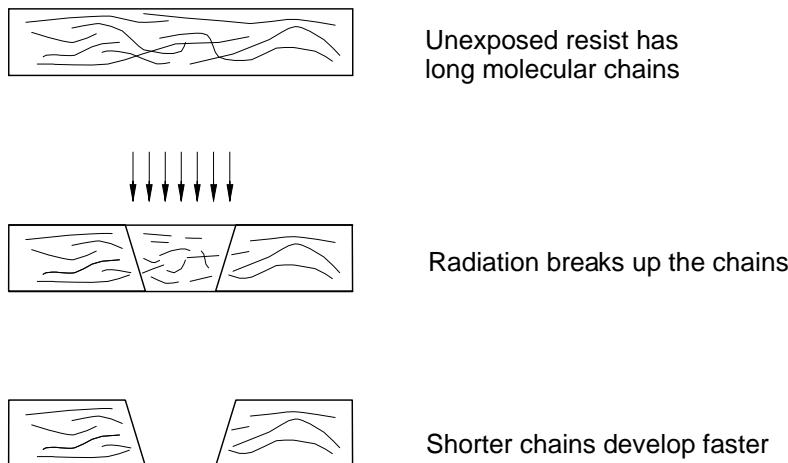


Figure 5.38 Chain scission of a one-component positive resist.

One-component negative resist systems—The one-component negative resist system works on crosslinking polymer chains upon radiation, as seen in Fig. 5.39. The crosslinked chains become harder to dissolve, thus the resist behaves like a negative resist. Moreau showed the relationship of the negative resist contrast γ to its polydispersity in single- or multicomponent systems:

$$\gamma = \left(\frac{2}{E_g} \right) \exp(-B^2) \quad (5.11)$$

and

$$B \equiv \frac{M_w}{M_n}, \quad (5.12)$$

where E_g is the gel point dose,³⁹ M_w is the weight-average molecular weight, and M_n is the number-average molecular weight.

Two-component positive resist systems—A two-component positive resist system consists of a base polymer and a sensitizer. The popular diazoquinone-sensitized novolak (DQN)⁴⁰ resist has a base polymer that is highly soluble in an alkaline developer. The sensitizer inhibits dissolution by as much as two orders of magnitude, as shown in Fig. 5.40. After exposure, the sensitizer turns into a dissolution enhancer that can further increase the dissolution rate by another two orders of magnitude, also shown in Fig. 5.40. The function⁴¹ of each component in the DQN resist is shown in Fig. 5.41. The diazoquinone sensitizer, in the presence of actinic light and water found in the resist layer, produces a base-soluble indenecarboxylic acid and nitrogen gas. The indenecarboxylic acid is responsible for increasing the dissolution rate at the exposed areas. The imaging of the DQN resist is shown in Fig. 5.42.

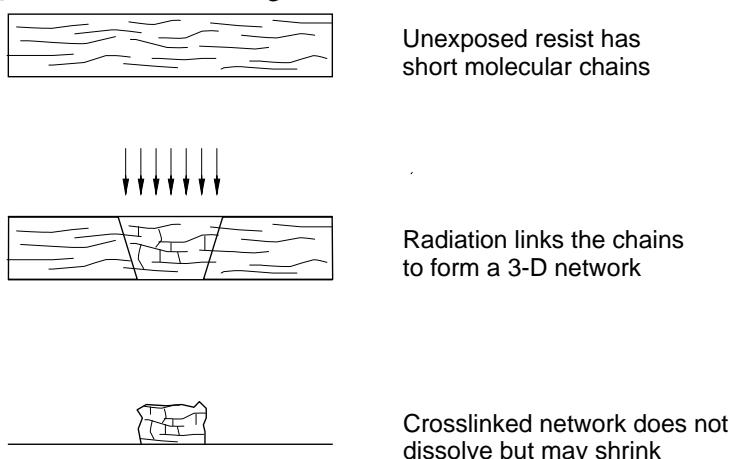


Figure 5.39 Crosslinking of a one-component negative resist.

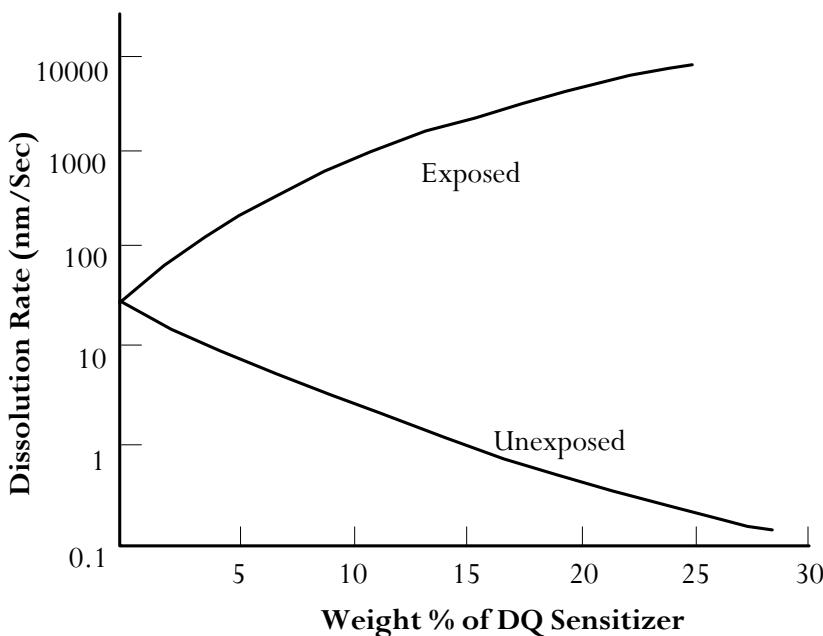


Figure 5.40 Effect of sensitizer concentration on DQN resist. [From D. Meyerhofer, "Photosolubility of diazoquinone resists," *IEEE Trans. Electron. Dev.* **27**(5), pp. 921–926 (1980). © 1980 IEEE.]

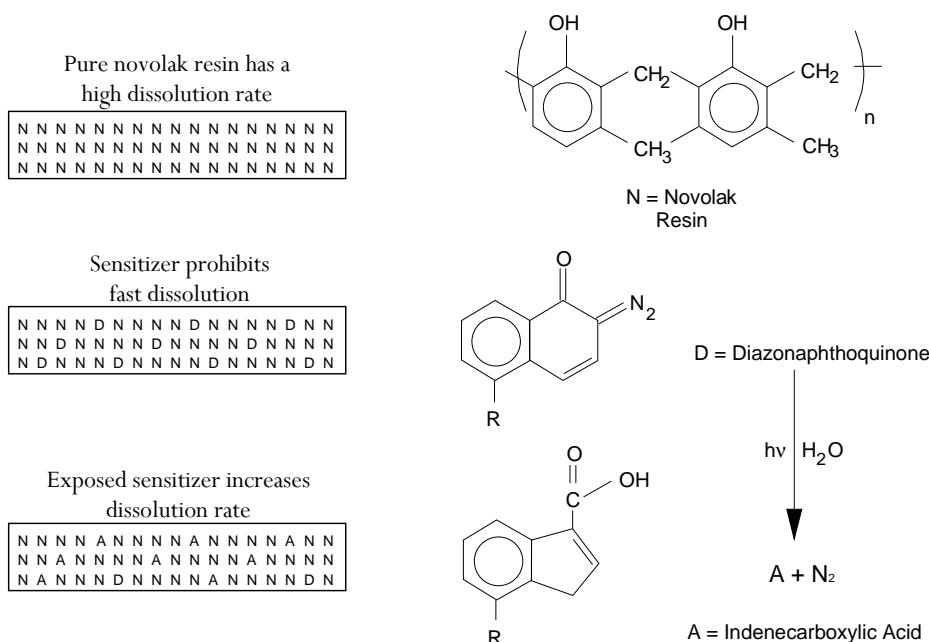


Figure 5.41 Function of each component in a DQN resist.

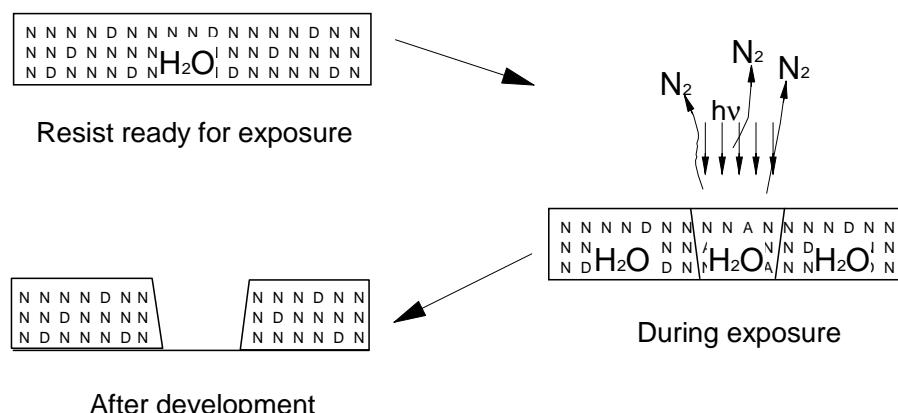


Figure 5.42 Imaging of a DQN positive resist.

Two-component negative resist systems—The two-component negative resists are typified by those developed by Kodak, such as KTFR, KMER, and KPR. The matrix resin of KTFR is cyclized poly(cis-isoprene), which is a cyclized rubber. The sensitizer is an azide, such as bis(aryl azide). Upon actinic radiation, the azide releases nitrogen to form an extremely reactive intermediate, possibly a nitrene, which eventually crosslinks the resin. Figure 5.43 shows crosslinking of this two-component negative resist. Consult Thompson⁴² and Moreau⁴³ for additional details. This type of resist is at least an order of magnitude more sensitive than the DQN resist. There is good adhesion, and the crosslinked product is very durable. However, it does not work very well for high-resolution imaging because the resist swells while it is being developed and shrinks afterwards. The resist image profile is not only reduced in height, but is also wrinkled. In the worst case, the resist line can take a zigzag shape due to the tremendous stress it must handle. This type of resist is very popular for low-resolution applications that call for high sensitivity and good adhesion.

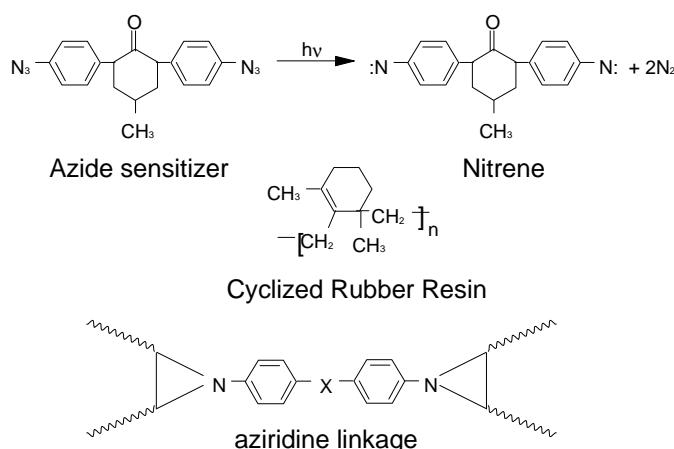


Figure 5.43 Crosslinking of a two-component negative resist.

Chemical amplification systems—The chemically amplified resist (CAR) system is the ultimate scheme for higher sensitivity, thus higher wafer throughput. In principle, it needs to be photoinitiated to produce just a single proton, which can be used again and again to generate products consisting of new protons. The proton acts like a catalyst that facilitates chemical reactions while it does not consume itself. Therefore, the sensitivity can be orders of magnitude higher than that of nonamplified systems. In modern CAR systems, the reaction is quenched to prevent the unexposed area from being thoroughly converted. The tertiary-butoxycarbonyl (t-BOC) deep-UV resist system⁴⁴ is given here as an example of chemical amplification, as shown in Fig. 5.44. The actinic radiation gives rise to protons from the onium salt sensitizer. These protons react with the base resin to turn it from t-BOC, a poly(p-t-butyloxycarbonyl-oxystyrene) (PBOCST), into p-hydroxystyrene (PHOST) and two volatile compounds, namely carbon dioxide and isobutene. The conversion is most appreciable during the postexposure bake (PEB) at temperatures above 180° C. The proton remains after the reaction. Hence, it can catalyze other unreacted t-BOC molecules and, in principle, be extensively used. PHOST is polar, whereas t-BOC is nonpolar. A polar solvent removes PHOST, the exposed part, making a positive resist system. A nonpolar solvent makes a negative resist by removing the opposite. Figure 5.45 shows the imaging of t-BOC from exposure to developing.

An important aspect in imaging CAR systems is the diffusion of the acid (proton) as it multiplies during the PEB. Diffusion and the quenching of it have an effect on the sensitivity of the CAR. More diffusion increases the

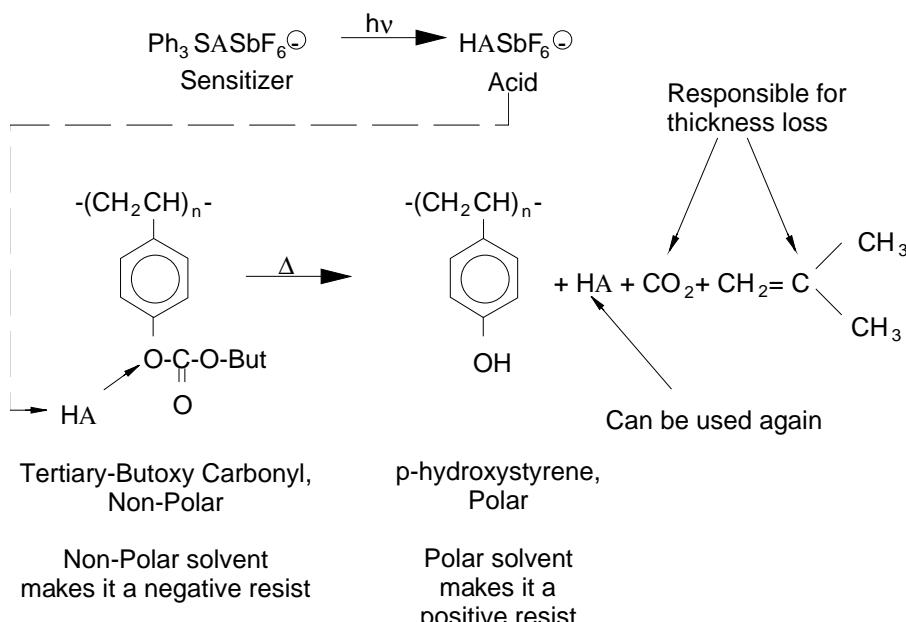


Figure 5.44 Photochemistry of a t-BOC chemically amplified resist.

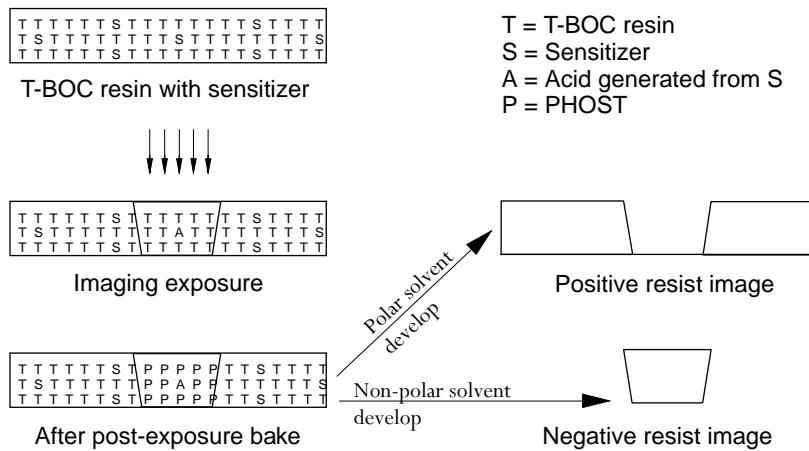


Figure 5.45 Imaging of a t-BOC chemically amplified resist.

amplification ratio, which raises the sensitivity of the CAR. Critical dimension (CD) can be affected by diffusion. The line-edge roughness (LER) in the resist image is also affected by diffusion. More diffusion tends to increase the LER.

Image-reversal systems—The DQN^{45,46} resist can be reversed into a negative-tone resist without the problems normally associated with conventional two-component resists, such as swelling, shrinking, low dry-etch resistance, and solvent developer. An image-reversal system is desirable for high-resolution imaging whenever a negative resist is called for, such as in situations discussed above in Sec. 5.6.1.1.

The working principle of image reversal is shown in Fig. 5.46. The novolak resin is turned into indene carboxylic acid just as in the positive imaging of DQN resists. However, due to a premixed base in the resist, the indene carboxylic acid is turned into indene during the PEB. Indene is insoluble in alkaline developers for DQN resists. One only needs to blanket expose the entire wafer to convert the unexposed resist into indene carboxylic acid to be soluble in the alkaline developer. The resist dissolution characteristics are now reversed to a negative tone. The imaging of a reversed DQN resist is depicted in Fig. 5.47.

5.6.1.3 Imaging configurations

The resist system can be classified in terms of how an image is formed in the resist:

Bulk imaging—This is the most common type of imaging. The resist is exposed and developed through its entire thickness. The exposure process contributes to the developed resist profile, and the DOF of the imaging system must support the thickness of the resist in addition to other components in the DOF budget,

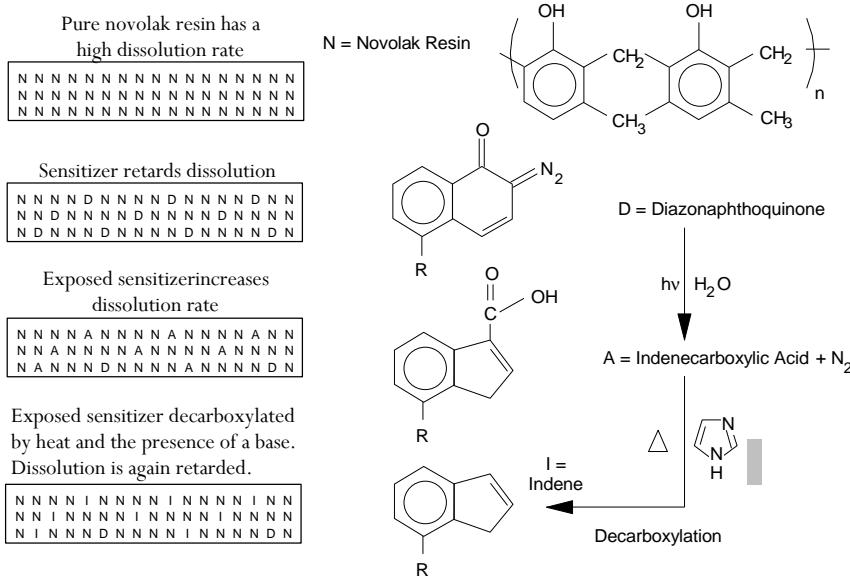


Figure 5.46 Function of each component in an image reversal of DQN resists.

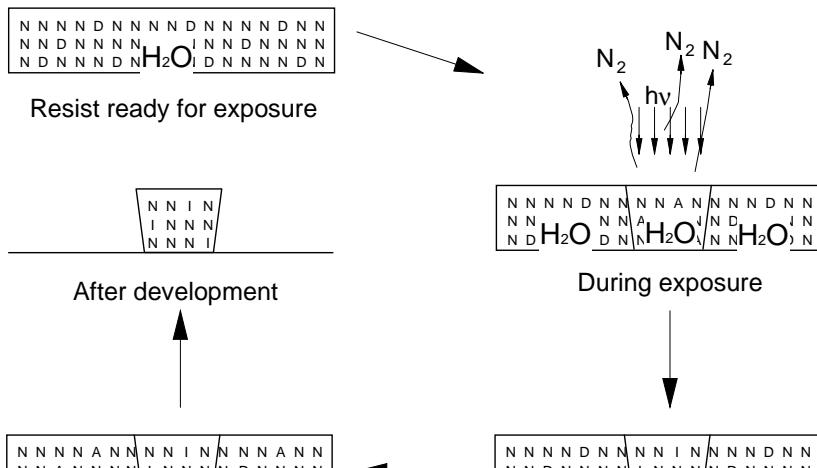


Figure 5.47 Imaging of a reversed-DQN resist.

such as lens field curvature, wafer flatness, leveling, wafer topography, and focusing errors. The impact of the resist on DOF will be discussed in detail in Sec. 7.3.

Top-surface imaging (TSI)—Only the top surface of the resist is exposed and developed, which makes for better planarization and takes up less of the DOF budget. The image is transferred to the bulk of the resist layer by another pattern-

transfer technique to produce a thick resist image with a vertical profile, so that the resultant resist image is suitable for the subsequent semiconductor delineation processes exemplified in Fig. 5.33, wherever a thick and high-aspect-ratio resist image is needed. One important motivation for using TSI systems is to isolate the topography on the wafer from producing a large exposure variation in the resist.

An example of the TSI system^{47,48} is depicted in Fig. 5.48. Starting with a single layer of highly absorptive resist, light penetrates only the top part of the resist layer. This way, no reflected light is produced, thus isolating the latent image from wafer topography. The resist is then subjected to a silylation process that selectively incorporates silicon into the exposed part of the resist. This sets up for the subsequent O₂ reactive ion etching (RIE) process that removes only the resist in the unsilylated part. Because the final resist image is void of resist in the unexposed areas, this is a negative TSI system. A positive TSI system is possible. A more straightforward way to produce a positive TSI system would be to crosslink the latent image just as in the image-reversal process, to be followed with the silylation process, as shown in Fig. 5.49.

Multilayer—Similar to TSI systems, an important motivation to use multilayer systems is to prevent the topography on the wafer from producing a large exposure variation in the resist.

There are two or three layers of material. The top layer is a photosensitive layer. In a three-layer system, after delineation of the image in the top layer, the image is transferred to the middle layer that serves as a mask to delineate the bottom layer. A two-layer system uses the top layer as the photosensitive layer as well as the masking for the bottom layer. In multilayer systems, the top imaging layer is thin so that it is easy to image and does not take up a significant part of the DOF budget. The bottom layer is sufficiently thick for planarization over topography on the wafer as well as for the subsequent semiconductor delineation processes exemplified in Fig. 5.33. Consult Ref. 49 for an extended coverage of multilayer resist systems.

Figure 5.50 shows a three-layer RIE-transfer resist system.⁵⁰ The bottom layer is a hard-baked resist that is desensitized and planarized. The middle layer is a thin inorganic layer that is resistant to O₂ RIE. The top layer is a photoresist suitable for the exposing wavelength. After normal exposure and the development process to delineate the top-imaging layer, the resist image is transferred to the middle layer by etching. An O₂ RIE step completes the delineation of the resist system.

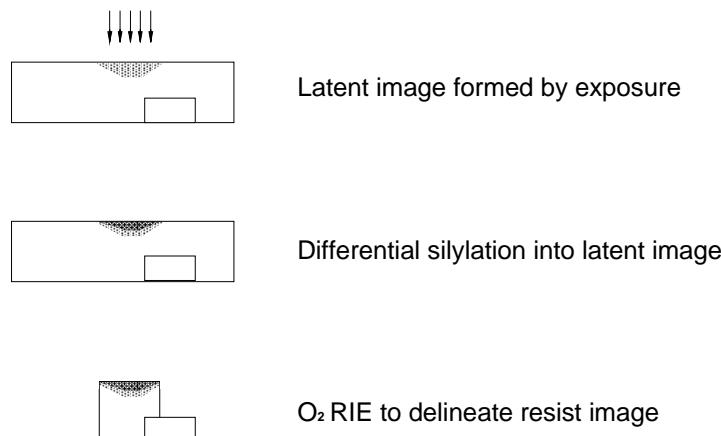


Figure 5.48 DESIRE negative top-surface-imaging resist system.

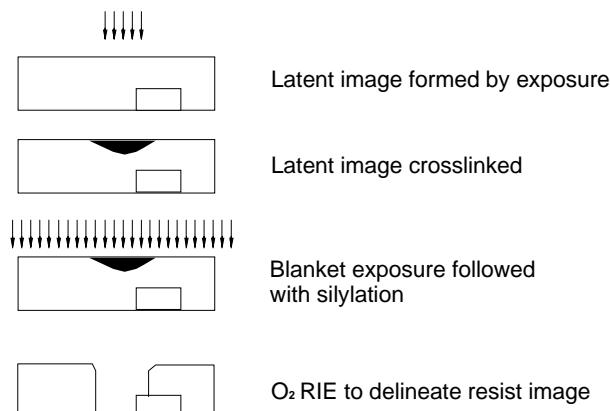


Figure 5.49 A positive top-surface-imaging resist system using image reversal.

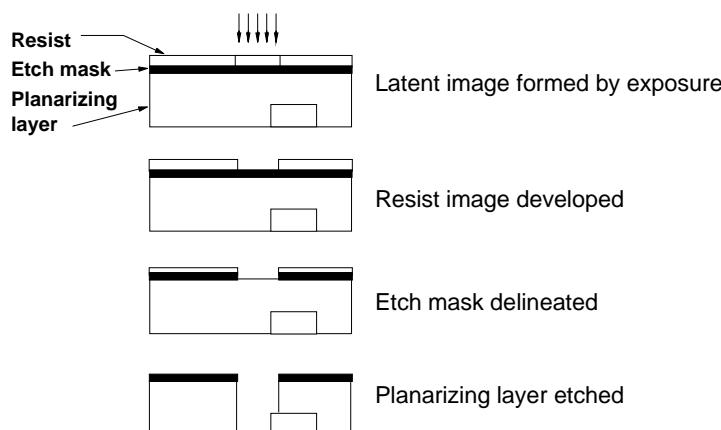


Figure 5.50 A three-layer RIE-transfer resist system.

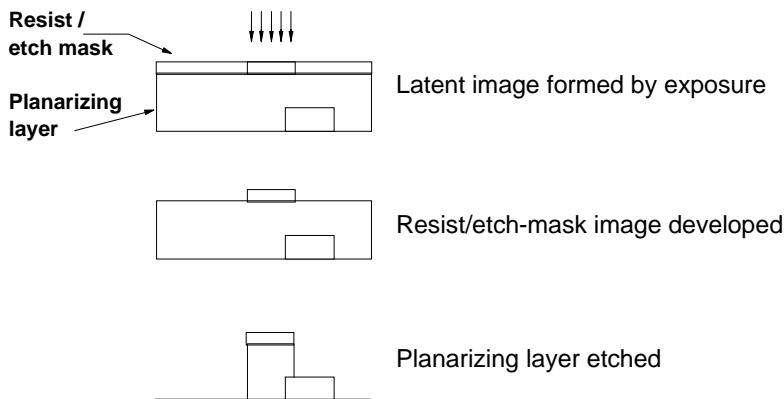


Figure 5.51 A two-layer RIE-transfer resist system.

A two-layer RIE-transfer resist system^{51,52} is shown in Fig. 5.51. The top-imaging layer contains an inorganic material, such as Si or Ge_xSe_{1-x}, and it is a good O₂ etch mask. After photodelineation of the top layer, the image is transferred to the bottom layer by O₂ RIE.

Note that O₂ RIE is used as a generic term for high-selectivity polymer etching. A trace of another gas is often added to help improve the process. For example, a very small amount of fluorine-containing gas is included to clean up possible inorganic residue. This can reduce the etch-rate ratio between the polymer and the inorganic surfaces.

The transfer process need not be RIE. When the feature size to be transferred is larger than 0.5 μm, a deep-UV blanket exposure masked by a delineated top or middle layer produces a well-controlled proximity-printed image in the bottom planarizing layer. Of course, the bottom layer must be a deep-UV resist. Wavelengths shorter than deep UV can be used to obtain a higher resolution or aspect ratio. The minimum feature and aspect ratio can be predicted with Eq. (2.1) to define the resolution limit of this system.

Figure 5.52 shows the schematic of a three-layer phototransfer resist system. The three-layer phototransfer system is not popular because there is a simpler two-layer phototransfer system,^{53,54} shown in Fig. 5.53. The AZ1350 resist is used as the top-imaging layer for near- or mid-UV exposure. This resist is highly absorptive in deep UV, making it a good deep-UV mask.¹⁶ PMMA, a well-known deep-UV^{55,56} and e-beam⁵⁷ resist, is used for the bottom layer. The top layer acts like a conformable contact mask that is attached to the wafer. This is why it is called the portable conformable mask in Refs. 53 and 54.

As long as the feature size and aspect ratio are supported by the blanket-exposing wavelength, the phototransfer system is preferred because of process simplicity and a low defect level. For features far below 0.5 μm, the RIE-transfer system must be used. The choice between TSI and multilayer RIE-transfer systems depends on process control and cost. Improvements to antireflection coatings for bulk imaging resist systems continue to be made. There is not an urgent need to use TSI or multilayer systems.

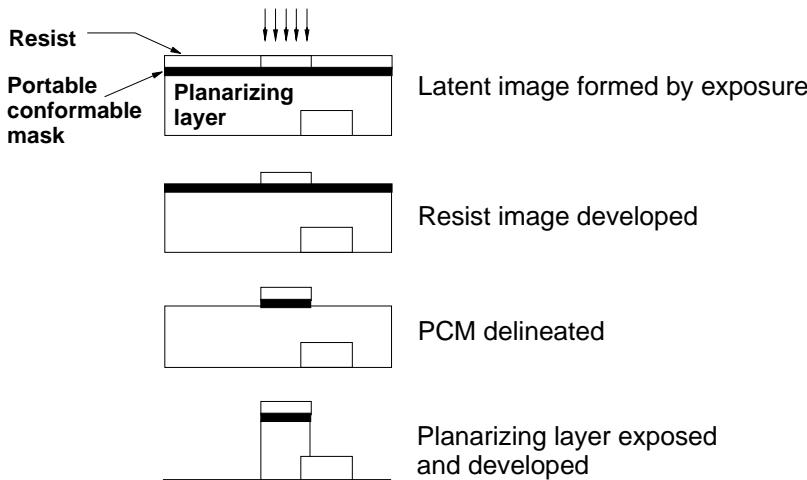


Figure 5.52 Three-layer phototransfer resist system.

Contrast enhancement—Contrast enhancement is a two-layer phototransfer system that uses near UV to expose both the top and the bottom layers.⁵⁸ The top layer is highly absorptive and bleaches as it is exposed, thus it also serves as an in-situ portable conformable mask. Therefore, contrast enhancement is a one-exposure system. It also develops in a single step because the top layer readily dissolves in the aqueous developer of the bottom layer, which is a regular novolak resist. As soon as any part of the pattern is bleached, it allows light to penetrate for even more bleaching. This nonlinear process enhances the effective contrast of the resist system. Unfortunately, it also tends to favor larger features that bleach earlier, enhancing proximity effects. Figure 5.54 shows the working principle of the contrast-enhancement system and the enhancement of proximity effects.

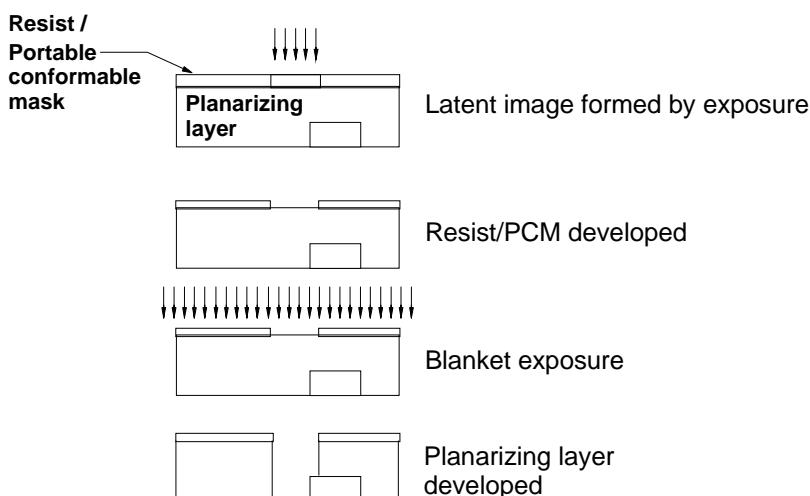


Figure 5.53 Two-layer phototransfer resist system.

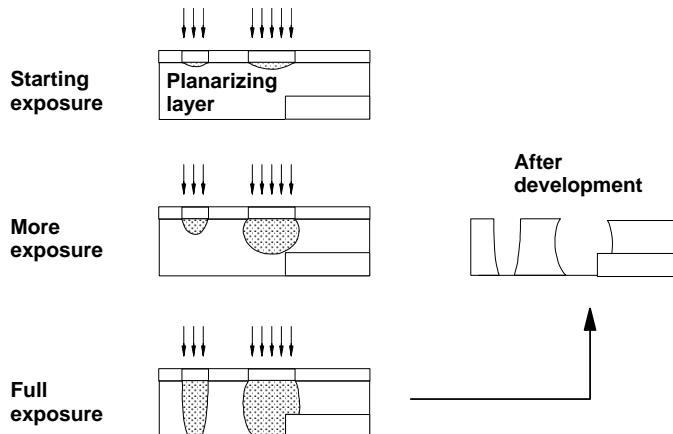


Figure 5.54 Contrast enhancement resist and proximity effects.

5.6.2 Light interactions with a photoresist

5.6.2.1 Wavelength compression

As seen in Fig. 5.17, when light enters a higher-refractive-index medium, the wavelength is compressed by the index of refraction as seen in

$$\lambda_2 = \left(\frac{n_1}{n_2} \right) \lambda_1 \quad (5.13)$$

and

$$D_2 = \left(\frac{n_1}{n_2} \right) D_1. \quad (5.14)$$

This is a desirable effect because the optical thickness D_2 of the resist film is now smaller as given in Eq. (5.14), allowing the resist thickness to take up a smaller portion of the DOF budget.

5.6.2.2 Light absorption

A certain amount of light must be absorbed by the resist in order for it to acquire the energy to induce chemical changes. If there is too much absorption, the exposure light cannot penetrate the resist film, and it would be difficult to clear the resist in the unwanted parts with the resist developer. Even before that takes place, the resist profile would suffer severely. A quantitative optimization will be presented in Chapter 6. Here, the effects of resist absorption are shown graphically.

5.6.2.3 Resist bleaching or dyeing

The resist may become more or less absorptive during exposure. The former is the dyeing phenomenon, the latter is bleaching. Bleaching is preferred because it

increases the contrast of the resist image. However, the amount of dyeing or bleaching in workable resist systems is small, unless something is designed specifically for it, such as the contrast-enhancement resist system.

5.6.2.4 Resist outgassing

Certain resists produce gas substances during exposure. For example, the diazoquinone-sensitized novolak resists release nitrogen gas during exposure, as shown in Fig. 5.40. In principle, outgassing can disturb the homogeneity of the refractive index in the vicinity of the resist surface that is critical for imaging. However, no ill effect has been reported, implying the amount of outgassing is negligible for imaging. In the long run, if the released material accumulates on the lens surface, it must be cleaned. Usually, the specification for resist outgassing that contaminates the lens surface is more stringent than for bubbles in the resist. When the medium between the front element of the lens and the resist surface is a liquid, as in an immersion-exposure system to enhance DOF, the perturbation to the homogeneity of the medium would be more severe and must be taken into consideration. This will be discussed in Chapter 7.

5.6.2.5 Multiple reflections

When light is incident on an interface, it is refracted and reflected, unless the refractive indices of the two media are perfectly matched. Therefore, it undergoes multiple reflections and refractions at the resist-air and resist-wafer interfaces, as depicted in Fig. 5.55. In practice, there is a stack of thin films between the wafer and the resist. Each interface induces its own reflection and refraction, making it very complicated to illustrate graphically. With only two interfaces shown in Fig. 5.55, light bounces back and forth until the intensity is too low to be of concern.

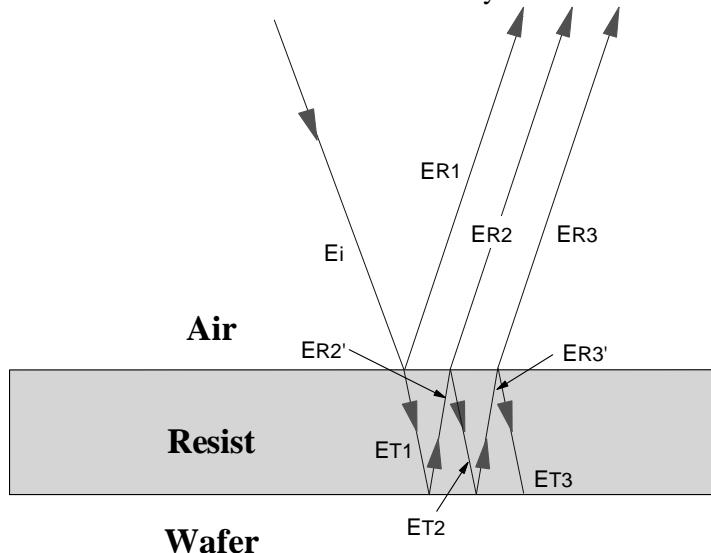


Figure 5.55 Multiple reflections and refractions of light in a photoresist.

Standing waves—Because of the interference of these light beams, standing waves are created, as illustrated in Fig. 5.56. Let's assume the electric field of the incident light E_i is a plane wave perpendicular to the resist-wafer interface. Then

$$E_i = e^{inkz}, \quad (5.15)$$

where n is the refractive index of the photoresist. $K \equiv 2\pi/\lambda$, where λ is the wavelength in vacuum and z is the longitudinal distance that the incident light is traveling. $z = 0$ refers to the resist-wafer interface. The electric field of the reflected beam E_r is represented by

$$E_r = R e^{-inkz}, \quad (5.16)$$

where R is the complex reflection coefficient of the resist-wafer interface. When both beams are present, the total electric field E in the photoresist is

$$E = E_i + E_r = (1 - R)e^{inkz} + 2R \cos nkz. \quad (5.17)$$

There are two components in E : the first component, represented by the first term, is a propagating wave just as E_i or E_r . The second component, represented by the second term, is a standing wave. This wave stands still. It does not propagate in either direction. The resist exposure dosage is proportional to the intensity I :

$$I = |E|^2 = (1 + R^2) + 2R \cos 2nkz. \quad (5.18)$$

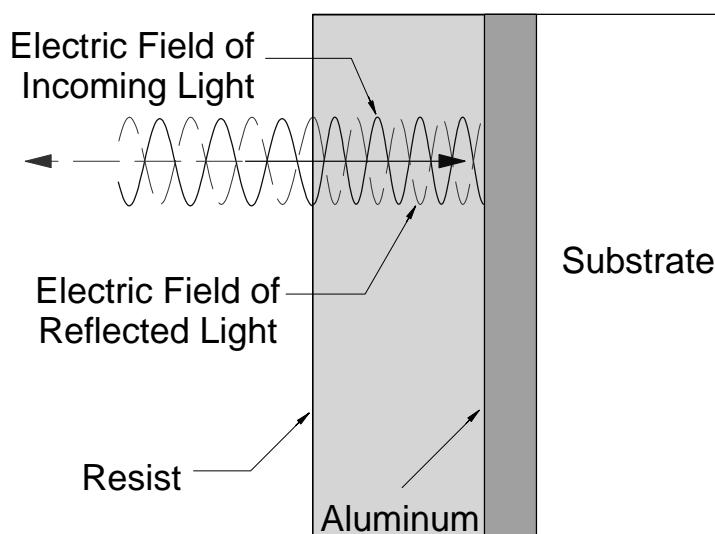


Figure 5.56 Standing waves in a photoresist on a metal layer.

Therefore, the periodicity P of the standing wave in the developed resist image is

$$P = \frac{\lambda}{2n}. \quad (5.19)$$

When the bottom surface of the resist does not reflect light, such as in the case of a perfectly matched substrate, $R = 0$, and E and I are uniformly distributed. The resist image is at its best-possible profile. When R approaches -1 , such as in the case of resist on aluminum, which is a good conductor and highly reflective, I varies from the two extremes, 4 and 0. At the resist-wafer interface, $z = 0$. E and I both become zero. Hence, no matter how intense the incoming beam is, as long as the reflecting beam is from a highly conducting and reflective surface, the incident and reflected fields cancel each other, resulting in insufficient exposure at the interface. With positive resists, the nodes of the standing wave block further development in the resist. At best, there is, at the resist-wafer interface, resist residue that must be ashed away. Before the advent of ashing, higher-concentration developers were used to increase the dissolution rate of the unexposed resist. The tradeoff is reduction of the resist-development contrast. With negative resists exposed with strong-standing waves, the resist loses adhesion at the interface because it is not exposed and remains highly soluble. No adhesion promotion scheme can fix the problem unless the resist is crosslinked at the interface.

When the standing waves are not sufficiently severe to produce the above problems, they seem to be only cosmetically undesirable. This applies to isotropic and anisotropic etching, lift off, and ion implantation, as seen in Fig. 5.57, where the standing waves are drawn at one side of the resist profile for comparison. In isotropic etching, the most critical location for linewidth control is at the resist-to-etch-layer interface. The resist edge position here is strictly dependent on the refractive indices of these two media and is independent of whether there is a tooth-shaped profile. For anisotropic etching with high etch selectivity, the arguments are similar to those of isotropic etching. When the selectivity is not so high and the resist is consumed during etching, the effect of standing waves is similar to that of a straight but more oblique profile drawn from one extreme of the standing wave at the bottom and the other extreme of the standing wave on top, as shown in the Fig. 5.57. For ion implant applications, the argument is similar. For lift-off application, it is obvious that the tooth shape of the standing waves does not play any role in linewidth control. Methods to control standing waves include PEB, low-contrast developer, antireflection coatings, surface imaging, and multilayer resist systems.

Eliminating standing waves with a PEB was first reported by Walker.⁵⁹ The standing waves in a diazoquinone novolak resist are eliminated with a 10-min PEB at 100° C, as shown in Fig. 5.58. The working principle is that the high longitudinal refractive-index gradient in the latent image is averaged out by diffusion at the PEB temperature. However, diffusion also takes place laterally.

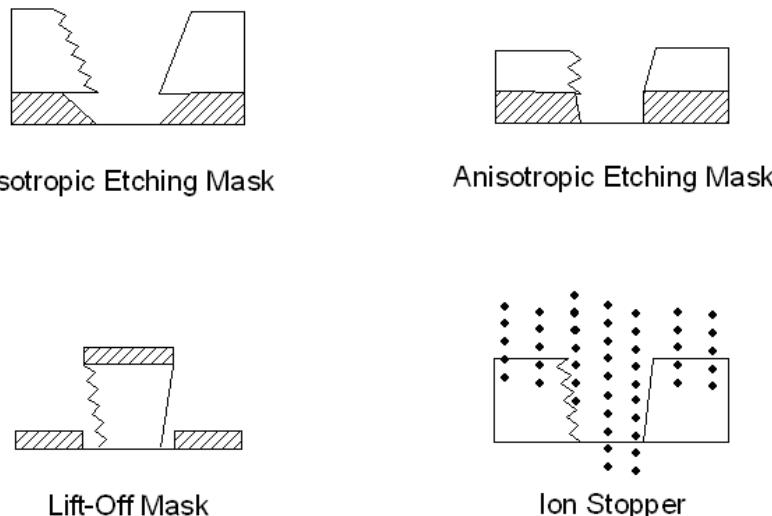


Figure 5.57 The effect of standing waves to resist applications.

Removing the high gradient in the longitudinal direction also results in a reduction of the gradient in the latent image in the lateral direction. Hence, the resist profile is slightly more overcut, which can be seen in the figure. A lower image gradient also leads to smaller exposure latitude. Fortunately, the longitudinal standing-wave-induced gradient is much higher than that of the lateral image edge. Otherwise, PEB may carry too much tradeoff. Diffusion is less effective in reducing the longitudinal gradient at the resist-wafer interface because the direction of diffusion is now restricted to the resist side.

Other than for cosmetic reasons, using a low-contrast resist developer can reduce the skin effect and loss of adhesion from substrate reflection. A high-contrast developer does not dissolve the unexposed positive resist; therefore, it makes the skin effect pronounced. When a developer can remove unexposed positive resist, the development contrast is dropped, but skins can be more easily removed. This is yet another tradeoff, as was the case of PEB.

The best method to eliminate standing waves is to remove substrate reflection, which calls for using an antireflection coating at the resist-wafer interface, which we will discuss in Sec. 5.7. Surface imaging or multilayer resist systems were discussed in Sec. 5.6.1.3.

Stray light—Multiple reflections from the resist and the underlying films contribute to stray light in the imaging system, just like the reflection from the chromium area of the mask discussed in Sec. 5.3.3.3 above. Referring back to Fig. 5.15, one readily sees that film interfaces not in the conjugate plane of the mask reflect light to the chrome area to be reflected back, which reduces image contrast. This type of stray light is best treated with a reduction of mask reflectivity. Optimizing resist and film thicknesses to minimize the total reflection is best left for controlling the exposure tolerance given later in Sec. 5.7.

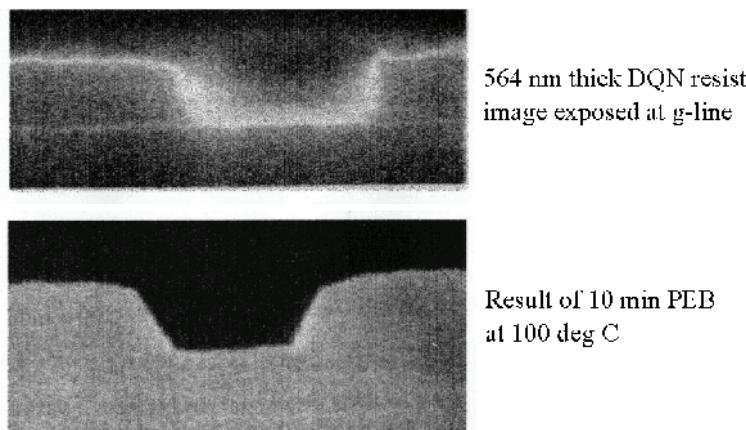


Figure 5.58 Standing wave elimination by postexposure bake (© 1975 IEEE).⁵⁹

Exposure nonuniformity—In addition to standing waves, multiple reflections in the resist contribute to exposure variation as a function of resist thickness. Figure 5.59 shows reflectivity from the resist-air interface taken from a poly-Si/SiO₂ film stack. The reflectivity can change by 26% within a thickness range of 50 nm. Therefore, the resist exposure also changes by that range because the resist-exposing light is the remainder of the incident light (after subtracting light that reflects from the air-resist interface) and light that leaves the resist to enter the thin-film stack and substrate. The latter is generally small in comparison to the reflected and exposing light, as will be seen in Sec. 5.7.

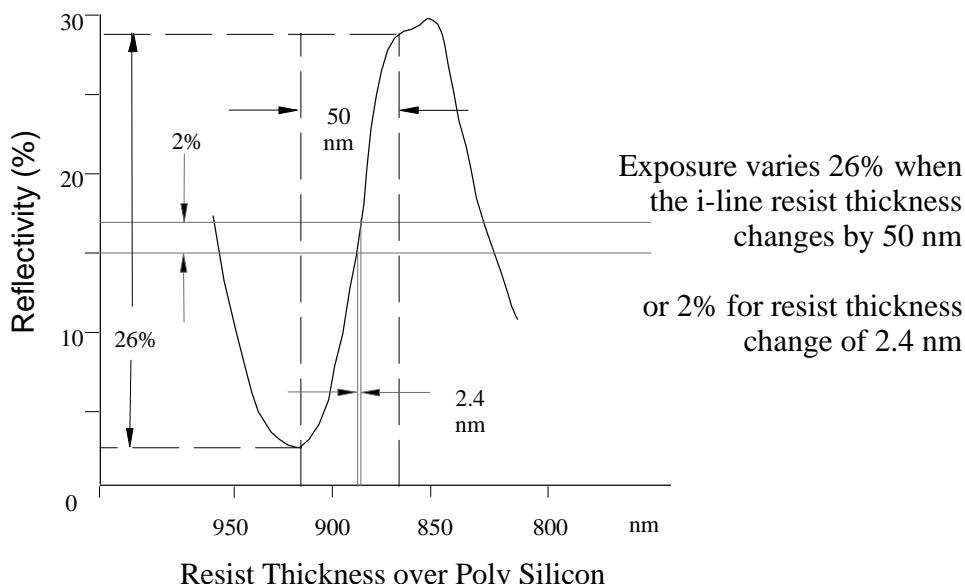


Figure 5.59 Reflectivity versus resist thickness.

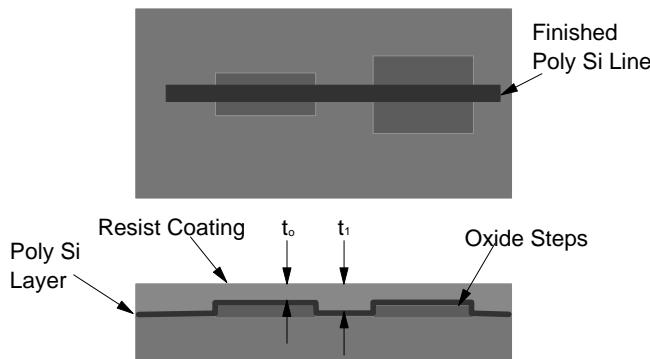


Figure 5.60 Resist thickness variation over topography.

Generally, the resist-coating thickness can be controlled to within a few nanometers and is not an issue. The primary cause of exposure nonuniformity is resist coverage over topography, as shown in Fig. 5.60.

An exposure variation of 26% requires an imaging system that can support exposure latitude better than 29%. The extra margin is used to allow for nonuniform exposure from illumination and inaccuracy in controlling the exposure shutter. As was shown in Sec. 4.3.2 and subsequent examples, the exposure latitude needs to be contained to within 15%, preferably less than 10%. This requires close attention to the reflection variation produced by the resist and the thin-film stack. These topics are extremely important and covered in Sec. 5.7.

5.6.3 Profile of resist images

Figure 5.61 assumes an aerial image that is independent of the longitudinal position in the resist. Even so, the latent image is reduced because of resist absorption. This promotes an overcut profile in the developed resist image. With a positive resist, the overcut is even more pronounced, because when the developer starts to penetrate into the resist, the side of the resist image in the developed area is exposed to the developer just like the bottom. The upper side is subjected to the developer dissolution longer than the lower side. With a negative resist, the development effect compensates for the absorption effect in the profile of the resist image.

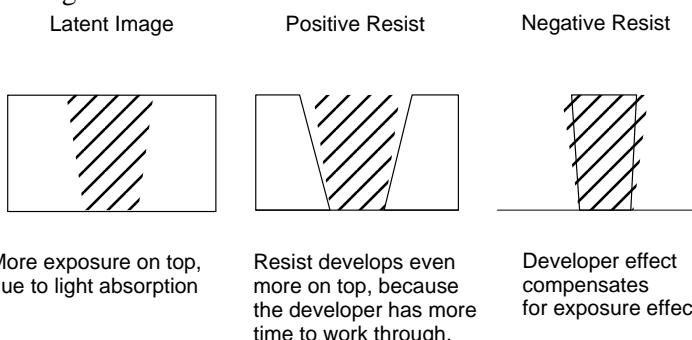


Figure 5.61 Profile of positive and negative resists.

In reality, the aerial image is not constant. Figure 5.62 shows the refracted and developed images of a $k_1 = 0.64$ line opening illuminated with $\sigma = 0.5$ at three defocal positions, as if a wafer with a coated resist is located at each of these three positions. The focal plane is marked and is in the vicinity of $k_2 = 1.8$. The resist thickness is 1 in k_2 units. The diffracted image is first simulated in air. When this image propagates into the resist, it continues to propagate in a new medium with n_r and α , where n_r is the real part of the complex refractive index n_c of the resist, and the imaginary part n_i is given in Eq. (5.20):

$$n_i \equiv \frac{\lambda}{2\pi} \alpha. \quad (5.20)$$

This image is strictly refracted. Reflections are ignored here. This refracted image is translated into a rate image and developed by simulation. The development contours are the developed surface at progressive development times. In addition to using Fig. 5.62 as a reference condition, an important observation can be made here. Even though the wafer is placed symmetrically off focus from the best focus, the developed images at the two defocal locations are quite different because the refracted images at the starting surface for the developer are different. In the negative defocus region, the refracted image starts as a double-peak deteriorated image, propagating into a single-focused image, whereas in the positive defocus region, the refracted image starts single peaked to degenerate into two peaks. This accounts for a focus shift when a resist is introduced. The loss in DOF in the negative region of defocus is made up with the gain in the positive region. Therefore, the net result in DOF is unaffected. The developed images flatten out in the last two positions because the developer has reached the wafer surface. In the first position, the resist is not completely developed, even though the same development time is used for all three cases.

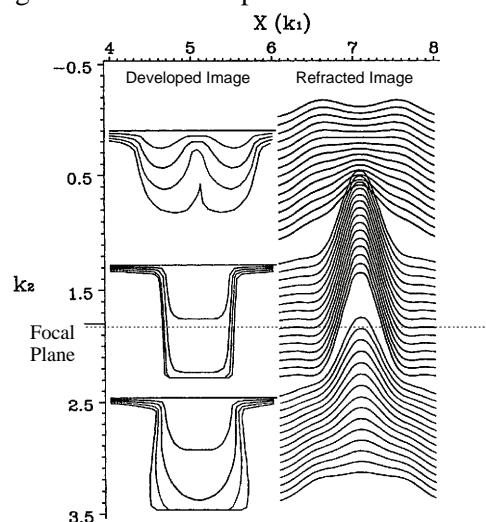


Figure 5.62 Refracted and developed images: $n = 1.5$ and $\alpha = 0$.

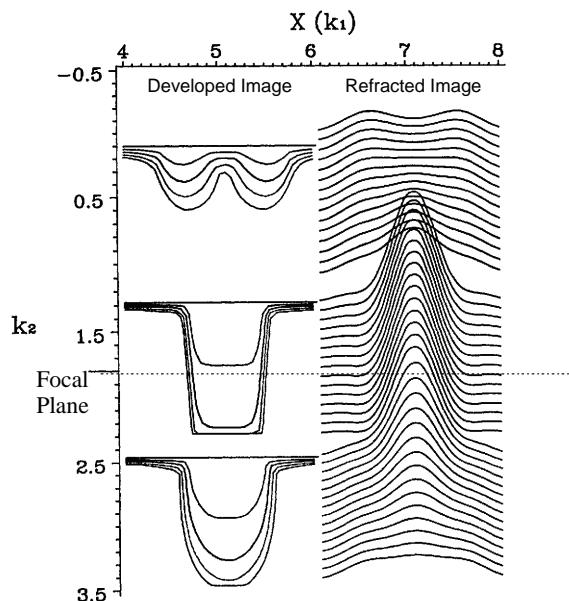


Figure 5.63 Refracted and developed images: $n = 1.5$ and $\alpha = 0.178 \mu\text{m}^{-1}$.

A small absorption is introduced, and its result is depicted in Fig. 5.63. In the negative defocus region, the resist is developed even less because less light reaches the bottom. In the focused region, the resist profile is slightly overcut, as predicted by Fig. 5.61. In the positive defocus region, the resist image has reversed its positive bias. This means that with a slightly smaller absorption coefficient, the resist image can be made identical to the focused image. Therefore, there is an optimum amount of absorption.

When the absorption coefficient is increased to $3 \mu\text{m}^{-1}$, the developed images do not clear the resist bottom in all three defocal positions, as shown in Fig. 5.64. However, this also means that reflection from the wafer substrate is completely annihilated. Though the resist is not suitable for bulk imaging, it is ideal for surface imaging. This comment is valid for resist thickness in the order of a micrometer. With resist thickness dropping to below 50 nm for nanolithography, $\alpha = 3 \mu\text{m}^{-1}$ is the absorption coefficient required for good imaging.

5.7 Antireflection Coatings

As pointed out in Sec. 5.6.2.1, controlling reflection from the air-resist interface is extremely important. There are many ways to accomplish this. The TSI or multilayer resist systems described in Sec. 5.6.1 are optically simple, but more processing steps are required, and they are not yet ready for large-scale manufacturing. This leads to the need to extend the use of single-layer resist systems with thin-film thickness engineering wherever the semiconductor device requirement is not violated. Extra polymer coatings that reduce reflection are also

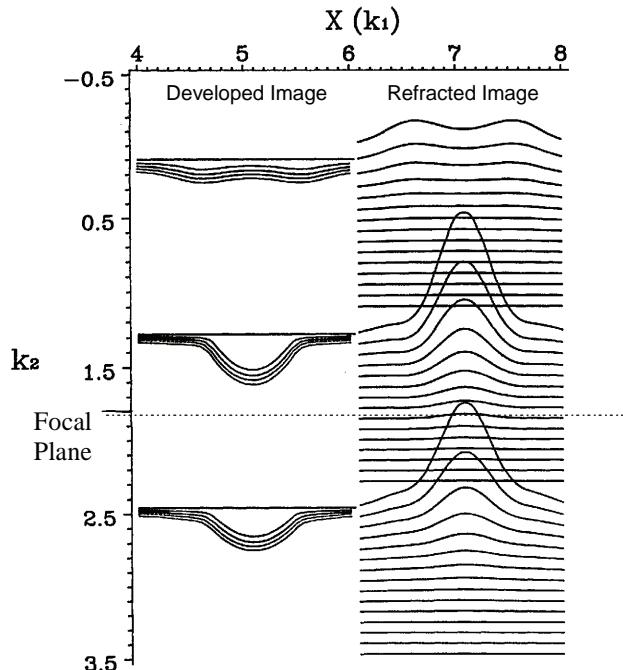


Figure 5.64 Refracted and developed images: $n = 1.5$ and $\alpha = 3 \mu\text{m}^{-1}$.

introduced. Hence, there are top antireflection coatings (TARC) and bottom antireflection coatings (BARC). The former is applied to the top of the resist layer, while the latter is applied to the resist-wafer interface. BARC can be a polymer film or an inorganic layer.

We shall use a simple case to understand antireflection coatings (ARC) analytically. Take the situation shown in Fig. 5.65. The resist is coated directly on the wafer. The incident beam, whose electric field E_i is reflected and transmitted at the air-resist interface, produces E_{rAR1} , and $E_{tAR1} \times E_{tAR1}$ is reflected at the resist-wafer interface, becoming E_{rRW1} , then reflected and transmitted at the resist-air interface, producing E_{rRA1} and E_{tRA1} . Hence, the total reflected electric field E_r is

$$E_r = E_{rAR1} + E_{tRA1} + E_{tRA2} + \dots \quad (5.21)$$

We now denote the reflection and transmission coefficients at each surface to be r_{AR} , t_{AR} , r_{RW} , r_{RA} , and t_{RA} , where r_{AR} is the reflection coefficient at the air-resist interface entering from air, t_{AR} is the transmission coefficient from air to resist at the air-resist interface, r_{RW} is the reflection coefficient at the resist-wafer interface entering from the resist, etc. Hence,

$$\frac{E_r}{E_i} = r_{AR} + t_{AR}r_{RW}t_{RA}\eta e^{i\delta} + t_{AR}r_{RW}r_{RA}r_{RW}t_{RA}\eta^2 e^{i2\delta} + \dots \quad (5.22)$$

where

$$\delta = \frac{4\pi}{\lambda} nh \cos \theta_R$$

is the phase introduced by each cycle of travel from the air-resist interface to the resist-wafer interface and back, n is the real part of the refractive index of the resist, h is the resist thickness, λ is the wavelength, and θ_R is the angle that the refracted beam makes with the normal of the reflecting surface. $\eta = e^{-2h/\cos\theta_R}$ is the attenuation induced by each cycle of travel, and α is the absorption coefficient of the resist. Summing up the infinite series in Eq. (5.22) and recognizing that $r_{AR} = -r_{BA}$, $t_{AR} = t_{RA}$ and $(n_A \cos \theta_i / n_R \cos \theta_R)$ lead to

$$\frac{E_r}{E_i} = \frac{r_{AR} + \left(\frac{n_R \cos \theta_R}{n_A \cos \theta_i} t_{AR}^2 + r_{AR}^2 \right) r_{RW} \eta e^{i\delta}}{1 + r_{AR} r_{RW} \eta e^{i\delta}} \quad (5.23)$$

and

$$\frac{E_r}{E_i} = r_{AR} + \frac{t_{AR} r_{RW} t_{RA} \eta e^{i\delta}}{1 - r_{RA} r_{RW} \eta e^{i\delta}}. \quad (5.24)$$

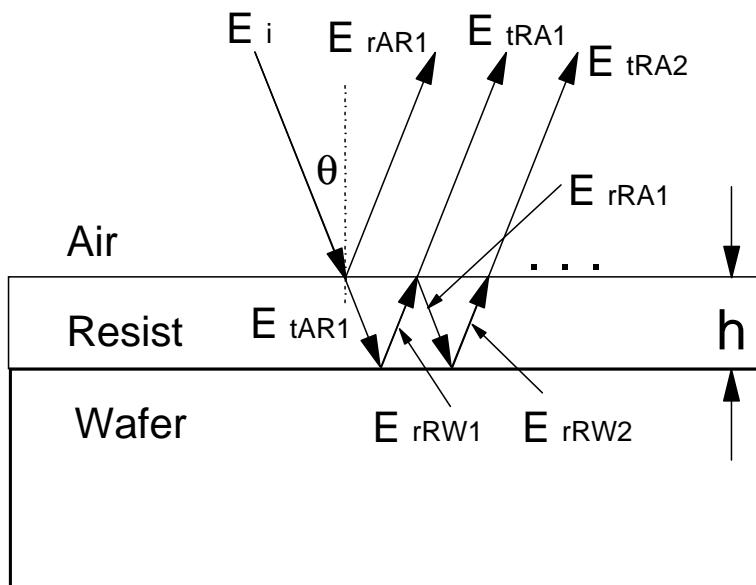


Figure 5.65 Multireflections in a single-resist layer coated directly on a plain wafer substrate.

If there is BARC or TARC, even if it is single layered or multilayered, the above formula is still valid in its present form. We only need to interpret r_{AR} , t_{AR} , r_{RW} , t_{RA} , and r_{RA} properly. That is, for example, t_{AR} is the ratio of the electric field at the resist top to the electric field at the air bottom, even if TARC is sandwiched in between. However, to find their numerical values in such a case, we need to trace reflection and transmission at all related interfaces, which becomes very tedious as the number of layers is increased.

In the following, we will adopt a different approach. By employing the matrix formulation, which is derived by considering total fields inside each layer, the effect of multiple reflections at each interface is automatically taken into account, which greatly simplifies the problem. In matrix formulation, there are two matrices that govern the propagation of waves inside the thin-film stack. One is the boundary matrix $D_{1,2}$, which relates electric field amplitudes at the two sides of an interface between, for example, layers 1 and 2:

$$\begin{bmatrix} E_1^+(d_1) \\ E_1^-(d_1) \end{bmatrix} = D_{1,2} \begin{bmatrix} E_2^+(0) \\ E_2^-(0) \end{bmatrix} \quad (5.25)$$

with

$$D_{1,2} = \frac{1}{t_{1,2}} \begin{bmatrix} 1 & r_{1,2} \\ r_{1,2} & 1 \end{bmatrix}, \quad (5.26)$$

where E_j^+ is the field amplitude of an electromagnetic wave propagating in the $+z$ -direction, with E_j^- in the $-z$ -direction. Note that we have used the coordinate system that the $+z$ -direction is normal to and points toward the thin-film stack. The z -coordinate at layer j is z_j and ranges from 0 at the top of the layer to d_j at the bottom of the layer. Figure 5.66 shows the nomenclature of the layers and the direction of the z -axis. Layer 0 denotes the resist layer. The negative indices indicate the layers above the resist, and the positive indices indicate the layers below it. The first layer is designated by the index $-m$, and the last layer is designated by the index n . The region above layer $-m$ is denoted by a , and the region below layer n is denoted by b .

Free propagation inside a layer, say layer 2, is governed by another matrix P_2 , called the propagation matrix:

$$\begin{bmatrix} E_2^+(0) \\ E_2^-(0) \end{bmatrix} = P_2 \begin{bmatrix} E_2^+(d_2) \\ E_2^-(d_2) \end{bmatrix}, \quad (5.27)$$

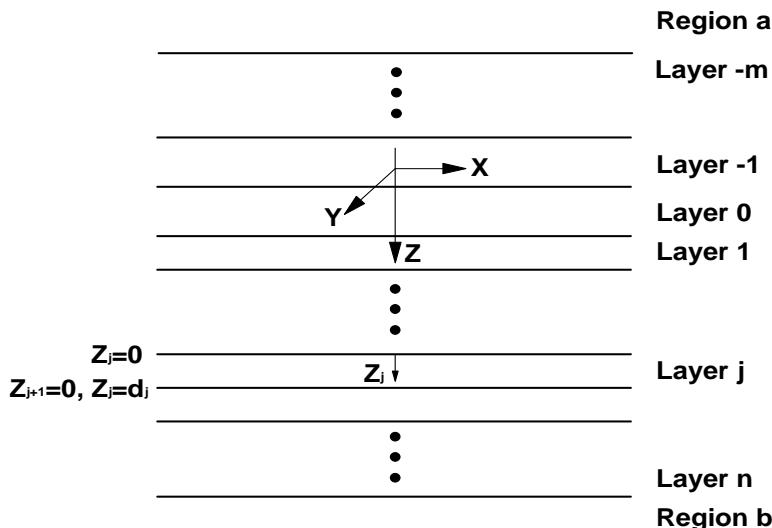


Figure 5.66 Coordinate system of the thin-film stack.

where

$$P_2 = \begin{bmatrix} e^{-i\delta_2} & 0 \\ 0 & e^{i\delta_2} \end{bmatrix}. \quad (5.28)$$

As will be shown later, parameters of the boundary matrix and the propagation matrix, such as $r_{1,2}$ and $t_{1,2}$, are functions of the wave vector of the related layers.

$$\delta_2 \equiv \frac{2\pi}{\lambda_2} d_2 = k_2 d_2 ,$$

where λ_2 is the wavelength in layer 2, and k_2 is the corresponding wave number.

With the two matrices above, we are now ready to derive reflectivity from the thin-film stack. First, consider wave propagation from layer a to layer b . We have

$$\begin{bmatrix} E_a^+ \\ E_a^- \end{bmatrix} = M \begin{bmatrix} E_b^+ \\ 0 \end{bmatrix}, \quad (5.29)$$

where E_a^\pm is the electric field at the boundary of region a to layer $-m$, E_b^+ is the electric field at the boundary of layer n and region b , and M is the matrix product of all of the dynamic and propagation matrices that govern the wave propagation through all of the layers between a and b .

$$\begin{aligned} M &= D_{a,-m} P_{-m} D_{-m,-(m-1)} \cdots D_{-1,0} P_0 D_{0,1} \cdots D_{n-1,n} P_n D_{n,b} \\ &= \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}. \end{aligned} \quad (5.30)$$

Note that in region b there is no wave propagating in the $-z$ -direction if the wave is incident from region a in the $+z$ -direction.

For notational convenience, rewrite $[D_{a,-m} \dots D_{-1,0}]$ as $M_{res top}^{-1}$ with

$$M_{res top} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad (5.31)$$

and $[D_{0,1} \dots D_{n,b}]$ as $M_{res bot}$ with

$$M_{res bot} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad (5.32)$$

where *res top* stands for resist top, and *res bot* stands for resist bottom.

Consider the wave propagating in the $-z$ -direction from the resist top to region a :

$$\begin{bmatrix} E_0^+(0) \\ E_0^-(0) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} 0 \\ E_a^- \end{bmatrix}, \quad (5.33)$$

which leads to

$$r_{res top} \equiv \frac{E_0^+(0)}{E_0^-(0)} = \frac{A_{12}}{A_{22}}. \quad (5.34)$$

The field at the resist bottom is related to that in region b for a wave traveling in the $+z$ -direction as follows:

$$\begin{bmatrix} E_0^+(d_0) \\ E_0^-(d_0) \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} E_b^+ \\ 0 \end{bmatrix}. \quad (5.35)$$

Hence,

$$r_{res\ bot} \equiv \frac{E_0^-(d_0)}{E_0^+(d_0)} = \frac{B_{21}}{B_{11}}. \quad (5.36)$$

Reflectivity at the resist top, $r_{res\ top}$, and resist bottom from inside the resist, $r_{res\ bot}$, are two important quantities relating directly to the swing effect, which plays an important role in the CD control of optical lithography.

The reflectivity at the top surface of the entire thin-film stack with the incident wave from air is

$$r \equiv \frac{E_a^-}{E_a^+} = \frac{M_{21}}{M_{11}}. \quad (5.37)$$

From Eq. (5.29),

$$M_{res\ top} \begin{bmatrix} E_a^+ \\ E_a^- \end{bmatrix} = P_0 M_{res\ bot} \begin{bmatrix} E_b^+ \\ 0 \end{bmatrix} \quad (5.38)$$

and

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} E_a^+ \\ E_a^- \end{bmatrix} = \begin{bmatrix} e^{-i\delta_0} & 0 \\ 0 & e^{i\delta_0} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} E_b^+ \\ 0 \end{bmatrix}. \quad (5.39)$$

Using Eqs. (5.34), (5.36), (5.37), and (5.39),

$$r = \frac{-A_{21} + A_{11}e^{i2\delta_0}r_{res\ bot}}{A_{22} - A_{12}e^{i2\delta_0}r_{res\ bot}}. \quad (5.40)$$

So the reflectivity of the thin-film stack is expressed in terms of the matrix elements of $M_{res\ top}$ and the inside reflectivity at the resist bottom.

To calculate the electric field $E_j^\pm(z_j)$ of the TE wave inside each layer of the thin-film stack, we start from

$$\begin{bmatrix} E_a^+ \\ E_a^- \end{bmatrix} = \begin{bmatrix} 1 \\ r \end{bmatrix},$$

and then propagate it to any position of the thin-film stack by using the boundary and the propagation matrices. First, we will derive the formula for $E_0^\pm(z_0)$, the electric field of the TE wave inside the resist. It is to be used for obtaining the analytical expression of the Poynting vector. To derive $E_0^\pm(z_0)$, we will start with deriving $E_0^\pm(0)$.

Relate the electric field on top of the resist to that in region *a*:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} E_a^+ \\ E_a^- \end{bmatrix} = \begin{bmatrix} E_0^+(0) \\ E_0^-(0) \end{bmatrix}. \quad (5.41)$$

From Eqs. (5.34), (5.40), and (5.41),

$$E_0^+(0) = \frac{\det M_{res\ top}/A_{22}}{1 - r_{res\ top} e^{i2\delta_0} r_{res\ bot}} E_a^+. \quad (5.42)$$

For $E_0^-(0)$, start with the relationship of the electric field at the top of the resist to that in region *b*:

$$\begin{bmatrix} E_0^+(0) \\ E_0^-(0) \end{bmatrix} = \begin{bmatrix} e^{-i\delta_0} & 0 \\ 0 & e^{i\delta_0} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} E_b^+ \\ 0 \end{bmatrix}, \quad (5.43)$$

from which $E_0^+(0) = e^{-i\delta_0} B_{11} E_b^+$ and $E_0^-(0) = e^{i\delta_0} B_{21} E_b^+$, leading to

$$E_0^-(0) = r_{res\ bot} e^{i2\delta_0} E_0^+(0). \quad (5.44)$$

Finally,

$$E_0^+(z_0) = e^{i\zeta_0} E_0^+(0) \quad (5.45)$$

and

$$E_0^-(z_0) = e^{-i\zeta_0} E_0^- = e^{-i\zeta_0} r_{res\ bot} e^{i2\delta_0} E_0^+(0), \quad (5.46)$$

where $\zeta_0 = k_0 z_0$. We now have the analytic expressions for $E_0^\pm(z_0)$.

To evaluate the Poynting vector, the total electric field inside the resist is

$$\mathbf{E}_0 = \mathbf{E}_0^+ + \mathbf{E}_0^-. \quad (5.47)$$

The corresponding total magnetic field is

$$\mathbf{H}_0 = \frac{\mathbf{k}_0^+ \times \mathbf{E}_0^+ + \mathbf{k}_0^- \times \mathbf{E}_0^-}{\omega \mu_0}, \quad (5.48)$$

where $\omega = 2\pi f$, f is the frequency of the propagating wave, and μ_0 is the permeability of the resist. The Poynting vector inside the resist is $\mathbf{I}_0 = \mathbf{E}_0 \times \mathbf{H}_0$. Since the electric field and the magnetic field involved are time harmonic, one considers only the time-averaged Poynting vector in the z -direction:

$$\begin{aligned} \langle I_{0z}(z_0) \rangle &= \langle \mathbf{I}_0(z_0) \cdot \hat{\mathbf{z}} \rangle \\ &= \frac{1}{2\omega\mu_v} \left\{ \left[|E_0^+(z_0)|^2 - |E_0^-(z_0)|^2 \right] k_0^R - 2 \operatorname{Im} [E_0^+(z_0) E_0^{-*}(z_0)] k_0^I \right\}, \end{aligned} \quad (5.49)$$

where we have replaced μ_0 by μ_v , the permeability in vacuum. Note that k_0 is complex.

Now, we are going to derive the analytic expression of the swing curve. By substituting $E_0^\pm(z_0)$, obtained previously from Eq. (5.49), we obtain

$$\begin{aligned} \langle I_{0z}(z_0) \rangle &= \left[\begin{array}{l} \left(e^{-2k_0^I z_0} + e^{2k_0^I z_0} |r_{res\ bot}|^2 e^{-4k_0^I d_0} \right) k_0^R - 2 |r_{res\ bot}| e^{-2k_0^I d_0} \\ \sin(2k_0^R z_0 - 2k_0^R d_0 - \gamma_{res\ bot}) k_0^I \end{array} \right] \\ &\times \left| \frac{\det M_{res\ bot}/A_{22}}{1 - r_{res\ top} e^{j2\delta_0} r_{res\ bot}} \right|^2 \langle I_{az}^+ \rangle, \end{aligned} \quad (5.50)$$

where we have used the Poynting vector of the incident wave

$$\langle I_{az}^+ \rangle = \frac{1}{2} \operatorname{Re} (\mathbf{E}_a^+ \times \mathbf{H}_a^{+*}) \hat{\mathbf{z}} = \frac{1}{2\omega\mu_v} |E_a^+|^2 k_{az}^+. \quad (5.51)$$

To calculate energy absorbed inside the resist, we will use Poynting's theorem, which is the energy ρ absorbed per unit volume, per unit time, equal to the negative divergence of the Poynting vector at that point:

$$\begin{aligned}
\rho &\equiv -\nabla \cdot \langle \mathbf{I}_0 \rangle \\
&= \frac{1}{k_a} \left[\begin{aligned} &\left(2k_0^I e^{-2k_0^I z_0} - 2k_0^I e^{2k_0^I z_0} |r_{res\ bot}|^2 e^{-4k_0^I d_0} \right) k_0^R \\ &+ 4k_0^R |r_{res\ bot}| e^{-2k_0^I d_0} \cos(2k_0^R z_0 - 2k_0^R d_0 - \gamma_{res\ bot}) k_0^I \end{aligned} \right] \\
&\times \left| \frac{\det M_{res\ top}}{A_{22}} \right|^2 \langle I_{az}^+ \rangle,
\end{aligned} \tag{5.52}$$

where we only care about variation of the Poynting vector along the z -direction. With its z_0 dependence, the term $4k_0^R |r_{res\ bot}| e^{-2k_0^I d_0} \cos(2k_0^R z_0 - 2k_0^R d_0 - \gamma_{res\ bot}) k_0^I$ in the above equation stands for the standing wave inside the resist. The period p of the standing wave is found to be $p = \pi/k_0^R = \lambda_0/2$. Assuming that the standing wave term can be dropped after PEB, due to diffusion of the photo acid, and that $|r_{res\ bot}|^2 e^{-4k_0^I d_0}$ as well as $|r_{res\ top}| |r_{res\ bot}|$ are very small compared to 1, we obtain

$$\begin{aligned}
\rho &\approx \frac{1}{k_a} 2k_0^I k_0^R e^{-2k_0^I z_0} \left| \frac{\det M_{res\ top}}{A_{22}} \right|^2 \\
&\times \left[1 + 2 |r_{res\ top}| e^{-2k_0^I d_0} |r_{res\ bot}| \cos(\gamma_{res\ top} + 2k_0^R d_0 + \gamma_{res\ bot}) \right] \langle I_{az}^+ \rangle.
\end{aligned} \tag{5.53}$$

Considering ρ at the resist bottom (i.e., at $z_0 = d_0$), we have, from Eq. (5.53),

$$\rho = C e^{-\alpha d_0} \left[1 + 2 |r_r| e^{-\alpha d_0} \cos(\zeta d_0 + \gamma_r) \right] \langle I_{az}^+ \rangle. \tag{5.54}$$

If the resist is operating in its linear response region, then the change of the resist CD is $\Delta w \propto \rho$. Upon integration, the generalized swing curve, such as the resist CD as an explicit function of resist thickness, can be obtained. The optical parameters of the thin-film stack are implicit in r_{res} or $r_{res\ top}$ and $r_{res\ bot}$.

In the following, we will derive elements of the boundary and the propagation matrices. We will solve the wave vector inside each layer of the thin-film stack first. Referring to Fig. 5.66, the z -axis is normal to the thin-film stack, and the xy -plane coincides with the surface of the thin-film stack. Since only the plane-wave solutions of electric and magnetic fields are considered, we assume for each layer j , $\mathbf{E}_j^\pm = \mathbf{E}_{j0}^\pm e^{i\mathbf{k}_j^\pm \cdot \mathbf{r} - i\omega t}$ and $\mathbf{H}_j^\pm = \mathbf{H}_{j0}^\pm e^{i\mathbf{k}_j^\pm \cdot \mathbf{r} - i\omega t}$, where \mathbf{E}_j^+ (\mathbf{H}_j^+) propagates into the thin-film stack, and \mathbf{E}_j^- (\mathbf{H}_j^-) propagates away from the thin-film stack. To satisfy boundary conditions on the interface between two layers, say region a and layer 1, functional dependence of all waves on the interface

should be the same, such as $\mathbf{k}_a^+ \cdot \mathbf{r} = \mathbf{k}_a^- \cdot \mathbf{r} = \mathbf{k}_1^+ \cdot \mathbf{r} = \mathbf{k}_1^- \cdot \mathbf{r}$, if \mathbf{r} lies on the interface. Choosing $\mathbf{r} = \hat{\mathbf{x}}$, we have

$$\mathbf{k}_a^+ \cdot \hat{\mathbf{x}} = \mathbf{k}_a^- \cdot \hat{\mathbf{x}} = \mathbf{k}_1^+ \cdot \hat{\mathbf{x}} = \mathbf{k}_1^- \cdot \hat{\mathbf{x}}. \quad (5.55)$$

Similarly, choosing $\mathbf{r} = \hat{\mathbf{y}}$ results in

$$\mathbf{k}_a^+ \cdot \hat{\mathbf{y}} = \mathbf{k}_a^- \cdot \hat{\mathbf{y}} = \mathbf{k}_1^+ \cdot \hat{\mathbf{y}} = \mathbf{k}_1^- \cdot \hat{\mathbf{y}}. \quad (5.56)$$

Let \mathbf{k}_a^+ lie on the zx -plane, which makes it the plane of incidence: $\mathbf{k}_a^+ \cdot \hat{\mathbf{y}} = 0$. Equations (5.55) and (5.56) can be applied repeatedly from between region a and layer 1 to between layer n and region b . We have $\mathbf{k}_j^\pm \cdot \hat{\mathbf{x}} = \mathbf{k}_a^\pm \cdot \hat{\mathbf{x}}$ and $\mathbf{k}_j^\pm \cdot \hat{\mathbf{y}} = 0$ for each layer j .

Inside each layer j , the electric field \mathbf{E}_j^\pm (or the magnetic field \mathbf{H}_j^\pm) satisfies the homogeneous wave equation

$$\nabla^2 \mathbf{E}_j^\pm - \mu_j \epsilon_j \frac{\partial^2 \mathbf{E}_j^\pm}{\partial t^2} = 0. \quad (5.57)$$

Substituting plane-wave solutions into the above equation, we have

$$\mathbf{k}_j^\pm \cdot \mathbf{k}_j^\pm = \mu_j \epsilon_j \omega^2. \quad (5.58)$$

For each layer j , assuming $\mu_j = \mu_v$ (nonmagnetic), substituting $\epsilon_j = \epsilon_v N_j^2$ where $N_j \equiv n_j + i\kappa_j$ is the complex index of refraction, we have $\mu_j \epsilon_j \omega^2 = (2\pi/\lambda_v)^2 N_j^2$, where ϵ_v is the permittivity, μ_v is the permeability, and λ_v is the wavelength, all in vacuum. From Eq. (5.58), \mathbf{k}_j^\pm is also complex. Dividing \mathbf{k}_j^\pm into its real and imaginary parts, we have $\mathbf{k}_j^\pm = \mathbf{k}_j^{\pm R} + i\mathbf{k}_j^{\pm I}$.

In the following, we will solve \mathbf{k}_j^\pm inside each layer j . This is not possible in general. However, here we only consider the special case that the material in layer a is nonabsorbing. In this case, $\mathbf{k}_j^+ \cdot \hat{\mathbf{x}} = (2\pi/\lambda_v) n_a \sin \theta_a$, if the angle of incidence at layer a is θ_a . Recall that $\mathbf{k}_j^+ \cdot \hat{\mathbf{y}} = 0$, and we may assume that

$$\mathbf{k}_j^+ = (2\pi/\lambda_v) n_a \sin \theta_a \hat{\mathbf{x}} + (k_{jz}^{+R} + ik_{jz}^{+I}) \hat{\mathbf{z}}.$$

By employing Eq. (5.58), we have

$$\mathbf{k}_j^+ \cdot \mathbf{k}_j^+ = (2\pi/\lambda_v)^2 n_a^2 \sin^2 \theta_a + (k_{jz}^{+R} + ik_{jz}^{+I})^2,$$

from which $k_{jz}^+ \equiv k_{jz}^{+R} + ik_{jz}^{+I}$ can be solved, resulting in

$$k_{jz}^+ = (2\pi/\lambda_v) \sqrt{N_j^2 - n_a^2 \sin^2 \theta_a}.$$

Actually, there are two square roots for any complex number. For k_{jz}^+ , we have chosen $\text{Re}[k_{jz}^+] > 0$, indicating that it propagates into the thin-film stack. The equation governing k_{jz}^- is exactly the same. So, k_{jz}^- must correspond to another square root. In summary, we have

$$k_{jz}^\pm = \pm (2\pi/\lambda_v) \sqrt{N_j^2 - n_a^2 \sin^2 \theta_a}. \quad (5.59)$$

Now, we are ready to derive the boundary matrix and the propagation matrix. For the system considered, it is well known that the TE wave and the TM wave are two linearly independent eigenfunctions that can be employed to expand the most general solution of the system.

First, consider the TE wave, where the electric field is perpendicular to the plane of incidence. Consider the interface between layers 1 and 2, for example. With the tangential component of the electric field being continuous, we have

$$E_1^+ + E_1^- = E_2^+ + E_2^-. \quad (5.60)$$

With the tangential component of the magnetic field being continuous, we have $\hat{\mathbf{x}} \cdot \mathbf{H}_1^+ + \hat{\mathbf{x}} \cdot \mathbf{H}_1^- = \hat{\mathbf{x}} \cdot \mathbf{H}_2^+ + \hat{\mathbf{x}} \cdot \mathbf{H}_2^-$. From $\mathbf{H}_j^\pm = \mathbf{k}_j^\pm \times \mathbf{E}_j^\pm$, we have

$$k_{1z}^+ E_1^+ + k_{1z}^- E_1^- = k_{2z}^+ E_2^+ + k_{2z}^- E_2^-. \quad (5.61)$$

When $E_2^- = 0$, we may define $E_1^-/E_1^+ = r_{1,2}$ and $E_2^-/E_1^+ = t_{1,2}$. Thus, Eqs. (5.60) and (5.61) can be rewritten as

$$1 + r_{1,2} = t_{1,2} \quad (5.62)$$

and

$$k_{1z}^+ - k_{1z}^+ r_{1,2} = k_{2z}^+ t_{1,2}, \quad (5.63)$$

where we have used $k_{1z}^- = -k_{1z}^+$ and $k_{2z}^- = -k_{2z}^+$.

By solving Eqs. (5.62) and (5.63), we obtain

$$r_{1,2} = \frac{k_{1z}^+ - k_{2z}^+}{k_{1z}^+ + k_{2z}^+} \quad (5.64)$$

and

$$t_{1,2} = \frac{2k_{1z}^+}{k_{1z}^+ + k_{2z}^+}, \quad (5.65)$$

where k_{1z}^+ and k_{2z}^+ can be obtained from Eq. (5.59). In general, inward and outward propagating waves exist, i.e., $E_2^- \neq 0$. Nevertheless, we can also rewrite Eqs. (5.60) and (5.61) in terms of $r_{1,2}$ and $t_{1,2}$. In matrix form,

$$\begin{bmatrix} 1 & 1 \\ k_{1z}^+ & k_{1z}^- \end{bmatrix} \begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ k_{2z}^+ & k_{2z}^- \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}. \quad (5.66)$$

Simplifying and substituting Eqs. (5.64) and (5.65), we finally have

$$\begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} = \frac{1}{t_{1,2}} \begin{bmatrix} 1 & r_{1,2} \\ r_{1,2} & 1 \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}. \quad (5.67)$$

For the TM wave, by following exactly the same procedure, we obtain

$$\begin{bmatrix} H_1^+ \\ H_1^- \end{bmatrix} = \frac{1}{t_{1,2}} \begin{bmatrix} 1 & r_{1,2} \\ r_{1,2} & 1 \end{bmatrix} \begin{bmatrix} H_2^+ \\ H_2^- \end{bmatrix}, \quad (5.68)$$

where

$$r_{1,2} = \frac{(k_{1z}^+/\epsilon_1) - (k_{2z}^+/\epsilon_2)}{(k_{1z}^+/\epsilon_1) + (k_{2z}^+/\epsilon_2)} \quad (5.69)$$

and

$$t_{1,2} = \frac{2(k_{1z}^+/\epsilon_1)}{(k_{1z}^+/\epsilon_1) + (k_{2z}^+/\epsilon_2)}. \quad (5.70)$$

The propagation matrix only accounts for phase change of the plane wave during its propagation. Thus, it is in the same form as the TE and the TM waves. This is confirmed by Eq. (5.28): $\delta_2 = k_{2z}^+ d_2$.

5.8 Wafer

From the lithographer's point of view, the wafer is a support for the light-sensitive media, such as the photoresist. This support must be flat to a small fraction of a micrometer within the lens exposure field so that it does not take up a significant portion of the DOF budget. Second, the wafer surface should be smooth. Surface roughness not only affects the image quality by producing local image intensity variations and by increasing the level of stray light, it also reduces the signal-to-noise ratio of the alignment mark. Third, the support should not contribute significantly to light reflection. Fortunately, with the advent of antireflection coatings, the impact of wafer reflectivity is manageable.

Wafer flatness is measured when the wafer is clamped down. Therefore, the wafer chuck has an undeniable contribution to wafer flatness; this is discussed in Sec. 5.9. As the wafer diameter increases, wafer thickness follows, making it more rigid and more difficult for the chuck to flatten the wafer. The wafer itself increasingly must be flat, even when it is free standing.

In principle, tilting the wafer chuck can compensate for deviation from parallelism between the top and bottom wafer surfaces. However, it is a good practice to make the surfaces as parallel as possible, so that the tilting range of the wafer chuck can be used to fine tune the flatness of the wafer front surface within the lens field and not be consumed with correcting poor parallelism.

Because roughness is not an optical issue for the backside of the wafer, most wafers were not polished on the backside in the early days. However, in the subhalf micrometer regime, wafers are polished on both sides to optimize the flatness of the front side.

To register the orientation of the wafer so that its crystal axis can be correctly aligned to build the devices properly and to coarse align the wafer in rotation, either a notch or a flat is made on the wafer, as shown in Fig. 5.67. A notch takes up very little area on the wafer, but its sharpness tends to induce accidental fracturing of wafers. A flat is safer but consumes more wafer area and is slightly less accurate in orienting the wafer. To meet higher angular accuracy requirements for larger wafers, the length of the flat must be increased, consuming even more wafer area. As the wafer thickness increases, following the increase of wafer diameters, wafer breakage is less of a problem, thus favoring the notch. Hence, notches are increasingly popular.

5.9 Wafer Stage

The function of a wafer stage is to hold the wafer firmly while positioning it in three dimensions and orienting it in three axes of rotation. It must be capable of flattening the front surface of the wafer as each exposure area comes under the

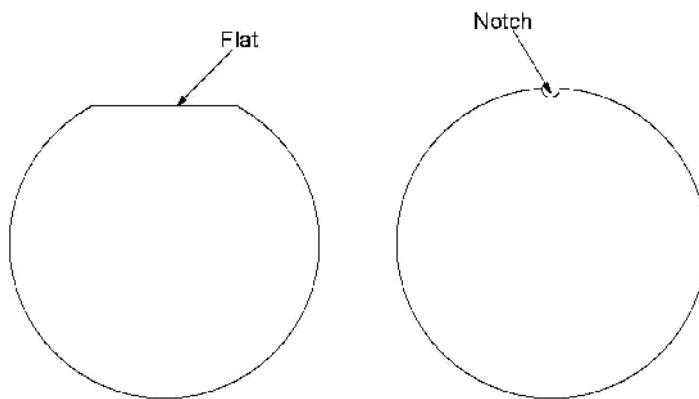


Figure 5.67 Flat and notch on wafer.

lens. The x and y movements step the wafer through each field position as well as align (in translation) the current masking level to the previously delineated masking level. The wafer stage tilts the wafer so that the largest flat area within the lens field is oriented perpendicular to the optical axis. It also rotates laterally for rotational alignment of the wafer. The z -movement is for focusing. Together, there are six degrees of freedom.

The wafer is held by vacuum to the chuck on the wafer stage. The contact surface of the chuck must be flat so that the bottom surface of the wafer is flattened. This area also must be minimized so that accidental particles attached to the bottom of the wafer do not induce flatness loss at the front surface of the wafer. Figure 5.68 shows two low-contact-area vacuum chucks for the wafer. Type A pulls the vacuum through the pins, whereas the pins in type B are simple pins that support the wafer. Though more difficult to make, the vacuum in type A does not deform the wafer. The front surface of the pins must be very flat to ensure planarization of the wafer bottom. However, in order to ensure that the front surface of the wafer is flat, regardless of the flatness of its bottom, the height of the pins may be adjusted.

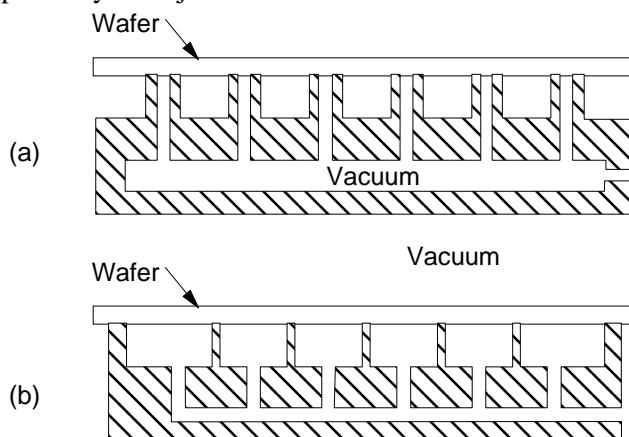


Figure 5.68 Low-contact-area wafer vacuum chucks.

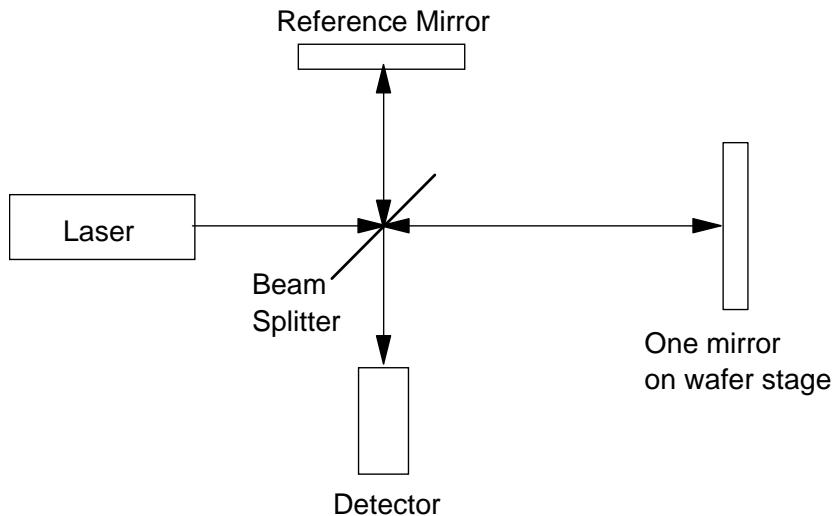


Figure 5.69 Laser interferometer.

Mirrors are mounted on the surfaces of the wafer stage to monitor its translation and rotation to control the six degrees of freedom. Figure 5.69 shows the schematic of a typical interferometer. Highly coherent light from the laser is split into two beams by the beamsplitter. One beam is reflected by a fixed reference mirror; the other beam is reflected by a mirror on the wafer stage. These two beams recombine at the beamsplitter. The detector collects their interference. As the wafer stage moves, the detected intensity changes from dark to bright and vice versa as a function of the relative position of the two mirrors. By precisely reading the intensity, the position of the moving mirror can be evaluated to better than $\lambda/64$.

5.10 Alignment System

The alignment system detects the location of the delineated masking layer, then directs the wafer stage to the desired position so that the exposed image correctly lines up with the delineated image. Instead of locating the delineated image in general, the alignment marks imbedded in that level are located. Special alignment-viewing optics are provided for this purpose. The accuracy required to locate the alignment mark is called, appropriately, “alignment accuracy.” A misleadingly similar but very different term is “overlay accuracy,” which includes the contribution from all overlay error sources, such as mask placement error, lens distortion, wafer distortion, and the tool-to-tool matching error of the exposure tools, in addition to the alignment error. Once the locations of the delineated alignment marks are determined, the lateral translation and rotation settings for the wafer stage are derived from them, referring to the position of the current mask level. Subsequently, the wafer stage is driven to these settings prior to exposure of the current masking level.

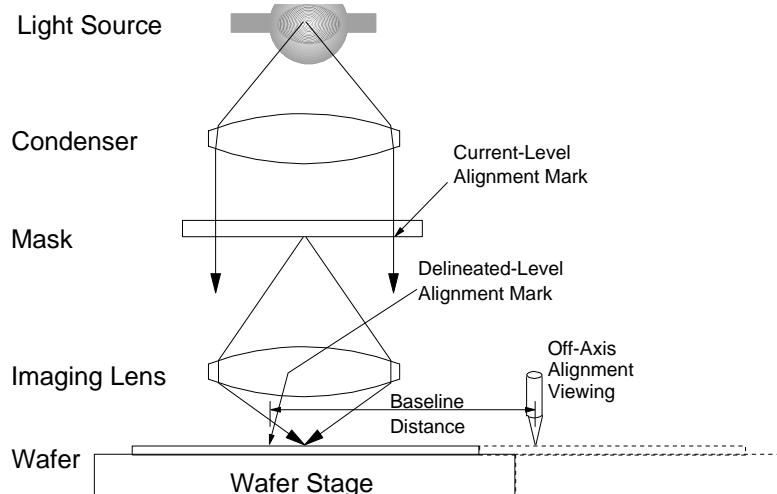


Figure 5.70 Off-axis alignment viewing.

5.10.1 Off-axis alignment and through-the-lens alignment

The off-axis alignment method is an indirect method that locates the alignment-viewing optics outside the optical path of the imaging lens so that the alignment mark of the delineated masking level is viewed at a location outside of the optical path of the current masking level, as depicted in Fig. 5.70. The distance between the alignment mark at the viewed position to its desired position for exposure is called the “baseline distance.” This distance must be included in the calculation of the desired lateral location and rotation of the wafer stage.

The through-the-lens alignment method is a direct method that incorporates the alignment-viewing optics in the optical path of the imaging lens in order to directly register the difference in the delineated and current masking levels, as depicted in Fig. 5.71.

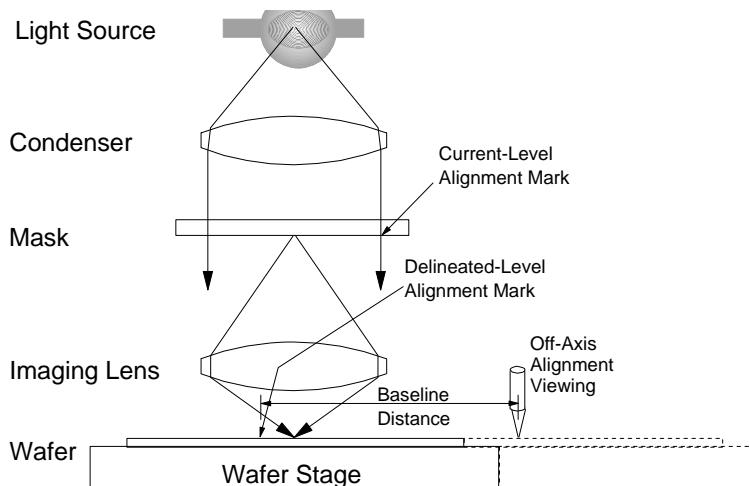


Figure 5.71 Through-the-lens alignment viewing.

Intuitively the through-the-lens method is better than the off-axis method because there is no baseline distance to consider. In reality, the following must be considered:

- (1) The viewing wavelength is difficult to decide. With actinic viewing, the alignment light can expose the photoresist, leading to unpredictable clearing of the resist in the vicinity of the alignment mark. In addition, the latent image of the exposed resist is a ghost image that can confuse the detection system, especially for sensitive detection schemes, such as dark-field alignment detection. When a nonexposing wavelength is used, the chromatic aberration must be corrected, but it is not always easy to do.
- (2) The alignment-viewing NA cannot exceed the NA of the imaging lens, thus capping the resolution capability of the alignment optics.
- (3) Unless the imaging lens is designed for polychromatic illumination, such as in catadioptric systems, it is not trivial to incorporate polychromatic alignment viewing. With monochromatic viewing, multiple reflections can create problems for alignment viewing just as in imaging the mask to the wafer.
- (4) Intrusion into the imaging optics makes the alignment-viewing optics more complicated than off-axis viewing. In addition, the intrusion may introduce imaging aberrations.

With off-axis alignment, the only problem is a baseline error that is a strong function of how well the wafer stage can be monitored. With interferometers, the wafer stage can be tracked in x , y , z , and three angular axes to nanometer accuracy. Therefore, high-NA polychromatic nonactinic optics, bright-field, dark-field, or phase detection can be used for off-axis alignment, whichever is most suitable to detect the alignment mark on the wafer.

5.10.2 Field-by-field, global, and enhanced global alignment

Field-by-field alignment means each field is aligned immediately prior to exposure. With global alignment, the entire wafer is aligned once for global translation and rotation, and then the exposure fields are stepped without further checking of the alignment marks. Enhanced global alignment is similar to global alignment except that a more careful inspection of the alignment marks is conducted to determine the positions for stepping the wafer.

Direct viewing facilitates field-by-field alignment but does not specifically call for field-by-field alignment. Global alignment using direct viewing has been done on some steppers and scanners. For field-by-field alignment, because each exposure field is tailored to each delineated field, one tends to think that this alignment method is the most accurate and that its lack of popularity is due to wafer throughput penalty. Actually, the field-by-field alignment accuracy is limited to alignment mark detection, but the global alignment accuracy can be

made much higher than the alignment mark detection limit by increasing the number of fields for global alignment. For example, let the 3σ alignment accuracy of alignment mark detection be 10 nm regardless of the alignment detection scheme, and the wafer-stage positioning accuracy be 2 nm. Also let 13 fields be aligned for global alignment. The field-by-field alignment accuracy in 3σ at the wafer is 10 nm, based on the ability to detect the alignment mark. For global alignment, the alignment accuracy is improved by the square root of the numbers of aligned fields. Hence, the accumulated alignment accuracy in 3σ is $10 / \sqrt{13} = 2.77$ nm. After incorporating the 2-nm wafer positioning accuracy quadratically, the resultant alignment accuracy is 3.4 nm.

Staying at the same alignment accuracy for field-by-field alignment, the only way to improve alignment is to repeatedly detect the same alignment marks n more times to produce a \sqrt{n} improvement. The throughput penalty would be even higher than just one alignment mark reading per field. Fortunately, enhanced global alignment combines the advantage of aligning over a large number of alignment marks without severe penalties.

5.10.3 Bright-field and dark-field alignments

Bright-field and dark-field alignments refer to the range of observed spatial frequency from the alignment mark image. With bright-field illumination, all spatial frequencies, from the 0th order to the highest one within the NA of the alignment imaging optics, are used, whereas dark-field illumination sends the 0th-order beam away. Only the other orders are observed. Figures 5.72 and 5.73 illustrate the principles of bright- and dark-field imaging.

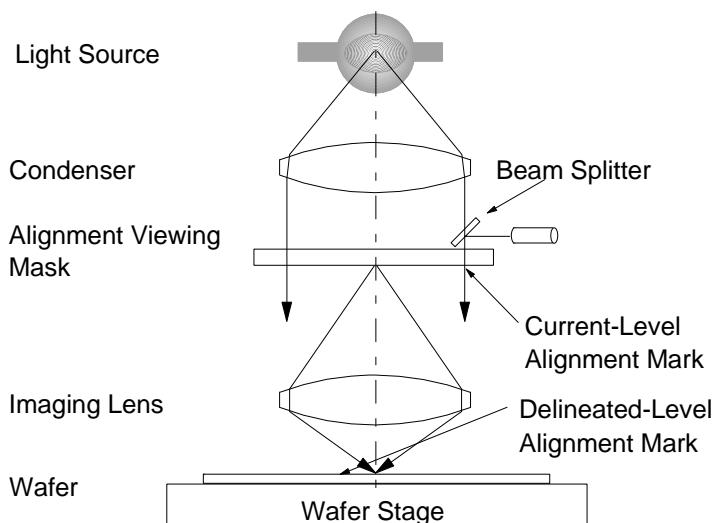


Figure 5.72 Bright-field imaging.

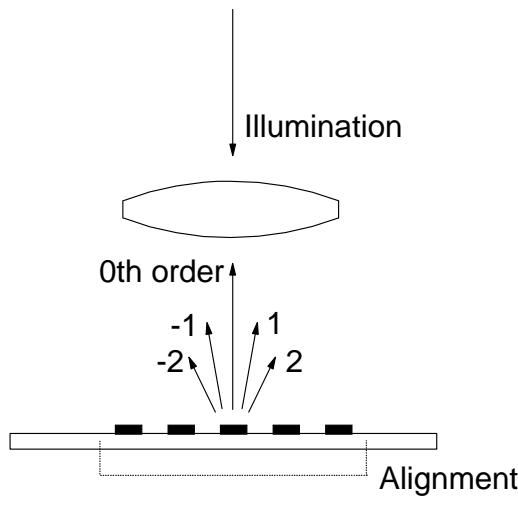


Figure 5.73 Dark-field imaging.

Because of the removal of the 0th-order beam, dark-field imaging significantly increases the signal-to-noise ratio of alignment viewing. It is less susceptible to multiple interferences, though not completely free from them. However, when aligning over grainy substrates, the grain boundary tends to show up with the alignment mark, confusing the alignment system. If the resist is exposed during the alignment process, the latent image can become a new detected image, resulting in confusion of the alignment system.

References

- 1 K. Jain and C. G. Willson, "Ultrafast deep UV lithography with excimer lasers," *Appl. Phys. B* **28**, p. 206 (1982).
- 2 K. Jain, C. G. Willson, and B. J. Lin, "Ultrafast deep-UV lithography with excimer lasers," *IEEE Electron. Dev. Lett.* **EDL-3**, p. 53 (1983).
- 3 K. Jain and R. T. Kerth, "Excimer laser projection lithography," *Appl. Optic.* **23**, p. 648 (1984).
- 4 V. Pol, J. H. Bennewitz, G. C. Escher, M. Feldman, V. A. Firtion, T. E. Jewell, B. E. Wilcomb, and J. T. Clemens, "Excimer laser-based lithography: a deep ultraviolet wafer stepper," *Proc. SPIE* **633**, p. 6 (1986).
- 5 P. Das and U. Sengupta, "Krypton fluoride excimer laser for advanced microlithography," Chapter 4 in *Microlithography—Science and Technology*, J. R. Sheats and B. W. Smith, Eds., Marcel Dekker, New York (1998).
- 6 Information on dispersion was provided by H. Feldmann at Carl Zeiss SMT AG.

- 7 I. H. Malitson, "Interspecimen comparison of the refractive index of fused silica," *JOSA* **55**, p. 1205 (1965).
- 8 R. Gupta, J. H. Burnett, U. Griesmann, and M. Walhout, "Absolute refractive indices and thermal coefficients of fused silica and calcium fluoride near 193 nm," *Appl. Optic.* **37**, p. 5964 (1998).
- 9 M. Daimon and A. Masumura, "High-accuracy measurements of the refractive index and its temperature coefficient of calcium fluoride in a wide wavelength range from 138 to 2326 nm," *Appl. Optic.* **41**, p. 5275 (2002).
- 10 J. H. Burnett, R. Gupta, and U. Griesmann, "Absolute refractive indices and thermal coefficients of CaF₂, SrF₂, BaF₂, and LiF near 157 nm," *Appl. Optic.* **41**, p. 2508 (2002).
- 11 P. Das and U. Sengupta, "Krypton fluoride excimer laser for advanced microlithography," Chapter 4 in *Microlithography—Science and Technology*, J. R. Sheats and B. W. Smith, Eds., Marcel Dekker, New York (1998).
- 12 R. Patzel, J. Kleinschmidt, U. Rebhan, J. Franklin, and H. Endert, "KrF excimer laser with repetition rates of 1 kHz for DUV lithography," *Proc. SPIE* **2440**, p. 101 (1995).
- 13 M. Born and M. Wolf, *Principles of Optics*, 6th Edition, Cambridge Univ. Press, p. 524 (1980).
- 14 B. J. Lin, "Optical projection printing threshold leveling arrangement," U.S. Patent 4,456,371 (1984).
- 15 C. A. Spence, "Method of optical lithography using phase shift masking," U.S. Patent 5,573,890 (1996).
- 16 B. J. Lin, "AZ1350J as a deep-UV mask material," *J. Electrochem. Soc.* **127**, pp. 202–205 (1980).
- 17 R. Winn, "Thin, optical membranes and methods and apparatus for making them," U.S. Patent 4,378,953 (1983).
- 18 P. Nehmiz, W. Zapka, U. Behringer, M. Kallmeyer, and H. Bohlen, "Electron beam proximity printing: complementary-mask and level-to-level overlay with high accuracy," *J. Vac. Sci. Tech. B* **3**, pp. 136–143 (1985).
- 19 B. J. Lin, "Phase shifting and other challenges in optical mask technology," *Proc. SPIE BACUS* **1496**, p. 54 (1990).
- 20 M. D. Levenson, N. S. Viswanathan, and R. A. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Trans. Electron. Dev.* **ED-29**, p. 1828 (1982).
- 21 T. Terasawa, N. Hasegawa, T. Kurosaki, and T. Tanaka, "0.3-micron optical lithography using a phase-shifting mask," *Proc. SPIE* **1088**, p. 25 (1989).
- 22 A. Nitayama, T. Sato, K. Hasimoto, F. Shigemitsu, and M. Nakase, "New phase shifting mask with self-aligned phase shifters for quarter micron photolithography," *IEDM Tech. Dig.*, p. 57 (1989).

- 23 H. I. Smith, E. H. Anderson, and M. L. Shattenburg, "Lithography mask with a π -phase shifting attenuator," U.S. Patent 4,890,309 (1989).
- 24 B. J. Lin, "The attenuated phase-shifting mask," *Sol. State Tech.* p. 43 (January 1992).
- 25 B. J. Lin, "The optimum numerical aperture for attenuated phase-shifting masks," *Microelectron. Eng.* **17**, p. 79 (1992).
- 26 H. Jinbo and Y. Yamashita, "0.2 μm or less I-line lithography by phase-shifting-mask technology," *IEDM Tech. Dig.* p. 825 (1990).
- 27 H. Jinbo and Y. Yamashita, "Improvement of phase-shifter edge line mask method," *Jpn. J. Appl. Phys.* **30**, p. 2998 (1991).
- 28 B. J. Lin, "Alternating rim phase shifting mask," U.S. Patent 5,403,682 (1995).
- 29 G. T. Dao, E. T. Gaw, N. N. Tam, and R. A. Rodriguez, "Method of fabrication of inverted phase-shifting reticle," U.S. Patent 5,300,379 (1994).
- 30 B. J. Lin, "Combined attenuated-alternating phase-shifting mask structure and fabrication methods therefore," U.S. Patent 5,565,286 (1996).
- 31 Y. T. Wang and Y. Pati, "Phase shifting circuit manufacture method and apparatus," U.S. patent 6,228,539 B1 (2001).
- 32 J. S. Wilczynski, private communication.
- 33 B. J. Lin, "New λ/NA scaling equations for resolution and depth-of-focus," *Proc. SPIE* **4000** (2000).
- 34 A. Suzuki, "Double telecentric wafer stepper using laser scanning method," *Proc. SPIE* **538**, p. 2 (1985).
- 35 W. Ulrich and R. Hudyma, "Development of dioptic projection lenses for deep ultraviolet lithography at Carl Zeiss," *J. Microlith. Microfab. Microsyst.* **3**, p. 87 (2004).
- 36 J. A. McClay, R. A. Wilklow, and M. Gregoire, "Polarization beamsplitter for photolithography," U.S. Patent 6,680,794 (2004).
- 37 F. Sporon-Fiedler and J. Williams, "Atmospheric pressure induced reduction errors in reduction stepper lenses," *Proc. SPIE* **538**, p. 86 (1985).
- 38 W. M. Moreau, *Semiconductor Lithography Principles, Practices, and Materials*, Plenum Press, New York (1988).
- 39 A. Reiser and E. Pitts, "Dynamics of copolymer solutions determined by using neutron spin-echo," *J. Photogr. Sci.* **29**, p. 87 (1981).
- 40 W. M. Moreau, *Semiconductor Lithography Principles, Practices, and Materials*, Plenum Press, New York (1988).
- 41 L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to microlithography," *ACS Professional Reference Book, 2nd Edition*, Oxford Univ. Press, New York, Sec. 3.3.2 (1994).

- 42 L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to microlithography," *ACS Professional Reference Book, 2nd Edition*, Oxford Univ. Press, New York, pp. 160–162 (1994).
- 43 W. M. Moreau, *Semiconductor Lithography Principles, Practices, and Materials*, Plenum Press, New York, Chapter 4 (1988).
- 44 L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to microlithography," *ACS Professional Reference Book, 2nd Edition*, Oxford Univ. Press, New York, pp. 212–232 (1994).
- 45 H. Moritz and G. Paal, "Method of making a negative photoresist image," U.S. Patent 4,104,070 (1978).
- 46 S. A. MacDonald, R. D. Miller, C. G. Willson, G. M. Feinberg, R. T. Gleason, R. M. Halverson, M. W. MacIntyre, and W. T. Motsiff, "Image reversal: the production of a negative image in a positive photoresist," *Proc. Kodak Microelectronics Seminar Interface '82*, p. 114 (1982).
- 47 F. Coopmans and B. Roland, "DESIRE: a novel dry developed resists system," *Proc. SPIE* **631**, p. 34 (1986).
- 48 F. Coopmans and B. Roland, "Enhanced performance of optical lithography using the DESIRE system," *Proc. SPIE* **633**, p. 262 (1986).
- 49 L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to microlithography," *ACS Professional Reference Book, 1st Edition*, Oxford Univ. Press, New York, Chapter 4 (1983).
- 50 J. M. Moran and D. Maydan, "High resolution, steep profile resist patterns," *J. Vac. Sci. Tech.* **16**, p. 1620 (1979).
- 51 M. Hatzakis, J. Paraszczak, and J. Shaw, "Double layer resist systems for high resolution lithography," *Proc. Int. Conf. Microlithog. Microcircuit Eng.* **81**, p. 386 (1981).
- 52 K. L. Tai, W. R. Sinclair, R. G. Vadimsky, J. M. Moran, and M. J. Rand, "Bilevel high resolution photolithographic technique for use with wafers with stepped and/or reflecting surfaces," *J. Vac. Sci. Tech.* **16**, p. 1977 (1979).
- 53 B. J. Lin, "Portable conformable mask—a hybrid near-ultraviolet and deep ultra-violet patterning technique," *Proc. SPIE* **174**, pp. 114–121 (1979).
- 54 B. J. Lin, "A molded deep-UV portable conformable masking system," *Polymer Eng. Sci.* **26**, p. 1112 (1986).
- 55 W. M. Moreau and P. R. Schmidt, "Photoresist for high resolution proximity printing," *Electrochem. Soc. Ext. Abstr.* **138**, p. 459 (1970).
- 56 B. J. Lin, "Deep-UV lithography," *J. Vac. Sci. Tech.* **12**, pp. 1317–1320 (1975).
- 57 I. Haller, M. Hatzakis, and R. Srinivasan, "High-resolution positive resists for electron-beam exposure," *IBM J. Res. Develop.* **12**, p. 251 (1968).

- 58 P. R. West and B. F. Griffing, "Contrast enhancement—route to submicron optical lithography," *Proc. SPIE* **394**, pp. 16 (1983).
- 59 E. Walker, "Reduction of photoresist standing-wave effects by postexposure bake," *IEEE Trans. ED-22*, p. 464 (1975).

Chapter 6

Processing and Optimization

Now that we have provided sufficient background on imaging and tools, the processes to use them and the optimization of these processes are given in this chapter. We start with optimization of the exposure tools, then move on to resist processing, k_1 reduction schemes, and finally the control of critical-dimension (CD) uniformity.

6.1 Optimization of the Exposure Tool

The optimization of an exposure tool involves its numerical aperture (NA), illumination setting, exposure/focus, depth-of-focus (DOF) budget, its components and monitoring of focus, as well as throughput optimization in field size and exposure routing.

6.1.1 Optimization of NA

Most modern exposure tools and steppers allow users to adjust the NA of the imaging lens because the NA cannot be preset in the factory when resolution is extended to the low- k_1 regime. The optimum NA^{1,2} is not the highest available setting on the tool, but rather a function of feature size and shape combinations, illumination, etc. It is important to set the NA to the best value at each given imaging environment to take advantage of the capability of the imaging system.

From Eq. (4.3), the resolution scaling equation, $\sin\theta$ must be sufficiently large to sustain the resolution of a given feature. However, Eq. (4.4) dictates that a large θ reduces DOF. Therefore, in a dry system where the space between the last lens surface and the resist is not immersed in a liquid, the NA of its imaging lens has a medium value whose DOF is maximum. Figure 6.1 shows DOF as a function of NA at $\sigma = 0.82$ for 0.20-μm lines separated by 0.20-μm spaces. The DOF is evaluated with the exposure-defocus (E-D) forest methodology covered in Chapter 4, based on 8% EL and CD bounds at ±15 nm. The DOF quickly rises from 0 to 1.04 μm at NA = 0.557, then gradually decreases as NA increases. Figure 6.2 depicts the E-D trees at the NA values evaluated. At NA = 0.455 and below, resolution is too low to support any E-D tree. An E-D tree can be constructed at NA = 0.487, but there is insufficient branch separation to support the E-D window specified for 8% EL. Then, the E-D trees become

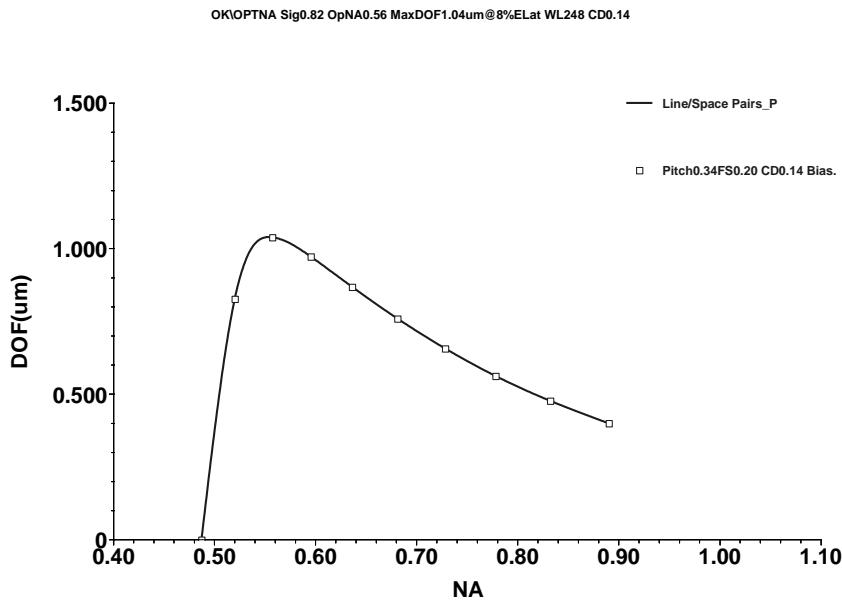


Figure 6.1 Optimized NA for DOF, 0.14- μm resist lines: 0.20- μm spaces, $\lambda = 248 \text{ nm}$, $\sigma = 0.82$, EL = 8%, $\text{CD}_{tol} = \pm 15 \text{ nm}$.

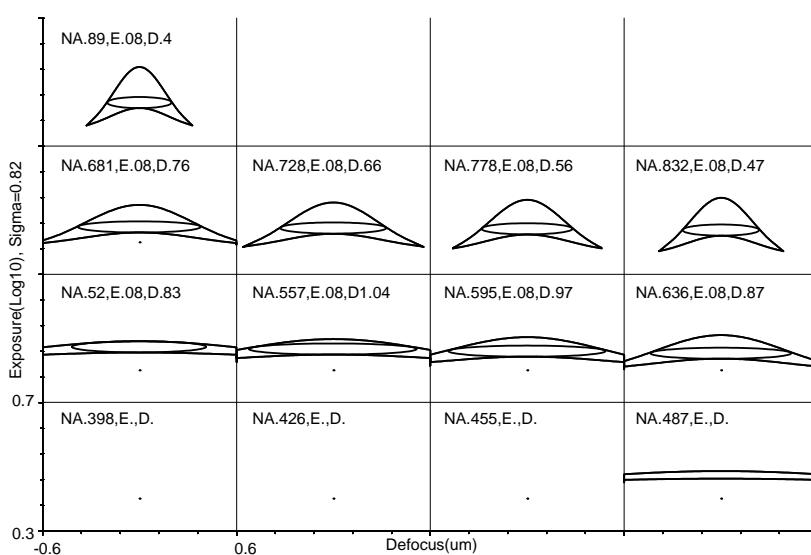


Figure 6.2 E-D windows of Fig. 6.1.

broader, trading off with increased curvature at the branches. The tradeoff favors DOF in the beginning, resulting in the maximum DOF at 1.04 μm . Then, it starts to favor EL at the expense of DOF, resulting in the gradual reduction of DOF at higher NAs.

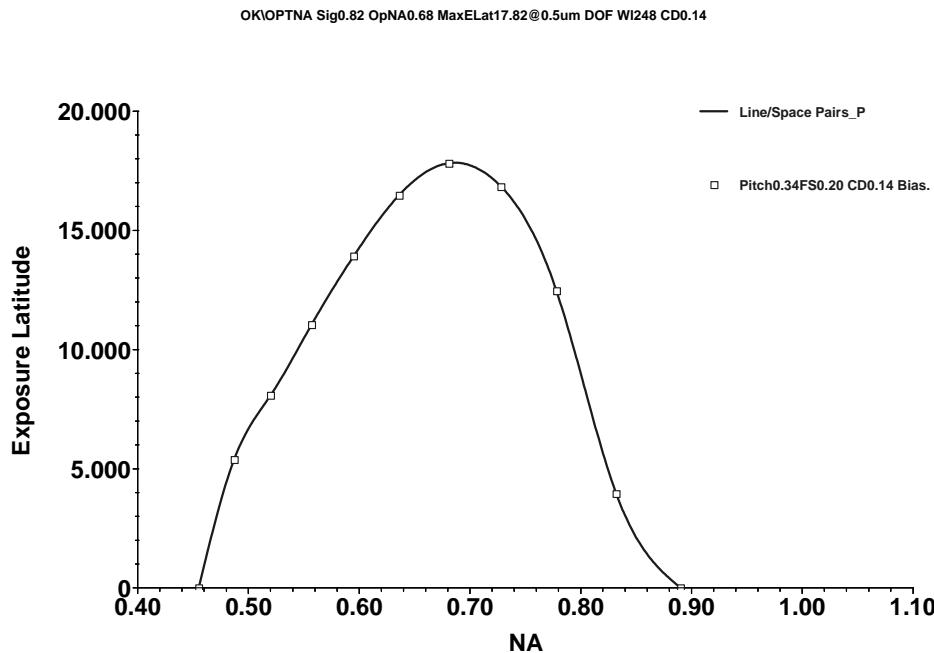


Figure 6.3 Optimized NA for EL, 0.14- μm resist lines: 0.20- μm spaces, $\lambda = 248 \text{ nm}$, $\sigma = 0.82$, DOF = 0.5 μm , $CD_{tol} = \pm 15 \text{ nm}$.

The existence of an optimum NA is not limited to extending DOF. There can be a different optimum NA for EL, as shown in Fig. 6.3, using the same feature and illumination condition as in Fig. 6.1, except that the DOF of the E-D windows is kept at 0.5 μm . The optimum NA for maximum DOF is 0.557, whereas optimum NA for maximum EL is 0.681. When NA is low, the image has a low contrast, leading to a small EL. When NA becomes larger, EL in the focal plane increases, but the curvature of the E-D trees and their shrinking in the defocal direction make it difficult to sustain a large EL at a given DOF, as seen in Fig. 6.4. At NA = 0.89, so much shrinking has occurred that no E-D window with DOF = 0.5 μm can fit into the E-D tree.

The NA can be optimized for the E-D area as shown in Figs 6.5 and 6.6. NA is now optimized at 0.636, which happens to be between the NAs for DOF and EL. The argument is similar. At low NA, a low image contrast leads to narrow E-D regions. At a high NA, image contrast is high in focus, but quickly vanishes.

Among DOF, EL, and E-D area, which should be optimized? One should first optimize for DOF at an EL that can be sustained by resist processing, such as using antireflection coatings, or multilayer resist systems. Typically, one takes the percentage exposure range derived from the resist swing curve within the range of topography in the given exposure field and combines it with the

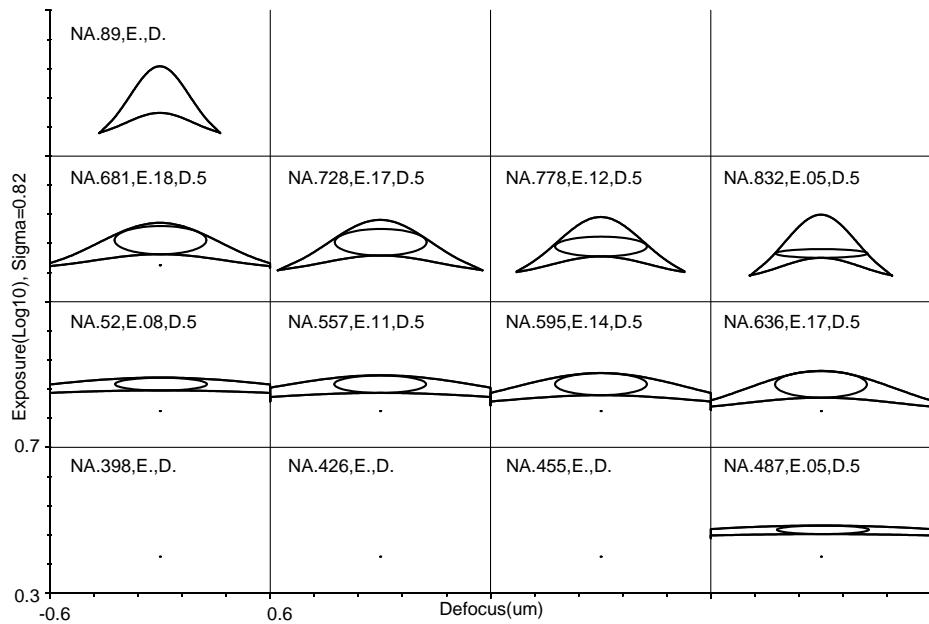


Figure 6.4 E-D windows of Fig. 6.3.

OKIOPNA Sig0.82 OpNA0.64 MaxEDA4.1 WL248 CD0.14248 CD0.14

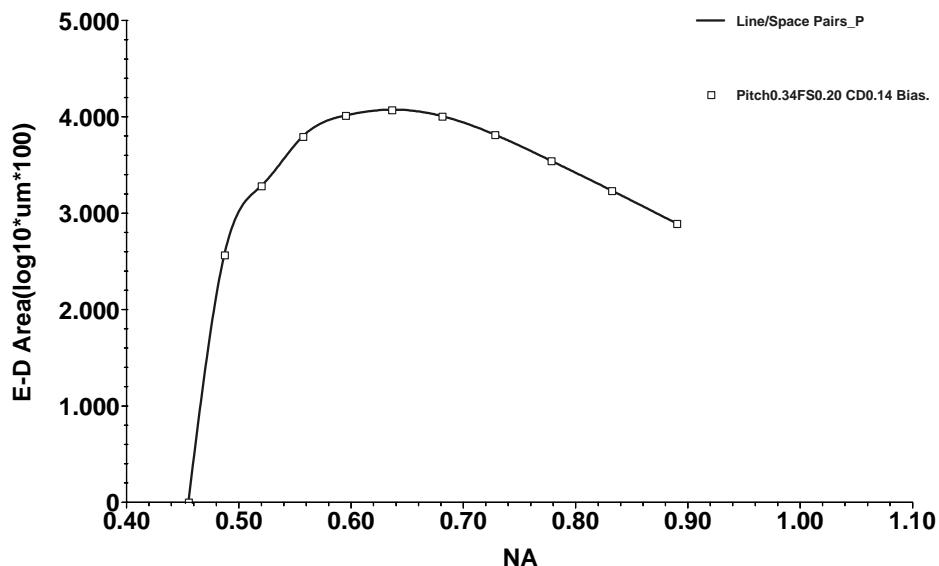


Figure 6.5 Optimized NA for E-D area, $0.14 \mu\text{m}$ resist lines: $0.20\text{-}\mu\text{m}$ spaces, $\lambda = 248 \text{ nm}$, $\sigma = 0.82$, DOF = $0.5 \mu\text{m}$, $CD_{tol} = \pm 15 \text{ nm}$.

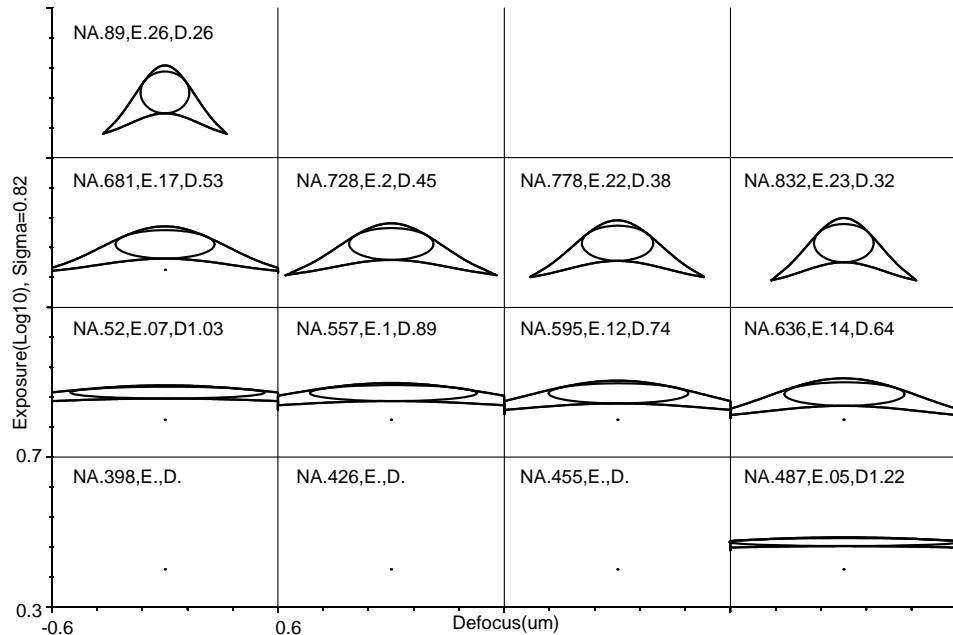


Figure 6.6 E-D windows of Fig. 6.5.

percentage illumination nonuniformity, as well as the percentage dosage control inaccuracy, to evaluate the EL. Usually, 8% is acceptable. If DOF at the optimum NA does not meet this requirement, this imaging system cannot handle the given task. If DOF at only one or two NA settings exceeds the requirements slightly, these NA settings should be used. If DOF exceeds the requirement at many NA settings within the NA adjustable range, then optimizing for EL should be attempted, because a larger EL helps CD control. If there are many NA settings exceeding DOF and EL requirements, then the NA that produces the largest E-D area should be selected, because this largest processing window is statistically easiest to meet; thus, a higher lithography-imaging yield is attained.

6.1.2 Optimization of illumination

The illumination settings on the exposure tool play a role in image performance that is equally important as the NA setting on the imaging lens. First, the type of illumination must be decided, namely disk, annular, or multipole. Once decided, the settings must be optimized, such as the σ setting for disk illumination (DKI), σ_{in} and σ_{out} for annular illumination, pole size, shape, and location for multipole illumination.

With DKI, the DOF, EL, and the E-D area obviously change as a function of σ . However, the existence of an optimum σ cannot be as easily visualized as using the resolution and DOF scaling equations to visualize the existence of an optimum NA. With off-axis illuminations, such as annular and multipole, the

existence of an optimum illumination angle is directly related to the working principle of these systems, as was given in Sec. 5.3.2.

Figure 6.7 shows the change in DOF as a function of the σ setting of DKI. The DOF is indeed dependent on σ . In this particular case, all features chosen favor a large σ , with the denser features optimizing at σ in the regime of 0.8. The optimum σ for the common E-D window is even more pronounced, centering at around 0.82.

Figure 6.8 is similar to Fig. 6.7, except that annular illumination with $\sigma_m = 0.67\sigma_{out}$ is used. The DOF peaks for the 1:1.5, 1:2, and 1:3 lines are even more obvious than in the case of DKI, and they are larger. Unfortunately, the DOF of common windows is smaller than in the former case. DOF can be improved with optical proximity correction, but it is limited by the features with larger spaces in both cases. Sec. 6.11 shows a way to further improve the DOF.

Following optimization of illumination, the NA can be optimized at this illumination setting, as shown in Fig. 6.9. In this case, the DOF is improved from 0.21 to 0.22 μm with NA changing to 0.74 from 0.7, which is only improved slightly. In principle, one can repeat the NA and σ optimization process alternately until the final optimum is reached. In the particular examples shown in Figs. 6.8 and 6.9, the convergence is fast. Figure 6.10 shows DOF as a function of NA and σ plotted together, using the situation studied in Fig. 6.1.

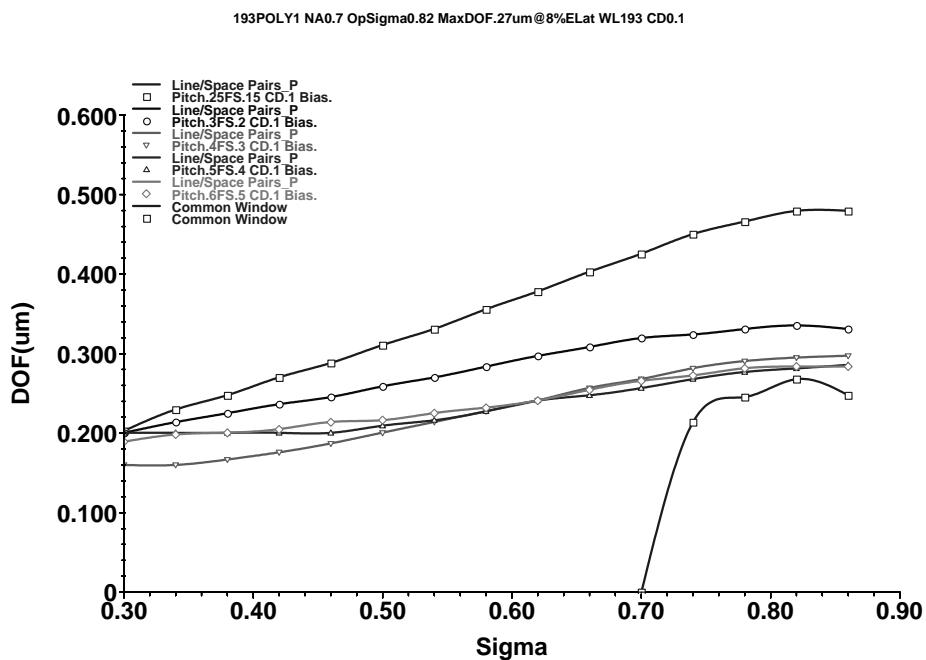


Figure 6.7 DOF as a function of σ for positive resist line-space ratios of 1:1.5, 1:2, 1:3, 1:4, 1:5, and their common window. The resist linewidth is 100 nm, $\lambda = 193$ nm, NA = 0.7, EL = 8%.

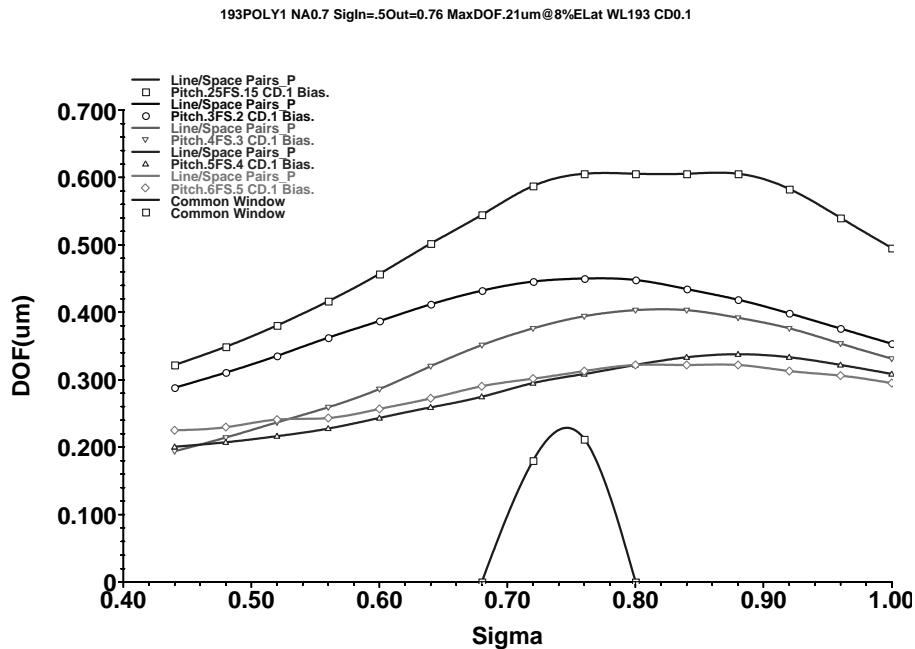


Figure 6.8 Same as Fig. 6.7, except annular illumination with $\sigma_{in} = 0.67\sigma_{out}$ is used.

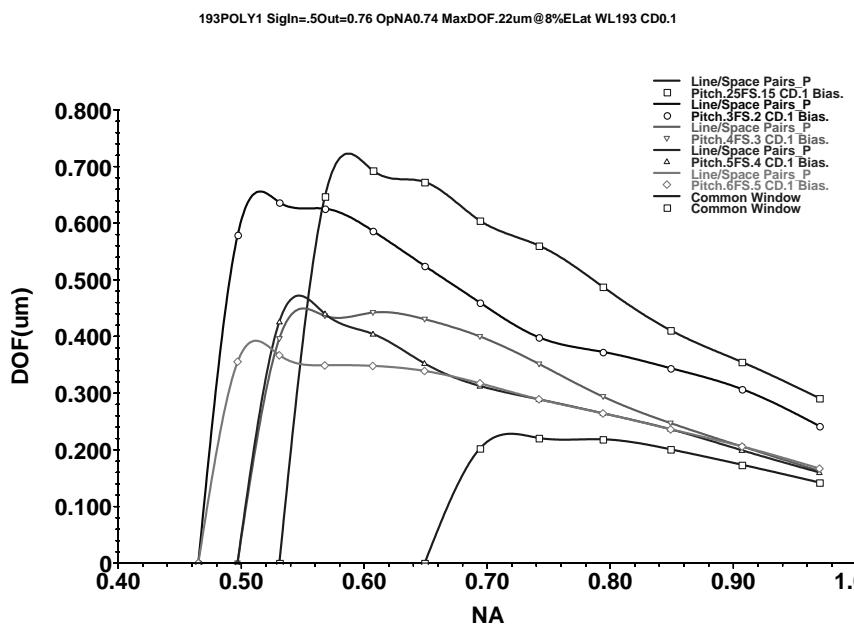


Figure 6.9 Same as Fig. 6.8, except that NA is optimized at the σ settings evaluated from Fig. 6.8, $\sigma_{in} = 0.5$ and $\sigma_{out} = 0.76$.

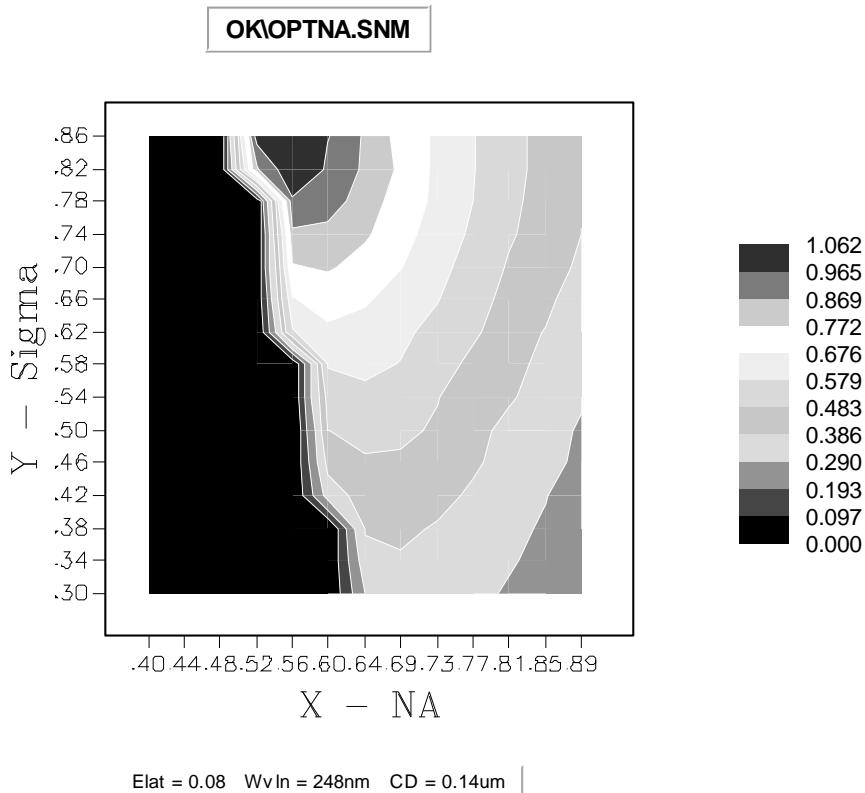


Figure 6.10 DOF as a function of NA and σ for the feature in Fig. 6.1.

6.1.3 Exposure and focus

Exposure dosage and focus are the two most important field-adjustable parameters. They must be carefully and correctly set. Sections 4.32 and 4.53 covered the need to define the EL and DOF together in an E-D window, because these two quantities are mutually dependent. The center of exposure and focus is the center of the E-D window. However, which E-D window should one use in a processing environment: the E-D window of a predominant representative feature, the common E-D window of all or some pertinent features, or the E-D window of one or a set of test features placed in every mask set?

First, let us examine the difference between the usual practice and the E-D window approach. Intuitively, one finds the best focus from a typical feature with reasonably decent exposure. On a product wafer, setting the center of exposure and focus is more complicated because the proximity environment must be taken into account. Figure 6.11 shows the E-D trees and windows of a 140-nm line with 120, 160, 200, 300, and 400-nm spaces, delineated with 193-nm light, using NA = 0.6 and annular illumination $\sigma_{in} = 0.3$ and $\sigma_{out} = 0.6$. The EL is set at 5%, and CD tolerance is ± 9 nm. The exposure and focal center of each feature combination and those of the common E-D tree are clearly different from each

other. The numerical values of the exposure and focal center, as well as the DOF, are listed in Table 6.1. In this case, the exposure center of the E-D window is within 0.01 mJ/cm² from that of the common window and the focal center, only 2 nm higher. Instead of monitoring all of the features, one can monitor the 140:400-nm line-space pairs and adjust the centers accordingly. This particular line-space ratio is the gating feature for a combination of the five. Therefore, its E-D behavior is close to that of the common window. Sometimes numerous feature combinations gate the common window; any one of them can be selected with the centers adjusted for the evaluated difference in exposure and focus.

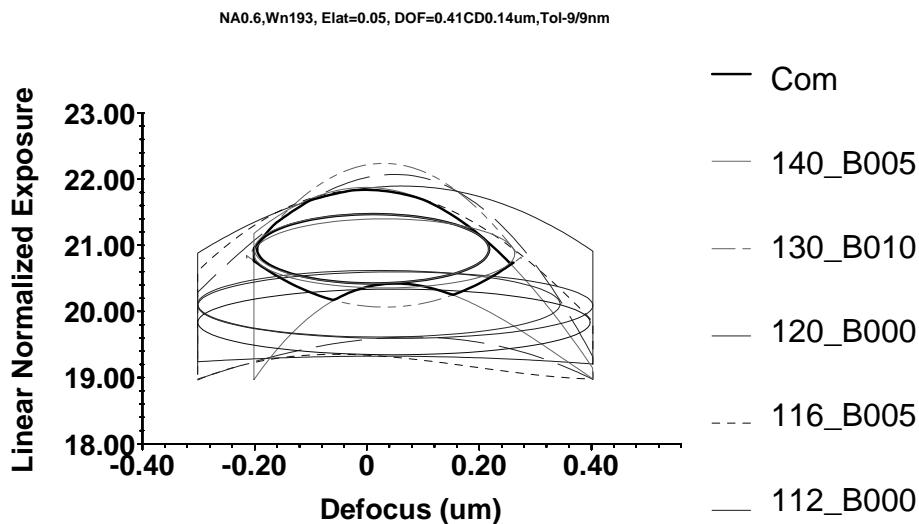


Figure 6.11 E-D windows of 140-nm lines with 120, 160, 200, 300, and 400-nm spaces. $\lambda = 193$ nm, NA = 0.6, and $\sigma = 0.6/0.3$. EL = 5% and $\Delta CD = \pm 9$ nm.

Table 6.1 Exposure center, focal center, and DOF of the five feature combinations in Fig. 6.11.

Feature combination	Exposure center	Focal center	DOF
Common window	20.94 mJ/cm ²	11 nm	413 nm
140 : 120 nm	20.09 mJ/cm ²	50 nm	704 nm
140 : 160 nm	19.83 mJ/cm ²	48 nm	699 nm
140 : 200 nm	20.10 mJ/cm ²	21 nm	646 nm
140 : 300 nm	20.87 mJ/cm ²	31 nm	465 nm
140 : 400 nm	20.94 mJ/cm ²	9 nm	418 nm

6.1.4 DOF budget

The DOF budget can be looked at in two ways: DOF budget provided and DOF budget consumed. The provided budget is determined by imaging parameters such as wavelength, bandwidth, NA, illumination, feature shape, feature size, feature combination, CD tolerance, phase and intensity on the mask, multiple reflections, and EL, as was elaborated in Chapter 4. The DOF budget is consumed by the imperfection of the components in the imaging train, such as mask flatness, mask tilt, mask topography, mask chucking error, proximity effects, lens aberrations, lens heating, illumination nonuniformity, resist thickness, thickness nonuniformity (in the TARC, resist, BARC, and the thin-film stack on the wafer), topography, wafer nonflatness, wafer tilt, wafer chucking error, and focusing error.

The provided budget is best allocated in at least two imaging media, such as the coupling medium from the lens to the resist and the resist medium. This is particularly important with immersion lithography in which the refractive index of the coupling medium exceeds one. An extensive treatment is given in Chapter 7. We will simply build this discussion on that work. $\text{DOF}_{\text{diffrac}}$, defined later in Sec. 7.3.3, provides the DOF budget. $\text{DOF}_{\text{required}}$, defined in Sec. 7.3.4, consumes the DOF budget. We now discuss the imperfections that constitute $\text{DOF}_{\text{required}}$.

6.1.4.1 Components in $\text{DOF}_{\text{required}}$

Mask contributions—Contributions from the mask include its flatness, roughness, topography of the absorber or phase shifter, mask tilt due to nonparallelism, chucking, and scanning imperfections. All vertical imperfections are reduced by $1/M^2$ for the DOF budget at the wafer side. Given $M = 4$ for scanners, the 1/16 reduction factor is not so effective in modern-day lithography. For example, the $\text{DOF}_{\text{diffrac}}$ consumed by the resist thickness allows an available DOF of only 100 nm for the polysilicon layer of the 32-nm logic node with a 45-nm half pitch. The best mask flatness presently commercially available is 500 nm. The chromium absorber thickness is 200 nm, while the phase shifter in quartz at the 193-nm wavelength is also on the order of 200 nm. The equivalent vertical variation on the wafer is now on the order of 57 nm, more than half of the available DOF budget.

Subsequent mask processing can affect the flatness of a mask blank. For example, the stress on the mask is changed after the absorber is delineated. It is also affected by mounting the mask pellicle.

Fortunately, the static imaging field of a scanner is 26×8 mm instead of the entire square area, as in a stepper. It is possible to actively compensate for mask imperfections by adjusting the tilt of the mask dynamically during scanning. We often see improvements in the order of 25%.

Lens effects—The lens effects that consume the DOF budget are generally lumped into image-plane deviation (IPD). Basically, IPD contains the variation of the Zernike Z_4 term at different field positions. Field curvature, astigmatism,

and coma all contribute to consumption of the DOF budget. Some of them can be artificially avoided. For example, assigning the critical dimensions in only one direction can alleviate astigmatism-induced DOF deficiency. It is known that some manufacturers align their critical gates in one dimension. Of course, reducing the effect of astigmatism is only one of the reasons to align the gates. The aberrations of a hot lens when the exposure tool is running at its full capacity are different from those evaluated when the lens is in a testing environment. The effect can be as large as 30 nm.

Illumination nonuniformity has a direct consequence on the DOF budget. This is obvious from E-D trees. Larger EL is required to accommodate the illumination nonuniformity, trading off DOF. When the illumination wavelength drifts, the change of wavelength causes focus shift.

Thicknesses—The resist thickness consumes a major part of the DOF budget. Take the 32-nm node as an example again. The resist thickness required is on the order of 100 nm. It consumes $100/1.75$ nm of the DOF budget in a dry exposure system, where 1.75 is the refractive index of the resist. In a water-immersion system, the resist consumes $100 \times 1.44/1.75$ nm of the DOF budget, where 1.44 is the refractive index of water at the 193-nm wavelength. For the reason behind multiplying with the refractive index of the coupling medium, refer to Eq. (7.10).

Film nonuniformity consumes the DOF budget physically and optically. Nonuniformities in the TARC, resist, BARC, and thin-film stack on the wafer dislocate the vertical position of the imaging layer. Optically, these thickness nonuniformities perturb multiple reflections on the interfaces, creating exposure nonuniformity, and lead to consumption of the DOF budget. Topography affects exposure and focus uniformity. In principle, chemical mechanical polishing (CMP) should have flattened the topography. However, at the 90-nm node, CMP residual topography can be on the order of 50 to 100 nm.

Wafer nonflatness, wafer tilt, and wafer chucking error obviously consume the DOF budget. For sub-100-nm lithography, double-sided polished wafers are mostly used because polishing the backside improves its flatness, so that the front surface is less distorted when the wafer is chucked down. Choi reported a 30% reduction in focus variation with a double-sided polished wafer.³

Similar to mask flatness, the wafer can deteriorate from its original flatness after repeated patterning and film deposition, especially with film stress caused by thermal cycles. CMP helps but does not completely eliminate it.

Fortunately, active compensation of wafer flatness imperfections can be performed on the scanner just like compensation for mask-flatness imperfections. Hence, SEMI proposes site flatness quality requirements (SFQR) to characterize wafer flatness that are site based, front-surface referenced, use a least squares reference plane within the site, and report the range of flatness within the site. The size of the site should be chosen as the size of the exposure slit, such as 26×8 mm.

In fact, when performing a scanning exposure, the exposure slit moves continuously across a field and levels dynamically. Hence, a point P in a field is

exposed by all points of the line segment traced out by P through the exposure slit. As pointed out by ASML,⁴ to be correlated more closely to the residual focus error after leveling, it is more appropriate to use moving average (MA), defined as the average of the residual defocus of all points exposing P .

$$\langle \Delta z_s(x_s) \rangle = h_s^{-1} \int_0^{h_s} \Delta z_s(x_s, y_s) dy_s , \quad (6.1)$$

where Δz_s is the residual defocus after leveling when P is exposed by $P_s(x_s, y_s)$ of the exposure slit, and h_s is the slit height. The leveling is optimized to take care of the local topography inside the region of the exposure slit.

Another important component of DOF consumption is focusing error. Focusing error is the error committed by the focusing sensor and the vertical positioning of the mask, the lens, and the wafer on the exposure tool. Once the best focus is determined, the focusing sensor and the positioning devices must consistently maintain each exposure at this best focus. Failure to do so results in focusing error.

Focus sensing can be done several ways: namely, optical sensing, air gauge sensing, and capacitive gauge sensing, as shown in Figs. 6.12, 6.13, and 6.14, respectively. Each of these sensing methods has advantages and disadvantages. Optical sensing uses an obliquely incident beam on the wafer surface at grazing incidence. The beam is detected at the other end. Any change in the vertical position or tilting of the wafer changes the position of the detected beam. The air gauge is a small tube placed in proximity to the wafer surface. Compressed air coming out of the tube changes pressure as a function of the closeness of the wafer surface. Monitoring the pressure change detects the vertical position of the wafer. The capacitive gauge places a small capacitive probe near the wafer surface, and the capacitance changes as a function of the vertical position of the wafer. The latter is obviously affected by the electrical property of the films and features on the wafer. For the air gauge, rushing air has a thermal effect on the wafer. It is also a contamination source. Optical sensing can be affected by the optical property of the films on the wafer. Grazing incidence tends to minimize this effect. Hence, all scanner suppliers utilize optical sensing. Before the shake up of optical-exposure tool suppliers, there were GCA, Perkin Elmer (which later became SVGA), Ultratech, Optimetrix, Censor, and Electromask. All three types of sensor have been adopted by one supplier or another. Currently, due to stringent focus-control requirements, the air gauge is used to supplement optical focus sensing.

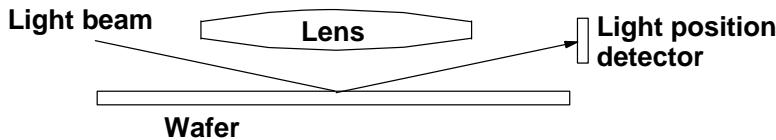


Figure 6.12 Optical focus sensing.

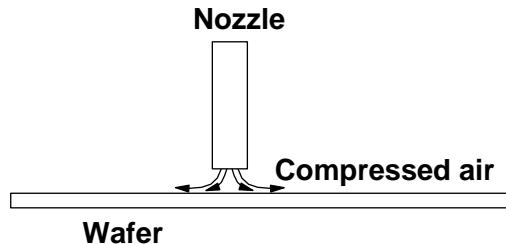


Figure 6.13 Focus sensing using air gauge.

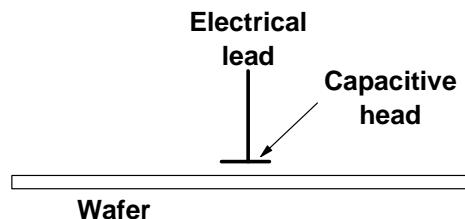


Figure 6.14 Focus sensing using capacitive gauge.

6.1.4.2 Focus monitoring

Focus monitoring builds on focus sensing. The wafer image is monitored for correctly setting the best focus. One must first define the best focus and then determine the way to evaluate it.

The best focus is the vertical position that keeps a given set of critical features within the largest EL in their common E-D window. One must make sure that all of the parameters that determine the provided DOF budget are used to set the best focus. They are wavelength, bandwidth, NA, σ , feature shape, feature size, feature combination, CD tolerance, phase and intensity on the mask, proximity correction, multiple reflections, and EL, as listed above in Sec. 6.1.4. If the Bossung⁵ curve is used, the best focus can only be evaluated on one critical feature. This is often insufficient. Figure 6.15 shows the E-D trees of 0.5- μm line-space pairs and a 0.5- μm isolated line opening, imaged at $\lambda = 365 \text{ nm}$, NA = 0.48, and $\sigma = 0.5$. The center of DOF for the former feature is at -231 nm, and that for the latter feature is at 308 nm. Taking the common E-D window of the two features, the center of DOF is at 154 nm. They are quite different. Frequently

CD=.5um,CDtol=-10/10%,Elat=5.3%,DOF=1.,\ls\m11\66m11_50\ln\66ln_50

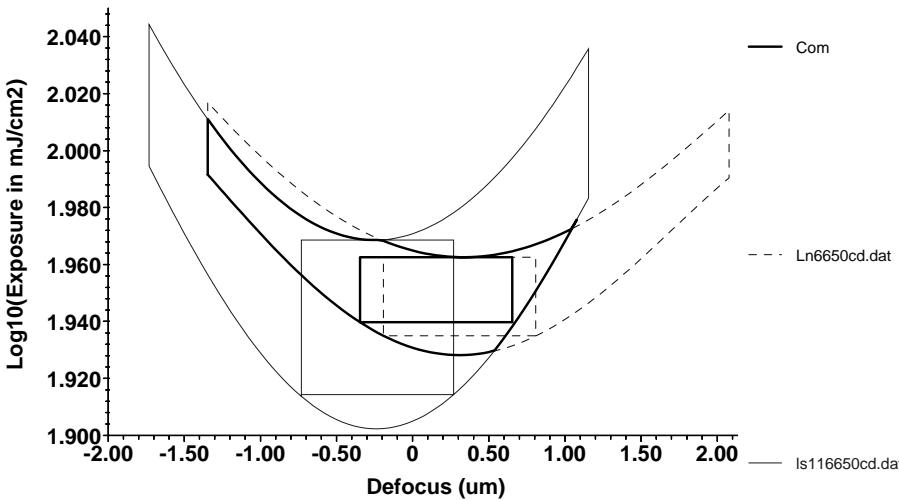


Figure 6.15 Determination of the best focus for two 0.66- k_1 features: line-space = 1:1, and the isolated line is $\lambda = 365$ nm, NA = 0.48. The center of DOF for the lines and spaces (LS) is -231 nm; the isolated line opening is 308 nm; the common E-D window is 154 nm.

monitoring the critical features can be tedious and unproductive. For daily focus monitoring, one should identify certain test patterns that can precisely indicate whether the best focus of the critical features is met. For example, line-end shortening (LES) is very sensitive to defocus and can be a good test pattern to monitor focus.

Ausschnitt⁶ made attempts to facilitate focus monitoring with a model-based approach by first calibrating a relationship of CD versus exposure dosage and focus, then finding the corresponding focus through the measured CD. By using LES as ΔW , the change of CD, the sensitivity to focus can be made even higher. With a proper arrangement of mask features, such LES can be measured by overlay metrology tools. This method has been commercialized.⁷

Since CD depends on exposure dosage E and focus F , E and F should be solved simultaneously. So, we need to measure the CD of at least two kinds of patterns. In Ausschnitt's original approach, W as a function of E and F is assumed to be of the following form:

$$W[k] = a_0[k] + a_1[k]E + (a_2[k] + a_3[k]E)F^2, \quad (6.2)$$

where k is used to indicate the pattern type, and $a_0[k]$, $a_1[k]$, $a_2[k]$, $a_3[k]$ are to be characterized by a series of Bossung curves. Then, if the CD of two kinds of patterns is measured, E can readily be solved by eliminating F , such as in

$$\frac{W[1] - a_0[1] - a_1[1]E}{a_2[1] - a_3[1]E} = \frac{W[2] - a_0[2] - a_1[2]E}{a_2[2] - a_3[2]E}. \quad (6.3)$$

F follows once E is obtained. However, since the response of CD to defocus is mainly quadratic, the sign of defocus cannot be determined unambiguously. Additional marks are added to solve the defocus sign problem. In addition to the sign problem, the center of focus for different patterns may not be identical. Equation (6.2) should be modified as below, with Eq. (6.1) adjusted accordingly. If the center of focus for the two chosen patterns are the same, then E and F can be solved by Eqs. (6.2) and (6.3). If not, E and F can also be solved since we know the center of focus of the two chosen patterns. However, the expression of the analytical solution is much more complicated. This method is only for focus monitoring, not for accurate determination of the center of focus of a group of patterns.

$$W[k] = a_0[k] + a_1[k]E + (a_2[k] + a_3[k]E)(F - F_0[k])^2. \quad (6.4)$$

The introduction of a phase-shifting mask (PSM) can solve the sign problem.⁸ In its normal implementation, such as in an alternating phase-shifting mask (AltPSM), the phase difference of two phase-shifters beside a dark line is π to maximize the DOF. However, it is found that if the phase difference is not exactly π , then the image of the dark line will be displaced laterally when there is defocus. A clever implementation of this idea realizes the phase-shift focus monitor (PSFM), as shown in Fig. 6.16. By adding phase shifters next to overlay marks, one can convert Δz in defocus to Δx (or Δy) in overlay errors. By choosing a phase difference of $\pi/2$, the sensitivity, defined as $\Delta x/\Delta z$, can be maximized. Since the metrology tool has now changed from scanning-electron microscope for CD to optical microscope for overlay, the measurement precision is greatly improved.

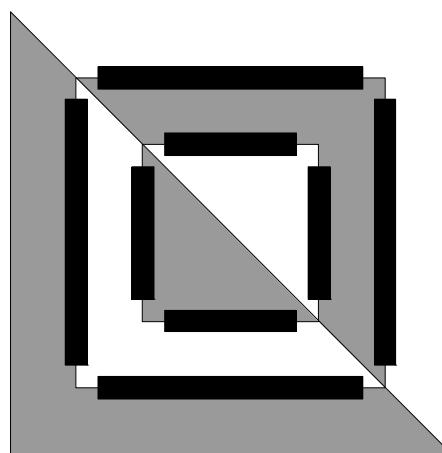


Figure 6.16 PSFM pattern.

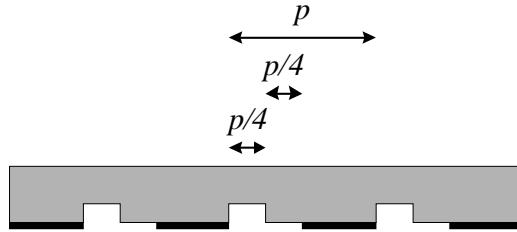


Figure 6.17 PGFM pattern.

The sensitivity can be further enhanced by using the phase grating.⁹ The corresponding methodology is called the phase-grating focus monitor (PGFM), illustrated in Fig. 6.17. The sensitivity is reported to be 3.03^{-1} for ArF scanners, more than two times higher than that of a PSFM at 7.14^{-1} under the same processing conditions.¹⁰ Since it is easier to handle the theoretical aspect of PGFM, we use it to illustrate the underlying principle of the focus-monitoring method by PSM.

Assuming the incident light is a plane wave corresponding to the conventional illumination of $\sigma = 0$, the amplitude $A(f)$ of various diffraction orders is calculated by the Fourier transform of the mask transmission and phase distribution $M(x) = 1$, from $x = 0$ to $p/4$, then $M(x) = e^{i\xi}$, from $x = p/4$ to $p/2$, and $M(x) = 0$ for the remaining range of the period.

$$\begin{aligned}
 A(f) &= \int_{-\infty}^{\infty} M(x) e^{-i2\pi fx} dx \\
 &= \sum_{n=-\infty}^{\infty} \int_{np}^{np+\frac{p}{4}} e^{-i2\pi fx} dx + \sum_{n=-\infty}^{\infty} \int_{np+\frac{p}{4}}^{np+\frac{p}{2}} e^{i\xi} e^{-i2\pi fx} dx \\
 &= \left(1 + e^{i\xi} e^{-i2\pi f \frac{p}{4}} \right) \int_0^{\frac{p}{4}} e^{-i2\pi fx} dx \sum_{n=-\infty}^{\infty} \left(e^{-i2\pi fp} \right)^n \\
 &= \sum_{m=-\infty}^{\infty} \left(1 + e^{i\xi} e^{-i\pi \frac{m}{2}} \right) \frac{p}{4} e^{-i\pi \frac{m}{4}} \text{sinc}\left(\frac{m}{4}\right) \frac{1}{p} \delta\left(f - \frac{m}{p}\right),
 \end{aligned} \tag{6.5}$$

where

$$\text{sinc}(z) = \frac{\sin(\pi z)}{\pi z}$$

and

$$\sum_{n=-\infty}^{\infty} \left(e^{i2\pi f} \right)^n = \sum_{m=-\infty}^{\infty} \delta(f - m).$$

$A(f)$ is nonzero only when $f = m/p$, and m is an integer. This indicates that there are discrete diffraction orders. The factor $(1 + e^{i\xi}e^{-itm/2})$ in Eq. (6.5) can be used many ways. When $\xi = -\pi$, this is the situation of PSM: the factor is 0, 1 + i , and 1 − i for $m = 0$, 1, and −1, respectively, assuming only three lowest diffraction orders are collected by the lens aperture. PGFM uses $\xi = -\pi/2$; the factor becomes 1 − i , 0, and 2 for the same respective values of m .

The wafer image is also a grating. Since the PGFM image is formed by the 0 and −1 orders (instead of the +1 and −1 orders), the image propagates obliquely along the optical axis. The intensity distribution of the grating image is given by

$$I = |A_0|^2 + |A_{-1}|^2 + 2 \operatorname{Re} [A_0 A_{-1}^* e^{i(Gx + \Delta qz)}], \quad (6.6)$$

where we used $I = |\mathbf{E}_0 + \mathbf{E}_{-1}|^2$, $\mathbf{E}_0 = A_0 e^{i\mathbf{k}_0 \cdot \mathbf{r}} \hat{\mathbf{y}}$, $\mathbf{E}_{-1} = A_{-1} e^{i\mathbf{k}_{-1} \cdot \mathbf{r}} \hat{\mathbf{y}}$, $\mathbf{k}_0 = k \hat{\mathbf{z}}$, $k = 2\pi/\lambda$, $\mathbf{k}_{-1} = \mathbf{k}_0 - \mathbf{G} - \Delta \mathbf{q}$, $\mathbf{G} = G \hat{\mathbf{x}}$, $G = 2\pi/p$, $\Delta \mathbf{q} = \Delta q \hat{\mathbf{z}}$, $\Delta q = k - \sqrt{k^2 - G^2}$. A_0 and A_{-1} denote the amplitude and phase of the diffraction orders corresponding to $f = 0$ and $-1/p$, respectively. The vectors \mathbf{E}_0 , \mathbf{E}_{-1} , \mathbf{G} , $\Delta \mathbf{q}$, and the angle θ are depicted in Fig. 6.18.

Let ϕ_0 be the phase of A_0 (i.e., $A_0 = |A_0| e^{i\phi_0}$, etc.). Then by requiring $Gx + \Delta qz + \phi_0 - \phi_1$ to be a constant, the sensitivity of PGFM ($\Delta x/\Delta z$) is $\Delta q/G$. Note that $G/k = \sin\theta$, where θ is the angle between the 0 and −1 order beams. Hence, the sensitivity of PSGM is $\tan(\theta/2)$. The maximum possible sensitivity that can be achieved theoretically is $\tan(\sin^{-1}NA/2)$ of a dry lens, or simply the tangent of half of the aperture angle for an immersion or a dry lens.

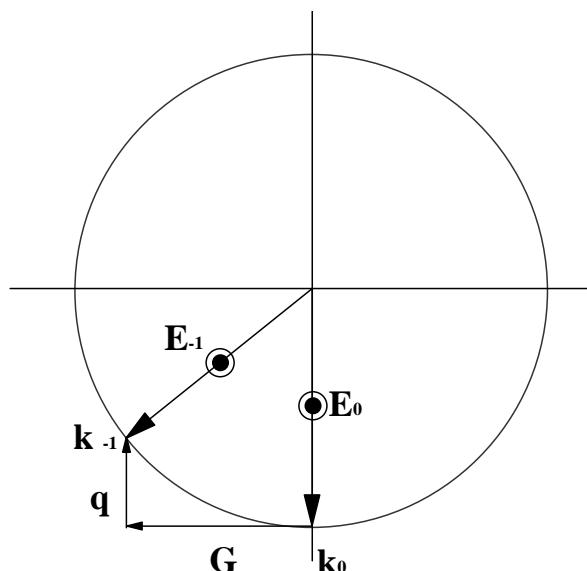


Figure 6.18 The vectors \mathbf{E}_0 , \mathbf{E}_{-1} , \mathbf{G} , \mathbf{q} , and the angle θ .

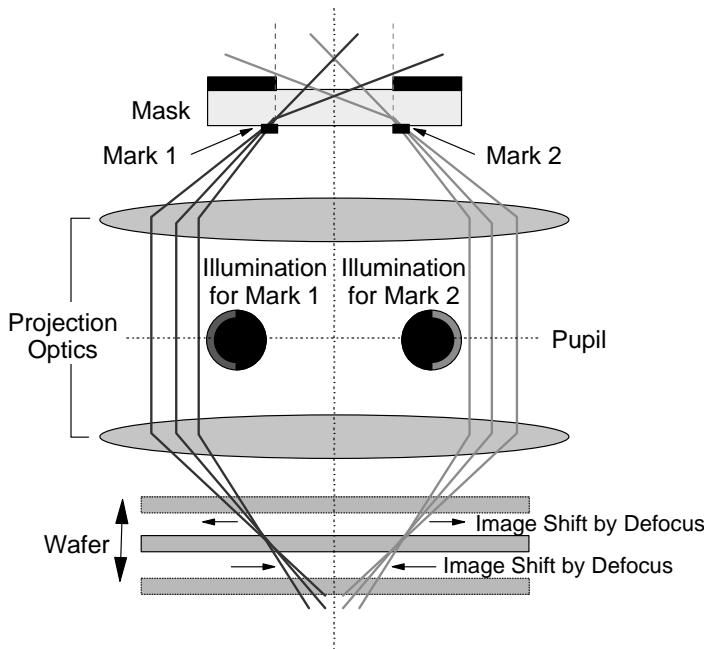


Figure 6.19 Marks on a double-sided chrome mask for one-beam imaging.

In addition to PSFM and PGFM, single-beam off-axis illumination (SB-OAI) to achieve the highest degree of image obliquity has been proposed.¹¹ This can be realized with an opening at the upper side of a double-sided chrome mask, along with the conventional overlay mark at the lower side of the same mask, located with its center directly beneath the edge of the opening. Therefore, the overlay mark sees only one half of the employed illumination, as shown in Fig. 6.19. Note that PSM is not needed. The overlay mark is much larger than the wavelength. It is applicable to ArF, KrF, i-line, and any other wavelengths used in microlithography. In addition, because the methodology is impervious to lens aberrations, the focus response is nearly linear. Due to the fact that a small portion of the projection lens is utilized, the effect of lens aberrations can be embedded in the calibration curve.

PFSM, PGFM, and SB-OAI call for specific illumination; the former two require small- σ DKI, and the latter requires OAI. They cannot be used with production masks with a wide variety of illumination conditions. Optical CD (OCD) metrology, often referred to as scatterometric CD (SCD) metrology, can measure not only the lateral dimension, but also the profile of the grating image in terms of the side-wall angle (SWA), which changes monotonically with defocus near the best focus. Hence, it is promising to realize an in-situ BIM, highly precise, polarity unambiguous, one-shot focus-monitoring metrology, provided that there is sufficient room on the mask to accommodate the OCD test pattern.

Presently, the best focus-monitoring method is to use part of a length-sensitive pattern to identify the value of F in Eq. (6.4). This method allows

monitoring of the best focus for a combination of critical patterns in the real layout, λ , bandwidth, NA, σ , and OPC settings imaged with the production mask on production wafers.

6.1.5 Exposure tool throughput management

The throughput of an exposure tool is very important because it has a direct impact on the economy of manufacturing. The per-exposure cost is usually calculated from the cost of the tool divided by the available hours of the tool within the depreciation period, and by the wafers-per-hour throughput:

$$\text{Cost of Ownership} = \frac{P}{N \times U \times A \times W}, \quad (6.7)$$

where P is the capital cost of the tool, N is the number of hours within the depreciation period, U is utilization, A is availability, and W is wafer-per-hour throughput.

A major component of wafer throughput is the exposure time. For a stepper, it is the time when the reticleⁱ and the wafer are static and the shutter is open. For a scanner, strictly speaking, the exposure time is the time during which the reticle and wafer are scanned in synchronization and the shutter is open. However, time is needed for the stages to accelerate to the full scanning speed before the shutter is opened, then decelerate in preparation for wafer-stage stepping. For convenience of calculation, the time during which both the reticle and the wafer stages are moving is classified as the exposure scanning time. The time during which the reticle remains stationary while the wafer is stepped into a new position for scanning is classified as the stepping time.

Figure 6.20 shows the wafer stage position, velocity, and acceleration of an early stepper in the 1980s. Note that for modern steppers, the stage velocity can be two orders of magnitude higher, and acceleration is in the m/sec² regime. For the early stepper, the wafer stage is accelerated from a static position to reach a constant velocity, then decelerated to zero velocity. For modern steppers, the acceleration does not stay constant; the jerk is in the order of hundreds of m/sec³. For scanner operation, both the reticle and the wafer follow similar motions, except for mx speed magnification at the reticle stage, where m is the reduction ratio of the image lens. To optimize the scanner exposure time, v must be high. It is inversely proportional to the width of the scanning slot and the illumination intensity. Also, a must be very high, such that the time during acceleration and deceleration is a small portion of the time. For stepper operation, it is more important to take the wafer stage to its exposure position as quickly as possible. The total time of movement must be minimized with the best combination of stage acceleration and velocity.

ⁱ The word “reticle” is often used for a mask used in an exposure tool. Mask is a more generic term. These words have been used interchangeably by many authors.

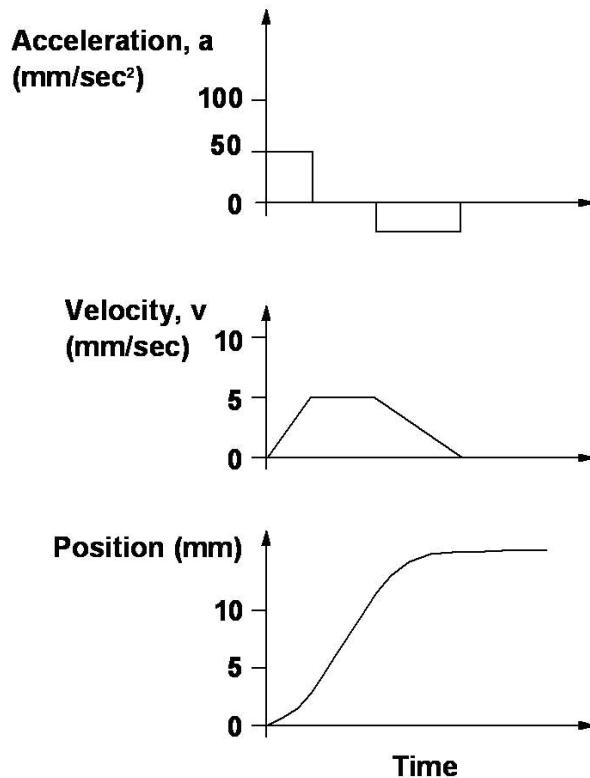


Figure 6.20 Position, velocity, and acceleration of an early wafer stage.

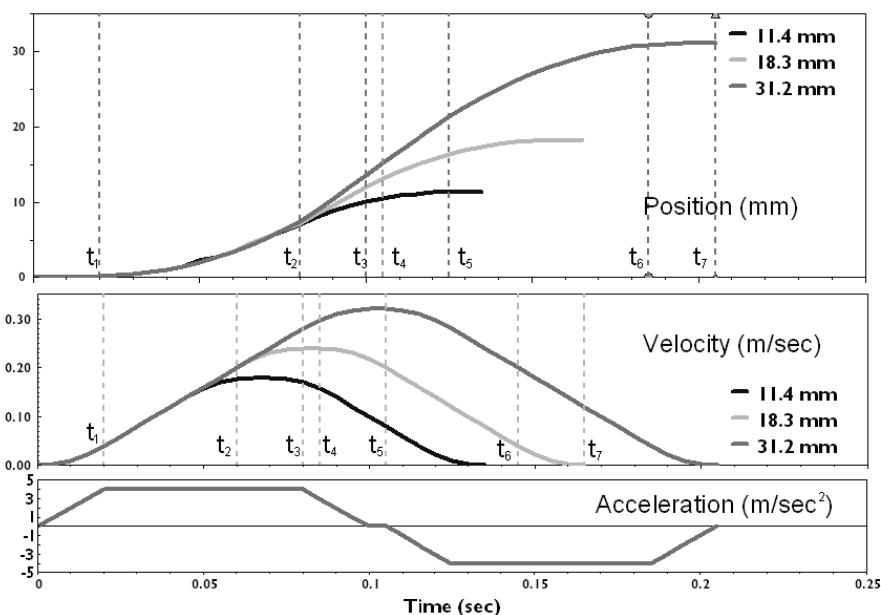


Figure 6.21 Position, velocity, and acceleration of a modern wafer stage.

Figure 6.21 shows the stage position, velocity, and acceleration of a deep-UV stepper. Its maximum stage velocity is $v = 0.32$ m/sec, maximum acceleration is $a = 4$ m/sec², and jerk is $h = 200$ m/sec³. The curves for three stepping distances that the stage traveled—11.4 mm, 18.3 mm, 31.2 mm—are shown. The equations used are as follows:

When $0 \leq t \leq t_1$,

$$\begin{aligned} a &= ht, & a_1 &= ht_1 \\ v &= \int_0^t a dt = \frac{1}{2} ht^2, & v_1 &= \frac{1}{2} ht_1^2 \\ s &= \int_0^t v dt = \int_0^t \frac{1}{2} ht^2 dt = \frac{1}{6} ht^3, & s_1 &= \frac{1}{6} ht_1^3. \end{aligned} \quad (6.8)$$

When $t_1 \leq t \leq t_2$,

$$\begin{aligned} a &= ht_1 = a_1 = a_2 \\ v &= v_1 + ht_1(t - t_1), & v_2 &= v_1 + ht_1(t_2 - t_1) \\ s &= s_1 + v_1(t - t_1) + \frac{1}{2} ht_1(t - t_1)^2, & s_2 &= s_1 + v_1(t_2 - t_1) + \frac{1}{2} ht_1(t_2 - t_1)^2. \end{aligned} \quad (6.9)$$

When $t_2 < t \leq t_3$,

$$\begin{aligned} a &= ht_1 - h(t - t_2), & a_3 &= ht_1 - h(t_3 - t_2) \\ v &= v_2 + ht_1(t - t_2) - \frac{1}{2} h(t - t_2)^2, & v_3 &= v_2 + ht_1(t_3 - t_2) - \frac{1}{2} h(t_3 - t_2)^2 \\ s &= s_2 + v_2(t - t_2) + \frac{1}{2} ht_1(t - t_2)^2 - \frac{1}{6} h(t - t_2)^3, & & \\ s_3 &= s_2 + v_2(t_3 - t_2) + \frac{1}{2} ht_1(t_3 - t_2)^2 - \frac{1}{6} h(t_3 - t_2)^3. & & \end{aligned} \quad (6.10)$$

When $t_3 < t \leq t_4$,

$$\begin{aligned} a &= 0, & a_4 &= 0 \\ v &= v_3, & v_4 &= v_3 \\ s &= s_3 + v_3(t - t_3), & s_4 &= s_3 + v_3(t_4 - t_3). \end{aligned} \quad (6.11)$$

When $t_4 < t \leq t_5$,

$$\begin{aligned}
 a &= -h(t - t_4), & a_5 &= -h(t_5 - t_4) \\
 v &= v_4 - \frac{1}{2}h(t - t_4)^2, & v_5 &= v_4 - \frac{1}{2}h(t_5 - t_4)^2 \\
 s &= s_4 + v_4(t - t_4) - \frac{1}{6}h(t - t_4)^3, & s_5 &= s_4 + v_4(t_5 - t_4) - \frac{1}{6}h(t_5 - t_4)^3.
 \end{aligned} \tag{6.12}$$

When $t_5 < t \leq t_6$,

$$\begin{aligned}
 a &= -h(t_5 - t_4) = a_5 = a_6 \\
 v &= v_5 + a_5(t - t_5), & v_6 &= v_5 + a_5(t_6 - t_5) \\
 s &= s_5 + v_5(t - t_5) + \frac{1}{2}a_5(t - t_5)^2, & s_6 &= s_5 + v_5(t_6 - t_5) + \frac{1}{2}a_5(t_6 - t_5)^2.
 \end{aligned} \tag{6.13}$$

When $t_6 < t \leq t_7$,

$$\begin{aligned}
 a &= a_6 + h(t - t_6), & a_7 &= a_6 + h(t_7 - t_6) \\
 v &= v_6 + a_6(t - t_6) + \frac{1}{2}h(t - t_6)^2, & v_7 &= v_6 + a_6(t_7 - t_6) + \frac{1}{2}h(t_7 - t_6)^2 \\
 s &= s_6 + v_6(t - t_6) + \frac{1}{2}a_6(t - t_6)^2 + \frac{1}{6}h(t - t_6)^3, \\
 s_7 &= s_6 + v_6(t_7 - t_6) + \frac{1}{2}a_6(t_7 - t_6)^2 + \frac{1}{6}h(t_7 - t_6)^3,
 \end{aligned} \tag{6.14}$$

where t_1 , t_2 , etc. mark the time that a changes. The stepping time at various step sizes is calculated by the above equations and shown in Fig. 6.22.

The 31.2-mm stepping distance in Fig. 6.21 is the minimum distance that the maximum velocity is reached. Time must be allowed to decelerate the stage to zero velocity for exposure. Actually, this stepping distance already exceeds the

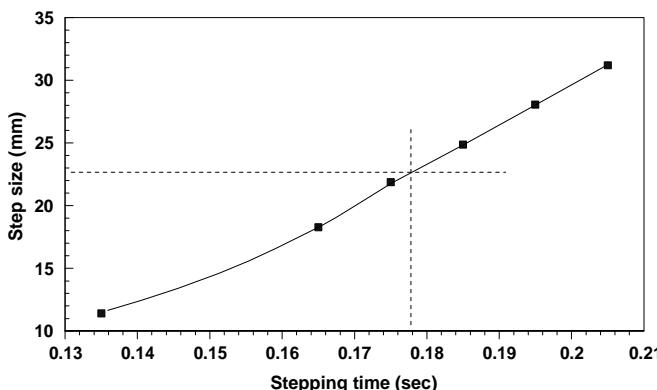


Figure 6.22 Calculated step size versus stepping time.

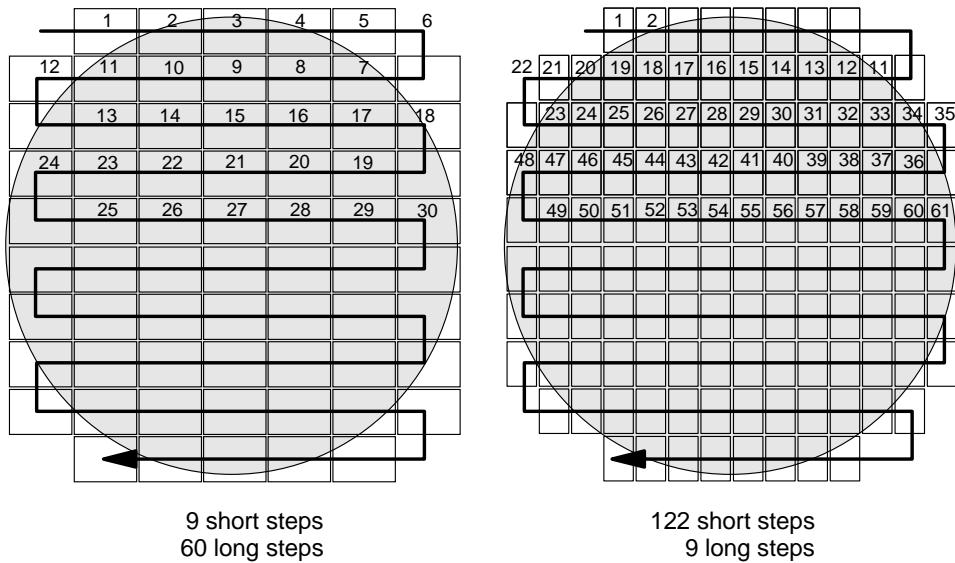


Figure 6.23 Exposure steps from different field sizes.

field size of the imaging lens at 22 mm^2 . For the 18.3-mm stepping distance, only 0.24 m/sec is reached before the stage must be decelerated. It is even shorter for the 11.4 mm distance of 0.18 m/sec. This means that a larger step size makes the total wafer-stepping time shorter. The average speed of the three stepping distances is 0.152 m/sec, 0.111 m/sec, and 0.0846 m/sec, respectively. Hence, for high throughput, the area covered in each exposure step should be as large as possible. Figure 6.23 shows a wafer exposed with a $11.4 \times 18.3 \text{ mm}^2$ area and a $22.8 \times 18.3 \text{ mm}^2$ area. The former area requires 112 steps of a 11.4-mm distance and 9 steps of a 18.3-mm distance. The total stepping time for this wafer is 17.955 sec. With the double-area field, the total stepping time is only 12.165 sec.

Next, let us look at the proper orientation of a rectangular exposure field for throughput. Even though the average speed stepping through the shorter side is lower than through the longer side, the total stepping time depends on the distribution of long and short steps. Figure 6.24 shows the same wafer layout exposed in two different routing sequences. The routing in Fig. 6.24(a) consists of 13 short steps and 116 long steps, whereas the routing in Fig. 6.24(b) consists of 122 of the former and nine of the latter. The total stepping time for (a) is 21.03 sec, and for (b) it is 17.955 sec.

Figure 6.25(a) shows¹² the velocity and acceleration of the wafer and the reticle stages specific for scanner operation. Jerk is introduced to speed up acceleration in order to minimize the acceleration and deceleration times. Figure 6.25(b) shows the change of direction of the wafer stage after each exposure. It also shows the wafer stepping in x taking place while the wafer and reticle stages are still decelerating in y after the scanning exposure. This way, the wafer-stepping time is absorbed in the scanning time.

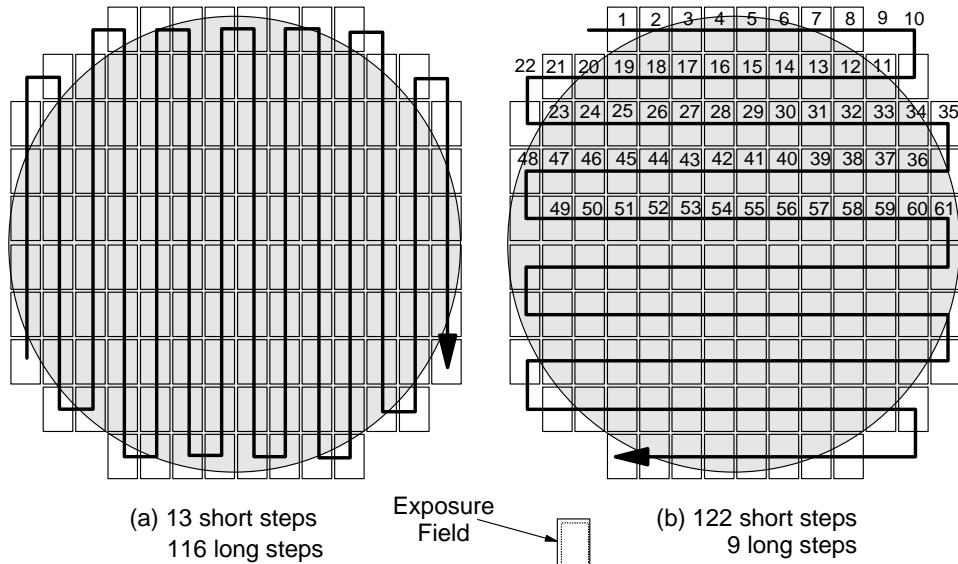


Figure 6.24 Two routing schemes.

Because the lens is a very expensive part of the exposure tool, letting it idle while the wafer stage moves away for alignment and wafer flatness measurement is wasteful. To get around this problem, tools with two wafer stages are built. One wafer stage moves under the lens for exposure, while the other stage moves in the metrology area outside of the lens to measure the position of the alignment marks and the topography of the wafer surface. The key challenge of such a twin-stage system is to keep the two stages under laser interferometer surveillance without interruption.^{13,14} Figure 6.26 shows how such a twin-stage system works. Wafer 1 is characterized while wafer 2 is exposed. After exposure, both wafer stages are moved all of the way toward the top edge of the characterization area so that wafer stage 2 can move out along the loading/unloading line for wafer unloading. Stage 1 now moves to the exposure position. While it is stepping for exposure, stage 2 is loaded with a wafer and comes into the characterization position to start topography and alignment measurements. Once the wafer is characterized, the laser-beam surveillance is unbroken. At least one vertical beam and two horizontal beams take turns to monitor the position of the wafer stage until the wafer stage is moved out for unloading.

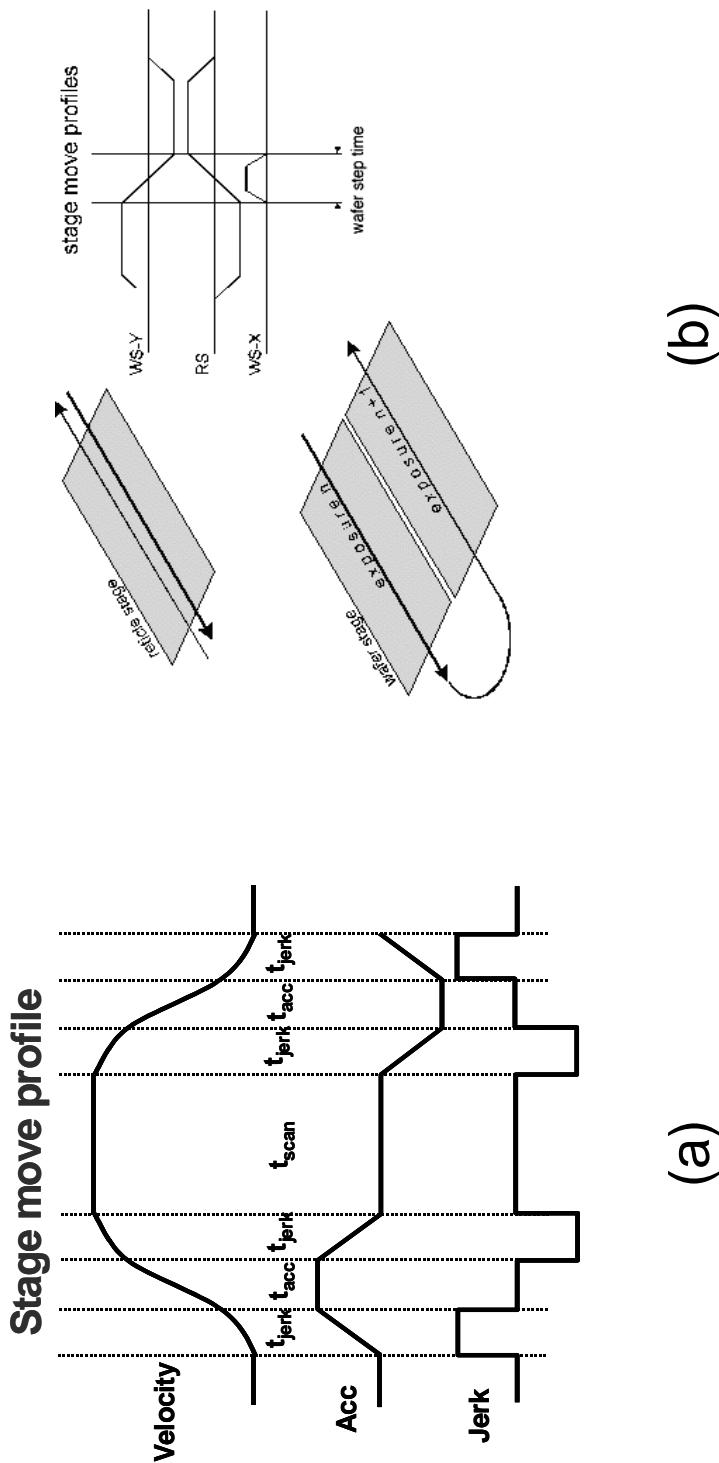


Figure 6.25 Velocity and acceleration of the step-and-scan wafer and reticle stages.

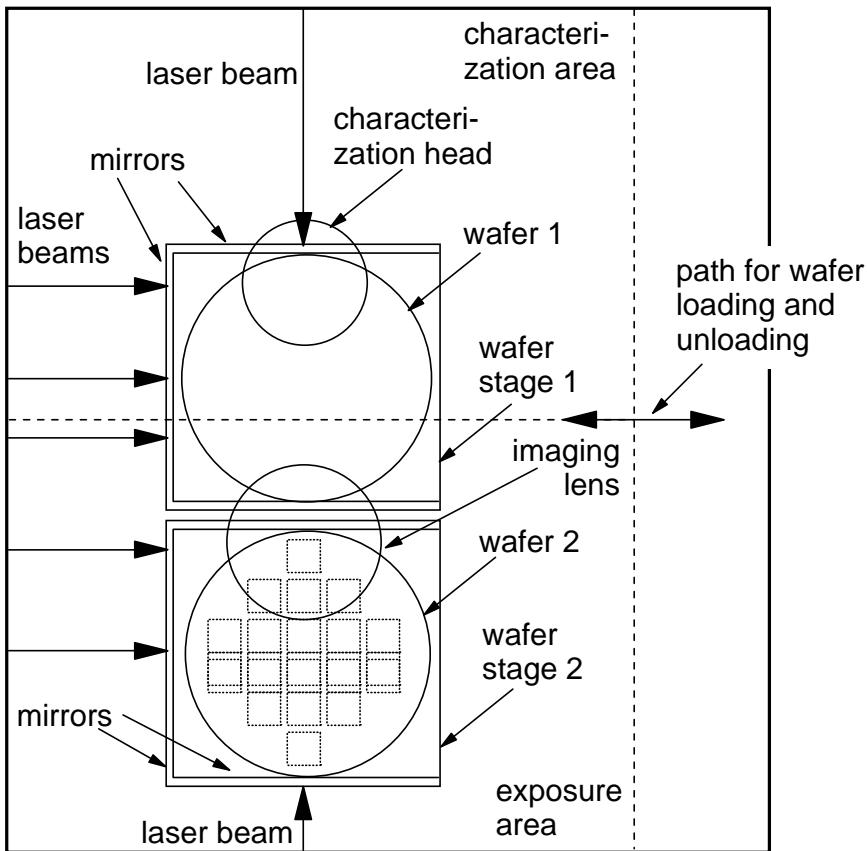


Figure 6.26 Twin-stage exposure station.

6.2 Resist Processing

The optimization of resist processing includes resist coating, various types of resist baking, and resist developing.

6.2.1 Resist coating

The most popular way to coat the photoresist is by spin coating. The wafer, after being primed with an adhesion promoter, is held by vacuum on a spin chuck. Resist is applied to the wafer while it is stationary or slowly spinning. After a sufficiently large puddle is formed on the wafer, the chuck starts accelerating to a constant spin speed between several hundred and several thousand revolutions per minute (RPM). Spinning stops after the resist thickness and uniformity stabilizes. Then, the wafer is sent for prebaking. Edge bead removal is usually performed while the wafer is still spinning. We will discuss edge beads further in Sec. 6.2.1.4.

A good resist coating has the following properties: high thickness accuracy and uniformity, good surface planarity, high adhesion to the substrate, and low defects.

6.2.1.1 Defects

The first step in achieving low defects is to have a thoroughly clean substrate surface. It is best to protect the surface from contamination instead of cleaning it prior to the coating process. If there is a nonwetting particle on the surface, it will induce a pinhole in the resist. A particle that is coated over usually shows comet tails toward the circumstance of the wafer, if a spin-coating process is used. It perturbs local exposure because of multiple reflections off the nonuniform resist surfaces. The particle itself introduces a different reflectivity from that of the substrate. If the particle stays on after resist development and stripping, it blocks the ensuing etching.

6.2.1.2 Resist adhesion

The photoresist polymer generally does not adhere well to Si-based substrates and many other types of solid films used in the semiconductor industry. Even though the theory of adhesion is out of the scope of this book, we will shed some light on surface tension. The surface tension of the resist and that of the substrate cannot differ by 20 dyne/cm² or more in order for the resist to adequately wet the wafer.¹⁵ Good wetting is necessary for good adhesion, but does not absolutely dictate it. We can also reason that water-wetting hydrophilic surfaces adhere well to other hydrophilic surfaces, but not to water-repelling hydrophobic ones, and vice versa. This is supported by the observation of the Langmuir-Blodget technique,^{16,17} in which the coating is built up by pulling the substrate on its side from a monolayer film suspended in a tank. By repetitively pulling and dipping the substrate from and into the tank, the monolayer wraps itself back-to-back and front-to-front to build up a substantially thick coating. One side of the monolayer is hydrophobic while the other side is hydrophilic, suggesting that the same type of surface is more likely to adhere to each other.

The wafer surface must be treated with an adhesion promoter before resist coating. Hexamethyldisilazane (HMDS) is a popular adhesion promoter. It is most effectively applied in a heated vacuum chamber using an inert carrier gas. Applying HMDS in the liquid phase has also been practiced, but it is less effective than gas-phase application primarily because it is easier to elevate the reaction temperature in the gas phase.

6.2.1.3 Resist thickness

The spun resist thickness is related to the polymer concentration C , molecular weight M , and spin speed ω by the following equation:¹⁸

$$thk \propto \frac{C^\beta M^\gamma}{\omega^\alpha}, \quad (6.15)$$

where β and γ are usually close to 2, and α is close to 0.5. Therefore, the resist concentration is usually used to change the resist thickness grossly and the spin speed is used to fine tune it.

6.2.1.4 Resist uniformity

The resist coating process is generally considered planarizing rather than conforming. This is true only for small features on the wafer. If there are large features, the resist over these features has an identical thickness to the resist over the wafer background, as shown in Fig. 6.27. Let the resist thickness over a planar wafer be a , then thickness a is maintained on large planar regions on the wafer. With a high feature density, the situation is similar. Thickness a is maintained from the top of the topography. Unfortunately, the resist thickness is now $a + b$ at the bottom of the topography, where b is the height of the topography. At a medium feature density, the resist is partially planarizing while its thickness on the top of the topography is less than a . For a totally isolated small feature, the resist thickness at its top may be only slightly larger than $a - b$.

The presence of topography on the wafer further complicates film planarity from radial resist pile-up against protrusions. The most significant impact of these pile-ups is alignment mark detection errors.

Resist significantly accumulates at the edge of the wafer. This is generally called “edge bead.” Its thickness can be many times more than normal, usually a few millimeters. Edge bead is undesirable in many ways. The extra resist tends to go under the wafer near the edge, making it difficult to attain uniform contact for wafer chucking or heat transfer. In addition, the extra thickness cannot support good imaging at the edge of the wafer. It also prolongs the resist drying and stripping times. In addition, it tends to contact the wall of wafer carriers or robotic wafer grips. Some resist may break loose and become a defect source. For full-wafer contact printing, a thick edge bead prohibits a uniform full-wafer contact to the mask. Edge beads can be removed in a modern resist coater. A small knife-edge jet of solvent is applied to the edge of the wafer while it spins to remove the resist at the edge of the wafer.

Striation in the resist film is exemplified by irregular radial lines that are easily seen by examining the resist surface in a reflection. It is usually caused by a rapid loss of solvent, either due to the wrong solvent or airflow management that can affect film uniformity in many other ways.

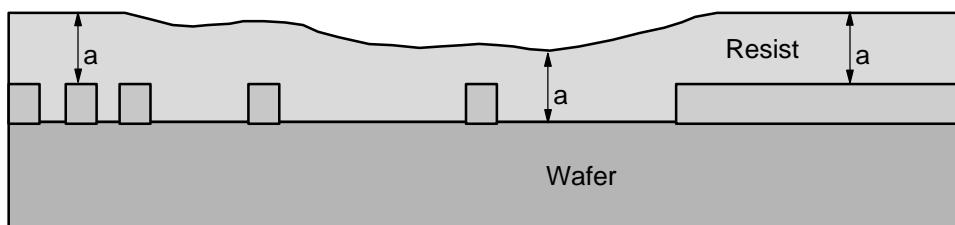


Figure 6.27 Resist planarization over topography.

6.2.1.5 Saving resist material

Resist coating is a very wasteful process. The puddle formed on the wafer prior to spinning is usually about half of the diameter of the wafer, and its thickness is on the order of 1 mm. Most of the material is discarded by the centrifugal force. Only a layer on the order of 1 μm is left for imaging. For imaging at a sub-0.1- μm resolution, the resist thickness can be 100 nm or less. Take a 200-mm diameter wafer, the material applied to the wafer is about 8 ml, and the resultant material on the wafer is about 32 nl. Only 1/250 is used. Many techniques have been employed to reduce the initial applied thickness to a few ml. There are still two orders of magnitude of waste. It is difficult to cover the entire wafer completely if the applied diameter is further reduced. Reducing the applied thickness results in a nonuniform spun thickness because the resist tends to form a high concentration layer, or even a skin at the top, before spinning takes place.

There are many reasons to save the resist material. In addition to the initial expense, there are costs associated with transportation, storage, and disposal of the excess material. The extra burden on the environment alone justifies a more efficient technique to coat the resist.

Many methods have been tried to deliver a distributed and thin coating to the wafer. Ausschnitt et al.¹⁹ use a vapor deposition technique to deposit a uniform layer of resist material on the wafer. However, it is not easy to simultaneously control the thickness and uniformity. Spraying can deposit a thin layer of polymer material to a surface, just as in spray painting. However, the uniformity is totally inadequate to meet semiconductor-manufacturing specs. Spraying followed by spinning suffers a solvent evaporation problem much worse than that in puddle-and-spin.

However, a spray-and-spin-in-vapor (SASIV) technique^{20,21} overcomes the above problems. The wafer is kept in a solvent-rich environment when the resist is sprayed on the wafer to achieve a thickness slightly larger than the target thickness. Excess material is spun off while still in vapor, which is removed after the resist planarizes over the topography while the wafer is stationary. This can result in savings by a factor of 50 or more over conventional methods and improve uniformity, especially for the resist piled up on protruding topography. Figure 6.28 shows the SASIV concept, which has been experimentally verified, as shown in Fig. 6.29. Using simple laboratory equipment, a 20X resist savings has been demonstrated, as shown in Fig. 6.30.

The essential sequence of SASIV is as follows:

- Induce solvent vapor to the wafer ambient.
- Spray resist on the wafer to a thickness slightly larger than the spun thickness in vapor.
- Wait 30 seconds until the stress in the resist induced by the spray is released in vapor.
- Spin for the desired resist thickness in vapor.

- Hold the wafer stationary in vapor to level the piled-up resist.
- Remove vapor.
- Bake wafer as usual.

The stress-releasing delay can be eliminated to enhance wafer throughput by using a proximity-dispense-in-vapor (PROXIV) technique.²² Instead of spraying, the resist is gently spread on the wafer by flowing through small nozzles under slight hydraulic pressure while the wafer chuck is slowly rotating. A solvent vapor environment is kept to prevent the thin resist layer from drying until it is spun off to reach the desired thickness.

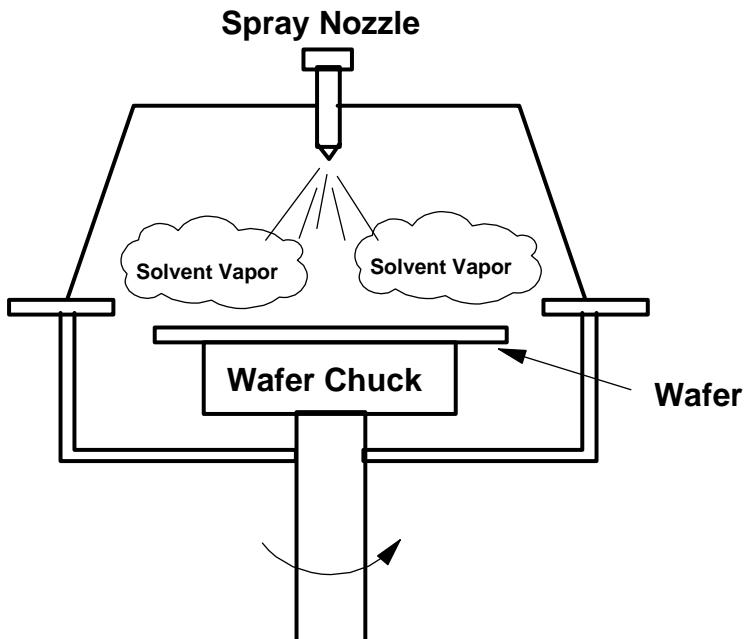


Figure 6.28 The SASIV technique.



Figure 6.29 The experimental apparatus for SASIV. (a) Vapor chamber closed to introduce vapor and resist. (b) Vapor chamber opened to load/unload wafer.

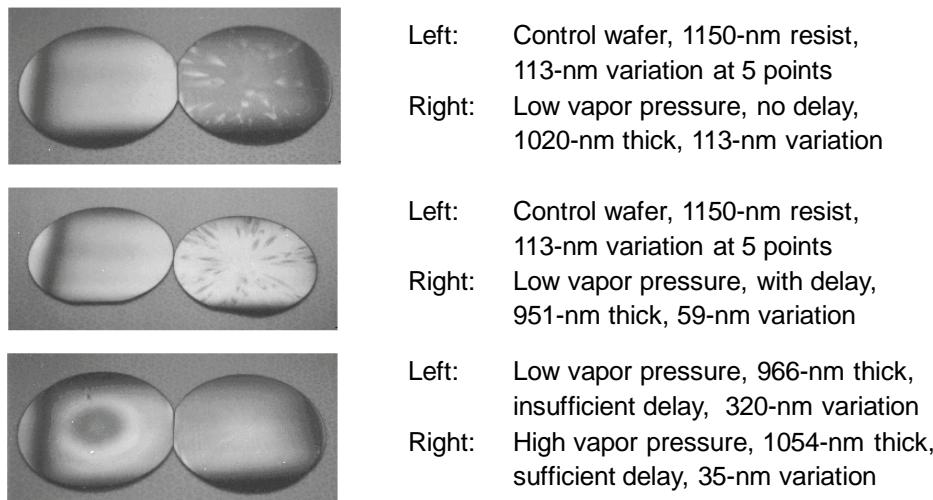


Figure 6.30 Resist thickness uniformity on wafer with/without vapor/delay. 1 ml of resist was needed to puddle a 3-in. wafer without SASIV. 0.05 ml was used with SASIV.

6.2.2 Resist baking

Baking serves many purposes in resist processing. The precoating bake improves adhesion by driving out surface-absorbed water at the wafer. The pre-exposure bake removes the resist solvent after coating, as well as fine tunes the photosensitivity. Diffusion of the photoactive compound at the postexposure bake before development helps to eliminate standing waves in the resist while diluting the lateral image contrast. With a chemically amplified resist system, the postexposure bake activates the catalyst reaction to convert the resist at the exposed areas. The postdevelopment bake hardens the resist to improve its adhesion in preparation for the substrate delineation processes to follow. Figure 6.31 shows the effect of temperature on a typical DQN resist.

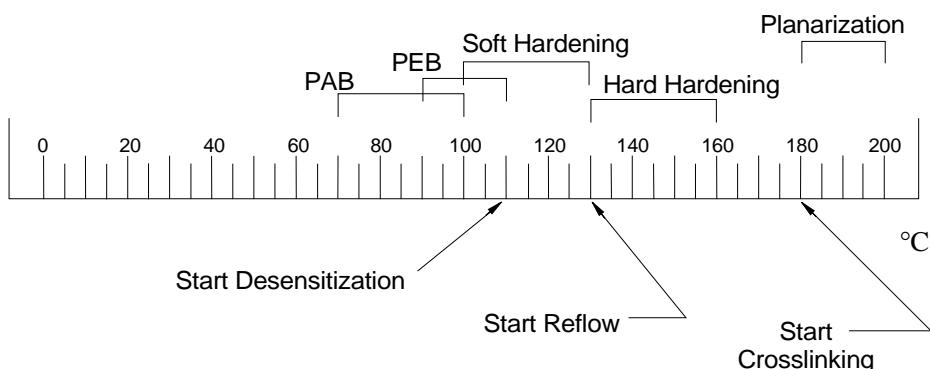


Figure 6.31 Effect of temperature on a typical DQN resist.

6.2.2.1 Precoating bake

The precoating bake should not be performed separately. Right after a SiO₂ film is grown from an oven (a metal film from the evaporator or sputtering machine, etc.), it is most desirable to prime the wafer with an adhesion promoter before the wafer is exposed to any water vapor. This not only ensures the freedom from water adsorption, but also eliminates a processing step.

6.2.2.2 Postapply bake (pre-exposure bake)

The most important function of the postapply bake (PAB) is to remove the photoresist solvent after spin coating. The baking temperature must be sufficiently high to remove the solvent in a reasonably short time. However, high temperatures may deactivate the photoactive compound and reduce the photoresist sensitivity. The development contrast of the resist is also affected. Figure 6.32 shows the resist thickness remaining as a function of PAB temperature and time.²³ The resist development rate at the same exposure dosage is reduced when a higher PAB temperature or a longer PAB time is used. However, the development contrast is improved with the higher PAB temperature.

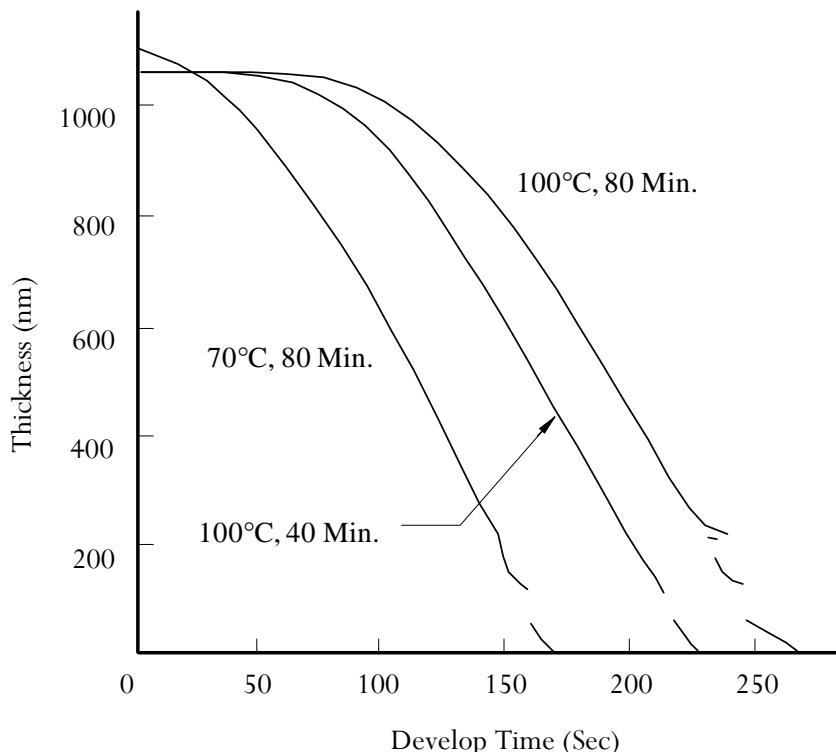


Figure 6.32 Effect of PAB to DQN resist-developing characteristics. (Reproduced with permission of *IBM Journal of Research and Development*.)

6.2.2.3 Postexposure bake

Initially, a postexposure bake (PEB) was used to remove the standing wave in the resist latent image, as shown in Fig. 5.58. When feature sizes are large compared to wavelength, removal of the standing waves is mostly cosmetic. With feature sizes reduced with respect to wavelength, i.e., k_1 becoming smaller, the PEB becomes critical. It reduces the lateral image contrast while the diffusion of the photoactive compound takes place in the longitudinal and lateral dimensions. With the advent of chemically amplified resists, PEB becomes a necessity. Fortunately, most chemically amplified resists have a high γ value compared to DQN resists, as was shown in Fig. 3.38, to make up for the loss of lateral contrast.

The PEB is usually the most critical processing parameter to control. Figure 6.33 shows 30 E-D trees constructed from five different types of edge under five different process tolerances. The five edges are ee'', aa'', bb'', cc'', and dd'', respectively, taken from a big-small-feature and Corner (BigMaC) test mask,²⁴ depicted in Fig. 6.77. A detailed description of the BigMaC mask will be given in Sec. 6.3.2.2. The processing conditions of the 30 E-D trees center at:

- Resist thickness = 910 nm
- PAB = 90° C
- PEB = 100° C
- PEB time = 100 sec
- Development time = 60 sec

Five E-D trees from five features are constructed using this central condition. Another five E-D trees are constructed using a thickness of 900 nm, but the other processing conditions are unchanged. Similarly, five more E-D trees are constructed with PAB = 92° C, PEB = 102° C, PEB time = 102 sec, and development time = 62 sec, respectively, resulting in the 30 E-D trees. The common E-D window of these trees is the actual usable E-D region for the feature combination and process tolerances. The two critical situations that bound the common E-D tree are (1) PEB = 102° C for edge ee'' as the top branch and (2) a 900-nm resist thickness for edge dd'' as the bottom branch. The PEB bound is more pronounced than the thickness bound.

To relax the PEB bound, one can pose a more restrictive condition on the PEB tolerance; for example, instead of a tolerance of $\pm 1^\circ$ C, reduce the tolerance to $\pm 0.5^\circ$ C or smaller. However, it may be difficult to control the baking temperature and uniformity at will. If the PEB tolerance must remain at $\pm 1^\circ$ C, then the position of edge ee'' can be adjusted to move the binding E-D tree upward for a larger window, as shown in Figs. 6.33 and 6.34. That is, optical proximity correction is used to compensate for PEB-induced errors. Gau et al.²⁵ went in a different direction—they used PEB to reduce the optical proximity effect.

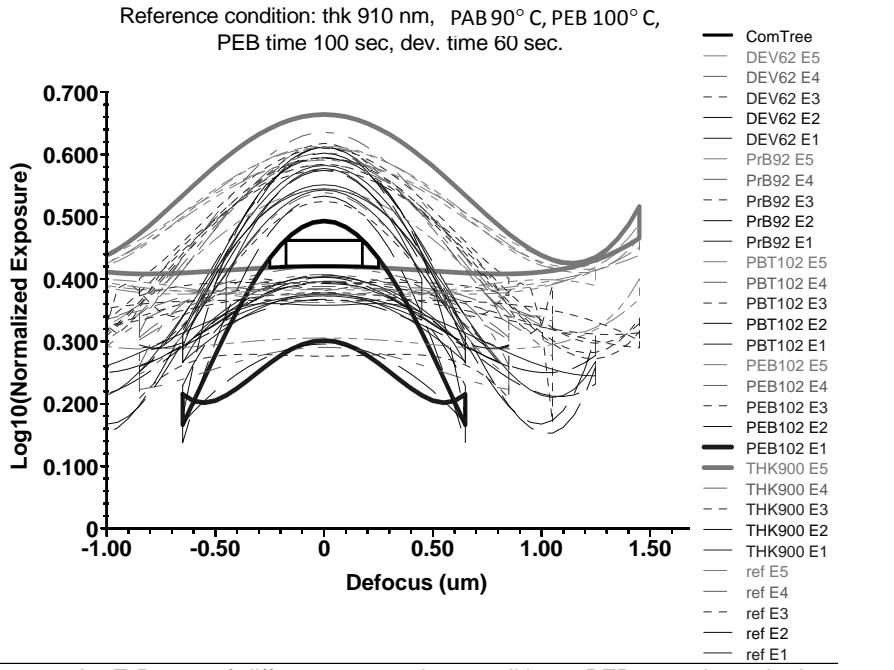


Figure 6.33 An E-D tree of different processing conditions. PEB102 edge 1 is the upper limit. THK900 edge 5 is the lower limit. The common E-D window is 10% EL and 0.35 μm DOF.

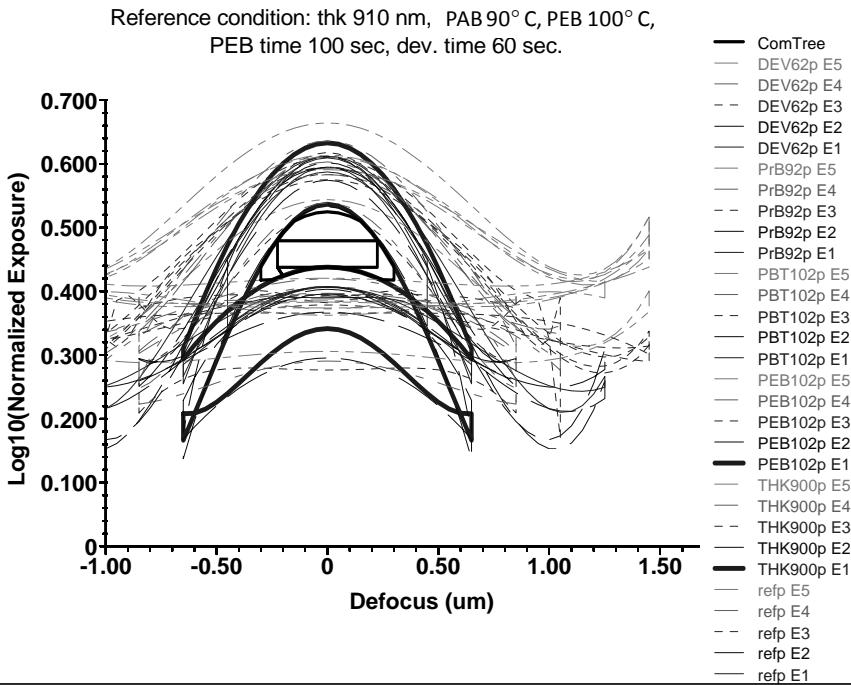


Figure 6.34 The PEB restriction in Fig. 6.33 eased with proximity correction. Edge 1 moved by 15 nm. The common E-D window is 10% EL and 0.45 μm DOF.

6.2.2.4 Hard bake

Hard bake is used after resist development to prepare the wafer for subsequent image-transfer processes. It serves to improve the adhesion of the resist and to harden it. The baking temperature is important. As it goes beyond the PEB temperature, the resist starts to desensitize. Further increases in temperature make the resist reflow. An even higher temperature crosslinks the novolak resin, rendering it chemically inert.

If the resist profile is to be preserved, the hard bake is usually carried out before the reflow temperature is reached. When a higher temperature is necessary, a hardening step must be included. Typical hardening steps consist of irradiation with deep-UV light²⁶ or treatment with a noneroding plasma.²⁷ These steps are effective for small features because either treatment hardens the skin of the resist image. When a wide resist line or a large area of resist is hardened, the treatment must be prolonged to ensure complete penetration throughout the entire thickness of the resist. This may be impractical for a resist thicker than 2 µm, especially for applications in fabricating magnetic read-write heads for disk memory. A mold-hardening technique^{28,29} completely hardens the resist in the bulk. The hardened resist can be taken to complete carbonization at temperatures exceeding 500° C in a nitrogen environment, without distorting the resist image profile. The mold-hardening process is shown in Fig. 6.35 and the result is shown in Fig. 6.36. After the DQN resist image is normally delineated, polymethylmethacrylate (PMMA) is spin-coated to cover the resist image completely, using a thickness larger than that of the resist image. This PMMA layer holds the DQN image to evenly distribute the stress induced by baking. After the DQN resist is completely hardened at 200° C and above, the PMMA layer can be removed with solvent. The hardened DQN resist is robust for any subsequent process, including further baking in nitrogen ambient to temperatures exceeding 500° C.

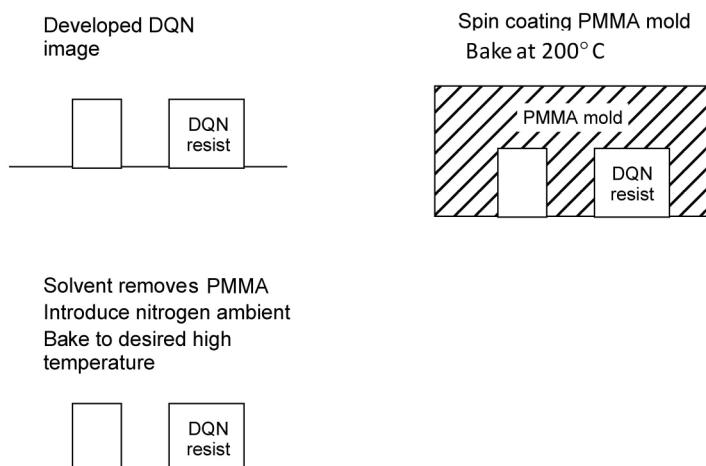


Figure 6.35 The mold-hardening process.

6.2.3 Resist developing

The resist-developing step is one of the most crucial steps in the resist-imaging procedure. If not carried out properly, the effort in getting the aerial and latent images would be wasted. The development process is governed by the resist dissolution rate, which is a function of resist sensitivity, absorption coefficient, exposure, and baking conditions. A typical rate versus exposure curve is seen in Fig. 6.37. It usually consists of low- and high-contrast regions. Complete depletion of the photoactive compound flattens the dissolution rate at extremely high exposures.

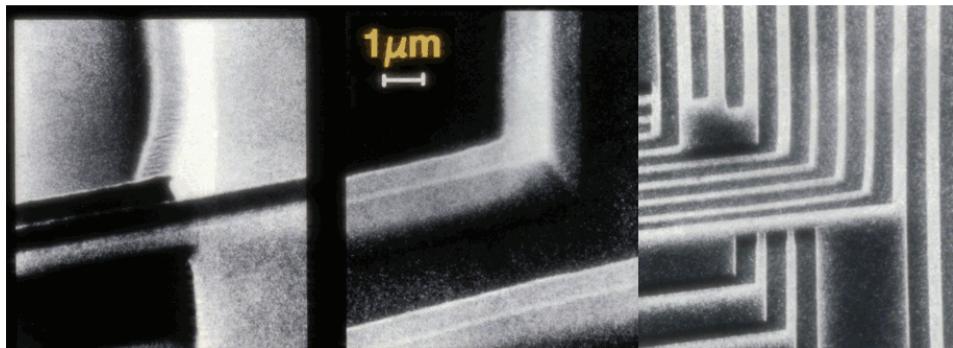


Figure 6.36 An AZ resist image after mold hardening.

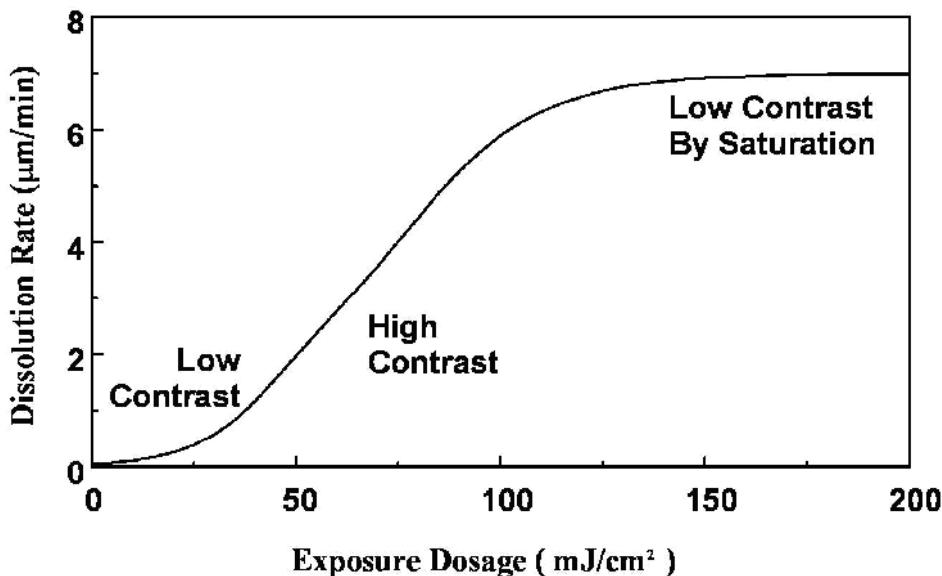


Figure 6.37 Bulk dissolution rate versus exposure of a typical DQN resist.

The developer removes the resist at the developer-resist interface according to the local dissolution rate. At the next instance, the developer is facing a new interface and continues to locally remove the resist according to the dissolution rate. This takes place in three dimensions in a way very similar to a diffusion process, except that diffusion has a new surface to diffuse into at each infinitesimal time increment. The development process is also likened to the wet-etching process, except that the latter is generally isotropic. The development phenomenon was illustrated in Figs. 5.59 through 5.62 by strictly applying the dissolution rate versus the new surface rule. An actually developed resist image is less optimistic because, in addition to the dissolution rate consideration, the wetting capability of the developer to the resist surface may decrease inside small openings, just as in partially developed contact holes or small line openings. The developer may also be locally depleted, depending on whether or not there is a large exposed area to be removed. This is a development proximity effect, similar to the etching proximity effect.

- *Immersion developing*—The wafer is entirely submerged in the developer. This is usually done with an entire batch of wafers. The developer environment is static. In some cases, the developer circulates in the developer tank to refresh. Nevertheless, circulation cannot overcome the problem of developer nonwetting. Incorporating a surfactant in the developer to reduce surface tension is desirable.

The immersion-developing process used to be the primary process in manufacturing. Due to the advent of track-processing systems, immersion development is giving way to the two other processes that are more adaptable to wafer-processing tracks.

- *Puddle developing*—A developer puddle is produced at the wafer in the developing station on a track. The developing station is similar to a resist-coating station using a wafer chuck that can spin to thousands of RPM. Instead of delivering resist to the wafer, the nozzle delivers developer instead. After the development endpoint is reached, the developer is spun off, and the wafer is spray rinsed with deionized water at medium speed. The wafer is completely spun dry at high speed to complete the process.

Puddle developing is the track-adaptable form of immersion development. Thus, developer nonwetting and development-proximity effects still apply. The differences between immersion and puddle developments are:

1. Each wafer must be processed separately, increasing the total processing time of a wafer batch.
2. Puddle development wastes developer because the puddle is spun off and is not reused.
3. The development-proximity effect cannot be overcome with circulation as in the case of immersion development.

- **Spray developing**—The wafer is developed in a developing station on the wafer track just as in puddle development. However, the developer is sprayed vigorously onto the wafer while the chuck is spinning at high speed, thus refreshing itself. If the spray has sufficient energy, surface tension can be reduced, making better wetting of the developer to the resist. Hence, spray developing has the ingredients of high-quality developing. The tradeoff of spray development is its excess use of developer, which creates cost and chemical-disposal problems.

Because of the tradeoffs of these three development methods, there is no clear choice. Each must be adopted by optimizing the balance of performance versus economy. The need for good surfactants is obvious. One important requirement of the surfactant, in addition to chemical compatibility to the developer and the resist, is that it can be completely removed, leaving no trace residue to affect the performance of the finished electronic device. Mariott³⁰ reported the resist contrast in terms of exposure versus developing time, as shown in Fig. 6.38. One can see that spray developing produces the highest development contrast, followed by puddle developing.

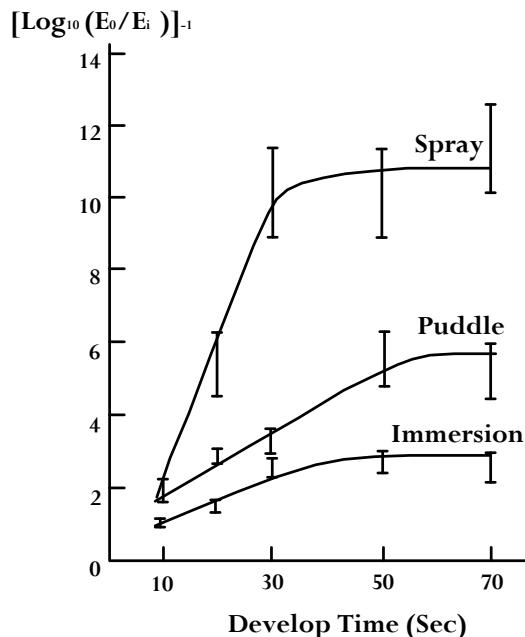


Figure 6.38 Effect on resist contrast by method of development. (Image courtesy of V. Mariott.)

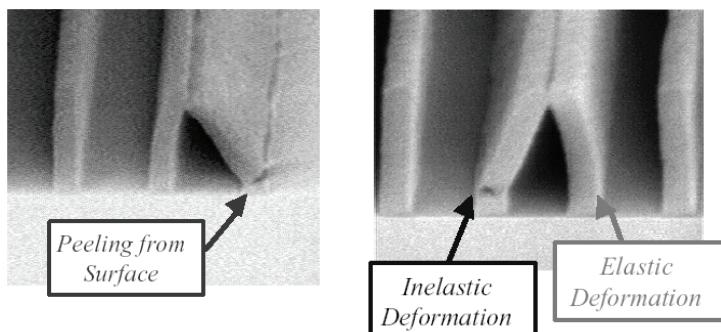


Figure 6.39 Resist collapse during liquid removal.

6.2.4 Aspect ratio of resist image

After developing the resist during the drying process, liquid is removed from the resist image. As it leaves the resist image, the capillary force pulls adjacent features together.³¹ If the adhesion of the resist to the substrate is insufficient, the resist peels. Alternately, if the resist is not sufficiently rigid, it bends or collapses by elastic and inelastic deformations, respectively, as shown in Fig. 6.39.

Note that only line pairs or the outside lines of a group collapse. The capillary pull on the inside lines is balanced at both sides of the lines; thus, the forces cancel out. The pressure causing the resist collapse is described by the following equation.³² This pressure P is asserted to the area $h \times l$, where h is the height of the resist and l is its length.

$$P = \frac{2\sigma \cos \theta}{d}, \quad (6.16)$$

where σ is the surface tension of the rinsing liquid, θ is the contact angle of the rinsing liquid, and d is the distance between two resist lines, as shown in Fig. 6.40. The pressure on the resist sidewall increases with the surface tension of the rinsing liquid and the sidewall angle of the rinsing liquid to the resist. The former is related to the liquid itself; the latter is related to the hydrophilicity of the liquid. When two resist lines are closer to each other, d is small. P becomes larger.

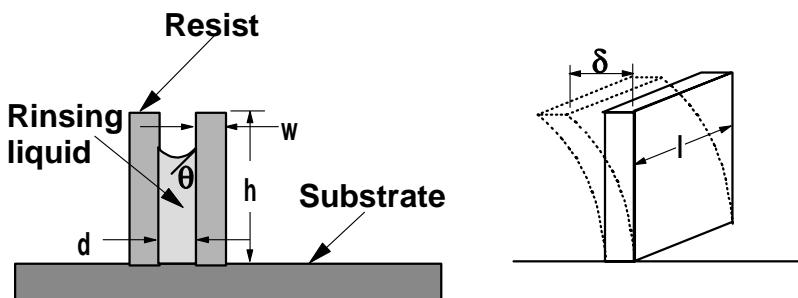


Figure 6.40 Contact angle and radius of curvature of rinsing liquid. (Image courtesy of The Japan Society of Applied Physics, reproduced from Ref. 30.)

The displacement δ of the resist is given by combining Eqs. (4) to (7) in Ref. 34:

$$\delta = \frac{3\sigma \cos \theta}{E \times d} \left(\frac{h}{w} \right)^3 h. \quad (6.17)$$

When δ exceeds a limit, the resist peels, breaks, or bends sufficiently to touch a neighboring line. It collapses. From Eq. (6.17), the onset of collapse is proportional to the surface tension and cosine of the contact angle, and inversely proportional to Young's modulus E . The onset is also proportional to the third power of the resist height-to-width aspect ratio modified by an additional dependence on the absolute value at the resist height.

It is very unusual that a resist image can exceed a 3:1 height-to-width aspect ratio without collapse. This ratio has been used as the rule of thumb to determine the resist thickness as a function of the minimum feature size. Therefore, the resist thickness decreases as the feature size is reduced for each technology node. The etching technology must be developed to handle the smaller resist thickness.

6.2.5 Environmental contamination

Chemically amplified resists (CAR) are extremely sensitive to alkaline-containing vapor in the environment. This is because very few protons are multiplied to many more protons to affect its dissolution characteristics during PEB. If the environment has trace amounts of alkaline-containing vapor, such as amine, after exposure and before PEB, any annihilation of protons is amplified. The time amount is on the order of parts per billion. Therefore, in the CAR processing area, heavy filtration of amines by charcoal filters must be implemented, requiring significant space and cost. In the early days of CAR development, the resist exposure-development characteristic was erratic and by no means manufacture-worthy until the root cause was identified³³ and solutions implemented.

6.3 k_1 Reduction

From the scaling equations presented in Sec. 4.1, the approaches to improving resolution and DOF are clear. For resolution, the approaches are to increase NA and reduce λ and k_1 . Equations (7.3) and (7.7) further point to the increase of the coupling medium's refractive index from the lens to the resist. Mathematically, this increase can be imbedded in NA and wavelength changes. For DOF, the scaling equations alert us to the tradeoff between resolution and DOF with changes in NA and λ . Reduction of k_1 is unique in the sense that it can improve resolution and DOF together. Because the imaging tool and imaging medium do not need to change, k_1 reduction is also potentially an economical solution. The tradeoff in k_1 reduction is the increase of the mask error enhancement factor (MEEF). Hence, k_1 is an indicator of the processing capability of a manufacturing

site and its potential to cut cost. However, a true measure of manufacturing prowess is k_1/cost . This way, one need not blindly pursue the lowest k_1 among all competitors; MEEF and yield are also included in the considerations. k_1 reduction is often referred to as the resolution enhancement technique (RET). However, the scope of k_1 reduction is much larger than resolution enhancement.

In the following sections, several k_1 -reduction techniques are covered, namely, phase-shifting masks, off-axis illumination, scattering bars, and optical proximity corrections. Several approaches can be taken before starting these k_1 -reduction techniques. Optimization of NA and illumination, covered in Secs. 6.1.1 and 6.1.2, must be used to realize the k_1 potential of a given imaging system before implementing k_1 reduction techniques. This is often done in conjunction with OPC.

Reducing reflections on the mask and the wafer restores lost contrast in the imaging system, thus facilitating k_1 reduction.³⁴ Reducing vibration in the imaging system also restores lost resolution.³⁵ These reductions are also covered in subsequent sections.

6.3.1 Phase-shifting masks

The working principle of many types of PSM and their combinations is given in Sec. 5.4.4. Here, the performance and optimization of PSM is included. AltPSM, attenuated phase-shifting masks (AttPSM), and chromeless PSMs have potential for manufacturing. They are included.

6.3.1.1 Alternating phase-shifting mask (AltPSM)

The AltPSM started the PSM movement. It is a strong PSM that can double spatial frequency and significantly improve the image contrast, characteristics that draw a lot of attention and studies. Its configuration and working principle are given in Sec. 5.4.4.2. However, AltPSM is the most difficult and expensive to practice. Figure 6.41 shows the ideal pattern to practice PSM: equally spaced lines within the coherence range and, similarly, evenly spaced holes. These structures lend themselves to straightforward phase assignments with no possibility of creating phase-shifting conflicts.

Difficulties and conflicts are depicted in Fig. 6.42. The brick layout in Fig. 6.42(a), once popular for dynamic random access memory (DRAM), has phase conflict between the ends of bricks along each row of bricks. If the phase of the bricks in a row is alternated, then there are conflicts between rows. One particular way to solve the conflicts is to assign 120-deg shifts instead of 180-deg shifts; however, image contrast is reduced. Mask fabrication also becomes more difficult. The π and disconnected- π in Figs. 6.42(b) and (c) have conflicts at the horizontal line. Even in a well-behaved line-space pattern, as shown in Fig. 6.42(d), if a light-field mask is used, redundant dark lines will be delineated at the phase-shifting edges that were inadvertently created. The uneven fork structure in Fig. 6.42(e) has an isolated opening that does not have any phase-shifting support, in addition to the conflicts in the π structure.

It is understandable that phase-shifting conflicts are unavoidable in arbitrary 2D patterns, because there is only one parameter at play in two degrees of freedom. Figure 6.43 demonstrates the phase-shifting conflicts in a metal-layer circuit pattern. On the leftmost side, the line openings are kept without conflicts, but when they degenerate into more complicated 2D patterns, conflicts arise.

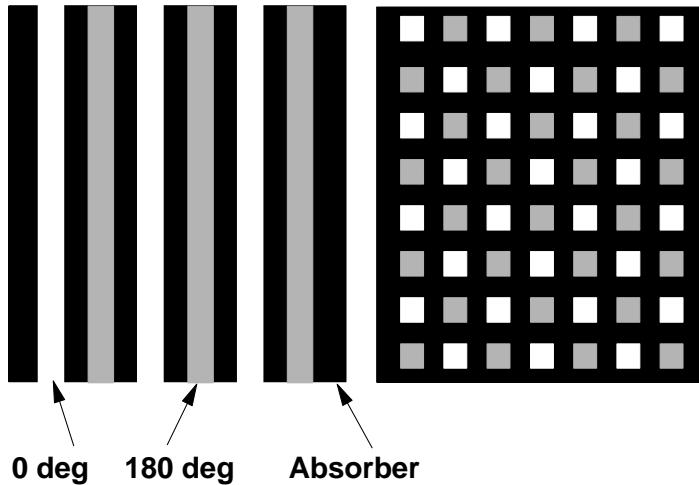


Figure 6.41 Ideal patterns for AltPSM.

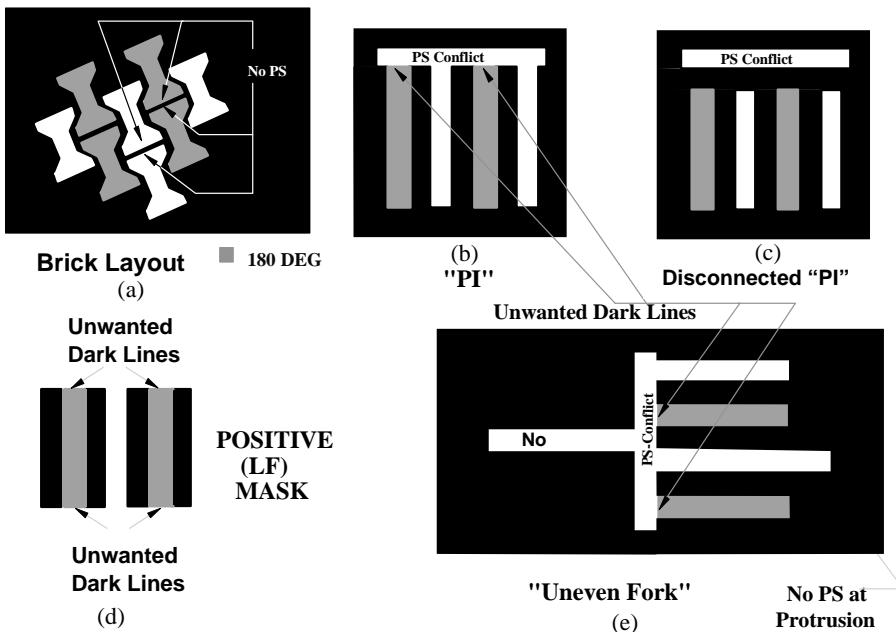


Figure 6.42 Difficulties and phase conflicts with AltPSM.

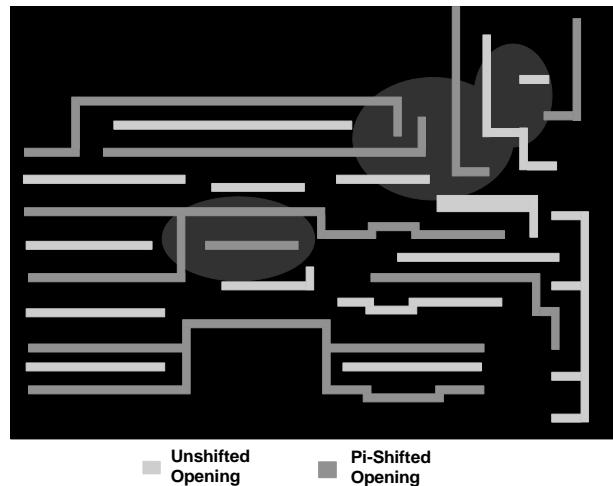


Figure 6.43 Dark-field alternating phase-shifting conflicts in a metal layer.

The mask of Figure 6.43 is a dark-field mask, which means the metal lines in the mask are transparent. They create openings in the positive resist. Electroplating patterns the metal lines. A subtractive process, such as etching, can be used to delineate the metal lines, so that a light-field mask can be used on a positive resist. This way, as shown in Fig. 6.44, the phase-shifting conflicts can be eliminated with two exposures. One of the exposures uses the layout and phase-shifting scheme shown, with phase shifting taking place at each opaque line. The edges resulting from the boundary between phase shifts can be trimmed with another mask that is illustrated in Fig. 6.45. The layouts in the two masks are superimposed in Fig. 6.46 to show the relative position of the edges of the two layouts. The trim patterns are intentionally placed to trim off some line edges in some cases. In other cases, they are intentionally placed slightly away from the metal line pattern, leaving a whisker metal line from the phase-shifting edge image. These are done to anticipate overlay errors between the two exposures.

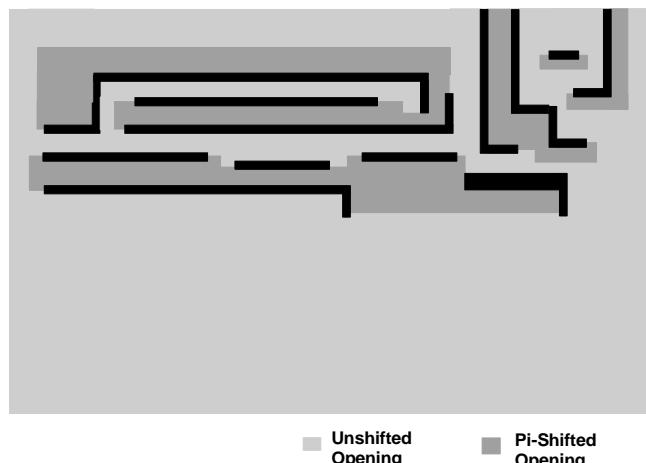


Figure 6.44 Bright-field AltPSM from the upper half of Fig. 6.43.



Figure 6.45 Trim mask complementing the one in Fig. 6.44 to remove the unwanted phase-shifting edges.

A double-exposure scheme³⁶ has been adopted by several companies for manufacturing the logic gate. This scheme defines the polylayer pattern in a BIM but with the gates enlarged, exemplified by mask 1 in Fig. 6.47. It superimposes mask 2, a covered-edge phase-shifting trim mask that looks similar to Figure 6.44. The gate length is now defined by the trim mask and can be much shorter, due to the high image contrast produced by the PSM, which results in narrower gates with better CD uniformity. The type of phase shifting is classified as covered-edge phase shifting, as discussed in Sec. 5.4.4.2. The effect of AltPSM is felt only when the phase-shifting edges get too close to each other.

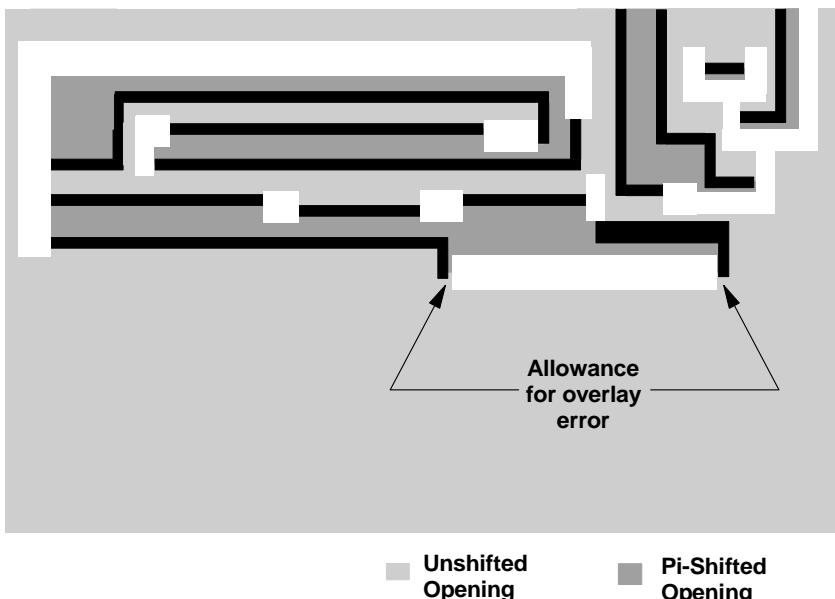


Figure 6.46 Two superimposed mask layouts in Figs. 6.44 and 6.45.

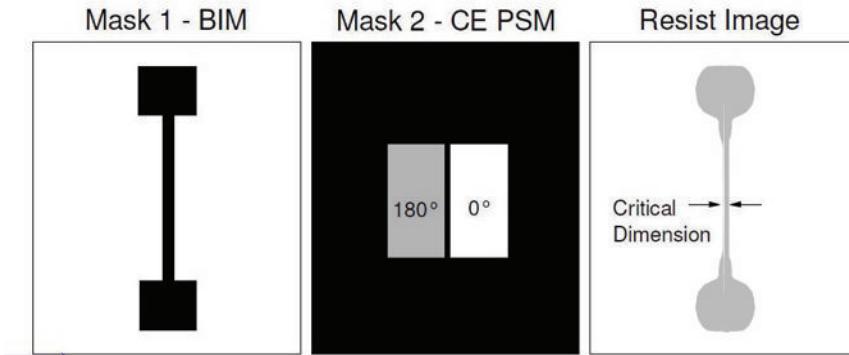


Figure 6.47 Covered phase-shifting edge with double exposure for the gate level.

The imaging performance of dark-field AltPSM is represented by the DOF versus NA plot in Fig. 6.48. Alternately, phase-shifted gratings with 1:1, 1:1.5, and 1:2 opening-to-space ratios, a DKI at $\sigma = 0.3$, a CD tolerance of $\pm 10\%$, and EL of 8% are used.

In addition to phase-shifting conflicts, this type of dark-field layout presents another problem for AltPSM. The shifted and unshifted openings are unevenly spaced, posing a proximity problem that is more severe than with BIM. The problem is depicted in the following figures. Take 0.35- μm equal-line-space pairs, using AltPSM at $\lambda = 365 \text{ nm}$; its DOF as a function of NA is shown in Fig. 6.49. It is based on 20% EL with a $\pm 10\%$ CD tolerance. The DOF is very large except when NA is above 0.65. By just introducing another set of line-space pairs, a 1:2 line opening to opaque space, the DOF drops to below 1 μm as shown in Fig. 6.50. Either feature supports a large DOF except when they must be imaged together without individually adjusting the print bias as seen in Fig. 6.51.

There is no reason that the mask features cannot be made larger or smaller than the target CD on the wafer by using biasing. Figure 6.52 shows the same set of features as Fig. 6.50 except that the 1:2 feature is now biased by 80 nm, making the line openings larger by 80 nm. Figure 6.53 shows that the E-D windows of the 1:2 feature are shifted lower toward the E-D windows of the 1:1 feature, making larger common windows. This is a preliminary form of optical proximity correction.

Optical proximity effect is not the only effect that can affect the imaging performance of AltPSM. Fabrication tolerance of PSMs also has a strong influence. Figure 6.54 shows the E-D windows of $k_1 = 0.46$ line-space pairs from BIM and AltPSM. The larger E-D window belongs to the AltPSM. The latter indeed performs better than the former. However, when the fabrication tolerance of $\Delta_{\text{phase}} = \pm 10 \text{ deg}$ and $\Delta_{\text{transmission}} = 100\text{--}90\%$ occurs, the common E-D window of four imperfect E-D trees superimposed on the perfect E-D tree is much smaller than just the perfect tree by itself, as shown in Fig. 6.55. Without even using the trees from combined transmission and phase errors, this common window is already smaller than the window of the BIM.

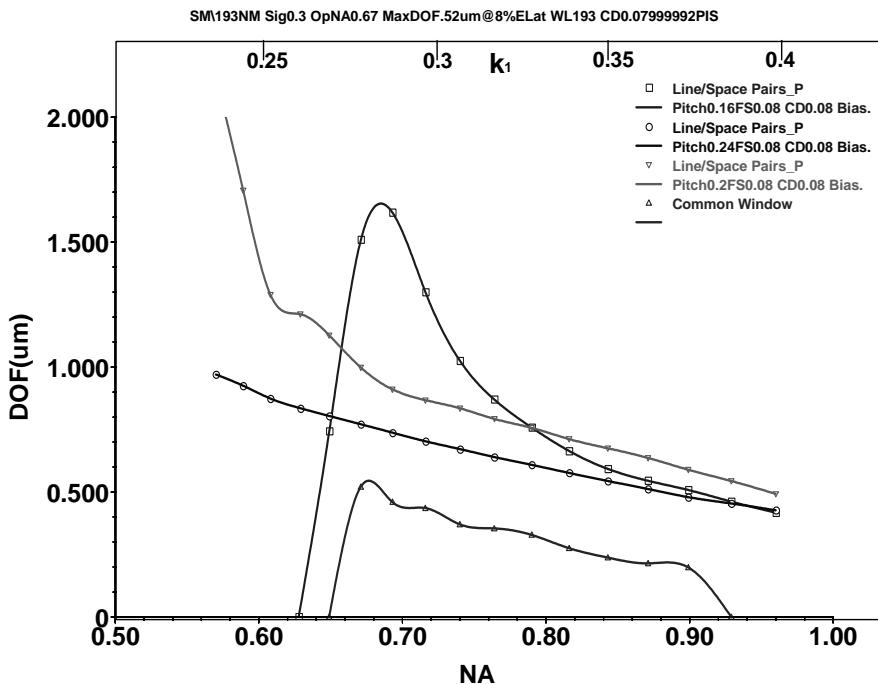


Figure 6.48 DOF versus NA for 80-nm 1:1, 1:1.5, and 1:2 line opening to opaque space at 193 nm for AltPSM.

210\LSAL Sig.30 OpNA.31 MaxDOF5.85um@20%ELat WL365 CD.35

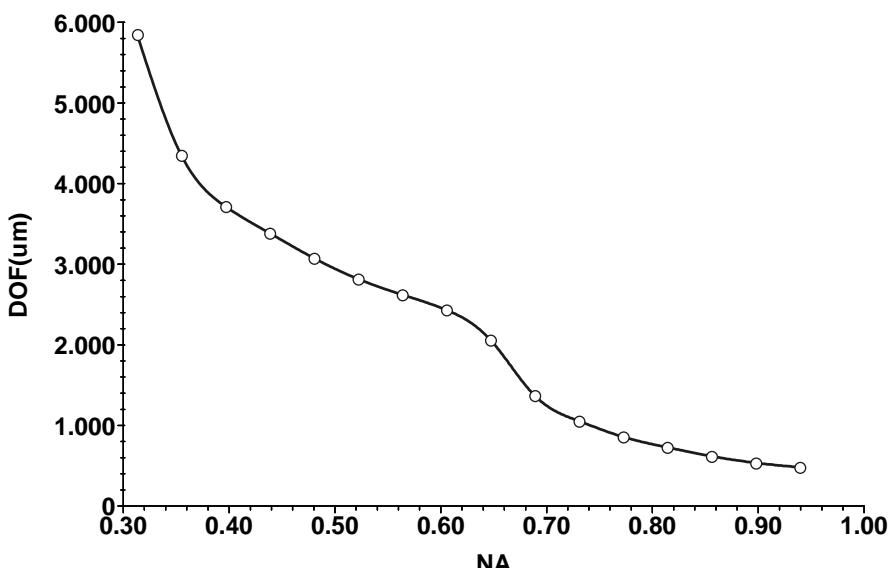


Figure 6.49 DOF versus NA for 0.35- μ m LS at 365 nm for AltPSM.

210\LSAL Sig.30 OpNA.63 MaxDOF.91um@15%ELat WL365 CD.35

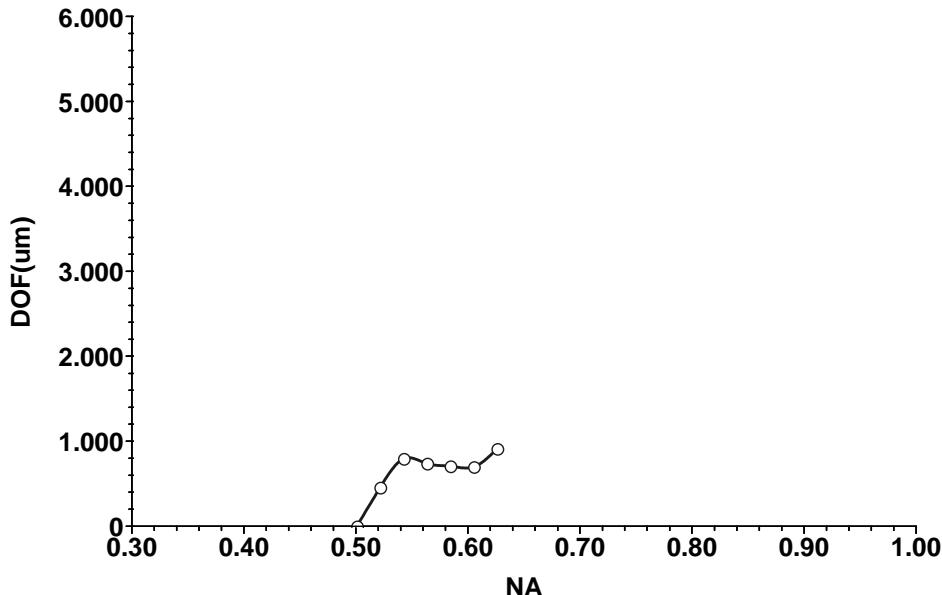


Figure 6.50 DOF versus NA for 0.35-μm 1:1 and 1:2 LS at 365 nm with AltPSM.

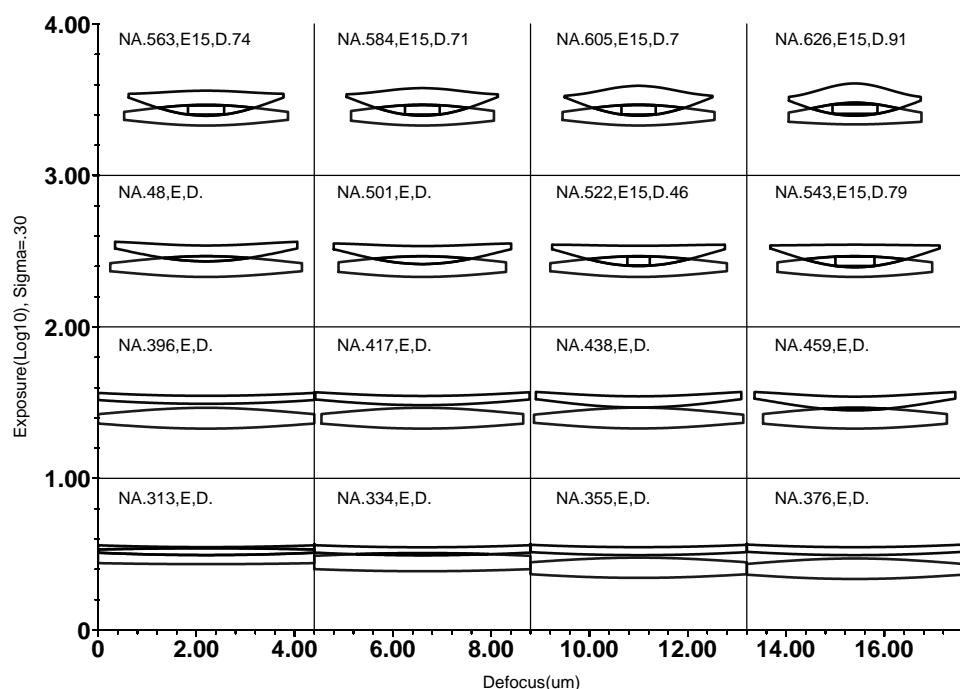


Figure 6.51 E-D windows of Fig. 6.50.

210\LSAL Sig.30 OpNA.52 MaxDOF2.28um@15%ELat WL365 CD.35

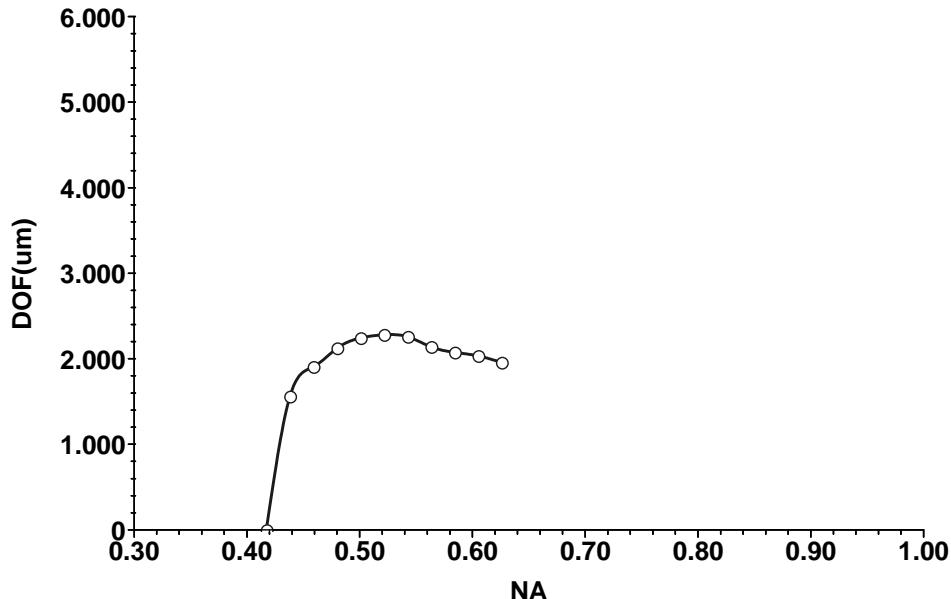


Figure 6.52 DOF versus NA for 0.35- μ m 1:1 and 1:2 LS at 365 nm AltPSM, with 80-nm bias for 1:2.

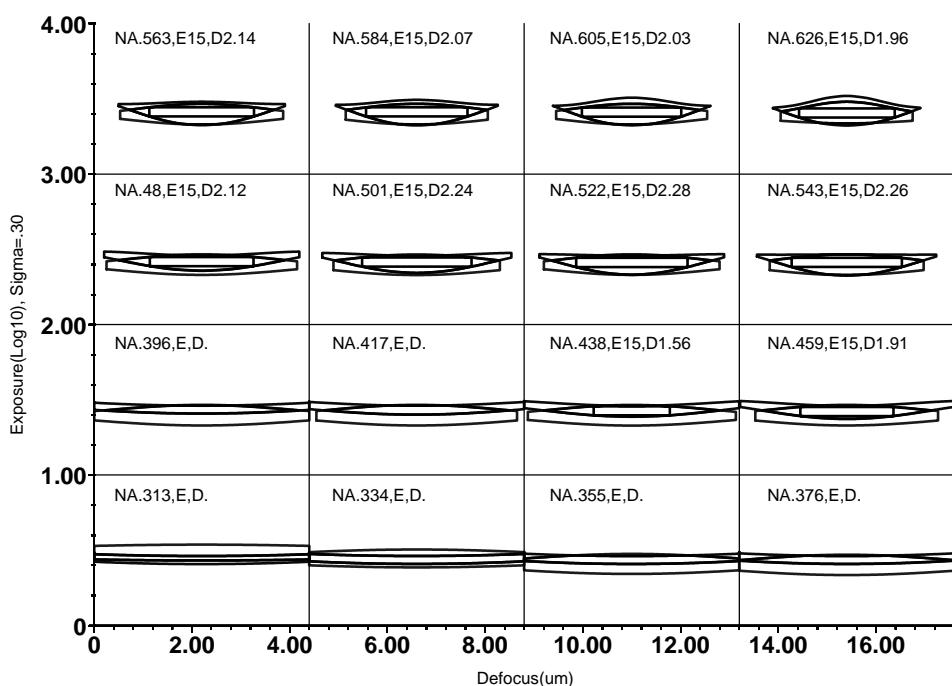


Figure 6.53 E-D windows of Fig. 6.52.

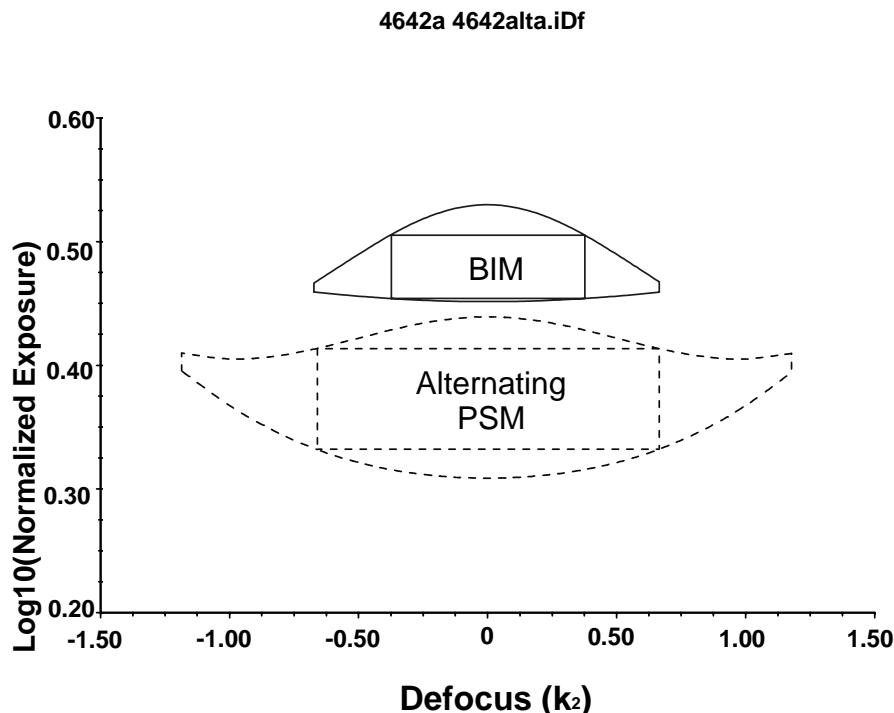


Figure 6.54 AltPSM improvement at $k_1 = 0.46$ and $\sigma = 0.42$.

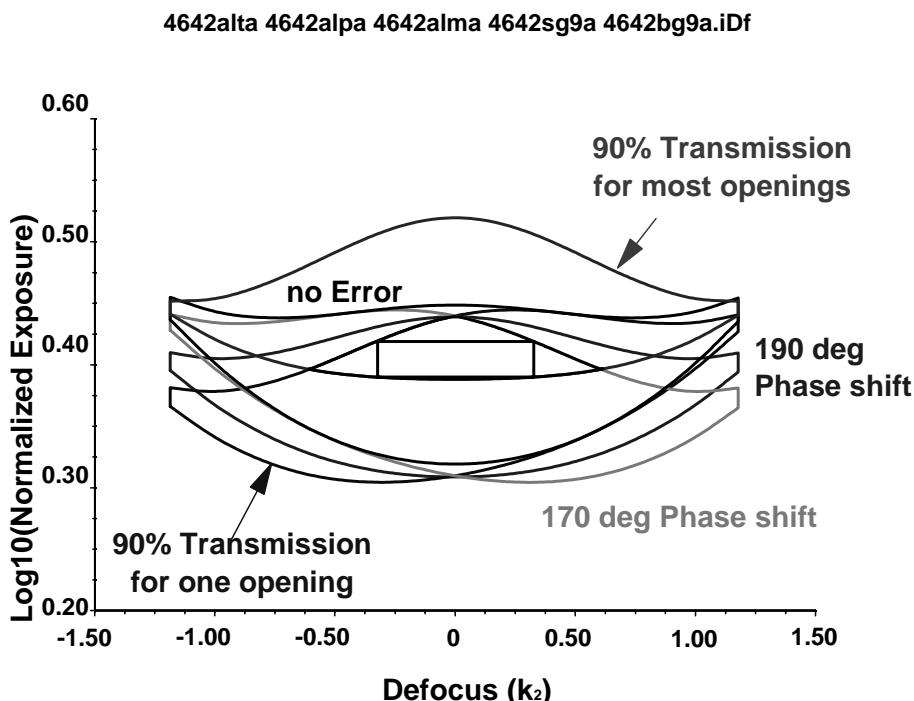


Figure 6.55 AltPSM with phase and transmission errors. Performance is worse than that of BIM.

6.3.1.2 Attenuated phase-shifting mask (AttPSM)

AttPSM is the most widely used PSM for manufacturing. Its configuration and working principle were given in Sec. 5.4.4.2. Its popularity is mainly due to the simple process and accommodation of an arbitrary mask pattern. Commercial mask blanks contain a MoSi absorber that provides the required 6% transmission and π phase shift. So, the delineation of the phase-shifting pattern is similar to that of BIM, namely, expose, develop, and etch. However, it is desirable to cover the area outside of the patterned area with chromium so that the reticle information for the exposure tool is on chromium instead of on a phase-shifting material. Therefore, commercial mask blanks come with a chromium layer above the MoSi layer. Needless to say, the resist layer is precoated on the chromium layer in the mask-blank factory just as in the case of BIM. The processing of such a composite blank is still simple. The chromium image can be used as the hard mask to delineate the MoSi image, followed by a second patterned exposure to remove the chromium at the phase-shifting areas. Because the continuity of the chromium layer is broken by the mask patterns, electrons can be trapped in the isolated conducting areas, thus producing charging effects. Hence, the second exposure uses an optical mask writer instead of e-beam.

The 6% transmission was defined in the early phase of AttPSM development.^{37,38} Too little transmission makes the phase-shifting effect insignificant. Too much of it increases stray light. Worse, it also produces ghost lines at the side lobe of the optical image.

Unlike AltPSM, which has the potential to double the resolution, AttPSM is a weak RET. It merely enhances the edge contrast. Fortunately, it can be applied to any arbitrary layout against the restriction from phase conflicts in AltPSM. In the early 1990s, while people were still overcoming the restriction of AltPSM, rim PSM was invented.³⁹ It immediately revived interest in PSMs. The lithography community dedicated much effort to developing rim PSM then, until the rediscovery of AttPSM.^{40,41} Figure 6.56 is an analysis of the imaging performance from the E-D methodology on AttPSM and rim PSM using an isolated line opening whose width ranges from $k_1 < 0.3$ to $k_1 > 0.7$. It shows that AttPSM attains a larger DOF with a smaller mask bias and requires less exposure from this wide range in k_1 . The three groups of DOF curve—SLR, SLR + ARC, and MLR—denote EL of 30%, 20%, and 10%, respectively. These exposure latitudes were used to construct the E-D trees and windows for the comparison. The rim main feature has a 0.18 k_1 bias against 0.1 k_1 with AttPSM. AttPSM also has the inherent advantage in process simplicity and process control.

Figure 6.57 shows the imaging performance of contact holes with these two types of PSM using just the 10% EL. AttPSM excels again in DOF, image intensity, and packing density. From Fig. 6.57, the rim PSM needs an additional mask area to accommodate the rim width of 0.15 k_1 , and thus has a severe impact on packing density. The use of AttPSM for contact holes is widespread.

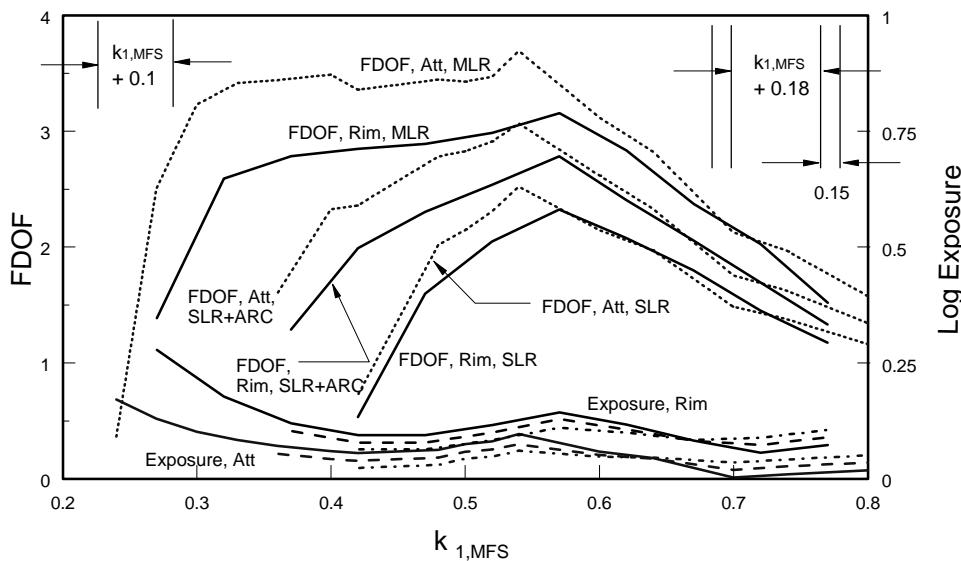


Figure 6.56 Analysis of imaging performance of AttPSM and rim PSM on holes: AttPSM -0.1 and rim PSM -0.18 bias.

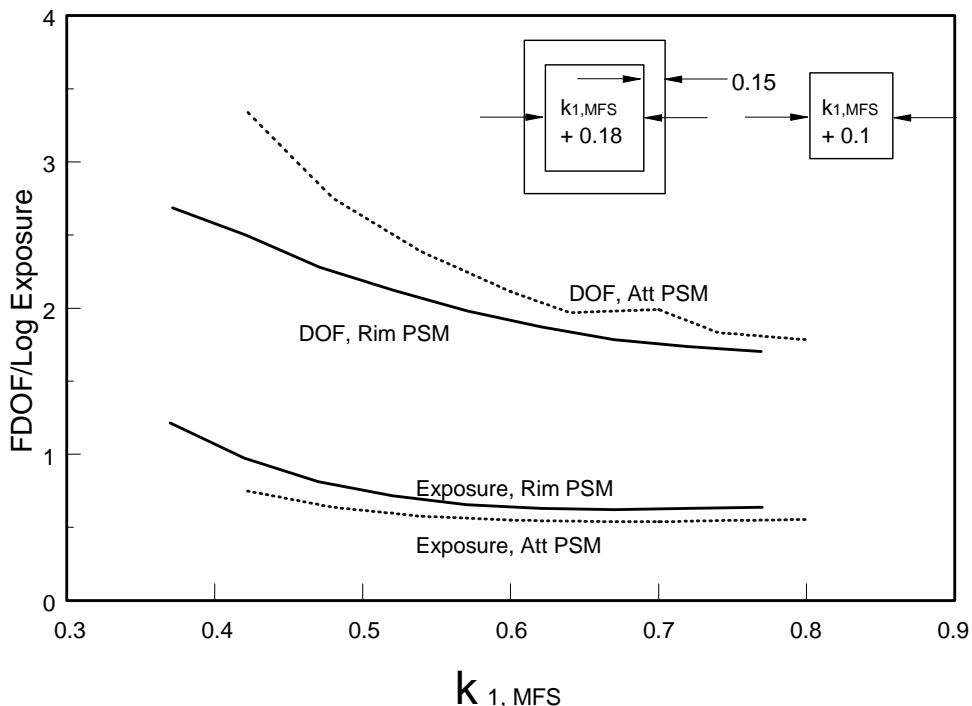
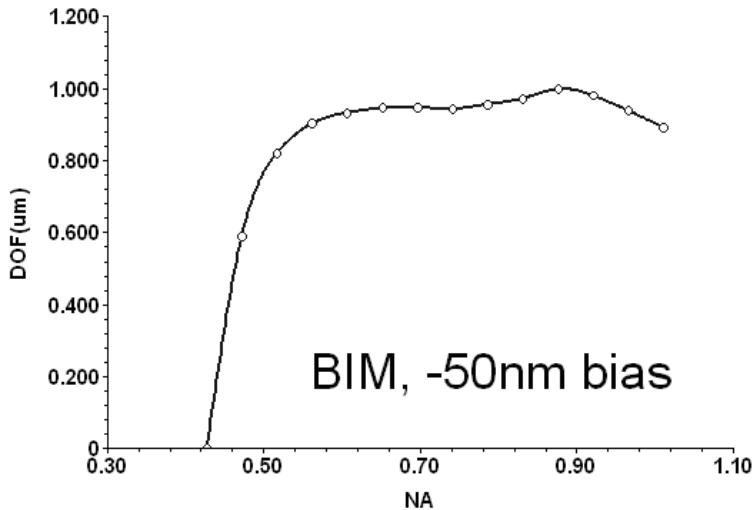


Figure 6.57 Analysis of AttPSM and rim PSM imaging performance on line openings: AttPSM -0.1 and rim PSM -0.18 bias.

A210\CT1 Sig.42 OpNA.88 MaxDOF1.um@20%ELat WL365 CD.35



A210\CT1 Sig.42 OpNA.52 MaxDOF1.15um@20%ELat WL365 CD.35

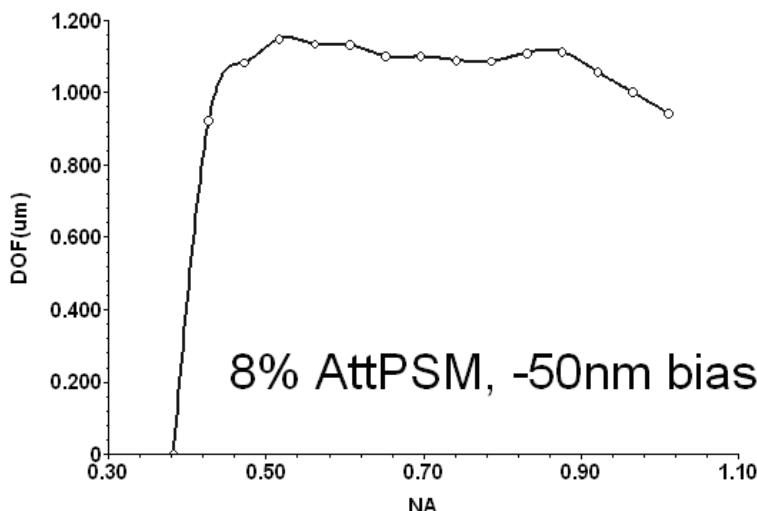


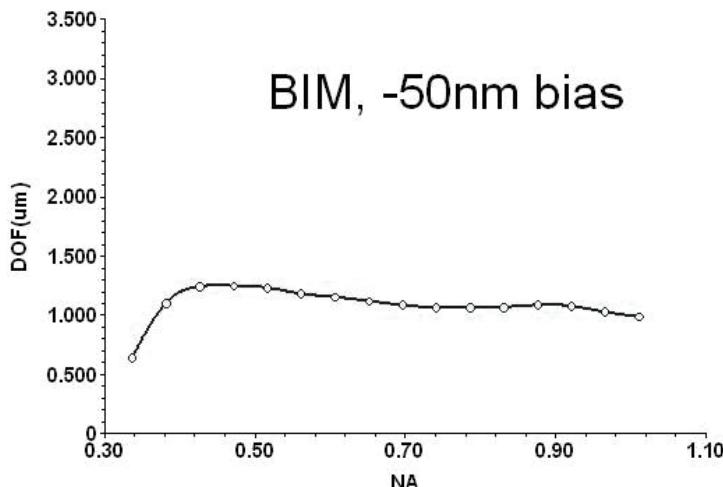
Figure 6.58 Comparison of 0.35- μm BIM and AttPSM holes in NA optimization, EL = 20%.

The improvement of contact-hole DOF with 8% AttPSM over BIM is shown in Fig. 6.58. The holes are 0.35 μm in size, and the wavelength is 365 nm. DKI at $\sigma = 0.42$ is used. Improvement is about 20% over the range of the NA evaluated. The -50-nm bias achieves a larger DOF in both AttPSM and BIM. The EL is set at 20%.

When the EL is changed to 10%, DOF improves in both types of mask. At NA = 0.34 corresponding to $k_1 = 0.355$, the improvement in DOF is dramatic. However, this is seldom taken advantage of for two reasons:

1. Using such low k_1 significantly increases MEEF, making CD control difficult.
2. The resist system does not have sufficient contrast to realize the aerial image.

A210\CT1 Sig.42 OpNA.47 MaxDOF1.25um@10%ELat WL365 CD.35



A210\CT1 Sig.42 OpNA.34 MaxDOF3.28um@10%ELat WL365 CD.35

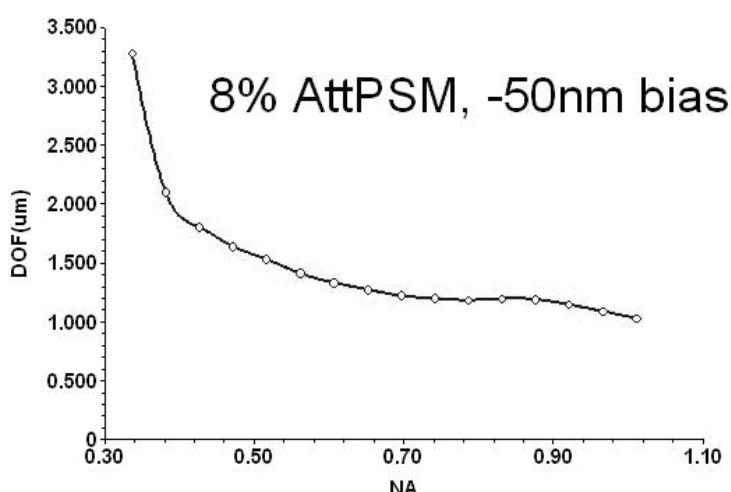
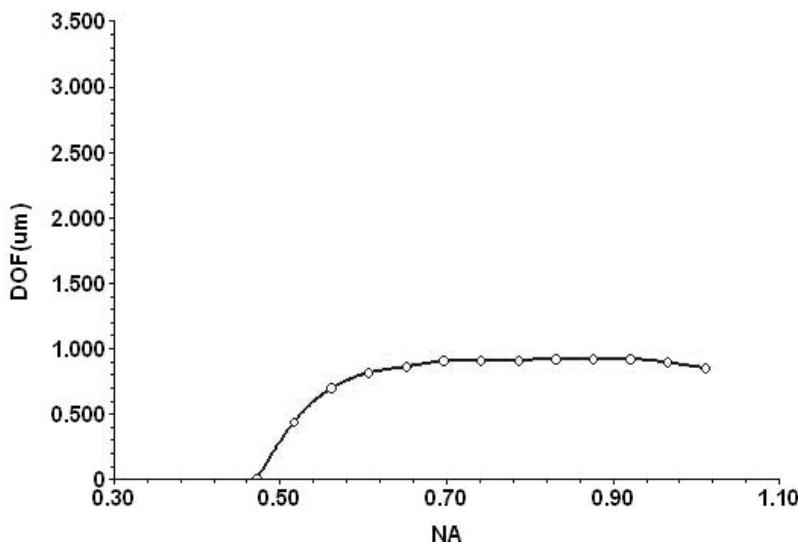


Figure 6.59 Same as Fig. 6.58, except EL = 10%.

Figure 6.60 turns the aerial image result into that of the resist image by using the lumped parameters. The parameters chosen for this resist are $\alpha = 1.12 \mu\text{m}^{-1}$, $\gamma = 5$, and a resist thickness of 770 nm, which is fairly typical. The extraordinary gain in DOF disappears with the resist image.

A210\CT1 Sig.42 OpNA.88 MaxDOF.92um@10%ELat WL365 CD.35



A210\CT1 Sig.42 OpNA.7 MaxDOF1.05um@10%ELat WL365 CD.35

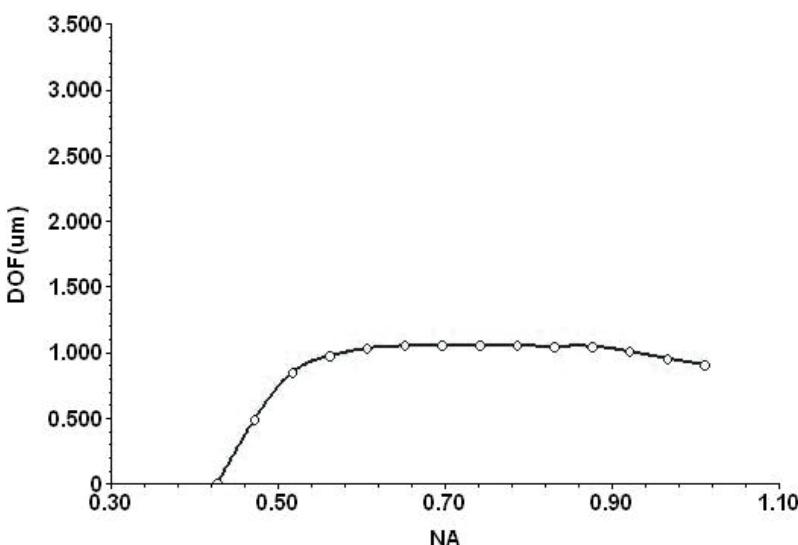
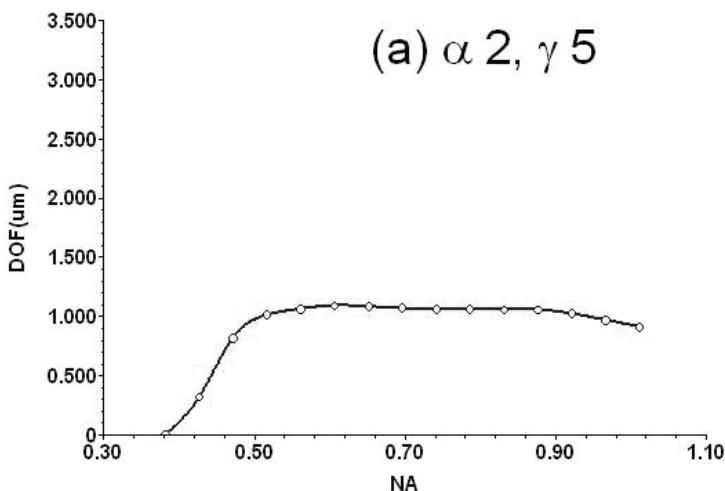


Figure 6.60 Same as Fig. 6.59 except lumped parameters are used to simulate a resist image instead of an aerial image. Resist thickness is 0.77 μm . $\alpha = 1.12$ and $\gamma = 5$.

Changing α to 2 does not seem to help, as seen in Fig. 6.61(a). However, when γ is set to 25, rarely achieved in commercial resist systems, the dramatic gain in DOF is restored, as seen in Fig. 6.61(b).

The AttPSM is often used with DKI at a low σ for contact holes. For line-space (LS) patterns, its best combined with OAI to enhance the DOF of closely packed lines and subresolution assist features to improve the isolated lines. More will be written on the LS application.

A210\CT1 Sig.42 OpNA.61 MaxDOF1.09um@10%ELat WL365 CD.35



A210\CT1 Sig.42 OpNA.34 MaxDOF2.83um@10%ELat WL365 CD.35

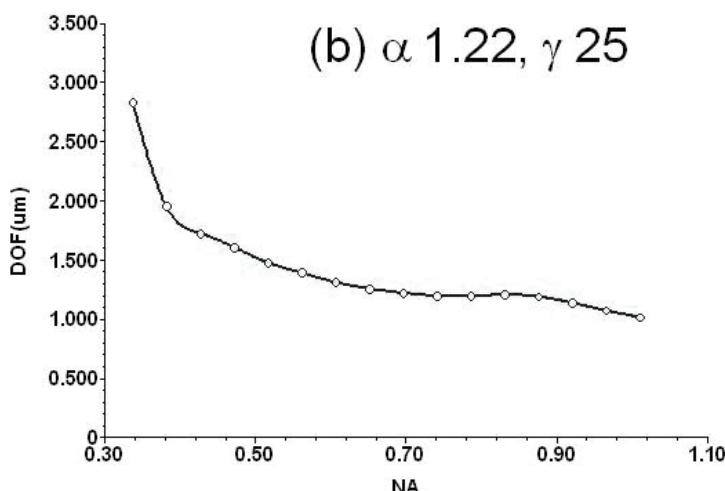


Figure 6.61 AttPSM part of Fig. 6.60 with α changed to 2 in part (a) and γ to 25 in part (b).

6.3.2 Off-axis illumination

Off-axis illumination^{42,43} (OAI) was proposed in the early 1990s to supplement or replace the phase-shifting mask technology. Its advantage is a lower cost, but it has its own limitations. In the following, we explain the working principle of OAI, provide an analysis, and simulate its imaging performance.

6.3.2.1 Working principle of OAI

6.3.2.1.1 Conceptual illustration

OAI takes advantage of spatial frequency shifting of a given object to improve resolution and DOF. It can best be illustrated with a simple grating object. Being periodic, the grating contains discrete spatial frequency components, namely, the 0th order, ±1st order, ±2nd order, etc. When the minimum feature is of interest, only the 1st-order frequencies are preserved so that the resolution potential of the imaging lens can be fully utilized. This situation is shown in Fig. 6.62(a), where DKI is chosen. The spatial-frequency spectrum consists of the vertically oriented 0th-order beam and the ±1st-order beams, whose angle is a function of the periodicity of the grating. A smaller periodicity produces large spatial frequencies, thus larger diffraction angles in the ±1st-order beams. When the feature size is too small, i.e., the spatial frequency too high, the angle of the diffracted beams becomes larger than the acceptance angle of the imaging lens, and the ±1st-order beams are rejected. Only the 0th-order beam passes. The image becomes a structureless uniform beam.

The electric field on a BIM is simply 1 and 0 at the transparent and opaque parts of the mask, respectively. The actual field is obviously much more complicated when exact electromagnetic diffraction is taken into account.⁴⁴ Here, the simple assumption suffices to explain the working principle of OAI. The intensity at the mask plane is proportional to the square of the simplified electric field and also 1 and 0, following the electric field completely. The intensity at the image plane is uniform, just as if no object is present, because the ±1st-order beams are not accepted by the lens. Only the 0th-order beam, which has no information content, passes through the lens. The intensity of the image is reduced because the energy of the rejected beams is not recovered.

Figure 6.62(b) shows the situation of a single collimated illumination beam obliquely incident on the mask, thus off axis. The three beams shown in Fig. 6.62(a) are now tilted by the incident angle of the illumination. When the angle is adjusted to make the 0th-order, and one of the 1st-order beams symmetrical with respect to the optical axis, the largest angular spread between the two beams is possible without being cut off by the acceptance angle of the lens; thus, the highest resolution is achieved. However, the other 1st-order beam is cut off, resulting in a lower exposing intensity.

In Fig. 6.62(c), two symmetrically opposed beams are used. When the angle of the illumination is optimized for a given periodic object, the 0th order of the

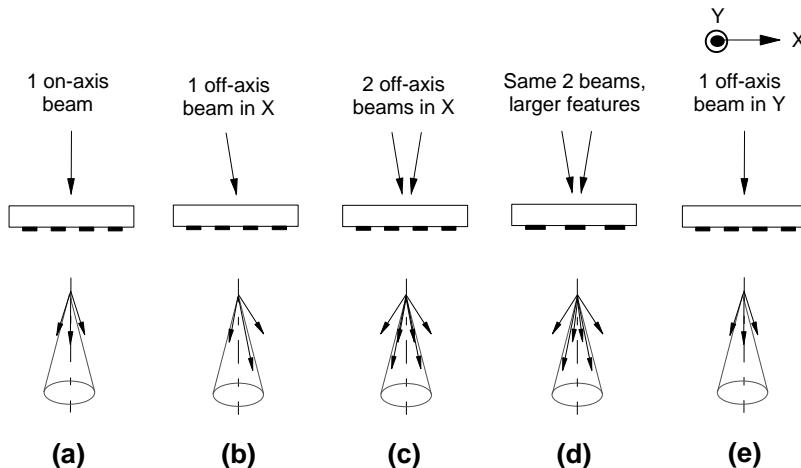


Figure 6.62 On-axis and off-axis illuminations affecting the 0th- and $\pm 1^{\text{st}}$ -order spatial-frequency vectors.

left beam coincides with the 1st order of the right beam, and the 0th order of the right beam coincides with the -1st order of the left beam, as shown in the figure. The image consists of a single frequency component and is well reproduced.

When objects of a lower spatial frequency, as shown in Fig. 6.62(d), are illuminated by the same set of beams, the spatial frequency vectors no longer coincide, producing additional spatial frequencies. These frequencies can induce a worse imaging characteristic than that from DKI.

Figure 6.62(e) shows that tilting the beam in x does not help to resolve the spatial frequencies in y .

6.3.2.1.2 3D illumination on 2D geometry

When 2D geometry is present, the OAI beams are directed along a conical surface to illuminate a ring area in the aperture of the imaging lens, as shown in Fig. 6.63(a). When only vertical and horizontal lines are present, the best OAI takes the form of four small illuminated areas at the intersection of the $\pm 45\text{-deg}$ lines with the ring, as shown in Fig. 6.63(b). This way, each beam is off axis to either the x - or the y -oriented lines on the mask. With 45-deg lines on the mask, only one of the two diagonal beam pairs is off axis to either aa' or bb' , as shown in Fig. 6.63(c). The other on-axis beam dilutes the gain in imaging performance produced by the off-axis beams. In addition, the gain in imaging performance for mask patterns with identical pitch but different orientations is inconsistent because the diagonal distance between the poles is larger than the vertical or perpendicular distances. The case of ring OAI in Fig. 6.63(a) is a continuous combination of off-axis and on-axis beams, plus the effect of changing optimizing pole distances. Therefore, its imaging performance is a compromise between DKI and OAI with diagonal quadrupoles. We have been discussing the above qualitatively. The actual imaging performance with partial coherent light must be simulated and compared carefully.

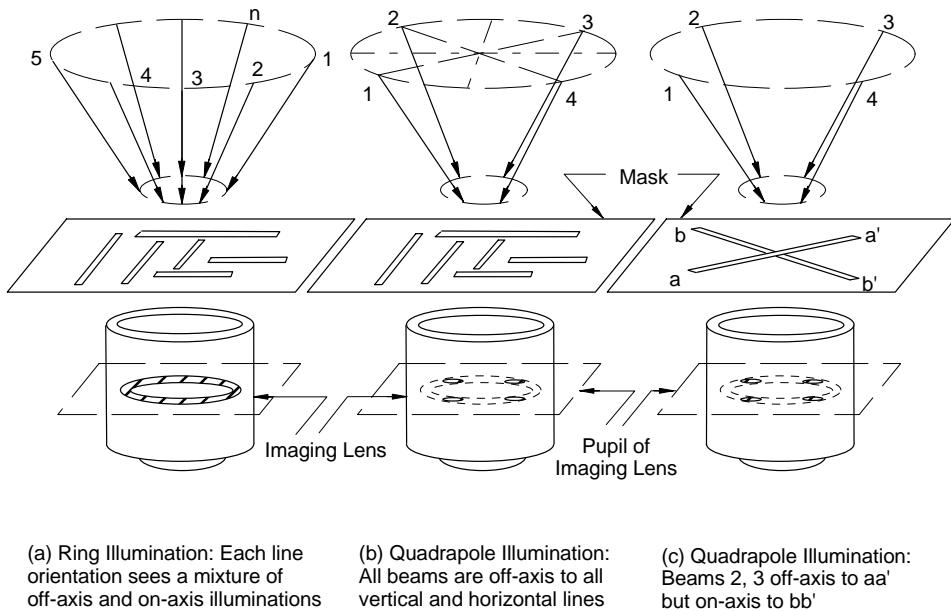


Figure 6.63 Illumination and pattern orientation.

DKI, ring illumination (RGI), quadrupole illumination (QRI), and a disk-OAI combination are shown in Fig. 6.64. DKI is on axis, characterized by the aperture ratio $\sigma \equiv \text{NA}_{\text{condenser}} / \text{NA}_{\text{imaging lens}}$. For ring OAI, the ring is characterized by σ_{in} and σ_{out} . For quadrupoles, the σ_{in} and σ_{out} terminology is still followed. The diameter of the poles can readily be derived from the two σ s. Angular spread ϕ is used to characterize the size of the fanned quadrupole. This type of quadrupole is often used for convenience because it can readily be formed from the basic RGI. The bull's-eye illumination combines a center disk with the quadrupole. It compromises the imaging performance of isolated and dense features to strike a desired balance. C-Quad illumination simply lays out the quadrupoles in a cross orientation. The dipole illumination is either the horizontally oriented poles in C-Quad or the vertically oriented ones. It is not shown.

What is the difference in the imaging performance of a quadrupole and C-Quad? Take a vertical line object: poles a and b are off axis, and so are poles c and d . Alternately, with a horizontal line object, poles a and d as well as b and c are off axis. Therefore, there are no on-axis poles to dilute the image contrast. However, the horizontal and vertical distance from the center of the poles to the x - or y -axis is only $\sigma/\sqrt{2}$, where σ is the distance from the center of the poles to the center of the lens pupil, whereas with C-Quad, the horizontal and the vertical distance of the poles is the full length of σ , despite contrast dilution from the perpendicular pair of poles. There is a tradeoff between resolution and contrast in these two types of quadrupole illumination.

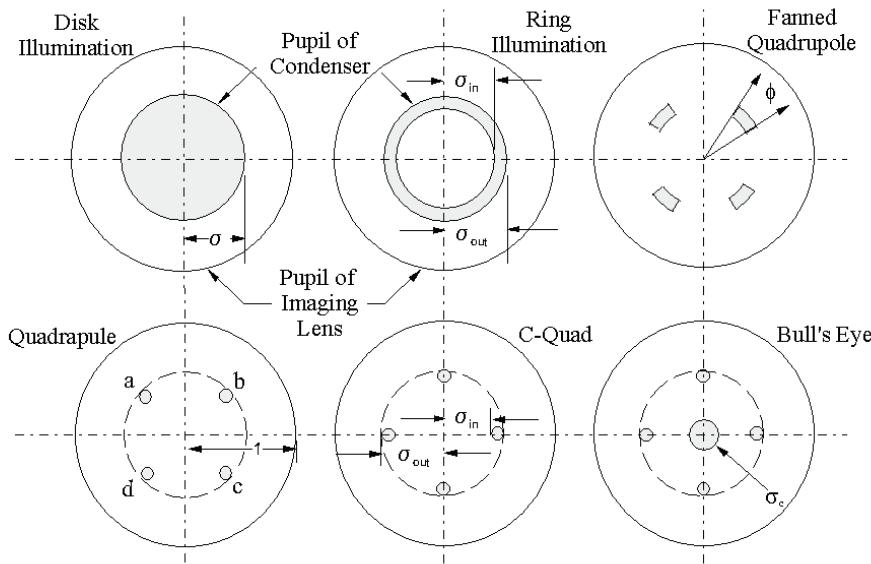


Figure 6.64 Illumination schemes.

Dipole illumination produces high resolution and contrast without compromise. However, if a 2D pattern is used, it must be split into two masks, with vertically oriented patterns exposed by the horizontally oriented dipole and the horizontally oriented patterns exposed by the vertically oriented dipole. This illumination setup is called double-dipole illumination (DPL).⁴⁵ Splitting of L-shaped patterns is illustrated in Fig. 6.65, where (a) and (b) are the two split patterns. Figure 6.65(c) is the target pattern, whereas Fig. 6.65(d) is the resultant pattern after combining the exposures from (a) and (b). If the pattern alignment is not perfect, the corners will have slight complications. Fortunately, the corner area has been known to be restricted for critical dimensions. One drawback of DPL is that two exposures reduce the throughput, thus increasing cost.

6.3.2.1.3 Analytic forms

In this section, OAI is analytically examined. We first derive the image electric field and its intensity distribution using DKI. A 1D object with a single spatial frequency is used for the study. From Eq. (3.67), with on-axis uniform illumination, the electric field through a BIM grating object of periodicity p and width of line opening w , as shown in Fig. 6.66, is

$$E_o(x) = 1, np - \frac{w}{2} \leq x \leq np + \frac{w}{2}, n = 0, \pm 1, \pm 2, \dots \\ = 0, \text{ elsewhere.} \quad (6.18)$$

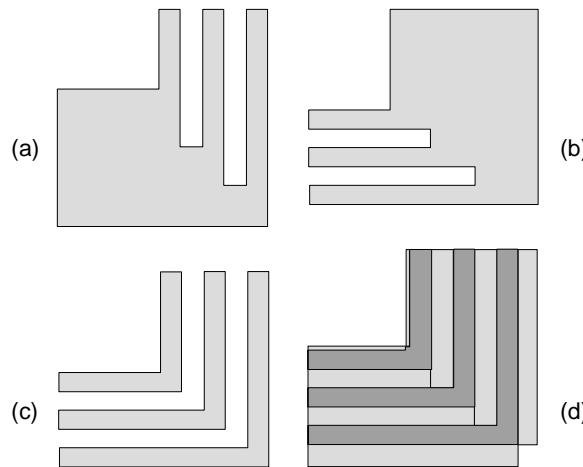


Figure 6.65 Splitting of L-shaped patterns for double-dipole illumination: (a) and (b) split patterns, (c) target pattern, and (d) pattern composed from superimposing the exposures from (a) and (b).

Its Fourier transformation normalized to w is

$$E_o(\xi) = \frac{\sin \pi \xi w}{\pi \xi w} \sum_{n=-\infty}^{\infty} \delta\left(\xi - \frac{n}{p}\right). \quad (6.19)$$

This is an infinite series of spatial frequencies in the object. The finite lens aperture cuts off the higher spatial frequencies to form the image, as we discussed in Chapter 3. Let us examine the case of $w = p/2$, where the lens only passes the frequencies $n = 0$ and $n = \pm 1$. Then, the electric field E_i of the image is

$$E_i(x) = 1 + \frac{2}{\pi} (e^{ifx} + e^{-ifx}) = 1 + \frac{4}{\pi} \cos fx, \quad (6.20)$$

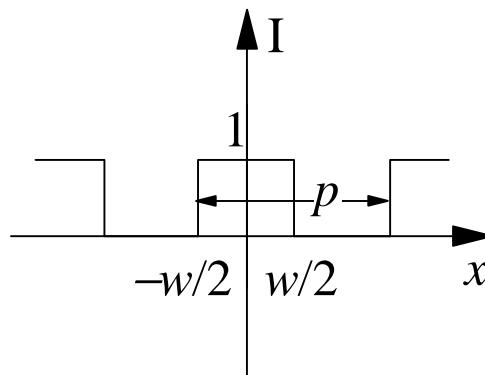


Figure 6.66 Transmitted intensity of DKI through a grating BIM. p is the periodicity of the grating and w is the width of each opening.

where $f \equiv 2\pi/p$. The corresponding intensity distribution $I_i(x)$ is

$$I_i(x) = \left(1 + \frac{4}{\pi} \cos fx\right)^2 = 1 + \frac{8}{\pi^2} + \frac{8}{\pi} \cos fx + \frac{8}{\pi^2} \cos(2fx). \quad (6.21)$$

Note that even though there is no higher diffraction order in the electric field, a double frequency term exists in the image intensity distribution. The 0th- and the 1st-order components in the electric field are shown in Fig. 6.67. The angle between the 0th- and the 1st-order beams is α , where $\sin \alpha = \lambda/p$.

Now, we change the illumination to off axis. Let the illumination make an angle $\pm\phi$ with respect to the z -axis. The electric field of the 0th-order beam is

$$E_{0\pm}(x) = e^{\pm ikx \sin \phi}, \quad (6.22)$$

where $k \equiv 2\pi/\lambda$. The image electric field becomes

$$E_i(x) = \left(e^{ikx \sin \phi} + e^{-ikx \sin \phi} \right) \left[1 + \frac{2}{\pi} (e^{ifx} + e^{-ifx}) \right], \quad (6.23)$$

where the terms in the last bracket are from Eq. (6.20). Substituting $fx = kx \sin \alpha$ and combining the exponential terms results in

$$E_i(x) = \cos(kx \sin \phi) + \frac{2}{\pi} [\cos kx(\sin \alpha + \sin \phi) + \cos kx(\sin \alpha - \sin \phi)]. \quad (6.24)$$

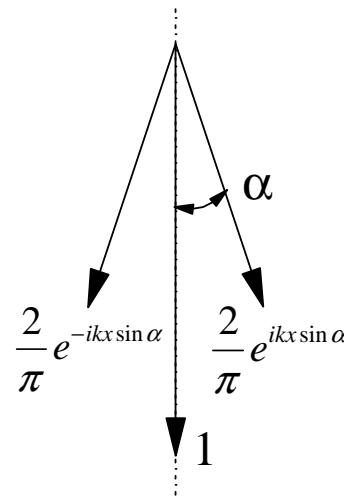


Figure 6.67 0th- and 1st-order beams in the image.

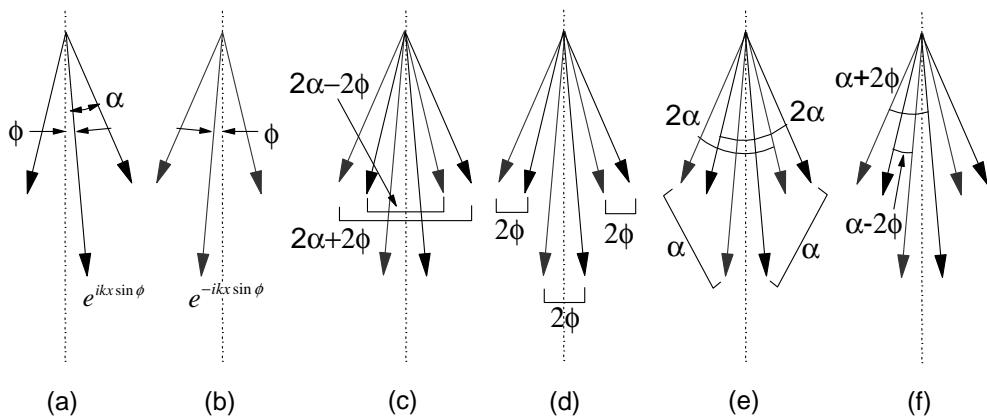


Figure 6.68 Off-axis illumination at incident angles $\pm\phi$.

The situation is depicted in Fig. 6.68. When only one off-axis beam illuminates the object, as in 6.68(a) and (b), only the corresponding term in the first bracket of Eq. (6.23) is used.

Note the difference between the angles ϕ and α . ϕ is the incident angle of the illumination beam, while α is the angle of the diffraction orders. They are independent of each other. The diffraction angle superimposed on the lens aperture angle θ is depicted in Fig. 6.69. With DKI, the 0th- and $\pm 1^{\text{st}}$ -order beams fall within $\pm\theta$; this is three-beam imaging. With OAI, only the 0th- and the 1st-order beams for incident angle ϕ and the 0th- and the -1st-order beams for $-\phi$ need be within $\pm\theta$; this is two-beam imaging.

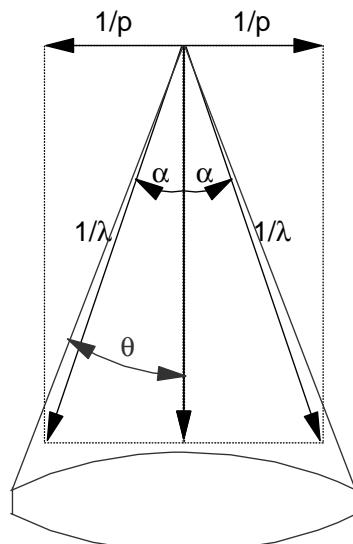


Figure 6.69 Relationship of diffraction angle α to λ , p , and aperture angle θ .

The image intensity distribution from Eq. (6.23) is

$$\begin{aligned}
 I_i(x) &= E_i^*(x) \cdot E_i(x) = \left[\cos^2(kx \sin \phi) \right] \left[1 + \frac{4}{\pi} \cos(kx \sin \alpha) \right]^2 \\
 &= \frac{1}{2} + \frac{4}{\pi^2} + \left(\frac{1}{2} + \frac{4}{\pi^2} \right) \cos(2kx \sin \phi) + \frac{4}{\pi} \cos fx + \frac{4}{\pi^2} \cos 2fx \\
 &\quad + \frac{2}{\pi} [\cos kx(\sin \alpha + 2 \sin \phi) + \cos kx(\sin \alpha - 2 \sin \phi)] \\
 &\quad + \frac{2}{\pi^2} [\cos 2kx(\sin \alpha + \sin \phi) + \cos 2kx(\sin \alpha - \sin \phi)].
 \end{aligned} \tag{6.25}$$

This represents a complicated mixture of spatial frequencies, such as ϕ , α , 2α , $\alpha \pm 2\phi$, and $2(\alpha \pm \phi)$. The term $(4/\pi)\cos fx$ is the desired image. All other terms make the image different. When $2\sin\phi = \sin\alpha$ (where ϕ is the ideal incident angle for a given grating periodicity), Eqs. (6.23) and (6.25) become

$$E_i(x) = 2 \cos \frac{fx}{2} \left(1 + \frac{4}{\pi} \cos fx \right) \tag{6.26}$$

and

$$I_i(x) = \frac{1}{2} + \frac{2}{\pi} + \frac{4}{\pi^2} + \left(\frac{1}{2} + \frac{4}{\pi} + \frac{6}{\pi^2} \right) \cos fx + \left(\frac{2}{\pi} + \frac{4}{\pi^2} \right) \cos(2fx) + \frac{2}{\pi^2} \cos(4fx). \tag{6.27}$$

Only the preferred spatial frequency and its harmonics are imaged. At any angles other than $2\sin\phi = \sin\alpha$, the image is not optimized. Note that the optimized angle is not $\phi = \alpha/2$.

6.3.2.2 Simulation results

Imaging performance of OAI systems from simulation will be given. Figure 6.70 shows the constant intensity contours using line-space pairs with 0.47 k_1 half pitch as the object. Physically, a 248-nm wavelength at 0.54 NA is used. The image from two types of illumination is illustrated: DKI at $\sigma = 0.8$ and RGI with $\sigma_{min} = 0.4$ and $\sigma_{max} = 0.8$. At zero defocus, the image contrast exhibited by contour density is higher. At a 0.5-μm defocus, the improvement from OAI is even larger.

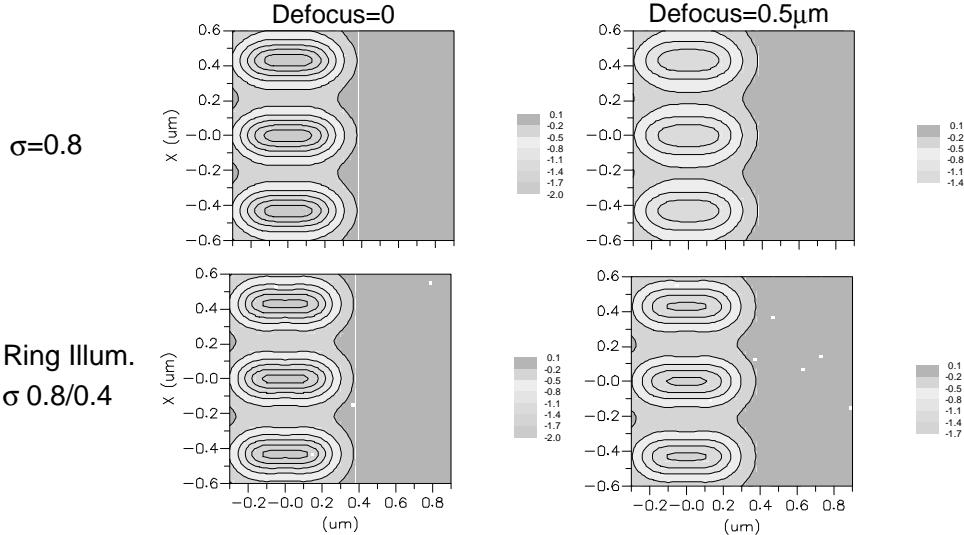


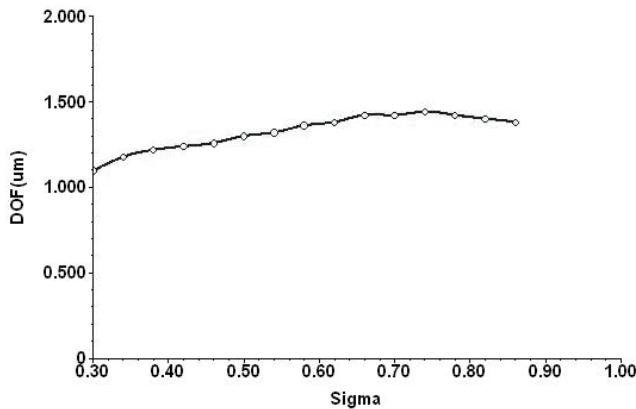
Figure 6.70 Constant intensity contours of disk and RGIs. Line-space pairs with $0.47-k_1$ pitch is studied. The physical parameters are $\lambda = 248$ nm and NA = 0.54.

The improvement in DOF is shown quantitatively in Fig. 6.71, still comparing DKI and RGI. Three $\sigma_{in}:\sigma_{out}$ ratios of RGI—namely, 1:3, 1:1, 2:3—are included. For DKI, σ approaches the optimum value around 0.8 slowly and insignificantly. Within the three RGIs, the narrowest ring, with $\sigma_{in}:\sigma_{out} = 2:3$, produces the most significant improvement in absolute DOF at the peak. The change of DOF as a function of σ_{out} is also more significant. In addition, the optimum σ_{out} shifts lower as the ring width narrows. At $\sigma_{in}:\sigma_{out} = 1:3$, the DOF versus σ_{out} behavior approaches that of DKI. The tradeoff to better imaging performance is higher coherence. Before OAI with a zoom lens was implemented, when a blocking aperture in the pupil plane was used to produce OAI, loss of exposure energy was also a tradeoff.

From Sec. 6.3.2.1, only one pitch is optimized for a given illumination angle. In Fig. 6.72, additional nonoptimized pitches are included to demonstrate the situation of simultaneously delineating multiple pitches. For the 1:1.5 line-space pairs, 950-nm DOF can be achieved at $\sigma_{out} = 0.84$. However, the DOF is limited to 470 nm by the features with larger openings. When all features are considered together, the common DOF is only 260 nm. Figure 6.73 shows the E-D windows that were used to produce Fig. 6.72. The windows for individual features are indeed very large. Only the common window suffers.

(a) Disk Illumination

0\LSMULT NA.6 OpSigma.74 MaxDOF1.44um@20%ELat WL365 CD.35



(b) Ring Illumination

0\LSMULT NA.6 SigIn=2/3 Out=.66 MaxDOF1.87um@20%ELat WL365 CD.35

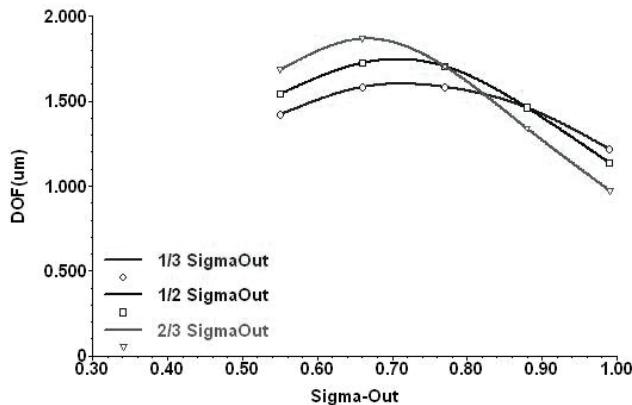
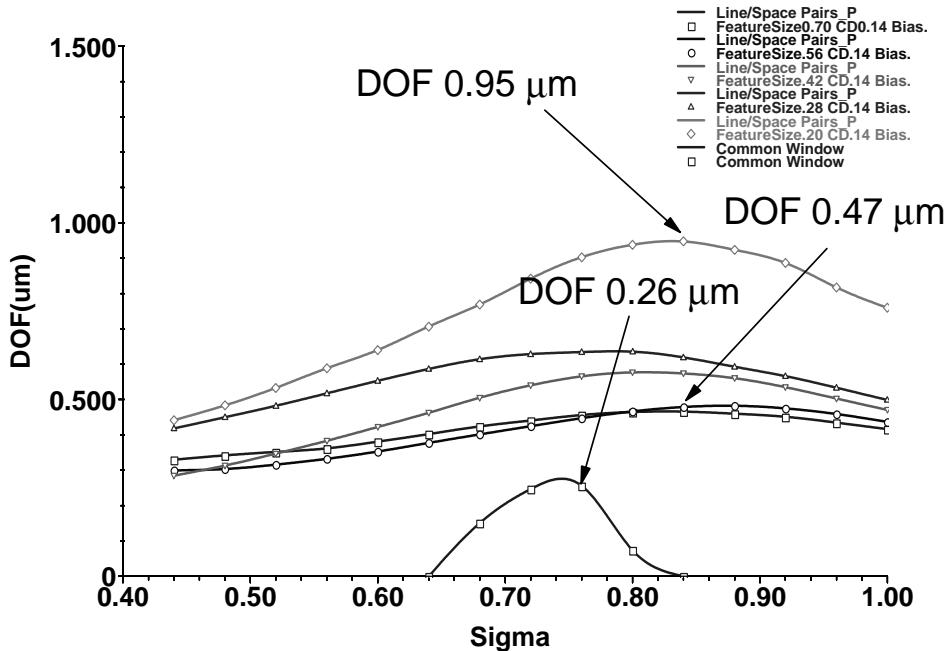


Figure 6.71 DOF versus σ_{in} and σ_{out} disk and ring illuminations, respectively, on a 0.35- μm 1:1 line-space pair, $\lambda = 365 \text{ nm}$, $E_{tol} = 20\%$, and $\sigma_{in}:\sigma_{out} = 1:3, 1:1, 2:3$.

In terms of feature shapes, Fig. 6.74 shows the E-D tree of line-space pairs, the line opening, opaque space, hole, and pedestal using DKI in Fig. 6.74(a) and QRI in Fig. 6.74(b). The line-space pairs with DKI have an extremely narrow and shallow E-D tree. Changing to AltPSM helps to enlarge the E-D tree, but not as much as QRI. With DKI, even though the E-D trees are very narrow, those for 1D features are close to each other. With QRI, they have wider separation. That is, the proximity effect is stronger. All features tend to straighten up with QRI, allowing for larger DOF.



Rectangular Window

Figure 6.72 DOF versus σ_{out} of 1:1.5, 1:2, 1:3, 1:4, and 1:5 line-space pairs with RGI. CD is opaque space at 140 nm. $\lambda = 248$ nm, NA = 0.65, EL = 8%, 6% AttPSM, $\sigma_{in} = 0.5$, and $\sigma_{out} = 0.76$.

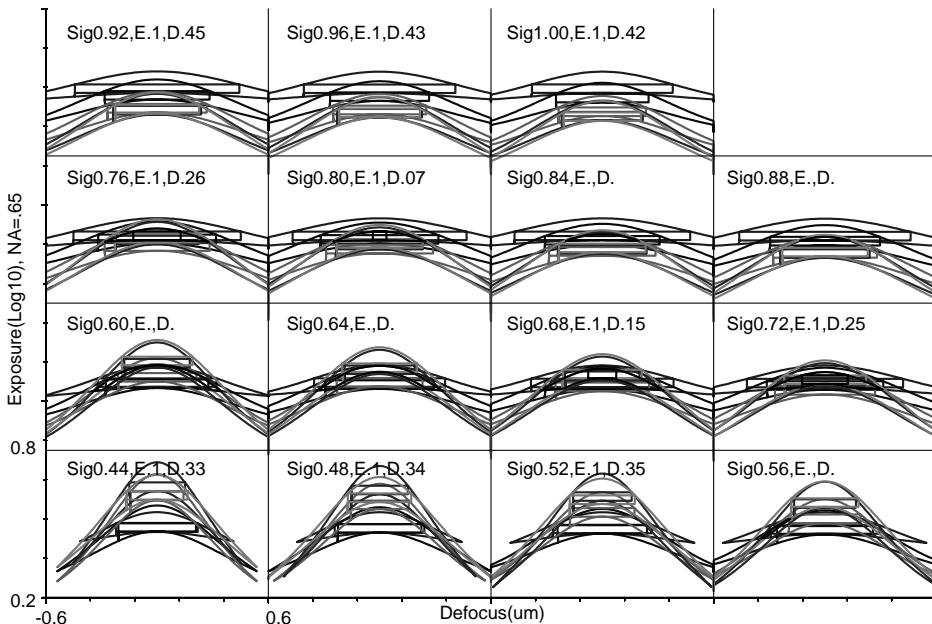


Figure 6.73 E-D windows of 1:1.5 to 1:5 LS with RGI from Fig. 6.72.

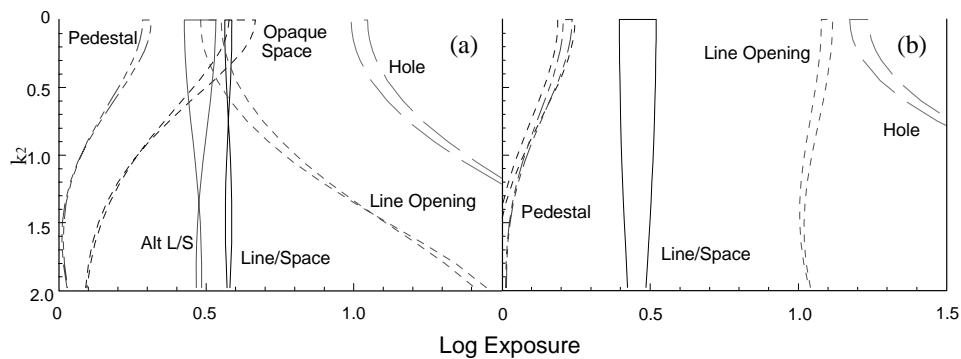


Figure 6.74 E-D trees for five $0.36 \lambda/NA$ representative features for: (a) DKI, $\sigma = 0.5$, and (b) QRI, four $\sigma = 0.04$ squares at the cross sections of 450 lines and the $\sigma = 0.94$ ring.

Using the same five features but at $0.48 k_1$, the effects of BIM, AttPSM, and two QRIs are compared. AttPSM makes the E-D tree of the openings wider and straighter. It also makes the trees wider for the opaque space and the pedestal, but it has more tilt. QRI with $\sigma_{in} = 0.44$ and $\sigma_{out} = 0.52$ does not improve the E-D trees over BIM or AttPSM with DKI, except for slight improvement of the LS pairs. The wider-spread quadrupoles produce extremely large DOF for the LS pairs at the expense of the image performance of other features. In this particular situation, AttPSM with DKI is preferred.

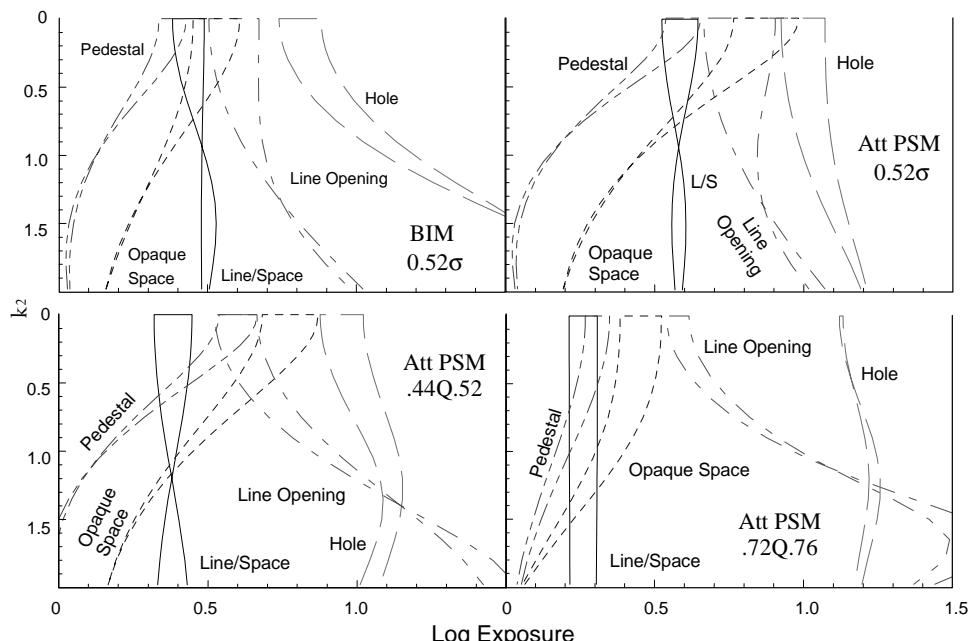


Figure 6.75 E-D trees of five features at $0.48 k_1$ with DKI and QRI from BIM and 6% AttPSM.

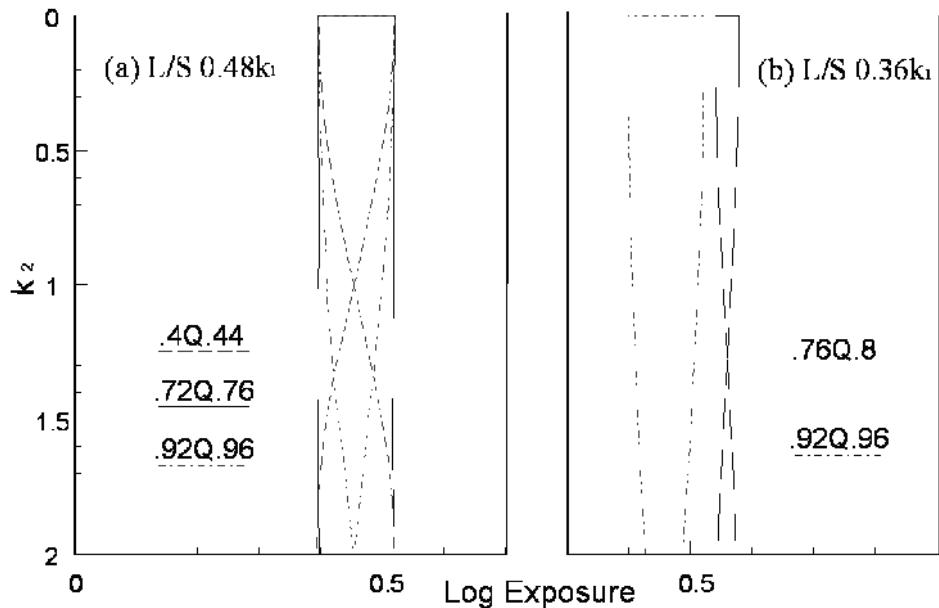


Figure 6.76 Change of E-D trees from different quadrupoles.

Figure 6.76 shows the E-D trees of different quadrupoles at (a) $k_1 = 0.48$ and (b) $k_1 = 0.36$ for line-space pairs. For 6.76(a), the quadrupoles are optimized at 0.72Q0.76. Larger or smaller σ s cannot achieve isofocus. For 6.76(b), because of the low k_1 , even 0.92Q0.96 cannot bring the image to isofocus.

A BigMaC mask shown in Fig. 6.77 serves as the test vehicle for 2D imaging performance. The minimum feature size (MFS) on this mask is $0.4 \lambda/\text{NA}$. At the middle, there are three MFS openings also separated by the MFS. The center opening ee'e''e'' is connected to a much larger $1.6 \times 3.2 \mu\text{m}^2$ opening, whereas the two openings abb''a and a'b'b''a''' on the sides have a 4:1 length-to-width ratio. Two more $0.4 \lambda/\text{NA}$ openings, cdd''c'' and c'd'd'''c''', are at the outside of the first three openings but are separated by $0.6 \lambda/\text{NA}$ from them. These two openings provide imaging information of an edge farther away than MFS and the other edge, in a completely dark background. There are inside and outside corners in large and small features on a BigMaC. The end of the two inner 4:1 openings are separated by $0.4 \lambda/\text{NA}$ to large openings ejk and its symmetrical counterpart in the other quadrants. The end of the two outer openings, cdd''c'' and c'd'd'''c''', are separated by the same distance but to small openings efj and its symmetrical counterpart in the other quadrants.

Figure 6.78(a) shows the diffracted image of a BigMaC at the focal plane, and $k_2 = 0.4$ for BIM and $\sigma = 0.4$. The solid line depicts the constant intensity contour at a -0.6 log exposure. The other contour is evaluated at a log intensity of -0.5 corresponding to 26% EL. The dotted line traces the $\pm 10\%$ linewidth

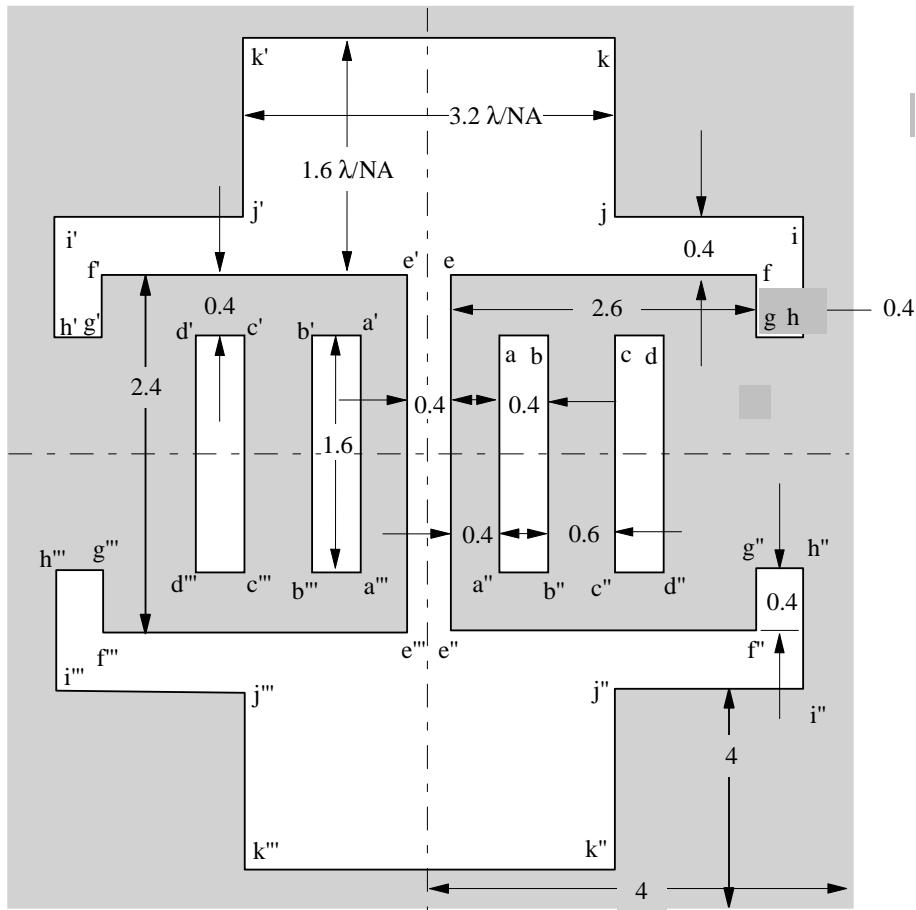


Figure 6.77 A big-small feature and corner-test mask (BigMaC).

control bracket of the original mask pattern. The log intensity level of the two image contours is chosen by centering them to the $\pm 10\%$ brackets. At such a low k_1 the diffracted image meets the linewidth control criteria only at limited areas, even with $k_2 = 0$, for example, small parts of edge bb'' , cc'' , ef , hi , ij , jk , and kk' with their symmetrical parts in the other three quadrants. Edge ee'' does not meet the criteria. The length of the openings $abb''a''$ and $cdd''c''$ are severely shortened. The latter shortens less because it is farther away from other features. Due to uneven edge misplacements, $abb''a''$ appears to have shifted to the left, whereas $cdd''c''$ shifts to the right. The bend $fghi$ is severely distorted. The centerline segment $ee'e'''e''$ suffers from ripples. When defocused to $k_2 = 0.4$, the image is even less useful. The focal and defocal images from a much lower illumination coherence corresponding to $\sigma = 0.8$ are shown in Fig. 6.78(b). The linewidth oscillation in the focal plane at the higher coherence is perceptibly reduced. Optimizing the exposure to produce an acceptable section of ee'' and aa'' results in losing the other sections, especially the arm $efghij$. The situation at $k_2 = 0.4$ is similar, though more severe.

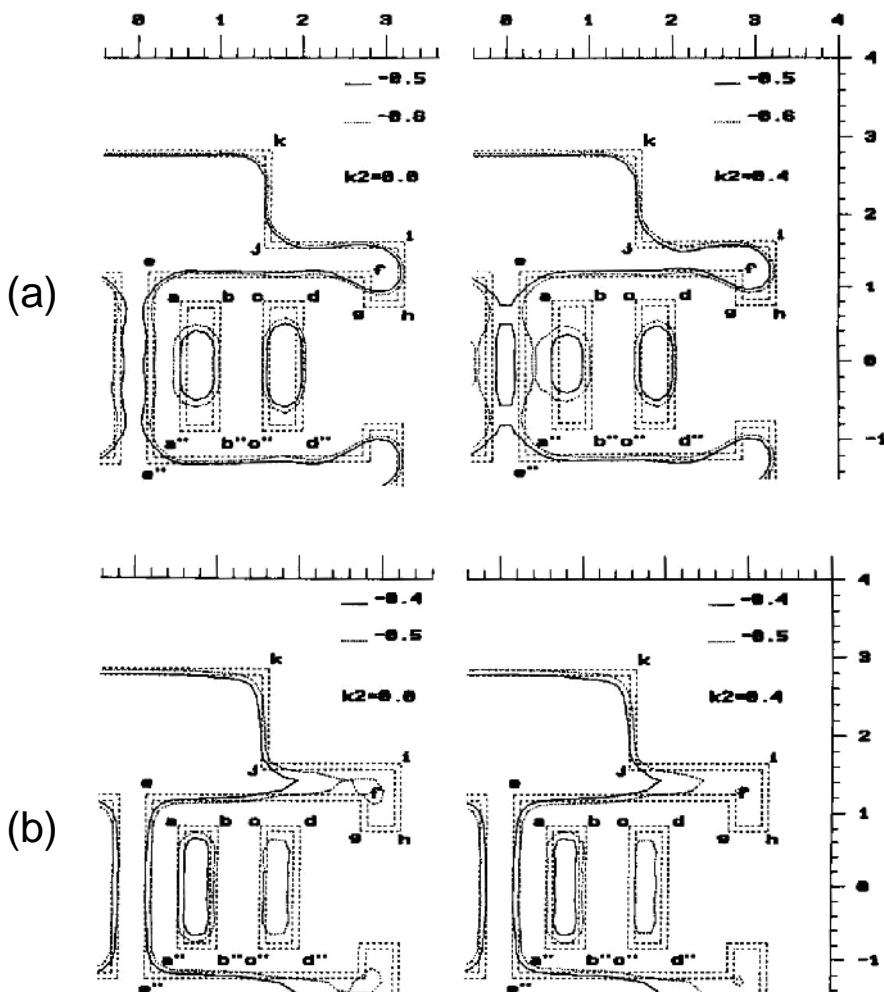


Figure 6.78 Constant intensity contours of the diffracted image from BigMaC using BIM
(a) $\sigma = 0.4$ and (b) $\sigma = 0.8$.

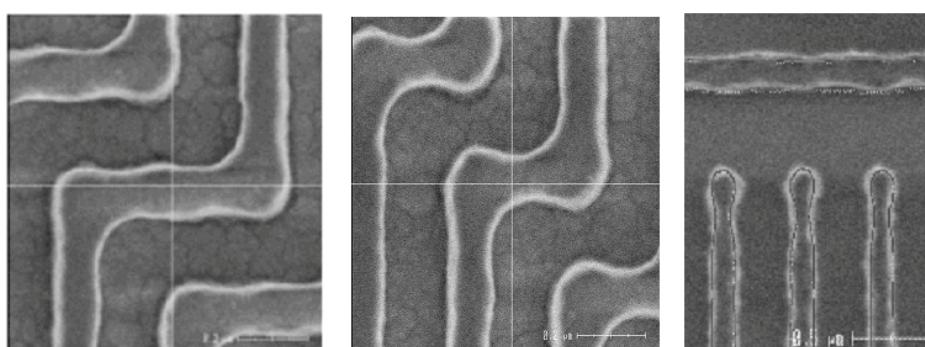


Figure 6.79 Diffracted image from the BigMaC BIM with RGI.

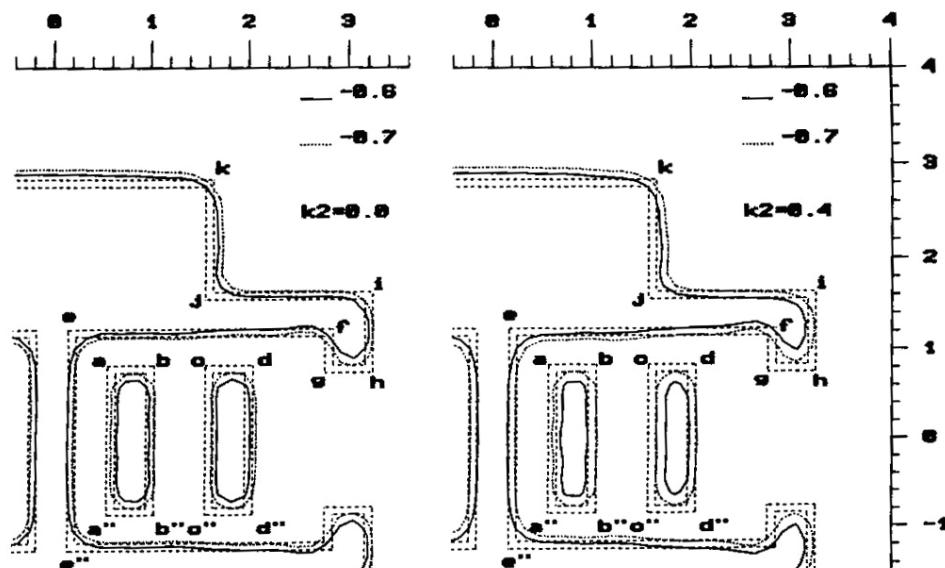


Figure 6.80 Ripples in the resist image in unoptimized OPC conditions.

Switching over to RGI, as shown in Fig. 6.79, the DOF is better than BIM at either σ . Pattern fidelity at the corners and line ends combine the good parts of BIM at either σ .

A severe problem with QRIs is ripples in the image, as shown in Fig. 6.80. They are typified with enlargement at the line ends and corners, followed with oscillation from wide to narrow into the center of the lines. The rightmost SEM micrograph shows superposition of the resist image with the simulated image to verify the fidelity of simulation. Ripples are most likely due to light interference of the two diffracted beams starting from the opposite poles. Unlike typical line-edge variations, OPC on ripples is not very effective. Even OPC optimized for ripples cannot fully remove them.

Ripples were predicted early on, from simulations using the BigMaC mask, as shown in Fig. 6.81. Ripples are severe along edge $k'kl..ee''$ and its symmetrical counterparts. The outer line opening $cdc''d''$ also suffers from ripples. The ripples are different from the waviness of edge ee'' in Fig. 6.78(a). The latter is confined only to the segment between e and e'' .

The image from Fig. 6.82 from AltPSM is not infected with ripples. The only deficiency here is the shortening of outer opening $cdc''d''$ due to the lack of a phase-shifting neighbor.

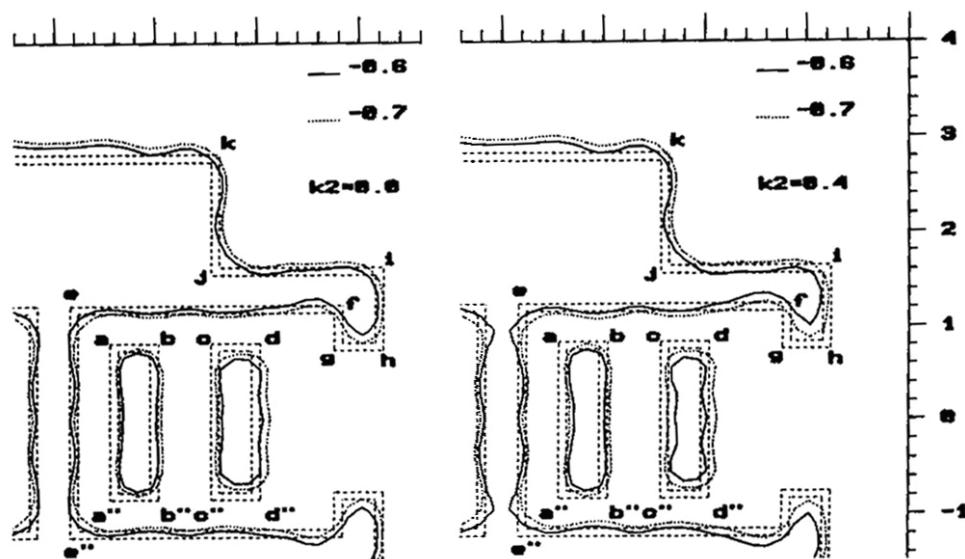


Figure 6.81 Diffracted image from the BigMaC BIM with QRI.

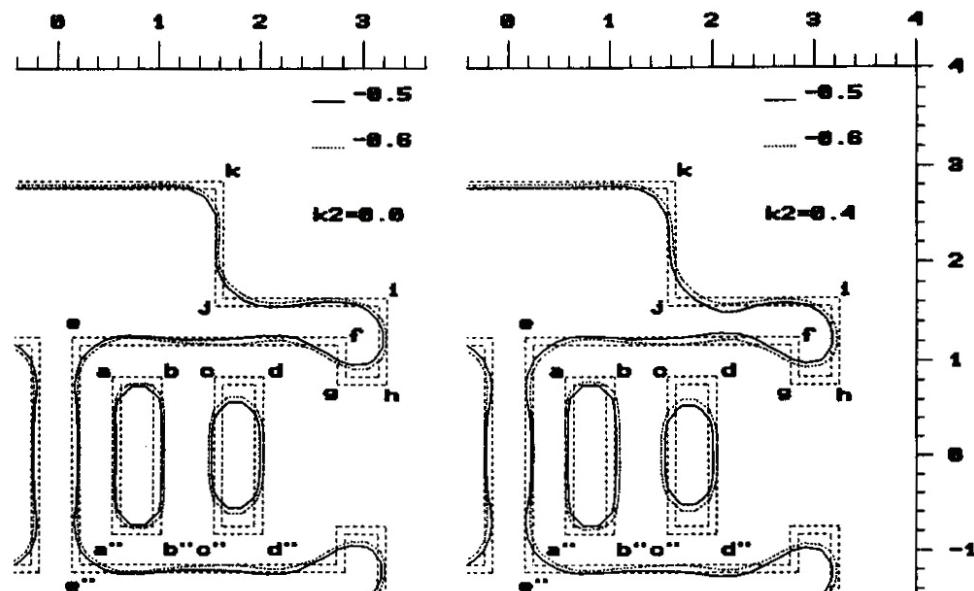


Figure 6.82 Diffracted image from the BigMaC AltPSM at $\sigma = 0.4$.

6.3.2.3 OAI and AltPSM

OAI and AltPSM behave similarly in many ways—namely, good isofocus behavior, the ability to double the resolution, and the inability to impact isolated features. But there are also differences to explore.

6.3.2.3.1 Comparison of imaging performance

A grating with $k_1 = 0.4$ line-space pairs is used to show the DOF improvement of OAI and AltPSM. Here, k_1 is used as a normalized lateral dimension related to a physical dimension W by $k_1 = W(\text{NA})/\lambda$. Similarly, k_2 is used as a normalized longitudinal dimension related to a physical dimension Z by $k_2 = Z(\text{NA})^2/\lambda$. Figure 6.83(a) shows the log intensity distributions and the E-D tree of 0.4σ DKI with a BIM, evaluated at $k_2 = 0, 0.2, 0.4, 0.8, 1.2, 1.6, 2, 2.4$, and 2.8 . At such a low k_1 , low image contrast and a shallow log slope in the intensity distributions are seen. The E-D branches are too closely spaced to support even 10% EL. Alternately phase shifting the mask produces a DOF of $0.8 k_2$ for 26% EL, as seen in Fig. 6.83(b). An even higher k_2 of 2.6 can be obtained by using optimized QRI. An isofocus situation is created, as seen by the identical log slope and edge location in the intensity distribution and in the almost horizontal E-D branches in Fig. 6.83(c).

6.3.2.3.2 Other considerations

Other than imaging performance, the most significant difficulty encountered with AltPSM is phase-shifting conflicts with arbitrary 2D patterns, as we discussed in Sec. 6.3.1.1. Shifting to a light-field mask to avoid phase-shifting conflicts necessitates a second exposure with a cutting mask to remove these lines. Hence, issues arise with overlay accuracy, reticle changing, throughput, and cost.

OAI is less expensive to use than AltPSM when the one-time cost of OAI is amortized over the cost of all of the AltPSMs. However, an inherent problem with OAI is that some of the illumination energy for high-resolution features is blocked by the finite acceptance angle of the lens NA. This not only leads to a longer exposure time, but also to an exaggerated proximity effect. The tradeoffs of the three RET are summarized in Table 6.2.

6.3.2.3.3 Combination of OAI and AltPSM

The working principle of AltPSM is restated here, and shown in Fig. 6.84(a), to show why AltPSM and OAI cannot be combined. For AltPSM, the phase of every other opening on the mask is shifted by 180 deg. The electric field behaves as if its spatial frequency is reduced to half of the original spatial frequency and with no 0th-order component. After passing through the lens, which would cut off the original frequency but not the reduced one, the frequency is doubled back into the resist image, which responds only to the intensity that is proportional to the square of the electric field amplitude. OAI cannot be combined with AltPSM to achieve an even higher resolution, as illustrated in Fig. 6.84(b). At the highest resolution achievable by either AltPSM or OAI, the off-axis beam turns one of the two phase-shifted beams out of the acceptance angle, while the remaining beam has a low spatial frequency, which cannot reconstruct the original frequency of the object.

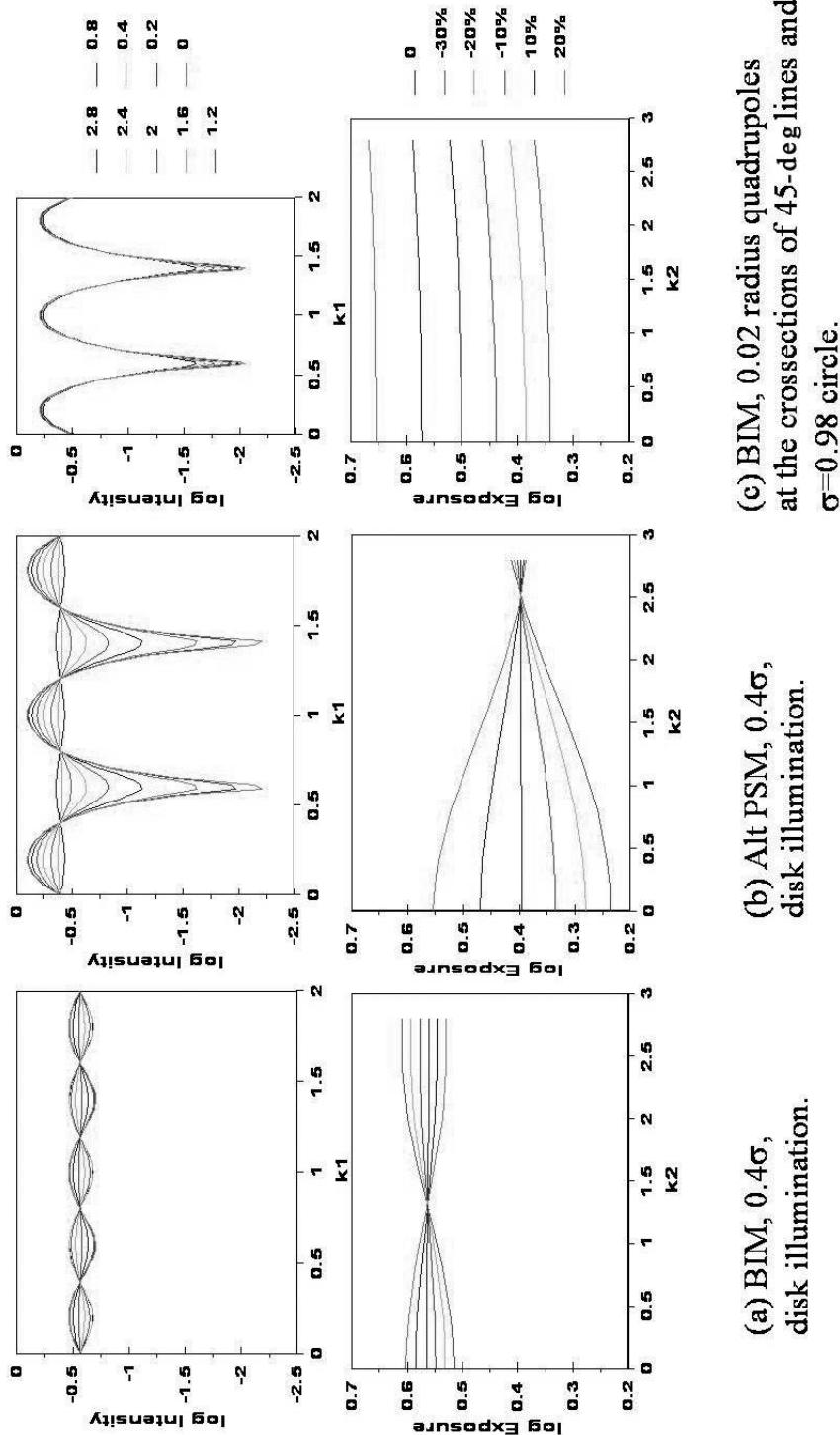


Figure 6.83 Comparison of BIM, AltPSM, and QRI for $0.4 k_1$ lines and spaces.

Table 6.2 Comparison of RGI, QRI, and AltPSM.

	RGI	QRI	AltPSM
Min. feature size	$0.4 k_1$	$0.3 k_1$	$0.3 k_1$
OPC required	Yes	Yes	Yes
Applicable patterns	High-density LS Any orientation	High-density LS x and y orientations	High-density LS Any orientation
Combination with low-density patterns	Use scattering bars	Use scattering bars	Combine with AttPSM
Cost	One-time cost of exposure tool	One-time cost of exposure tool	Extra cost for each mask

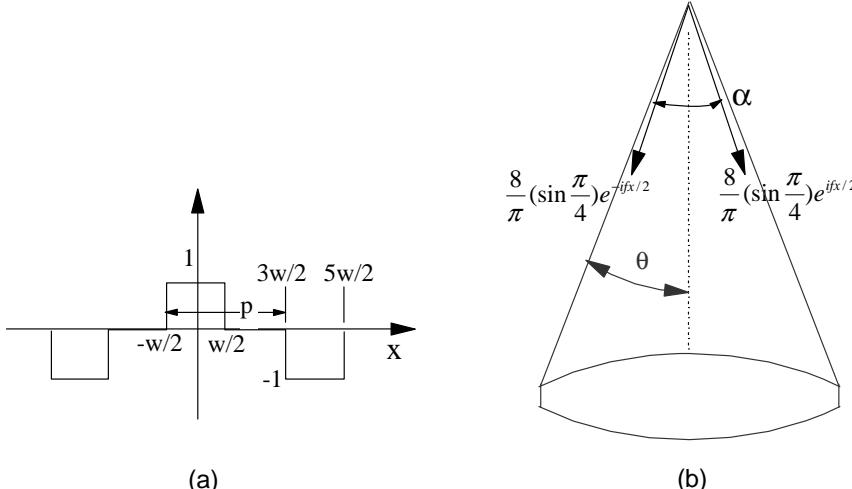


Figure 6.84 Electric field and spatial frequencies of DKI through AltPSM. OAI cannot tilt the two imaging beams for any gain in resolution.

Analytically, the electric field on the BIM shown in Eq. (6.18) is now split into two parts: the unshifted part $E_u(x)$ and the phase-shifted part $E_s(x)$. The latter is shifted by π .

$$E_u(x) = 1, \quad 2np - \frac{w}{2} \leq x \leq 2np + \frac{w}{2}, \quad n = 0, \pm 1, \pm 2, \dots \\ = 0, \text{ elsewhere.} \quad (6.28)$$

$$E_s(x) = -1, \quad 2np - \frac{3}{2}w \leq x \leq 2np + \frac{3}{2}w, \quad n = 0, \pm 1, \pm 2, \dots \\ = 0, \text{ elsewhere.} \quad (6.29)$$

$$E_0(x) = E_u(x) + E_s(x). \quad (6.30)$$

Following Eq. (6.19) by introducing an x translation in $E_s(x)$, the Fourier transformation of $E_0(x)$ normalized to w is

$$E_o(\xi) = \frac{\sin \pi \xi w}{\pi \xi w} \left(1 - e^{i2\pi p\xi}\right) \sum_{n=-\infty}^{\infty} \delta\left(\xi - \frac{n}{2p}\right). \quad (6.31)$$

Note that the delta functions are now centered on $n/2p$ instead of n/p . That is, the spatial frequencies are now reduced by a factor of 2. Also, $E_0(\xi) = 0$. Hence, there is no 0th-order component. When the lens NA limits n to ± 1 , the image electric field becomes

$$E_i(x) = \frac{8}{\pi} \left(\sin \frac{\pi}{4} \right) \left(e^{ifx/2} + e^{-ifx/2} \right). \quad (6.32)$$

There are only two beams symmetrically displayed with respect to the z -axis. Imposing OAI on AltPSM leads to

$$E_i(x) = \left[e^{ikx \sin \phi} + e^{-ikx \sin \phi} \right] \left[\frac{2}{\pi} \left(e^{ifx/2} + e^{-ifx/2} \right) \right]. \quad (6.33)$$

There is no 0th-order beam to be tilted. OAI either keeps the two beams within the lens acceptance angle or moves one beam out of it. Hence, OAI either does not change the resolution or loses it. The resolution cannot be doubled further. The situation is depicted by Fig. 6.84.

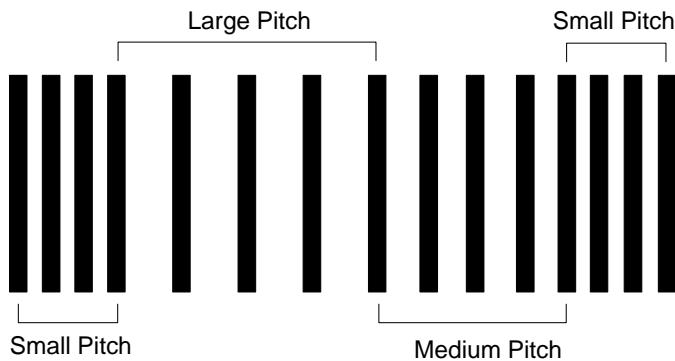
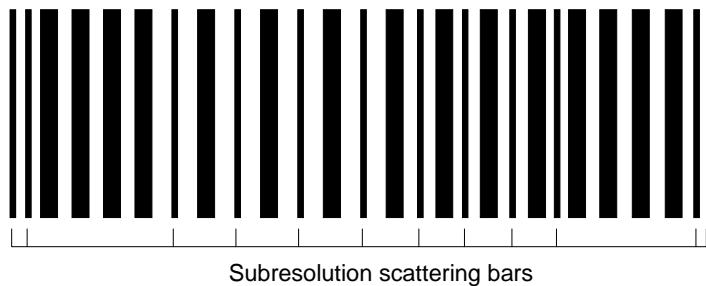
6.3.3 Scattering bars

From Fig. 6.85, it is desirable to pack the layout with dummy patterns to reduce the spread of pitches. The packing material should not be printed. Subresolution assist features serve the purpose. They are generally referred to as scattering bars (SB) or subresolution SB (SSB).

There is room in the medium and large pitches to insert dummy features. Figure 6.86 shows SSB inserted between features wherever space allows.

6.3.3.1 Imaging improvement from scattering bars

The width of SSB and their distance from the main feature can be fine tuned to optimize the imaging performance. A larger width provides a stronger pitch filling, but it must be optimized against SSB printing and allow sufficient space to insert the bars. Figure 6.87 shows the intensity distribution of a main feature surrounded by two scattering bars. The side lobes of the intensity distribution

**Figure 6.85** Pitch distribution.**Figure 6.86** SSB to reduce pitch dispersion.

have to be kept within an allowed limit to prevent its printing. The imaging improvement of SSB can be depicted with E-D windows, as shown in the same figure. When there is no SB, the DOF at $EL = 7\%$ is $0.6 \mu\text{m}$. With an SSB of typical width and spacing, the DOF is extended to $0.9 \mu\text{m}$. Further customizing of w and S brings the DOF to $1.26 \mu\text{m}$.

6.3.3.2 Complications

6.3.3.2.1 Restricted pitch

The OAI angle is usually set near the minimum pitch to gain DOF there. The DOF gradually decreases as pitch increases to an unacceptable level. As soon as space between features allows, SSBs are inserted. Normally, there is an ideal distance from the SSB to the edge of the feature that the SSB is intended to enhance. This is an edge SB (ESB). However, the distance is quite large, and DOF already drops beyond acceptability with a much smaller pitch. Central scattering bars (CSB) are placed at the center of two tight features, regardless of whether the position is optimal, as long as it does not create an unresolvable pitch or become part of the printed image. There may still be a small range of pitches whose DOF is so small that CSB cannot be used. This is the range of restricted pitch that should be avoided in the circuit design. Figure 6.88 shows

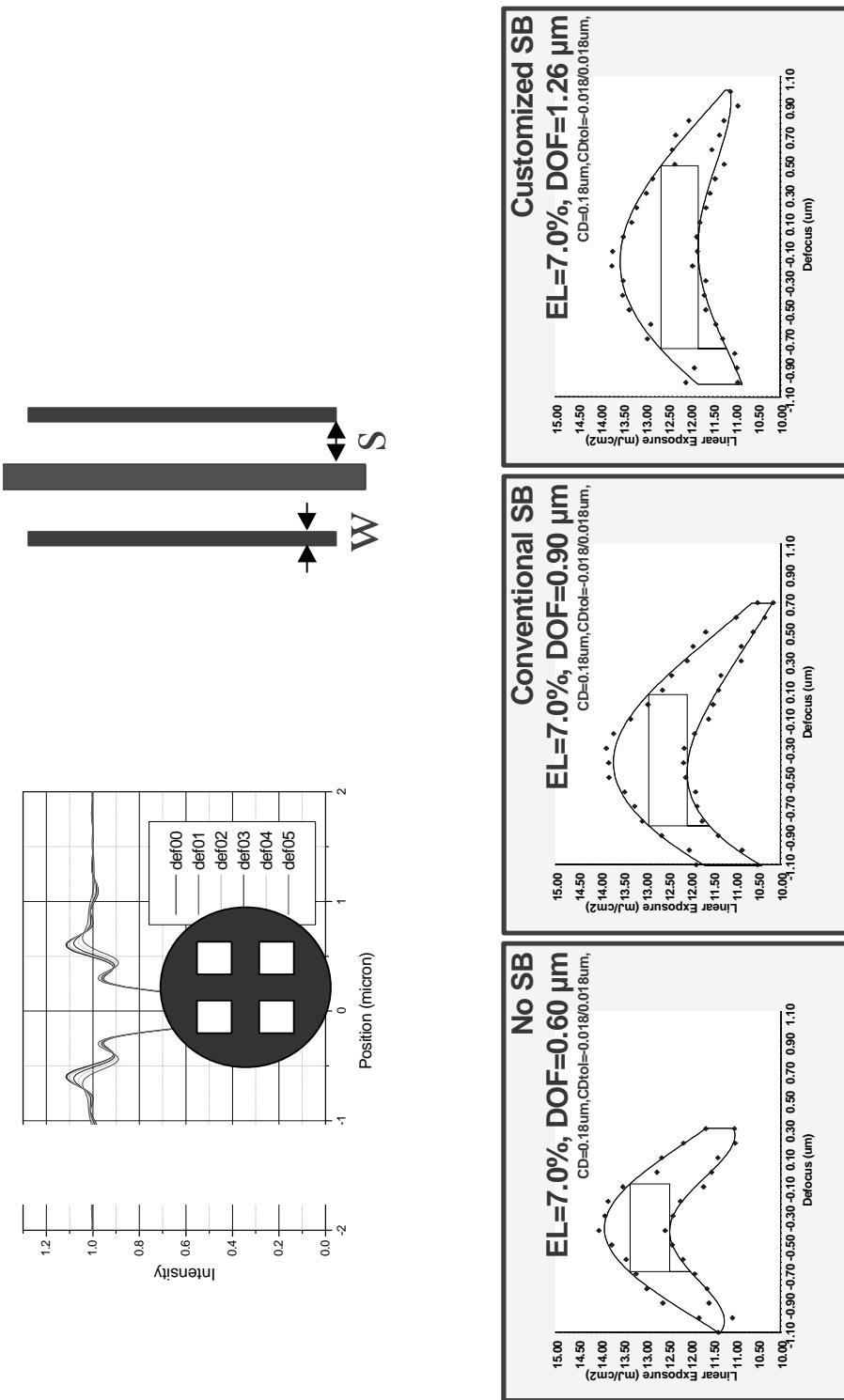


Figure 6.87 Scattering bars, intensity distribution, and E-D trees.

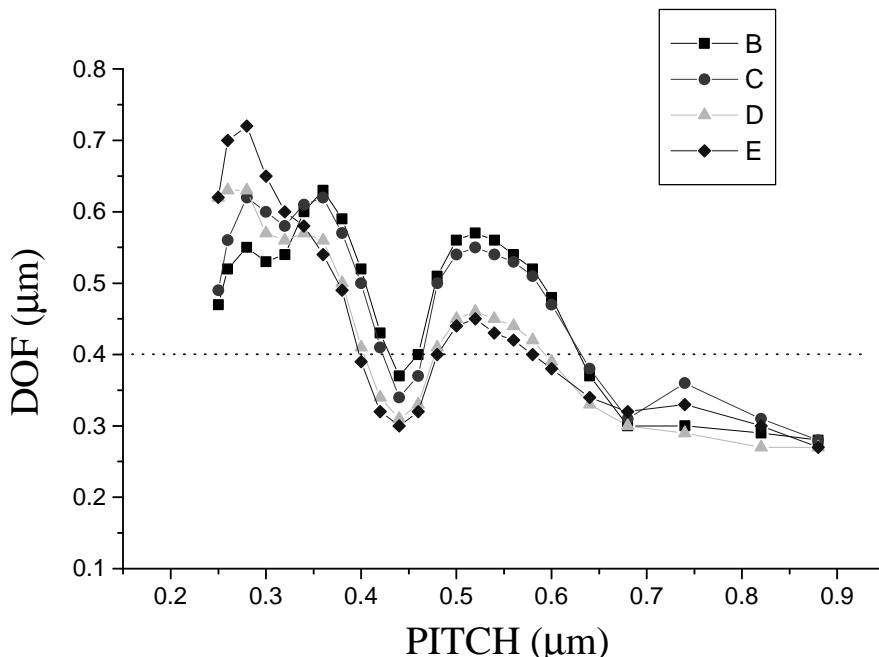


Figure 6.88 DOF versus pitch with QRI and SSB on AttPSM; $\lambda = 248$ nm, NA = 0.7, and EL = 8%.

the DOF of a range of pitches from AttPSM using four different QRI settings. The zones for CSB and ESB are marked. The desired 0.4-μm DOF line is drawn. The DOF falls below 0.4 μm in the pitch range between 0.4 and 0.48 μm. This is the range of the restricted pitch with AttPSM 0.55Q0.87, and SSB at a 248-nm wavelength, NA = 0.7, and EL = 8%. The other quadrupole settings are better. Note that DOF is less sensitive to changes in σ_{min} and that the only benefit of a larger σ_{max} is larger DOF at very small pitches.

6.3.3.2.2 2D features

Adding SSB to 2D features is neither as straightforward nor as effective as for 1D features. Figure 6.89 shows the design of two polygates with an enlarged portion for contact holes. The original design is perturbed at the edges corrected for optical proximity effects, which were covered in Sec. 6.1.1. Scattering bars are placed near the pattern edges, but cannot wrap around the features without breaking to prevent ghost images at SSB corners.

6.3.3.2.3 Mask-making concerns

The broken SSBs pose the risk of the resist falling off during mask making. Short SSBs are vulnerable to fall offs. When the SB is too narrow, resist fall off also happens. Therefore, the width of the SSB must be carefully optimized. A large width prints through, while a small width promotes fall off of the SSB resist during mask making.

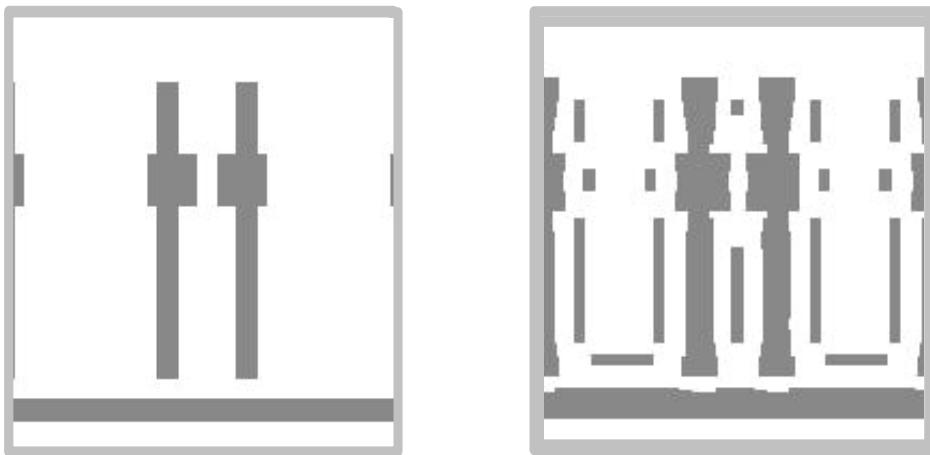


Figure 6.89 Scattering bars with optical proximity correction.

The small size of the SSB also poses a problem for mask inspection. Scattering bars should not be mistaken as defects by the mask inspection tool. The size of a printable defect usually falls in a similar dimension as that of SSBs. Therefore, the short scattering bars shown in Fig. 6.89 can be mistaken as defects.

6.3.3.2.4 Full-size scattering bar

By definition, the full-size scattering bar (FSB) has dimensions similar to the main features. Therefore, it is printable and must be removed with an additional trim mask. Is it worth the extra cost? Figure 6.90(a) is identical to Fig. 6.88, which illustrates QRI and SSB. Figure 6.90(b) displays a similar DOF versus pitch plot with identical imaging conditions, except that FSBs are used instead of SSBs. In the “no-SB zone” there is no change, which is expected. In the zone where SB can be inserted, FSB clearly outperforms SSB. However, it does not help the first group of restricted pitches. Figures 6.90(c) and 6.90(d) show the E-D trees for pitches of 260, 450, and 700 nm; the common E-D tree; and the corresponding E-D windows for SSB and FSB, respectively. The feature size is 113 nm, $CD_{tol} = \pm 9$ nm, $\lambda = 248$ nm, NA = 0.7, and EL = 8%. The E-D trees and corresponding windows are identical at pitches 260 and 450 nm, since no SBs are used. The E-D window for FSBs at $P = 700$ nm is larger than that for SSBs. The resultant common DOF is also larger for FSBs—480 against 390 nm. Whether it is worth the extra cost to use FSBs depends on whether the DOF required to support mask flatness, mask topography, lens field curvature, focusing tolerance, wafer leveling, wafer flatness, resist thickness, wafer topography, etc., can be met with the 390-nm DOF.

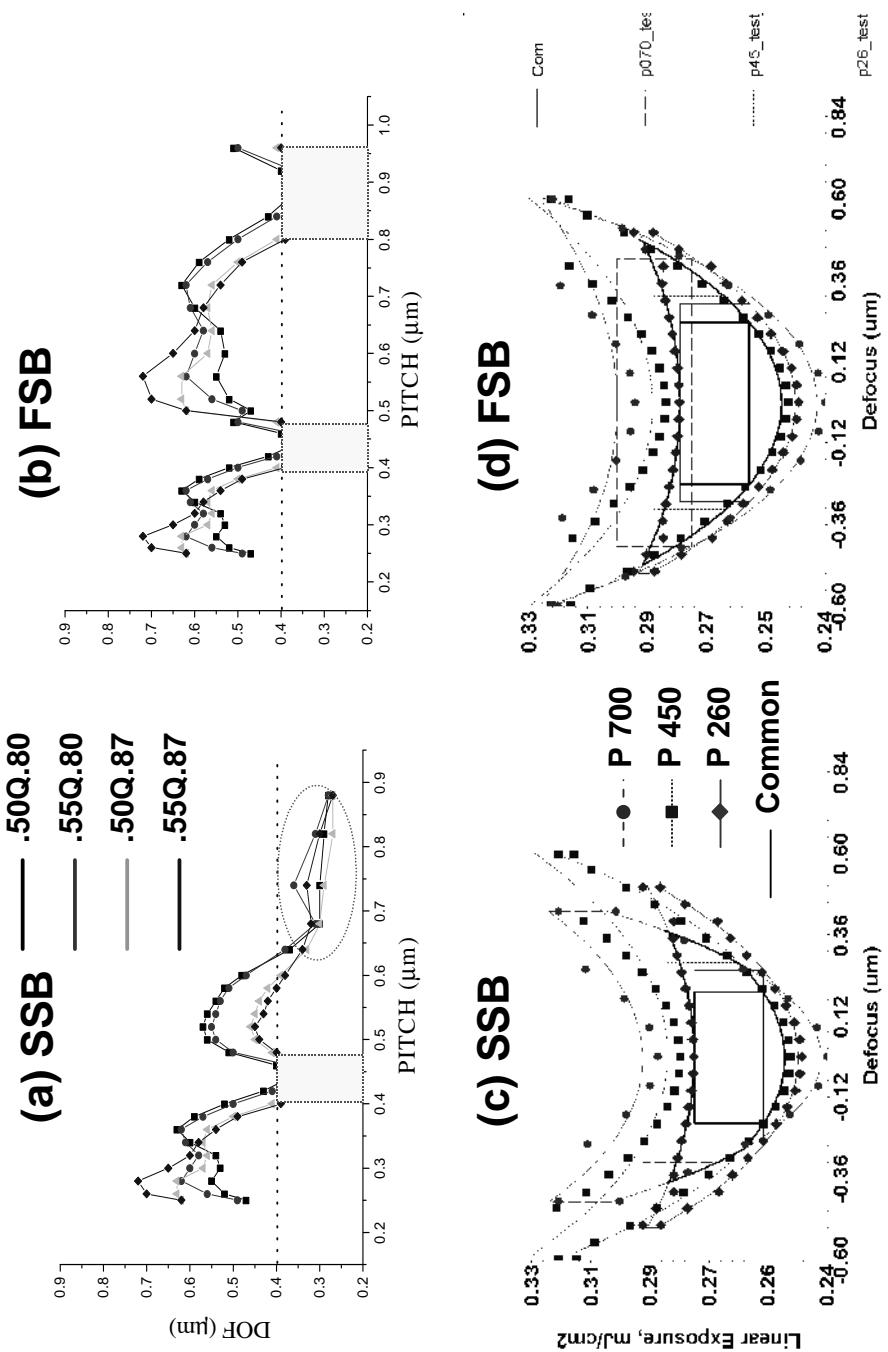


Figure 6.90 Imaging performance of SSBs and FSBs.

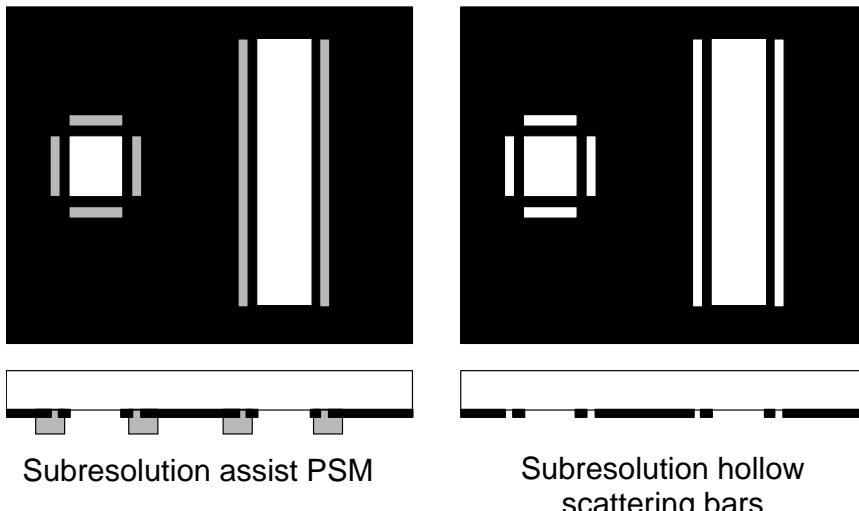


Figure 6.91 Subresolution hollow scattering bars and subresolution assist PSM.

6.3.3.3 Hollow subresolution scattering bars and subresolution assist PSM

Hollow SSB (HSSB) assisting opening features, such as holes and line openings, look very similar to the subresolution-assisted PSM. Figure 6.91 shows these two types of features. The only difference between them is phase shifting in the case of SA PSM. They share the same space limitation; in other words, there must be sufficient space for the assist feature. SA PSM helps isolated features more than HSSB. For spaces suitable for CSB, SA PSM is more vulnerable to printing through. The combination of OAI with SSB is well studied. Combining OAI with SA PSM requires more studies.

6.3.4 Optical proximity correction

When k_1 is large, all types of features can be imaged together with a large common E-D window. While k_1 is being reduced, the common window becomes smaller and smaller, even to the point of nonexistence. Note that optical proximity correction (OPC) is not a bona fide k_1 -reduction technique because it does not increase the image slope or double the spatial frequency like PSM or OAI. Similar to SSB and FSB, OPC is a low- k_1 enabler. After the k_1 -reduction techniques take care of the image slope and/or the spatial frequency, there may still be no common E-D window. OPC produces the common window.

There are many ways to bring the individual E-D trees together to improve the common E-D window, such as changing the exposure level of individual features, changing the feature size and pattern edge position, as well as distorting the pattern edge. Initially, attempts were made to avoid changing the mask pattern. Writing the mask already took a long time; further complicating the pattern was simply impractical. The pattern shapes started to be sorted. Early on, it was sufficient to only use holes in the contact hole layer. The choice was further narrowed down to single-size contact holes. Then, the range of contact

hole density was reduced. For line-space layers, no square or short rectangles were allowed, followed by restriction to single-sized CD, and then a limited range of pitches.

Line-end shortening is most problematic; it was extended on the mask for compensation, then enlarged with a hammer head. Pretty soon, complicated shape corrections were implemented. We will first introduce the cause of the proximity effect, then present the methods to correct it, from past to future.

6.3.4.1 The proximity effect

When feature size and pattern density approach the submicrometer level, the proximity effect sets in. There are proximity effects from optical imaging, from dry or wet etching, and chemical mechanical polishing. The optical proximity effect (OPE) is mostly due to diffraction, which induces a finite spread in the intensity distribution that can substantially influence the intensity distribution of close-by neighbors, as shown in Fig. 6.92. The diffracted intensity distribution of five 100-nm line openings in 100-nm spaces and one 100-nm line opening in a 500-nm space is simulated for defocus equal to 0 and 200 nm, using $\lambda = 193$ nm, NA = 0.75, and DKI is $\sigma = 0.8$. This corresponds to $k_1 = 0.39$. Because of the close proximity of the equal line-space pairs, the intensity level is raised. When the exposure threshold is chosen to be 0.3, the image linewidth of the equal line-space pairs is within the design limit, but that from the line with a large space is smaller, especially at defocus. Increasing the exposure by lowering the intensity threshold to 0.24 brings the defocus image of the line with a large space to the same width as the equal line-space pairs in focus, but the defocussed line-space pairs become wider.

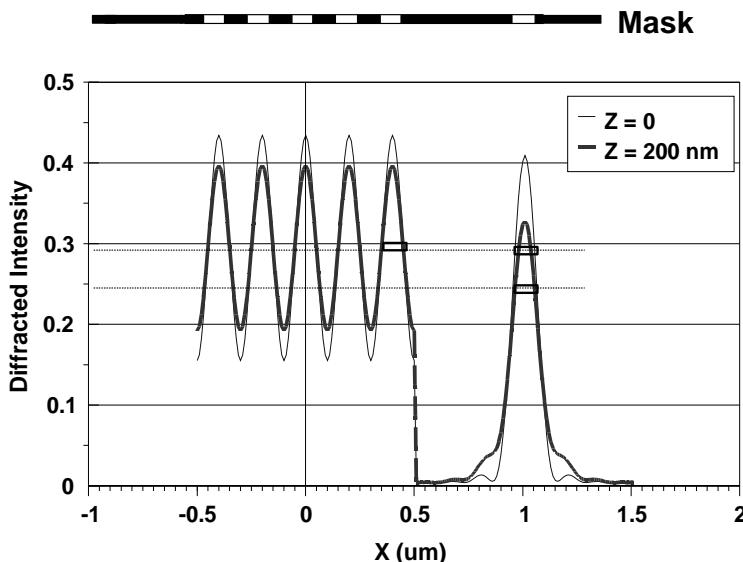


Figure 6.92 Proximity effects due to diffraction demonstrated with 100-nm openings at $\lambda = 193$ nm, NA = 0.75, and $\sigma = 0.8$.

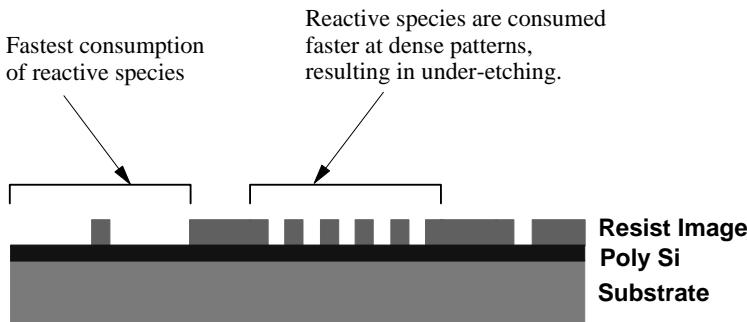


Figure 6.93 Proximity effects due to microloading during etch.

The proximity effect is not only optical; an e-beam has a finite spread in the resist because of forward and backward scatterings, resulting in a Gaussian distribution. Close-by neighbors are thus influenced. In dry or wet etching, closely spaced openings consume etchants more slowly than sparsely spaced openings. Figure 6.93 shows three levels of etchant consumption that lead to three levels of etch rate. Where the etchants are depleted sooner, etch rate is lower.

Microloading is also found in chemical mechanical polishing (CMP). Here, in addition to chemical depletion, the pressure difference due to the difference in pattern density is also a cause of CMP proximity effect.

Aside from 1D OPEs, there are also 2D OPEs, such as corner rounding and line-end shortening, as shown in Fig. 6.94. These 2D OPEs also result from the finite pupil size of exposure tools. Though rounding of corners is irremediable with OPC if the NA of the imaging lens is not allowed to increase, OPC is done for lithographic reasons other than pattern fidelity (e.g., making sufficient polyendcaps to avoid leakage or image areas for the landing of contacts or vias to reduce contact resistance). Figure 6.94 also shows other OPEs that are independent of being 1D or 2D. These are local CD changes, which when reaching the extreme, become missing patterns. Pattern touching is also a derivative of CD changes. In fact, the idea of OPC is to counter the local CD changes from OPE so that the pattern edges or areas meet circuit requirements.

6.3.4.2 Edge corrections

The most straightforward way to adjust the position of the image edge is to modify the pattern edge at the mask. Figure 6.95 shows how the mask pattern edge is modified to bring the edge of the wafer image as close to the desired position as possible. The line ends are enlarged with the so-called ‘hammerhead’ to maintain the linewidth and length. The inside corner is carved in while the outside corner is beefed up. The line edge facing a line end is moved slightly away from the line end. The two islands are enlarged to compensate for the reduced image size. Unfortunately, this particular correction did not anticipate the loss of one island to make it larger at the mask. Even though the loss is averted, the size of the two islands is not identical at the wafer image.

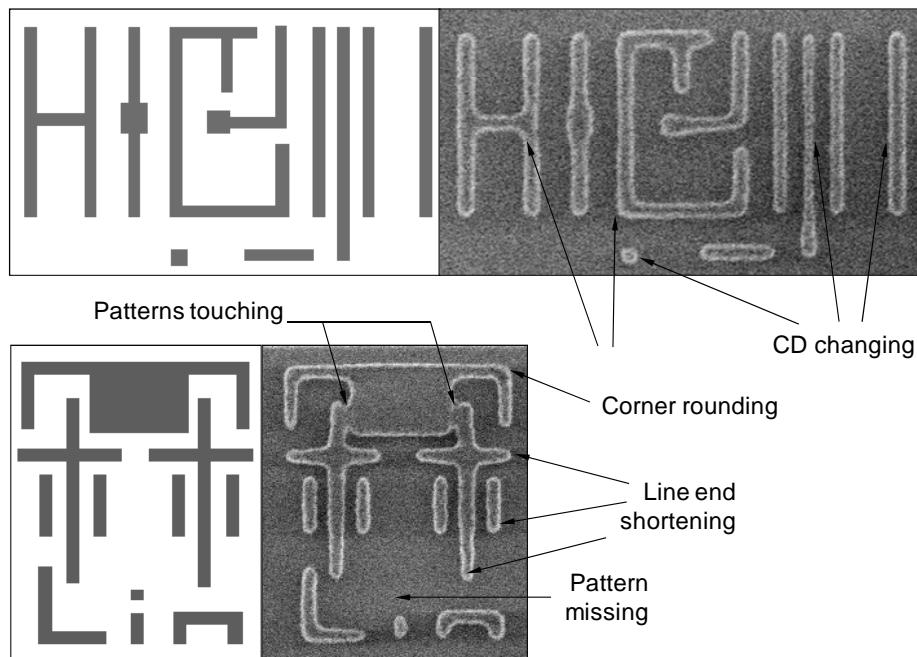
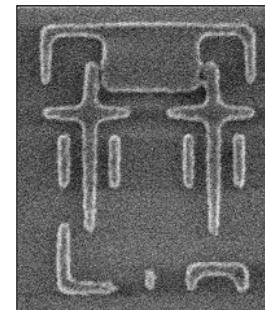
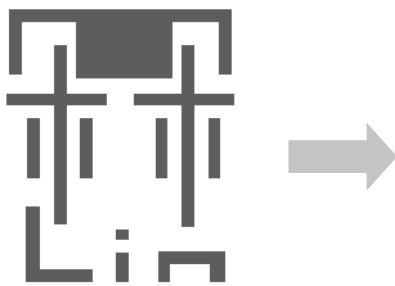


Figure 6.94 2D and other proximity effects.

Initial mask pattern



OPC mask pattern



Resist image on wafer

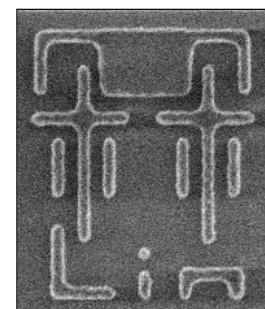


Figure 6.95 OPC by edge bias.

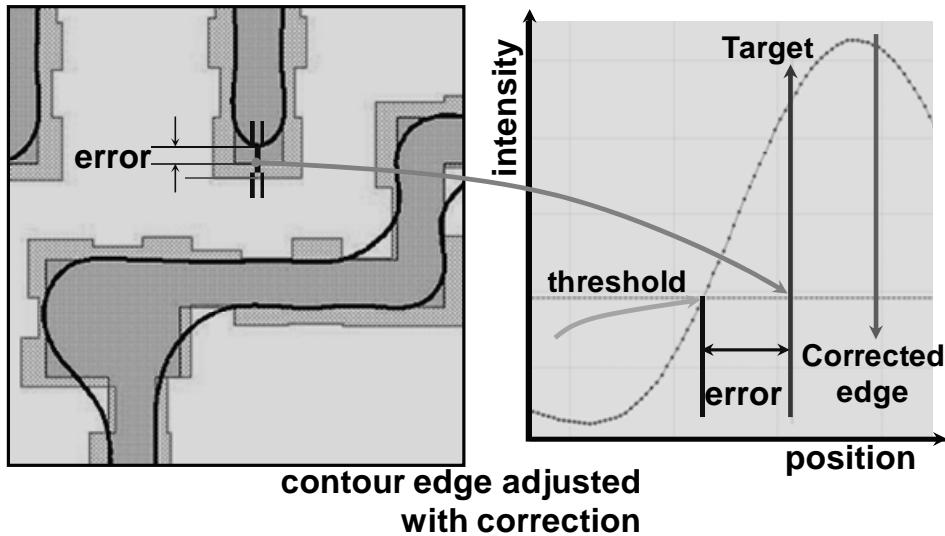


Figure 6.96 Correction by moving the edge to meet the threshold.

The basis of edge correction lies in relocating the edge of the mask pattern so that the image edge falls on the desired position. Figure 6.96 shows the direction in which to move the pattern edge.

The exact amount to move is related—but not equal—to the amount of edge error. There are three ways to assign the changes at the mask pattern edge: rule-based, model-based, and manual fine tuning. The latter is suitable only for highly repeating cells, such as those found in memory devices and circuits.

6.3.4.2.1 Rule-based OPC

Rule-based OPC is done via experimentally determined OPC rules. Such rules include the bias table for 1D patterns and serif or hammerhead rules for 2D patterns. With the help of design rule checking on CAD tools, OPC can be performed automatically. For correction efficiency, OPC rules are established by only considering the pattern itself or its nearest-neighboring patterns, thus it is only applicable for correcting loosely distributed ASIC or patterns with a fixed environment, such as memory cells. Figure 6.97 shows a bias table to dictate the change of features according to their width and space and the distance from a neighboring feature. The lighter shades indicate addition to the feature; the darker shades indicate subtraction.

Rules can get quite complicated. Take the hammerhead rule, for example. Figure 6.98 shows that when the space between a line end and another feature is larger than $0.20\text{ }\mu\text{m}$, a hammerhead with the dimension shown should be added. However, the added hammerhead is not allowed to protrude into the $0.20\text{-}\mu\text{m}$ zone, resulting in the partial hammerhead at the rightmost part of the figure.

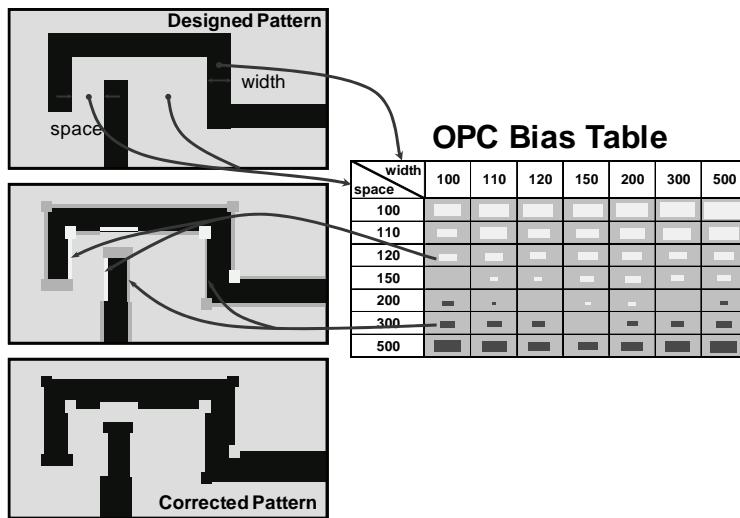


Figure 6.97 Rule-based OPC from a lookup table.

6.3.4.2.2 Model-based OPC

The model-based OPC predicts the wafer image from the mask pattern using an equation similar to Eq. (3.52). The model is a transfer function connecting the intensity at the mask plane to the intensity at the wafer plane:

$$I_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_o(x_o, y_o) H(x_i, x_o, y_i, y_o) dx_o dy_o . \quad (6.34)$$

Actually, I_i can also be governed by Eq. (3.47) or (3.53), depending on the degree of the illumination coherence. However, the functional connection is more

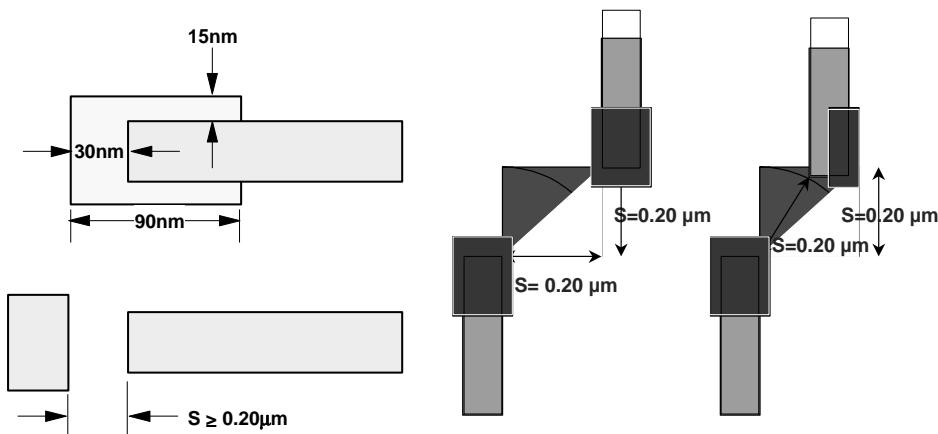


Figure 6.98 Hammerhead as example of rule-based OPC.

than just optical. The transfer function includes the effect of anything that can affect the final image. After the formation of the optical aerial image, the resist exposure, baking, and development behaviors take place, followed by etching. The transfer function in Eq. (6.34) must encompass all of these effects. Theoretically, the governing equation of each phenomenon must be used as follows:

- To transfer I_o through the aerial image stage
- To propagate through multiple reflections and absorptions in the resist
- To create the acid in the resist from the exposure
- To amplify the acid through PEB if the resist is a CAR
- To diffuse the acid during PEB
- To describe the time development behavior of the resist to produce the after-development image (ADI)
- To account for etching to transfer the resist image to the after-etch image (AEI).

At the aerial image generation stage, the governing equation can be Eq. (3.47), (3.52), or (3.53). All other governing equations are assumed to be in the form of Eq. (3.52). In reality, it is impractical to use so many stages of calculation due to the computation time and the lack of an exact analytical description of the physical phenomena. One would use an experimental result to establish an empirical model to transfer the mask image into the wafer image through Eq. (6.34). This is the starting point of model-based OPC.

Fitting kernels to the experimental data generates the model. As shown in Fig. 6.99, representative patterns with relevant points at the pattern edges are measured. The data is used to evaluate the kernels for OPE.

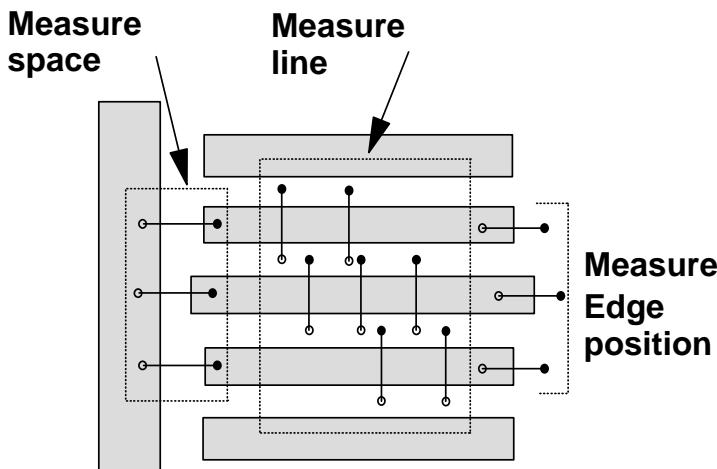


Figure 6.99 OPC test site. The measured values from the wafer image are used to evaluate the kernels.

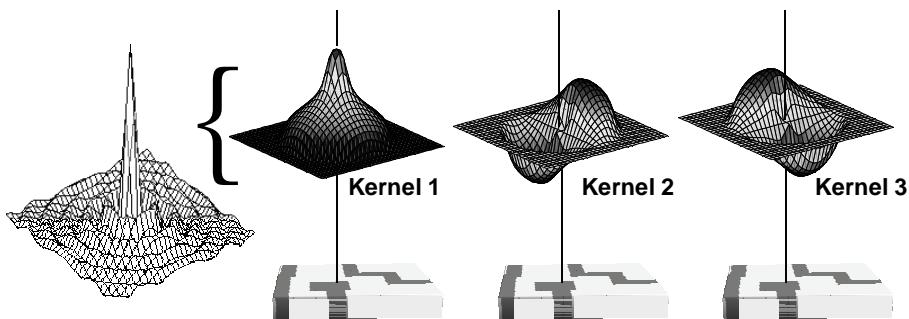


Figure 6.100 Model and kernels.

For lack of a theoretically derived closed-form expression, a set of eigenfunctions with adjustable coefficients is used to represent the transfer function H . These eigenfunctions are called kernels, as they are used as components of H in Eq. (6.34). Figure 6.100 shows a typical kernel and example of four kernels. The coefficients for the kernels are evaluated by fitting the kernels to experimental data. Hundreds to thousands of data points must be taken when using many high-order kernels, thus model fitting is a time-consuming process requiring very careful and accurate metrology.

The accuracy of the model is judged by the fit between the resist image and the model-generated image. Figure 6.101 shows the image-edge contour generated from the given model, superimposed on the resist image, confirming both the accuracy of the model and the fidelity of the SEM tool taking the resist image. Contour plots are time consuming. Other ways to determine the fitting accuracy is to plot the CD and edge intensity as a function of pitch for 1D patterns and plot line-end position as a function of separation.

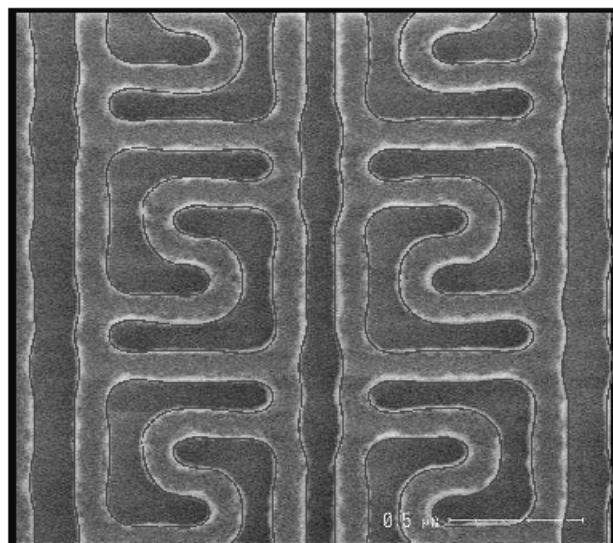


Figure 6.101 Model-generated edge contour superimposed on the resist image.

Once the model is established, the first-pass wafer image from the mask can be calculated and compared with the target image, which is usually the image edge position at selected target points. The mask pattern is adjusted to reduce the difference, calculated, and compared again until the difference is acceptable, as depicted in Fig. 6.102. The whole process starts with the layout from the designer, $I_o^0(x_o)$. It is used in two operations. First, the target point distribution $I_T(x_i)$ is generated from the given layout. The target points run along the pattern edges on the layout that will be compared with corresponding points in the wafer image $I_i^n(x_i)$, which is generated by applying Eq. (6.34) to $I_o^0(x_o)$ with the transfer function $H(x_o, x_i)$; this is the second operation on $I_o^0(x_o)$. The first corrected layout $I_o^1(x_o)$ is made based on the difference from the comparison, and it goes through the transfer function to generate $I_i^1(x_i)$ for comparison with $I_T(x_i)$ for the second round. If the comparison shows an acceptable difference, the new layout $I_o^1(x_o)$ is accepted as the optical proximity corrected layout for mask making. If the difference is not acceptable until $n = N$, then $I_o^N(x_o)$ is the accepted optical proximity corrected layout.

The target point is an important part of model-based OPC. Figure 6.103 shows how a pattern is dissected at its edges. There is one target point in each dissection. Note that the dissections and the target points need not be equally spaced. The edge section may move in or out depending on the difference between the location of the diffracted wafer image and the target image. In this figure, target point 1 in the first dissection is moved out. The amount of movement indicated by the correction vector is a function of the difference in the edge positions according to the model.

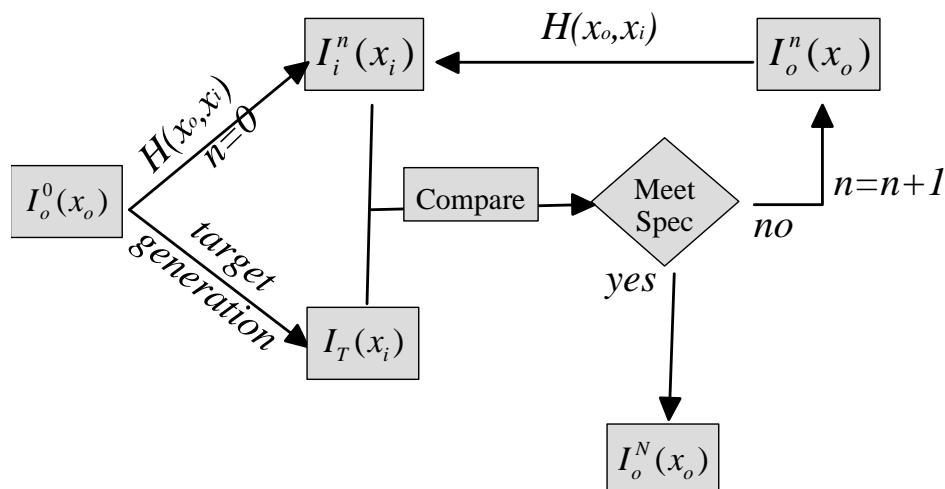


Figure 6.102 Block diagram showing the flow of model-based OPC.

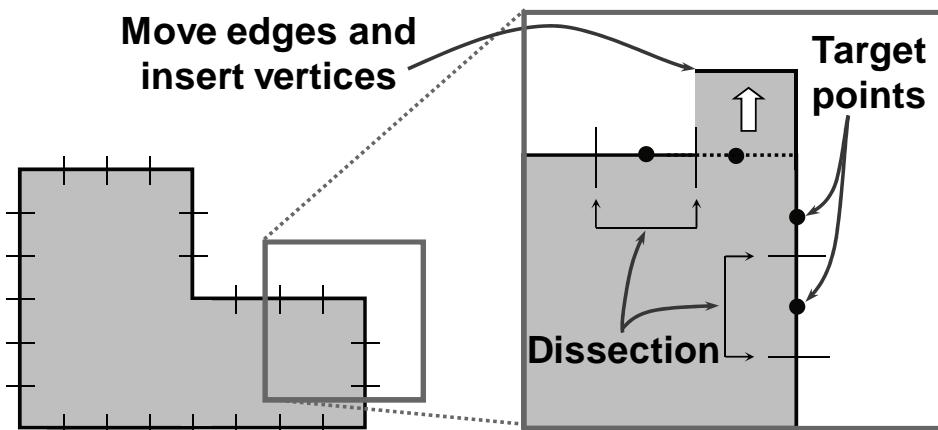


Figure 6.103 Dissections and target points.

Figure 6.104 shows the pattern from Fig. 6.103, the correction vectors applied to the entire pattern, and the resultant optical proximity corrected layout. The placement of the dissections and target points is very important. The wrong placement leads to slowly converging iterations. If converged, the OPC improvement may meet specs only at the target points, but not in between. Too many dissections prolong the cycle time. Days of running time at hundreds of central processing units (CPU) are often required for a complicated layout, even with optimally placed target points.

Reducing the OPC time while maintaining correction accuracy is very demanding. In every node advancement the pattern density increases by a factor of two. If the computation efficiency is not improved, the OPC time will increase by a factor of two. One way to reduce the OPC time is to take advantage of the hierarchy in the original design, as illustrated in Fig. 6.105. One seeks the identical patterns in the layout and corrects one pattern, which carries over to the other patterns. Unfortunately, the design hierarchy was established to improve

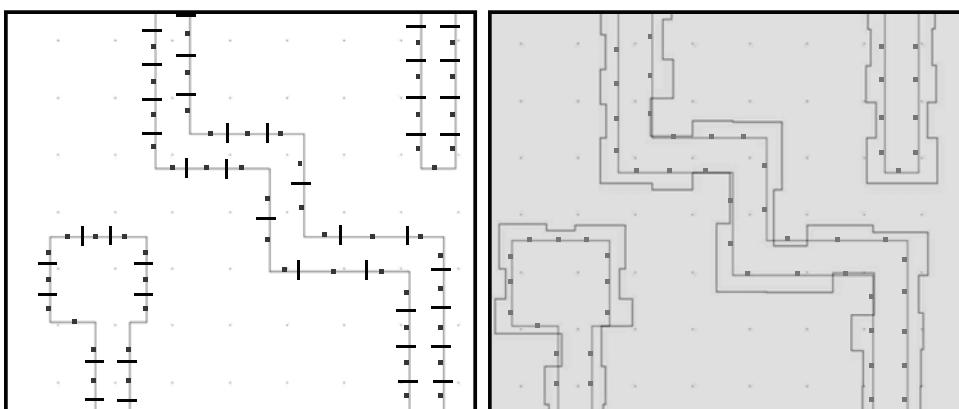


Figure 6.104 Correction vectors applied to an entire pattern.

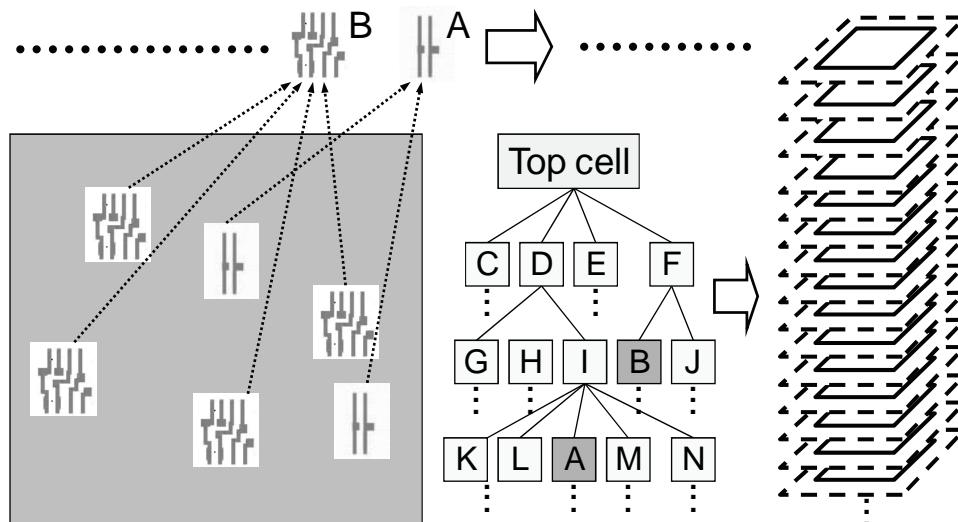


Figure 6.105 Use a hierarchy to easily correct identical patterns for all layouts.

the efficiency of design, not for OPC. A usable hierarchy for OPC must include the layout of the identical patterns within the OPE range. It also must be identical for all representative corrections. Also, the algorithm to handle the hierarchy for OPC is not yet mature. Hierarchical OPC is still not very dependable.

6.3.4.3 Local-dose OPC

So far, edge corrections have been covered. However, this is not the only correction method. One can correct for proximity effects by local dosage—optical proximity correction (LD-OPC) just as in the case of directwrite e-beam, where the beam current or dwell time can be set point-by-point. For optical projection, a local dosage variation is built on the mask. Figure 6.106 shows the combination of dense and isolated openings taken from Fig. 6.92. A gray-level mask having 12.9% absorption in the dense area is used to lower the intensity distribution of the dense line openings by 0.056. The openings in the image are now much closer in dimension to each other at the new threshold of 0.244 from 0.3 in Fig. 6.92. This technique is called threshold leveling.⁴⁶

A more sophisticated LD-OPC is shown in Fig. 6.107. Subresolution half-tone features are properly distributed on the mask to strengthen outside corners and to thin down inside corners. Line ends and contact holes are beefed up. All corrections are within the confinement of the original openings. This way, features do not run into each other as in the case of edge correction OPC. The distribution is automatically generated with an algorithm given in Ref. 47. Figure 6.108 shows a printed result from the half-tone corrected mask. A $0.58-k_1$ optical proximity corrected image is better than the $0.78-k_1$ uncorrected image.

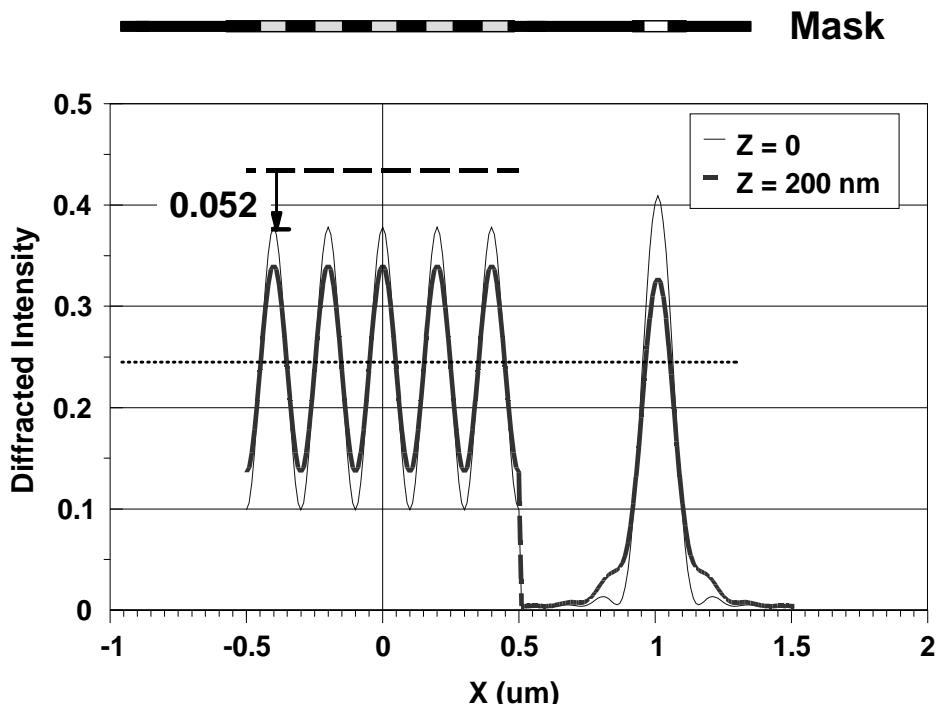


Figure 6.106 Threshold leveling.

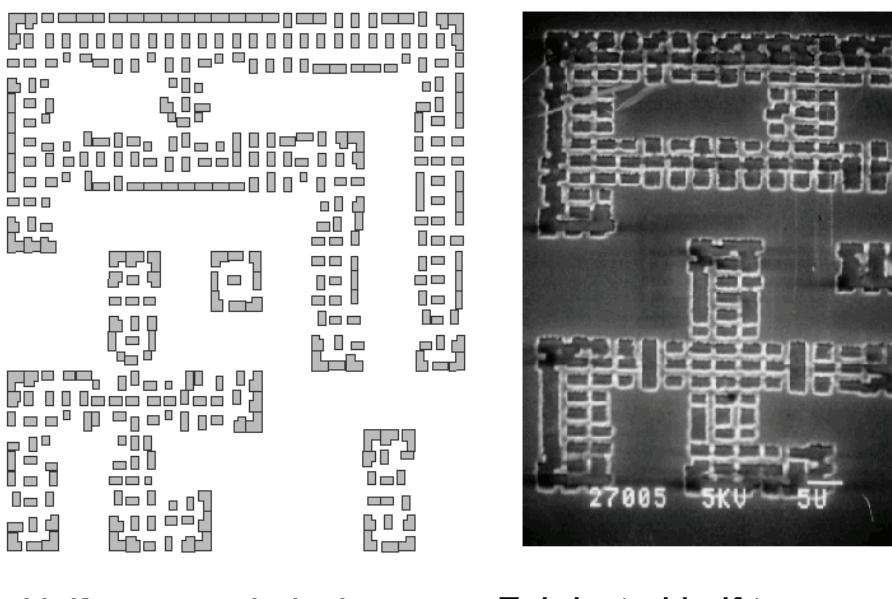


Figure 6.107 Half-tone mask used for local dosage correction.

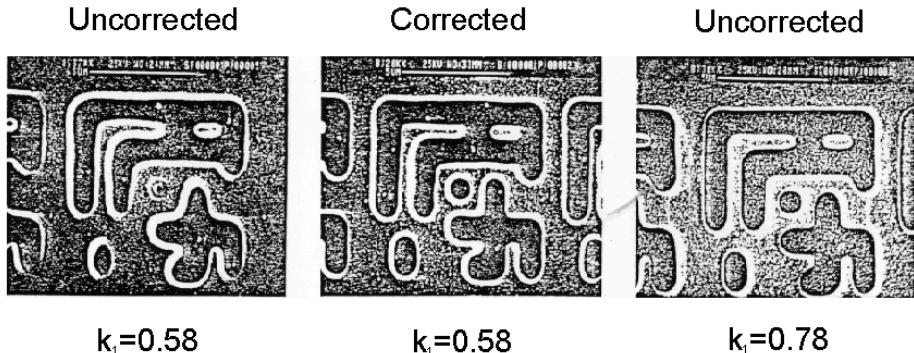


Figure 6.108 Half-tone LD-OPC results.

It is surprising to find the need of OPC, even with the relatively high- k_1 imaging condition. From ideal conditions, $k_1 = 0.58$ does not need much correction. However, during the early days of optical lithography, lens aberration and stray light were comparatively high. They tend to reduce the contrast of the optical image, resulting in equivalence to lower- k_1 images. LD-OPC restores the image to higher k_1 , restoring the lost resolution. Similarly, a stringent surface accuracy and smoothness requirement at the 13.5-nm EUV wavelength leads to severe stray light and difficult aberration correction. The equivalent k_1 in EUV images is reduced.

6.3.4.4 Full-depth OPC

Most OPC methods optimize the wafer image at the focal plane-optical proximity correction (FP-OPC) and take whatever improvement at the off-focus image. To get the most benefit of OPC, the entire range of defocus to be imaged should be considered. To make this point, Fig. 6.109 shows the E-D trees of three different resist thicknesses: 750, 820, and 960 nm. The feature is a 250-nm resist line delineated at $\text{NA} = 0.55$, $\lambda = 248 \text{ nm}$, and RGI $\sigma_{in}:\sigma_{out} = 0.4:0.8$. The E-D trees are evaluated at $\text{CD}_{tol} = \pm 10\%$ and $\text{EL} = 4\%$. The common E-D windows clearly do not line up with the center of the individual E-D trees. This means the center of focus is different for each resist thickness. These resist thicknesses occur simultaneously due to topography on the wafer. Full depth-optical proximity correction (FD-OPC) is performed by adjusting the bias of the features on the mask to center the E-D window while keeping the same 250-nm CD. The DOF is now $1.04 \mu\text{m}$ instead of $0.4 \mu\text{m}$.

Figure 6.110 shows another example of the need for FD-OPC. Equal line-space pairs and an isolated line opening, both at $k_1 = 0.66$, are imaged together. DKI at $\sigma = 0.5$ is used. $\lambda = 365 \text{ nm}$, $\text{NA} = 0.48$, $\text{CD} = 0.5 \mu\text{m} \pm 10\%$, and $\text{EL} = 5.3\%$. The trees have centers of focus at 154, -231, and 308 nm, as indicated in the figure.

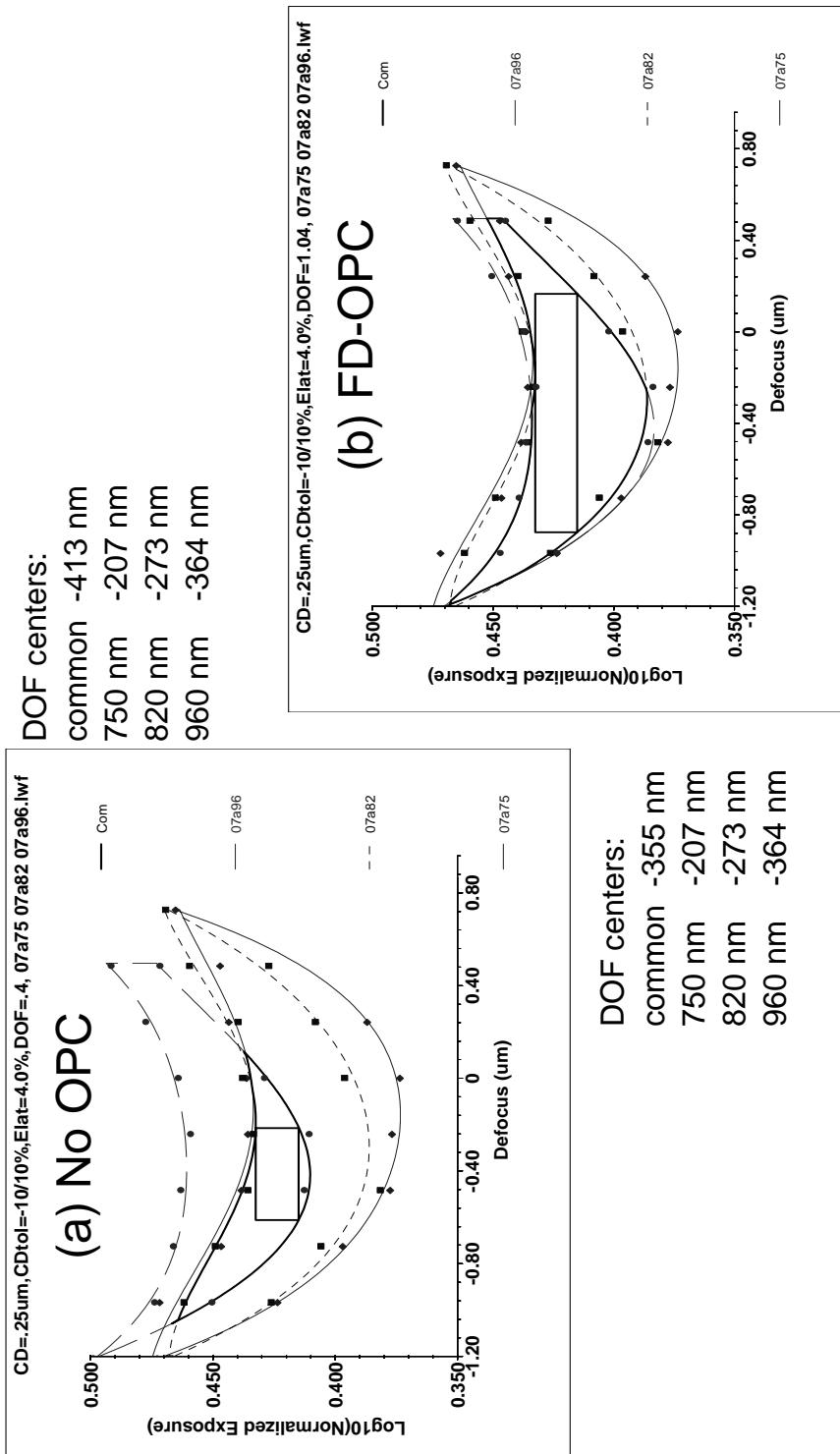


Figure 6.109 E-D window of three resist thicknesses. They do not center at "best focus."

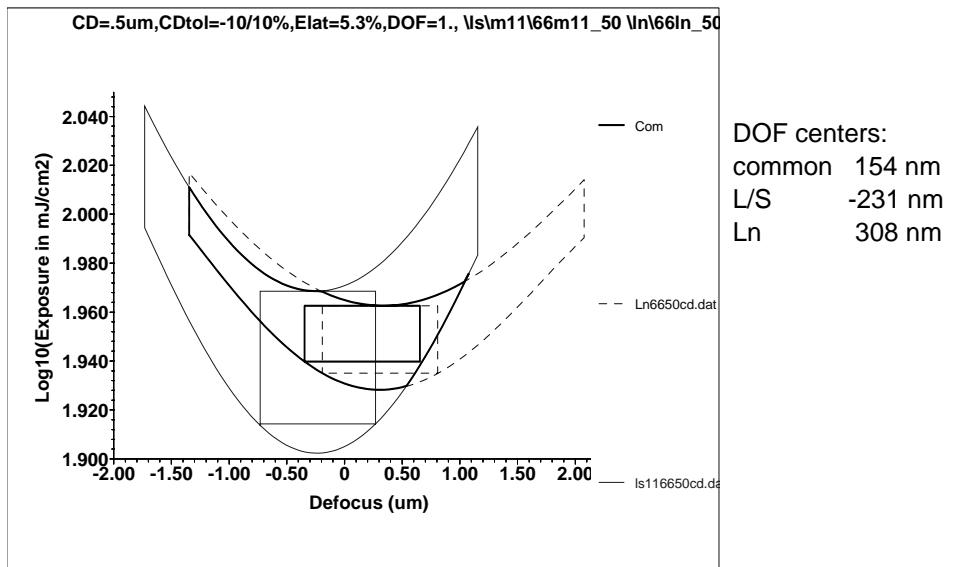


Figure 6.110 E-D window for two $0.66-k_1$ features: line-space = 1 and an isolated line. The separation of their center of focus is as large as 539 nm.

Performing FD-OPC is more time consuming. Instead of generating the wafer image at the focal plane, one must take as many image planes as necessary and evaluate the difference between each image and the target image. Adjustment of the mask pattern edge is now based on minimizing the difference in all of these image planes.

6.3.4.5 Correction to AEI

The electrical circuit performance is based on the transferred image from the resist image. It is usually the etched image. Therefore, OPC should be based on the AEI. However, it is not easy to extend model-based OPC to an AEI for the following reasons:

1. The range of microloading types of proximity effect is much larger than that from the optical kernels. For the latter, a few micrometers are sufficient to include in the 2D integration of Eq. (6.34), whereas the former case requires a range at least an order of magnitude larger, thus putting a tremendous burden on the OPC cycle time and cost.
2. During the development phase of a manufacturing technology, the lithography and the etching processes usually go through independent iterations. If the OPE model includes both lithography and etching, frequent reestablishment of the model is required, which is a heavy burden on time and resources.

The most popular practice to extend OPC to the AEI is to establish a model for the ADI for model-based OPC and to build a correction table from the resist image to the etched image for rule-based correction from the calculated ADI.

6.3.4.6 Hot-spot checking

After OPC, but before sending the data for mask making, it is preferable to perform a hot-spot check to fine tune the corrected mask pattern. The reason is that the OPC spec is based on the summation of the deviation from the target image at all target points. This addresses the fidelity of a single image edge but not the distance between nearby edges. There may be situations where two edges get too close. Even though the design rule before OPC addresses feature size and the distance between edges on the design, OPC changes them. The optical-proximity-corrected image must be examined. Figure 6.111 shows the formation of hot spots at the boundary of defocus and exposure. In the lower row, the hot spots identified by a hot-spot check are marked on the optical-proximity-corrected image and the resist image.

6.4 CD Uniformity

To accommodate more devices in an exposure field, feature size and the spacing between them should be as small as possible. However, since there are so many features in an exposure field and so many exposure fields in a wafer, critical-dimension uniformity (CDU) should also be good enough to ensure a reasonable device yield. In general, CDU becomes worse when CD becomes smaller. The purpose of process-tuning work is to find an operating point that compromises between a small feature size and CDU to ensure economically optimum device density versus device yield under current semiconductor manufacturing technology.

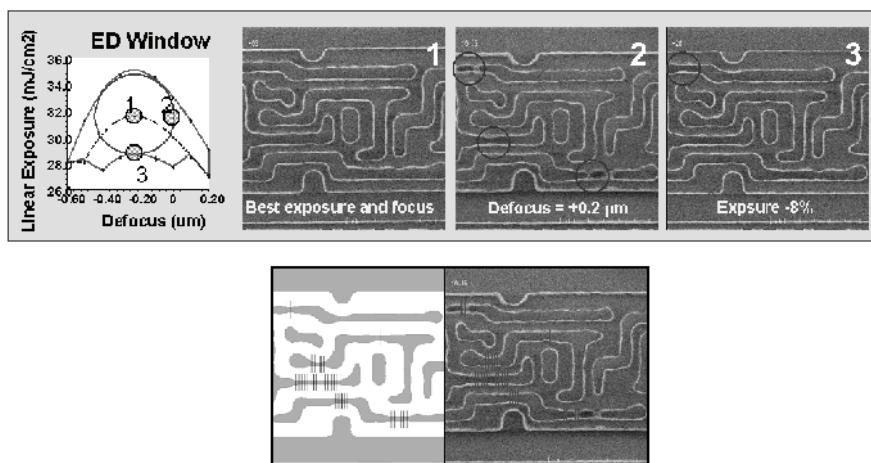


Figure 6.111 The series of resist images in the top row indicate hot spots near the boundary of defocus and exposure. The markings in the lower row show detected hot spots and the corresponding locations on the resist image.

CDU is a good indicator of the processing capability of a semiconductor factory. It is determined by two factors: the processing window and process fluctuations. The larger the processing window, the less sensitive CD is to process fluctuations. To enlarge the processing window, k_1 -reduction techniques, such as OAI, PSM, and low- k_1 enabling techniques, such as OPC, should be employed. To reduce process fluctuations, we need to take care of imperfections from masks, exposure tools, and processes.

We will perform an analysis to identify contributors to critical-dimension nonuniformity (CDNU) and their relative importance, so that we can find efficient ways for CDU improvement.

6.4.1 CDNU analysis

All imperfections in masks, exposure tools, and processes contribute to CDNU. In addition to lithography-related imperfections, previous processing steps, such as chemical vapor deposition (CVD) and chemical-mechanical polishing (CMP), and subsequent processing steps, such as etching, all have impacts to CDNU.

6.4.1.1 Linear model for CDU contributions

Before doing CDU analysis, we will introduce a linear model for CDU contribution that makes possible the decomposition of CDU into its contributors, based on a general observation. From these components, possible contributors to CDNU might be found.

Assuming there are only two contributors to CDNU, we have, for each CD denoted as W , in a specific CD distribution

$$W = \overline{\overline{M}}_A \overline{\overline{M}}_B W^0, \quad (6.35)$$

where $\overline{\overline{M}}_A$ and $\overline{\overline{M}}_B$ are two operators that stand for two independent contributors to CDNU, and W^0 is the CD when $\overline{\overline{M}}_A$ and $\overline{\overline{M}}_B$ are not present. For useful lithographic processes, $\overline{\overline{M}}_A$ and $\overline{\overline{M}}_B$ deviate only slightly from the perfect CD. That is, $\overline{\overline{M}}_A \approx \overline{\overline{I}} + \overline{\Delta M}_A$, etc., where $\overline{\overline{I}}$ is the identity operator. Substituting the relations into Eq. (6.35), only keeping terms to the first order of $\overline{\Delta M}_A$ or $\overline{\Delta M}_B$, we have

$$\Delta W \approx \overline{\overline{\Delta M}}_A W^0 + \overline{\overline{\Delta M}}_B W^0, \quad (6.36)$$

where $\Delta W = W - W^0$. In semiconductor manufacturing, total CD error tolerance is on the order of 10% of the target CD value. So, we may say that each contributor to CDNU amounts to only a small portion, say 5%, of the total CD

error tolerance. The error committed in Eq. (6.36), by neglecting higher-order terms, is at most 0.25%, only about 0.25 nm for the 90-nm node.

6.4.1.2 Geometrical decomposition

Since circuit patterns are first delineated on a mask and then exposed field-by-field on a wafer with current semiconductor manufacturing processes, CD at any location on a wafer should exhibit spatial characteristics of a field as well as a wafer—the so-called intrafield and interfield signatures, respectively. In the following, subscript B will be used to indicate interfield quantities, while subscript A represents intrafield quantities.

Extraction of the intrafield signature of the CD error distribution (i.e., CDNU on a wafer) can be accomplished by choosing a set of coefficients c_i such that

$$\sum_k \left[\Delta W(\mathbf{r}_k) - \sum_i c_i P_{i,A}(\mathbf{r}_{k,A}) \right]^2$$

is minimized. $\Delta W(\mathbf{r}_k) = W(\mathbf{r}_k) - W^0(\mathbf{r}_k)$, with $W^0(\mathbf{r}_k)$ being the target CD value at \mathbf{r}_k . In the above expression, the position vector \mathbf{r}_k is decomposed into its interfield and intrafield components, such that $\mathbf{r}_k = \mathbf{R}_{k,B} + \mathbf{r}_{k,A}$ where $\mathbf{R}_{k,B}$ is the vector from the wafer center to the center of the field where \mathbf{r}_k is located. Summation over k is through each location on the wafer where CD is measured. Summation over i is through the intrafield basis functions $P_{i,A}$. Since a field is rectangular in shape, the Legendre polynomials $L_\alpha(x_{k,A}) L_\beta(y_{k,A})$ are a good choice for $P_{i,A} = P_{(\alpha,\beta),A}$, where $\mathbf{r}_{k,A} = x_{k,A}\hat{\mathbf{x}} + y_{k,A}\hat{\mathbf{y}}$. The number of intrafield basis functions employed will depend on the number of sampling points in a field.

Once c_i are found, the intrafield signature of the CD distribution is obtained:

$$\Delta W_A(\mathbf{r}_{k,A}) = \sum_i c_i P_{i,A}(\mathbf{r}_{k,A}). \quad (6.37)$$

The interfield signature follows readily, such that for each \mathbf{r}_k ,

$$\Delta W'_B(\mathbf{r}_k) = \Delta W(\mathbf{r}_k) - \sum_i c_i P_{i,A}(\mathbf{r}_{k,A}). \quad (6.38)$$

Practically, we are only interested in the slowly varying part of the CD error distribution. The rapidly varying part generally results from edge roughness. It is a random error that is impossible to compensate. So, we also expand the interfield signature in terms of interfield basis functions just as we do for the intrafield signature, such as

$$\Delta W'_B(\mathbf{r}_k) = \Delta W_B(\mathbf{r}_k) + \Delta W_R''(\mathbf{r}_k), \quad (6.39)$$

with

$$\Delta W_B(\mathbf{r}_k) = \sum_j c_j P_{j,B}(\mathbf{r}_k), \quad (6.40)$$

where summation over j is over interfield basis functions $P_{j,B}$, and $\Delta W''_R$ is the residual after expansion. Since the wafer is of a circular shape, the Zernike polynomials are good choice for $P_{j,B}$. Note that the set of c_j is the same for all possible \mathbf{r}_k .

The above argument assumes that the intrafield signature for each field is the same. In real situations, this is not the case because of lens heating, lens transmission hysteresis, the transient effect of sensors, the focus difference between scanning up and scanning down, etc. We extract such field-to-field CD error variation ΔW_F by averaging $\Delta W''_R$ inside each field. Then, we obtain $\Delta W''_R(\mathbf{r}_k) = \Delta W_F(\mathbf{R}_{k,B}) + \Delta W_R(\mathbf{r}_k)$, where ΔW_R is the residual after complete CDNU geometrical decomposition.

In fact, the interfield signature from the scanner and from the track cannot be decomposed unambiguously. The reason we identify ΔW_F with $\Delta W''_R$ rather than $\Delta W'_B$ is based on our experience that the interfield contribution from the scanner is much smaller than that from the track.

In summary, the decomposition of each individual CD error in a CD error distribution is achieved.

$$\Delta W(\mathbf{r}_k) = \Delta W_A(\mathbf{r}_{k,A}) + \Delta W_B(\mathbf{r}_k) + \Delta W_F(\mathbf{R}_{k,B}) + \Delta W_R(\mathbf{r}_k). \quad (6.41)$$

The above decomposition is good for steppers. For scanners, one generally extracts further from the intrafield CD error signature the CD error signature in the slit (x) and in the scan (y) directions, denoted as $\Delta W_{A,slit}(x_{k,A})$ and $\Delta W_{A,scan}(y_{k,A})$, respectively. To obtain the CD error signature in these directions, for each x and each y , find the average CD error at all possible y and x positions. Figure 6.112 shows decomposition of CD error distribution ΔW into ΔW_A , ΔW_B , ΔW_F , and ΔW_R , as shown in (a), (b), (c), (d), and (e), respectively. Here, CDs of a specific mask feature in an array of $49(x) \times 13(y)$ inside each field are measured for all 67 fields on a wafer. The number of intrafield basis functions employed is 49×13 , i.e., the set of Legendre polynomials $L_\alpha(x_{k,A}) L_\beta(y_{k,A})$ with (α, β) from $(0,0), (0,1), \dots, (48,12)$. The number of interfield basis functions employed is 36, i.e., the set of Zernike polynomials $Z_j(\mathbf{r}_k)$ with j from 1 to 36. Increasing the number of interfield basis functions from 36 to 100 changes little of the interfield signature ΔW_B , e.g., 3σ of ΔW_B changes only by 0.01 nm. Note that all CD error signatures are plotted in terms of mosaic, with the color of each block determined by its local CD error value. No interpolation or smoothing in data or figures is made. The CD error signature in the slit/scan direction for the intrafield CD error signature in Fig. 6.112 is shown in Fig. 6.113.

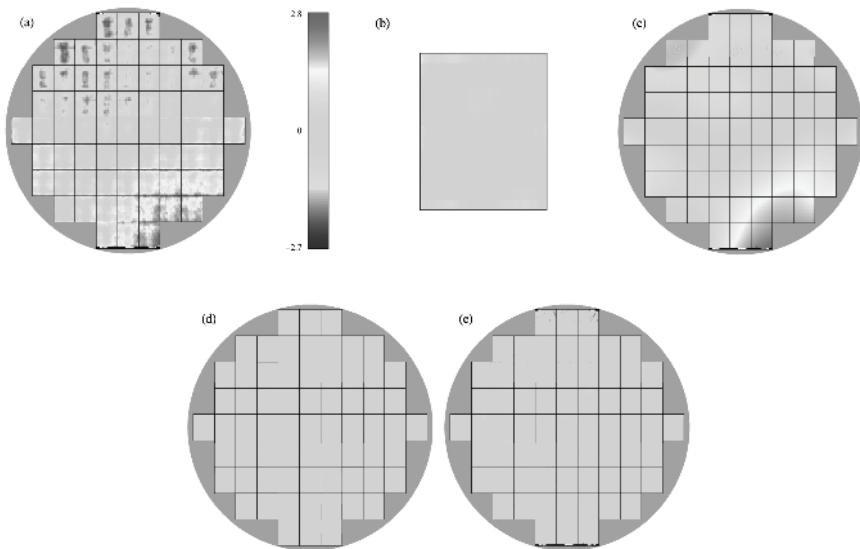


Figure 6.112 An example illustrating the geometrical decomposition of CD error distribution on a wafer. ΔW , ΔW_A , ΔW_B , ΔW_F , and ΔW_R , are shown in (a), (b), (c), (d), and (e), respectively.

One should not evaluate the intrafield CD signature ΔW_A or the interfield CD signature ΔW_B at $\mathbf{r} \neq \mathbf{r}_k$. This is because polynomials of very high degrees are employed to extract them, and therefore spurious oscillations may take place.

Despite the fact that interfield basis functions are not orthogonal to intrafield basis functions, CDU geometrical decomposition can be performed without ambiguity by following the above procedure to extract the intrafield component first. No iteration is needed.

Our method of finding the intrafield signature is through the projected image, and it is different from that proposed by Wong et al.,^{48,49} who average CD error at each intrafield location over all interfield locations, which is a good approximation if the average of the interfield signature at each intrafield point is negligible.

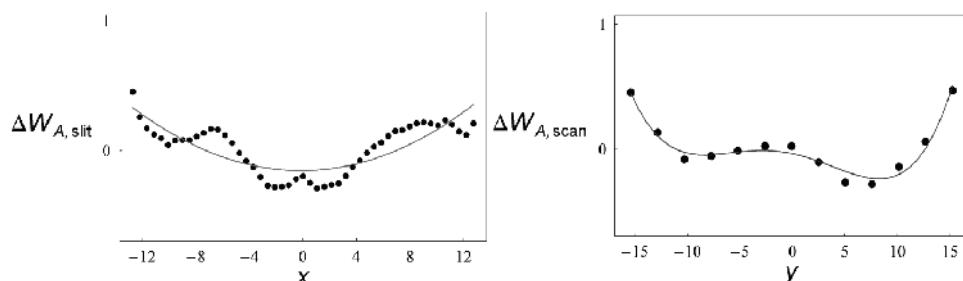


Figure 6.113 CD error signature in the slit (x) and scan (y) directions for the intrafield CD error signature of Figure 6.112.

Peter Vanoppen et al.⁵⁰ use a similar method to analyze the CD error distribution taken with electrical measurement. However, they employ polynomials for both interfield and intrafield basis functions.

Geometrical decomposition of the CD error distribution not only provides clues for contributors of CDNU but also provides the information necessary for CDU improvement through active CD error compensation, as will be detailed later.

6.4.1.3 Physical decomposition

Geometrical decomposition of CD error distribution can separate CDNU contributors with different spatial characteristics. However, there are many CDNU contributors that all lead to the intrafield or intrafield signature of CD error distribution. To separate them, we need to perform independent tests. With the help of physical models, CDNU can be further decomposed. This is physical decomposition.

Generally speaking, imperfections of the mask contribute to the intrafield signature, while imperfections of the track contribute to interfield signature. Imperfection of the scanner contributes to both.

For intrafield signature, CD errors on the mask, as well as illumination imperfection, lens aberrations, flare variation, and slit nonuniformity on the scanner are contributors. For interfield signatures, temperature variation on hot plates, coating nonuniformity, exposure variation between fields, scanning speed variation between fields, and field-to-field focusing variations are contributors.

To separate the contribution of the mask from that of the scanner and to find the MEEF under the employed process conditions, we need to do independent mask CD measurements. Then, the mask contribution to the intrafield signature is $\Delta W_B^{\text{mask}} = \text{MEEF} \times \Delta W^{\text{mask}} / M$, where ΔW^{mask} is the mask CD error, and M is the reduction ratio of the exposure tool.

For other contributors, such as lens aberrations, we need more deliberate models implemented in lithographic image simulators. This will be described in other sections.

As for the interfield signature, thickness nonuniformities from the wafer substrate—including the TARC, resist and BARC layers, temperature nonuniformity of the hot plates for postapply and postexposure bakes, development nonuniformity, polishing and etching nonuniformities—are all contributors.

According to the linear model, all components of the intrafield and the interfield signatures are just added together.

6.4.1.4 CDU summation

There are always arguments about how to sum different contributors to CDU—arithmetic or root-sum-square. The general rule is that if the errors are systematic, use an arithmetic sum; if the errors are random, then a root-sum-square should be applied to evaluate the standard deviation.

However, there are many contributors. Each contributor has its own functional dependence. For example, 1D proximity errors depend on pattern geometry or the pitch of the pattern. Mask-related contributors depend on intrafield locations. Track-related contributors depend on interfield locations. Aberration-related contributors depend on both pattern and intrafield location. The CDNU from all of these contributors certainly should be classified as systematic if their signature is fixed.

If we directly sum them up, the resulting CDU is too pessimistic. It means a pattern with the largest bias error, of the highest aberration sensitivity, located at an intrafield location with the largest lens aberration, and located at an interfield location with the highest temperature for PEB is equally probable to happen as a pattern with medium bias error, average aberrations sensitivity, at interfield and intrafield positions with an average lens aberration and average PEB temperature. It does not reflect the real situation in manufacturing.

For those CD errors that are to be directly summed, their contributors at least should be of the same functional dependence, i.e., with the same independent variable. So, for contributors with various functional dependences, it is more reasonable to sum them statistically, i.e., by root-sum-square, even if they are systematic in other senses.

Next, we discuss how to combine the CDNU of different kinds of patterns. Note that patterns with the same characteristics but at different intrafield/interfield locations are classified as the same kind of patterns. In the following, we will derive formulae that are very useful when calculating the 3σ of the total CD errors by combining CD errors from different contributors.

First, consider the simplest case of two groups of CD data. For group 1 and 2, respectively, the individual CD is denoted as $W_{1,i}$ and $W_{2,i}$; the total number of CDs as n_1 and n_2 ; and the mean CD as M_1 and M_2 . When they are regarded as a single group, with the individual CD denoted as W_i , the total number of CDs as n , the mean CD as M , its standard deviation σ can be found as follows:

$$\begin{aligned} n\sigma^2 &= \sum_{i=1}^n [W_i - M]^2 \\ &= \sum_{i=1}^n [(W_{1,i} - M_1) + (M_1 - M)]^2 + \sum_{i=1}^n [(W_{2,i} - M_2) + (M_2 - M)]^2, \quad (6.42) \\ &= n_1\sigma_1^2 + n_1(M_1 - M)^2 + n_2\sigma_2^2 + n_2(M_2 - M)^2 \end{aligned}$$

i.e.,

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n} + \frac{n_1(M_1 - M)^2 + n_2(M_2 - M)^2}{n}. \quad (6.43)$$

For K groups of data, the above formula can be extended as

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^n n_k \sigma_k^2 + \frac{1}{n} \sum_{k=1}^n n_k (M_k - M)^2. \quad (6.44)$$

If the number of CD sampling for each group is the same, we obtain

$$\sigma^2 = \frac{1}{K} \sum_{k=1}^K \sigma_k^2 + \frac{1}{K} \sum_{k=1}^K (M_k - M)^2, \quad (6.45)$$

i.e.,

$$\sigma^2 = \overline{\overline{M}}[\sigma^2] + \overline{\overline{\sigma^2}}[M], \quad (6.46)$$

where operators $\overline{\overline{M}}$ and $\overline{\overline{\sigma^2}}$ mean finding, for a collection of numbers, their mean and σ squared, respectively, or

$$(3\sigma)^2 = \overline{\overline{M}}[(3\sigma_k)^2] + \overline{\overline{(3\sigma)^2}}[M]. \quad (6.47)$$

In other words, the total 3σ squared is equal to the mean of 3σ squared plus the 3σ squared of the mean. Note that Eq. (6.47) is exact since it is derived without any approximation.

Also, in Eq. (6.47), the second term does contribute and cannot be neglected. A good example is to find the 3σ of the CDNU of one-dimensional through-all-pitch patterns. With Eq. (6.47), the impact of 1D OPC errors to the total CDNU can be easily calculated.

On the manufacturing floor, when we talk about CDNU, we generally do not mean the CDNU of a single pattern, but rather the total CDNU of a product, which includes CD error variations from within the field, within the wafer, wafer-to-wafer, and lot-to-lot of all critical patterns of that product. With Eq. (6.47) applied recursively, the total CDNU in terms of 3σ of a product can readily be found.

6.4.2 CDU improvement

CDU improvement is very important, since it is directly related to yield improvement, especially for low- k_1 optical lithography.

There are several ways to improve CDU. One way is to use a resolution-enhancement technique to increase the process window. Another is to improve the imaging performance of process tools, such as the track and the exposure tool.

Here, we will introduce yet another method, i.e., the method of active compensation, which utilizes the high controllability of exposure dosage on the exposure tool or temperature on the hot plates of the track. It is cost effective, quick to implement, and has little impact on the existing lithographic processes.

6.4.2.1 Active compensation with exposure tools

Modern scanners provide many ways to manipulate exposure at the interfield and intrafield levels. For interfield levels, the exposure at each field can be assigned to compensate for systematic interfield CDNU. This can be accomplished in many ways. One can adjust the scanning speed, the width of the scanning slot, the laser pulse rate, or the light attenuator in the illumination system on a field-by-field basis. For intrafield levels, the width of the scanning slot along its length and the scanning speed can be adjusted dynamically within each scanning field. In principle, the combined CD fingerprint from the mask, the track, and the etcher, etc., can all be corrected.

6.4.2.2 Active compensation on tracks

An example of track active compensation is a CD optimizer, which is used to correct interfield CDNU caused by PEB of the resist.

For the purpose of PEB, a multizone hot plate, called a chilling precision hot plate (CPHP), is commercially available. With a CPHP, the temperature of each heating zone can be independently set, allowing some flexibility when adjusting the temperature distribution of the hot plate.

However, for commercially available CARs, especially those for ArF exposure, the PEB sensitivity can be as high as $10 \text{ nm}^{\circ} \text{C}$. This means that the temperature uniformity should be better than 0.2°C to achieve interfield CDU of less than 2 nm. Adjusting the temperature distribution to achieve such uniformity is a difficult task. The main concern is that the thermal environment must be kept the same as that of a real product wafer.

Even if we can measure the entire temperature history during PEB in the correct thermal environment by using commercially available wireless sensor wafers, we still do not know how to optimize the temperature distribution because the link between CD and the temperature history is still not accurately known, e.g., by matching the temperature of different heating zones in the rising stage or in the steady state.

Hence, we propose a CD optimizer that directly uses resist as the temperature sensor, with its response to the temperature revealed in the final CD after the resist process. To prove that our method is feasible, we first expose and process wafers with the CPHP for PEB set at different temperatures. It is shown that the average CD for each wafer varies linearly with the PEB temperature in a sufficiently large temperature range. Figure 6.114 demonstrates the linearity on a wafer. Next, we find the response of each heating zone by measuring the baseline CD distribution W^0 of a wafer processed with the CPHP for PEB at the baseline setting. Additionally, for each heating zone i , we measure the perturbed CD

distribution W_i of a wafer processed with the temperature of only the heating zone i deviated from the baseline setting by ΔT_i .

After evaluating the response of each heating zone, we can subtract the baseline CD distribution from the perturbed CD distribution to obtain the basis function $P_i = W_i - W^0$. Repeating the procedure for each heating zone i produces the basis functions. Figure 6.114 shows the basis functions for a seven-zone CPHP designed for the ACT12 track from Tokyo Electron Ltd.

The CDNU is obviously $\Delta W = W^0 - W^t$, where W^t is the target CD distribution. To know how much temperature should be adjusted for each heating zone to correct the CDNU, we expand the CDNU in terms of the basis functions, i.e.,

$$\Delta W = \sum_i c_i P_i ,$$

where c_i is the expansion coefficient. If the temperature of each heating zone i is adjusted by $-c_i \Delta T_i$, the resulting CD distribution should be optimal under given degrees of freedom, i.e., the number of heating zones in the CPHP. The predicted CD distribution after correction is

$$W^c = W^o - \sum_i c_i P_i .$$

Note that W^0 , W_i , P_i , W^t , and W^c are all scalar fields, i.e., at each specified location on a wafer, there is a corresponding CD.

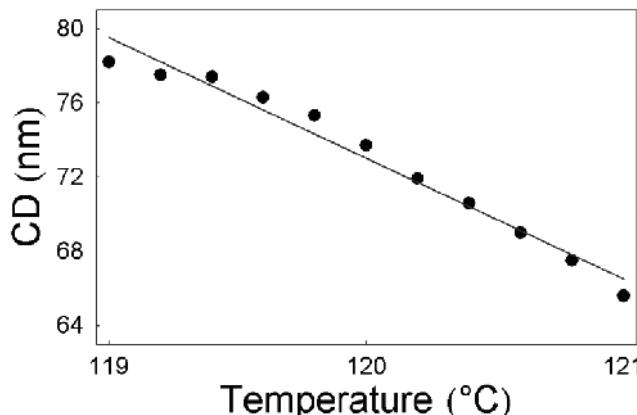


Figure 6.114 Average CD on a wafer versus temperature setting of the CPHP for the PEB process of that wafer.

The preceding method is illustrated with experimental results shown in the following figures. Figure 6.115 shows the average CD on a wafer as a function of the temperature setting of the CPHP used for PEB of that wafer. The experiment was done by increasing the temperature from 119°C to 121°C in 0.2°C steps. If the first two data points are ignored, the CPHP has the potential to achieve temperature controllability better than 0.1°C . The abnormal behavior of these points is probably due to data taken before thermal equilibrium was reached. Figure 6.116 shows the basis functions for a seven-zone hot plate. These functions were obtained by setting $\Delta T_i = 0.8^{\circ}\text{C}$ for each heating zone. Figure 6.116 shows the CDU improvement achieved with a CD optimizer. The original CD map is shown in (a), while the predicted CD map after correction is shown in (b). The actual corrected CD map is shown in (c). After correction, the average of the interfield CD distribution remains unchanged, while the 3σ is improved from 4.64 to 1.86 nm in (b) and 1.96 nm in (c).

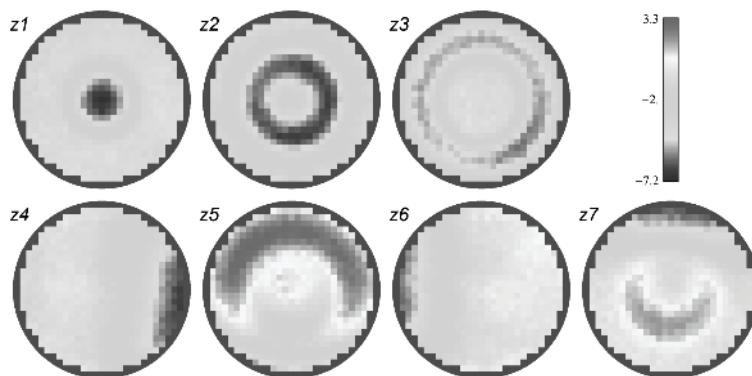


Figure 6.115 Basis functions for a seven-zone hot plate, obtained by choosing $\Delta T_i = 0.8^{\circ}\text{C}$ for each heating zone.

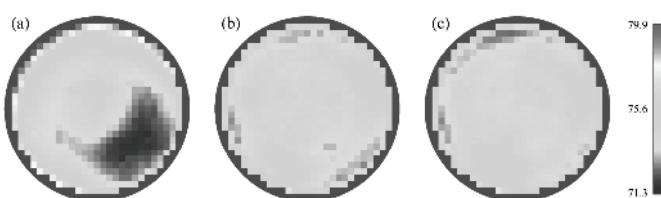


Figure 6.116 CDU improvement achieved by employing the proposed CD optimizer.

References

- 1 B. J. Lin, "The optimum numerical aperture for optical projection microlithography," *Proc. SPIE* **1463**, p. 42 (1991).
- 2 B. J. Lin "The optimum numerical aperture for attenuated phase-shifting masks," *Microelectron. Eng.* **17**, p. 79 (1992).
- 3 S. W. Choi, S. J. Lee, J. Shin, S.-G. Woo, H. K. Cho, and J.-T. Moon, "Influence of material on process focus budget and process window of 80 nm DRAM devices," *Proc. SPIE* **5752**, p. 570 (2005).
- 4 Private communications.
- 5 J. W. Bossung, "Projection printing characterization," *Proc. SPIE* **100**, p. 80 (1977).
- 6 C. P. Ausschnitt, "Distinguishing dose from defocus for in-line lithography control," *Proc. SPIE* **3677**, p.140 (1999).
- 7 KLA MPS reference.
- 8 T. A. Brunner, A. L. Martin, R. M. Martino, C. P. Ausschnitt, T. H. Newman, and M. S. Hibbs, "Quantitative stepper metrology using the focus monitor test mask," *Proc. SPIE* **2197**, p. 541 (1994).
- 9 H. Nomura, "New phase shift gratings for measuring aberrations," *Proc. SPIE* **4346**, p. 25 (2001).
- 10 Y. Shiode, H. Morohoshi, A. Takagi, K. Fujimaru, K. Mizumoto, and Y. Takahashi, "A novel focus monitoring method using double side chrome mask," *Proc. SPIE* **5754**, p. 303 (2005).
- 11 S. Nakao, J. Sakai, S. Maejima, A. Ueno, A. Nakae, S. Yanashita, K. Itano, H. Kozawa, A. Tokui, and K. Tsujita, "Simple and high sensitive focus monitoring utilizing an aperture on backside of photomask," *Proc. SPIE* **5040**, p. 582 (2003).
- 12 ASML scanner manual.
- 13 R. A. McEachem, M. S. Lucas, C. R. Simpson, "Lithography system using dual substrate stages," U.S. Patent 5,677,758 (1997).
- 14 B. J. Lin, "Step and repeat apparatus having enhanced accuracy and increased throughput," U.S. Patent 5,715, 064 (1998).
- 15 W. M. Moreau, *Semiconductor Lithography Principles, Practices, and Materials*, Sec. 13.2.2, Plenum Press, New York (1998).
- 16 A. Barraud, C. Rosilio, and A. Raudel-Texier, "Monomolecular resists—a new approach to high resolution electron beam microlithography," *J. Vac. Sci. Technol.*, **16**, pp. 2003–2007 (1980).
- 17 Ref. 15, Sec. 6.6.1, p. 304.

- 18 L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to microlithography," Chapter 4 in *ACS Professional Reference Book*, American Chemical Society, Washington D.C., p. 311 (1994).
- 19 C. P. Auschnitt and D. A. Huchital, "Coating apparatus," U.S. Patent 4,290,384 (1981).
- 20 B. J. Lin, "In situ resist control during spray and spin in vapor," U.S. Patent 5,366,757 (1994).
- 21 B. J. Lin and T. J. Cardinali, "Material-saving resist spinner and process," U.S. Patent 5,378,511 (1995).
- 22 B. J. Lin, "Material-saving resist spinner and process," U.S. Patent 5,449,405 (1995).
- 23 F. H. Dill and J. M. Shaw, "Thermal effects on the photoresist AZ1350J," *IBM J. Res. Dev.* **21**, 210 (1977).
- 24 B. J. Lin, "Off-axis illumination—working principles and comparison with alternating phase-shifting masks," *Proc. SPIE* **1927**, p. 94 (1993).
- 25 T. S. Gau, C. M. Wang, and C. M. Dai, "Strategy to manipulate the optical proximity effect by post-exposure bake processing," *Proc. SPIE* **3334**, p. 885 (1998).
- 26 J. C. Matthews, M. G. Ury, A. D. Birch, and M. A. Lashman, "Microlithography techniques using a microwave powered deep uv source," *Proc. SPIE* **394**, p. 172 (1983).
- 27 W. H.-L. Ma, "Plasma resist image stabilization technique (PRIST)," *IEDM Tech. Dig.* pp. 574–575 (1980).
- 28 F. S. Lai, B. J. Lin, and Y. Vladimirska, "Resist hardening using a conformable mold," *J. Vac. Sci. Tech. B* **4**(1), p. 426 (1986).
- 29 B. J. Lin, "Applications of the mold controlled profile (MCP) technique for resist processing," *Proc. SPIE* **771**, p. 180 (1987).
- 30 V. Mariott, "High resolution positive resist developers: a technique for functional evaluation and process optimization," *Proc. SPIE* **394**, p. 144 (1983).
- 31 K. Deguchi, K. Miyoshi, T. Ishii, and T. Matsuda, "Patterning characteristics of a chemically-amplified negative resist in synchrotron radiation lithography," *Jpn. J. Appl. Phys.* **31**, pp. 2954–2958 (1992).
- 32 T. Tanaka, M. Morigami, and N. Atoda, "Mechanism of resist pattern collapse during development process," *Jpn. J. Appl. Phys.* **32**, pp. 6059–6064 (1993).
- 33 S. A. MacDonald, W. D. Hinsberg, H. R. Wendt, C. G. Willson, and C. D. Snyder, "Airborne contamination of a chemically amplified resist. 1. Identification of problem," *Chem. Mater.* **5**, p. 348 (1993).
- 34 B. J. Lin, "Where is the lost resolution?" *Proc. SPIE* **633**, p. 44 (1986).

- 35 B. J. Lin, "Vibration tolerance in optical imaging," *Opt. Eng.* **32**, p. 527 (1993).
- 36 Y. T. Wang and C. Yagyensh, "Phase shifting circuit manufacture method and apparatus," U.S. Patent 5,858,580 (1999).
- 37 B. J. Lin, "The optimum numerical aperture for attenuated phase-shifting masks," *Microelectron. Eng.* **17**, p. 79 (1992).
- 38 B. J. Lin, "The attenuated phase-shifting mask," *Sol. State Tech.* p. 43 (Jan. 1992).
- 39 A. Nitayama, T. Sato, K. Hashimoto, F. Shigemitsu, and M. Nakase, "New phase shifting mask with self-aligned phase shifters for quarter micron photolithography," *IEDM Tech. Dig.* p. 57 (1989).
- 40 H. I. Smith, E. H. Anderson, and M. L. Shattenburg, "Lithography mask with a π -phase shifting attenuator," U.S. Patent 4,890,309 (1989).
- 41 B. J. Lin, "The attenuated phase-shifting mask," *Sol. State Tech.* p. 43 (Jan. 1992).
- 42 N. Shirashi, S. Hirukawa, Y. Takeuchi, and N. Magome, "New imaging technique for 64-M-DRAM," *Proc. SPIE* **1674**, p. 741 (1992).
- 43 M. Noguchi, M. Muraki, Y. Iwasaki, and A. Suzuki, "Subhalf-micron lithography system with phase-shifting effect," *Proc. SPIE* **1674**, p. 92 (1992).
- 44 B. J. Lin, "Electromagnetic near-field diffraction of a medium slit," *J. Opt. Soc. Am.*, **62**, pp. 977–981 (1972).
- 45 S. D. F. Hsu, N. Corcoran, and J. F. Chen, "Orientation dependent shielding for use with dipole illumination techniques," U.S. Patent 7,246,342 (2007).
- 46 B. J. Lin, "Optical projection printing threshold leveling arrangement," U.S. Patent 4,456,371 (1984).
- 47 B. J. Lin, A. M. Pelella, and A. E. Rosenbluth, "Lithographic process having improved image quality," U.S. Patent 4,902,899 (1990).
- 48 A. K. Wong, A. F. Molless, T. A. Brunner, E. Coker, R. H. Fair, G. L. Mack, and S. M. Mansfield, "Linewidth variation characterization by spatial decomposition," *J. Microlithogr. Microfabr. Microsyst.* **1**(2), p. 106 (2002).
- 49 A. K. Wong, A. F. Molless, T. A. Brunner, E. Coker, R. H. Fair, G. L. Mack, and S. M. Mansfield, "Characterization of linewidth variation," *Proc. SPIE* **4000**, p. 184 (2000).
- 50 P. Vanoppen, O. Noordman, J. Baselmans, and J. van Schoot, "Analysis of full-wafer/full-batch CD uniformity using electrical linewidth measurements," *Proc. SPIE* **4404**, 33 (2001).

Chapter 7

Immersion Lithography

Immersion lithography offers an attractive alternative to 157-nm and EUV lithography. Due to the tremendous obstacles associated with these exotic technologies, immersion lithography has become the last technique for the incremental extension of dioptric/catadioptric optical lithography, if not the last technique for photon lithography per se.

7.1 Introduction

The basic idea of immersion lithography is to fill up the air space between the last element of the imaging lens and the surface of the photoresist on the wafer, as shown in Fig. 7.1. With a higher-index medium, light of higher spatial frequencies can be coupled to the resist, improving the resolution. Liquid coupling also minimizes the reflection loss at the two affected interfaces. The classical interpretation of the improvement of resolution is that $\text{NA} \equiv n \cdot \sin\theta$ is increased with the change of refractive index n in the coupling medium. But, how does one visualize it? The wavelength in the fluid is reduced from the wavelength in vacuum by the formula

$$\lambda = \lambda_0 / n. \quad (7.1)$$

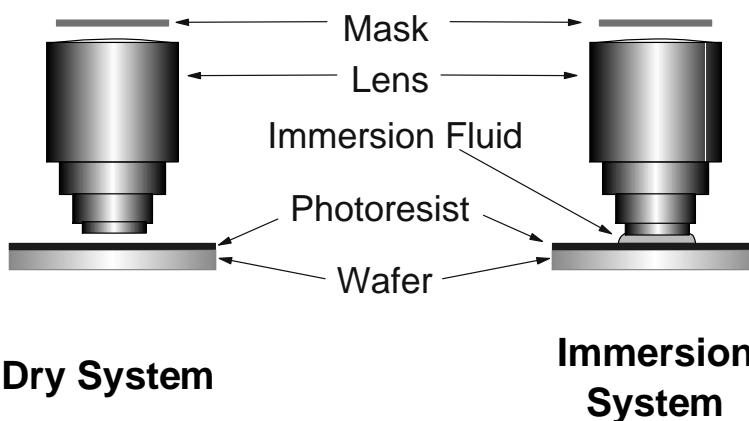


Figure 7.1 A schematic immersion system compared to a dry one.

From Eq. (4.3), the resolution¹ of a projection imaging system is given by

$$W = k_1 \frac{\lambda}{\sin \theta}, \quad (7.2)$$

where W is the resolution given in the size of a feature or a half pitch, λ is the wavelength in the medium of concern, and θ is the aperture angle, which is the angle sustained by the ray of the largest spatial frequency permitted by the optical system. When λ is reduced, W follows. Folding it to the classical interpretation,

$$W = k_1 \frac{\lambda_0}{n \sin \theta} = k_1 \frac{\lambda_0}{NA}. \quad (7.3)$$

Basically, NA acts as the threshold of a low-pass spatial-frequency filter anywhere along the optical path. Therefore, when the NA in the gating part of the optical train is increased, the NA in other parts of the optical train follows until the NA in the particular part becomes gating. For example, when the NA is gated by the coupling medium, even though the lens or the resist can support a larger NA, resolution is dictated by this gating NA. When the coupling medium is switched from air to a fluid, higher spatial frequencies are coupled to the resist to improve the resolution. Note that the wavelength in the resist is also reduced by its refractive index, which is usually the highest in microlithography imaging systems, ensuring that refractive index is not a gating factor.

Wavelength reduction in the medium caused by a change in refractive index is exhibited in the change of light velocity in the medium. The frequency of the light wave is not changed, as given by the well-known equation on photon energyⁱ

$$E = h\nu, \quad (7.4)$$

where E is the energy of a photon, h is Planck's constant, and ν is the frequency of the light wave. The energy of the light wave and thus the characteristics of the interaction between the materials are unchanged in any transparent medium. The mask, lens, and resist materials are illuminated with the same light frequency; even the light source only needs to emit at the same frequency. Hence, they do not must be treated differently. However, a change of the vacuum wavelength does change the frequency, and thus would require new materials.

The immersion system is better than a system with a shorter vacuum wavelength because one does not have to be concerned with developing new mask, lens, and resist materials. Purging or evacuating the atmosphere is not necessary either. The tradeoff is lower optical performance of the immersion system compared to a system using a dry wavelength identical to the immersed

ⁱ Consult any textbook in modern physics.

wavelength with the same NA. This will be explained in more detail in Secs. 7.3 and 7.4.

Hence, it is possible to extend existing λ and NA combinations by two to three nodes, e.g., from 90 nm to 65, 45, and 32 nm using a 193-nm vacuum wavelength and infrastructure, which fulfills the technology and economy requirements. However, there are many foreseeable challenges, including producing a fluid that has the required optical, mechanical, and chemical properties, designing and building immersion exposure systems, developing resists for immersion, etc. Even before considering the materials and systems requirements, the immersion system is quite different from the conventional dry system because of the high-index coupling medium and extremely high spatial frequencies in the resist. The latter makes optical imaging very sensitive to the polarization of imaging light.

Immersion microscopy has been well known and commercially available for decades. While the wavelength-reduction principle of the coupling medium and its function is similar to immersion lithography, the former is much easier. Once an inert transparent high-index fluid is found, and an immersion objective is fabricated, the challenge is over. For immersion lithography, meeting the first two requirements is much more difficult. The fluid must have a uniform and consistent refractive index. The wafer must move at a high speed with the fluid in place because the step-and-scan exposure tool has replaced the step-and-repeat tool as the frontrunner of the feature-size reduction movement. In addition to these two classical requirements, many other difficult challenges must be met.

Starting in the 1980s, there were proposals and demonstrations on immersion lithography. In 1989, Kawata et al.² demonstrated NA = 1.4 optical lithography using a 453-nm wavelength and a high-index oil on a commercial optical microscope. The smallest smooth resist line produced was a 230-nm isolated line. In 1999, Hoffnagle et al.³ extended the limit of interferometric lithography using the index-matching fluid cyclooctane, which has a refractive index of 1.51 at the 257-nm wavelength. Hoffnagle's work shows that a commercial, chemically amplified resist designed for 248 nm works in an immersion mode. Switkes and Rothschild^{4,5} have been reporting their findings on immersions fluids for 157- and 193-nm light since 2001. In 1987, Lin⁶ proposed extending the lithography field size by scanning with an on-axis slot and extending DOF with liquid immersion. He also discussed the advantages and concerns for such systems. The former pointed to modern step-and-scan exposure tools. The latter can extend the scanners by several more nodes.

Several publications, dealing mostly with polarization effects at high-NA imaging, contribute to the understanding of immersion lithography, especially when it is not used to simply extend DOF but, in a hyper-NA mode, to increase resolution. Flagello and Rosenbluth,⁷ (and later Flagello, Milster, and Rosenbluth,⁸) treated vectorial diffraction theory and high-NA imaging. Brunner et al.⁹ provided valuable insight on polarization effects and hyper-NA lithography. Lin¹ defined a nonparaxial scaling coefficient k_3 for high-NA

lithography and carried the scaling equations to include immersion applications. Much of Sec. 7.2 is taken from my previous work in Ref. 10.

7.2 Resolution and DOF

The resolution and DOF of an immersion system requires different considerations than a dry system. Due to the high-index coupling medium between the lens and the wafer, the wavelength in the medium reduces the incident and exiting angles of the imaging light change accordingly, and higher spatial frequencies are accepted and propagated. Resolution and/or DOF are improved.

7.2.1 Wavelength reduction and spatial frequencies

The wavelength changes as a function of the refractive index of the propagation medium, as indicated by Eq. (7.1). In a dry system, the wavelength changes inside each piece of the lens element, the resist, and the underlying thin films on the wafer. In an immersion system, the wavelength also changes in the coupling medium between the last element of the lens and the resist, as depicted in Fig. 7.2. Ultimately, the wavelength of the coupling medium reaches identity with the wavelength inside the resist. When an incident beam makes an angle to the optical axis, wavelength reduction induces a change in the beam angle according to Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 . \quad (7.5)$$

The change in the incident angle in the lens element, coupling medium, and the resist is illustrated in Fig. 7.3. Larger spatial frequencies for higher resolutions sustain a larger θ . We emphasize here that the NA in the recording medium is what matters. The coupling medium performs the function of coupling the large

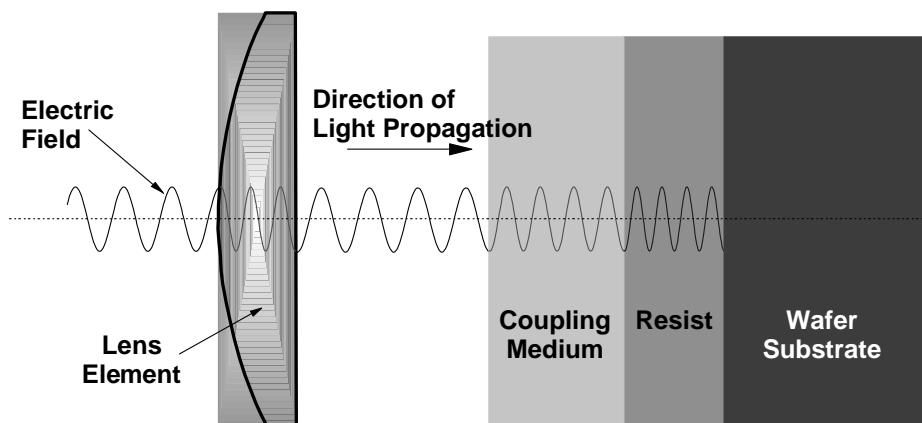


Figure 7.2 Wavelength reductions in an immersion system.

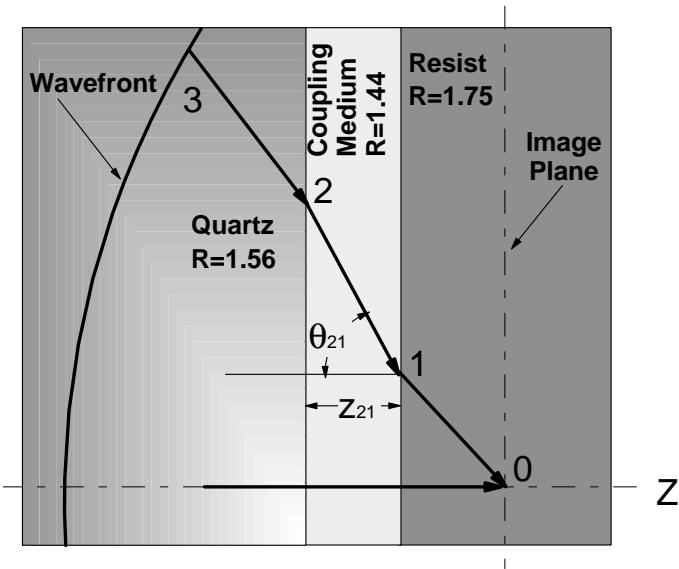


Figure 7.3 The angle of a ray undergoing changes according to Snell's law.

angles, which a medium with a lower refractive index cannot sustain, to the recording medium. For example, with NA equaling unity in vacuum and a refractive index of 1.75 for the resist, let 1 denote the coupling medium and 2 denote the resist. Equation (7.5) sets $\theta_2 = 35$ deg. There is ample room to increase θ_2 , but θ_1 is already saturated. However, with $n_1 = 1.44$, θ_2 increases to 55 deg. The NA is now 1.44 in the resist instead of just 1. The NA can be further increased by a coupling medium with a higher refractive index.

7.2.2 Resolution and DOF scaling equations

The resolution scaling equation is given in Eq. (7.2), and the equation for DOF was given in Eq. (4.4):

$$DOF = k_3 \frac{\lambda}{\sin^2 \frac{\theta}{2}}, \quad (7.6)$$

where k_3 is the DOF scaling coefficient for high-NA imaging. The other two parameters were defined by Eq. (7.2). Representing the equation in NA is less straightforward. One of several ways to present the DOF scaling equation is as follows:

$$DOF = k_3 \frac{n\lambda_0}{NHA^2} = \frac{\lambda_0}{n(1-\cos\theta)} = \frac{\lambda_0}{n \left(1 - \sqrt{1 - \frac{NA^2}{n^2}}\right)}, \quad (7.7)$$

where NHA is the numerical half aperture $n \cdot \sin(\theta/2)$. The first expression is taken from Ref. 1, and the latter two expressions are from Ref. 8. The former uses the form most similar to its paraxial counterpart and is easier to remember. The last form is expressed in terms of the invariant NA. Note that NHA is not an invariant.

7.2.3 Improving resolution and DOF with an immersion system

One application of immersion systems is to improve the DOF of a system whose resolution is adequate but marginal in DOF. In this case, the image-forming angle in the resist does not have to increase; ditto for the angle in the lens. The angle in the coupling medium is reduced, making focusing less sensitive to any physical longitudinal displacement of the wafer. Figure 7.4 shows an example using $\sin\theta_{resist} = 0.51$ and refractive indices 1.56, 1.44, and 1.75 for the lens material, water, and resist, respectively. Simple Snell's law calculations set $\sin\theta_{quartz} = 0.57$, $\sin\theta_{air} = 0.89$, and $\sin\theta_{water} = 0.62$. If the image were to be recorded in the coupling medium, Eq. (7.6) predicts a gain in DOF by 77%. Remember that λ in the equation is reduced in the immersion fluid; it is much larger than the DOF obtained by just multiplying with the refractive index because of the leverage of $\theta/2$ in the high-NA DOF scaling equation. Since the image is not recorded in the coupling medium, the DOF should be expressed in terms of DOF_{avail} , to be given in Sec. 7.3. Nevertheless, the gain in DOF in such a scheme is appreciable. Note that if the paraxial DOF scaling equation is used, the gain in DOF is just by the refractive index, as shown in the following analytic derivation:

$$\frac{DOF_{water}}{DOF_{air}} = \frac{\lambda_{water}}{\sin^2 \theta_{water}} \frac{\sin^2 \theta_{air}}{\lambda_{air}} = \frac{n_{water}}{n_{air}}. \quad (7.8)$$

Immersion systems also improve resolution beyond what a dry system using the same light frequency can provide. Figure 7.5 shows an immersion system that preserves the physical angle in the coupling medium, using the same distribution of refractive indices in Fig. 7.4. If the image were to be recorded in the fluid, then the resolution is improved by 44%, while DOF is reduced by 44%,

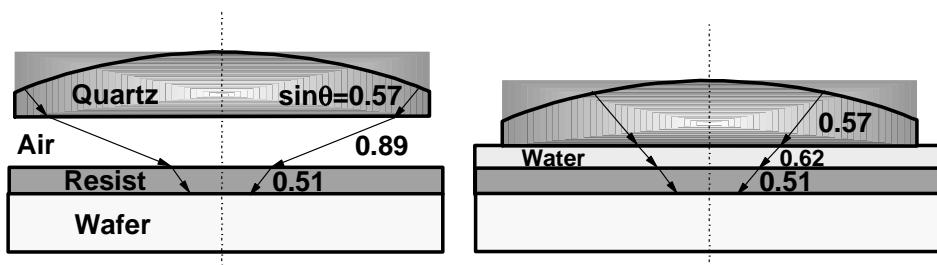


Figure 7.4 Preserving the incident angle in a resist to improve DOF. Assumptions: $n_{quartz} = 1.56$, $n_{water} = 1.44$, $n_{resist} = 1.75$.

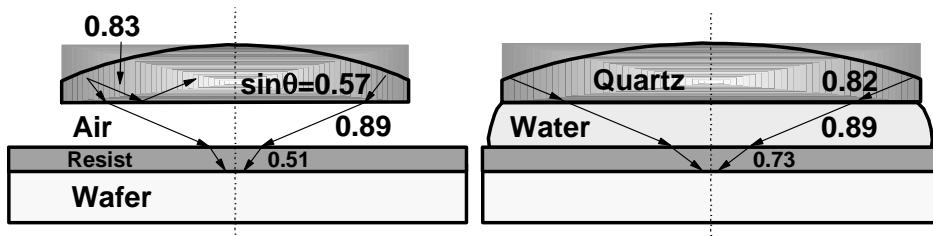


Figure 7.5 Preserving the physical angle in the coupling medium to improve resolution. Assumptions: $n_{\text{quartz}} = 1.56$, $n_{\text{water}} = 1.44$, $n_{\text{resist}} = 1.75$.

according to Eqs. (7.2) and (7.6). Similar to the statement for Fig. 7.4, since the image is not recorded in the coupling medium, the DOF should be expressed in terms of $\text{DOF}_{\text{avail}}$, to be given in Sec. 7.3 below. The actual imaging performance of the immersion system is slightly worse than that of an equivalent reduced-vacuum wavelength due to polarization-dependent stray light, as will be discussed in Sec. 7.4.

7.3 DOF in Multilayered Media¹¹

DOF is often taken from many different perspectives. There is the DOF dictated by diffraction, which supports all of the components in the DOF budget, such as the topography on the wafer, resist thickness, wafer and mask flatness, tilts, focusing error, image plane deviation, film nonuniformity, and scanning imperfections. The required DOF is a function of the budget components. The available DOF is the diffraction DOF with the resist thickness removed. This is the DOF available to the tool designer and process engineer for a given technology node after the resist-thickness requirement is locked in.

7.3.1 Transmission and reflection in multilayered media

The situation we are analyzing is shown in Fig. 7.6, which illustrates the last lens element in a projection lens; the coupling medium can be gas, fluid, solid, or just a vacuum. It also shows the resist contacting the coupling medium and the wafer substrate. We are accustomed to only examining the diffracted image in air. Immersion systems alert us of the need to consider the diffracted image in the resist, as well as in the coupling medium between the lens and the resist.

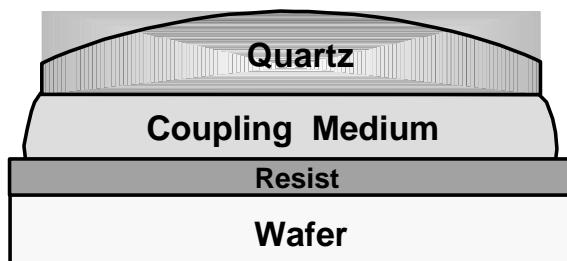


Figure 7.6 Lens, coupling medium, resist layer, and wafer.

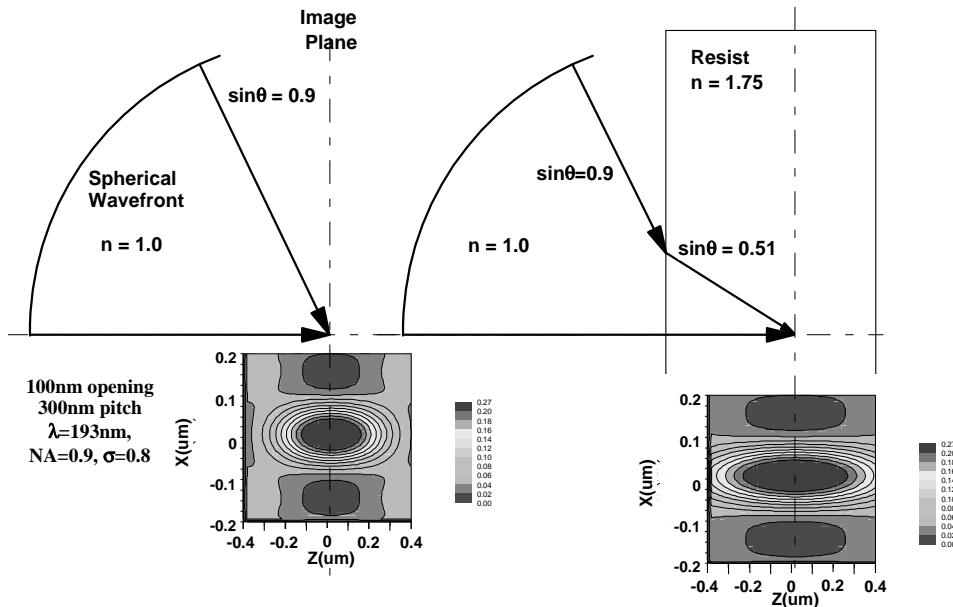


Figure 7.7 Diffracted image near focus with and without a resist medium.

Let us first look at the coupling medium that has a refractive index close to 1. The image space is split into two types. Figure 7.7 shows a simulated diffraction image completely in air, then completely in a thick-resist system. Here, 100-nm line openings 200 nm apart are projected through an $NA = 0.9$, $\lambda = 193\text{-nm}$ imaging lens using an axial illumination of $\sigma = 0.8$. The optical system is assumed to be diffraction limited. Hence, the wavefront in air is spherical. If we kept the spherical wavefront in air as shown, spherical aberration would automatically be incorporated as soon as the thick resist layer is introduced.¹² However, in the simulation, we assume that the lens is corrected for spherical aberration induced by the resist. This is equivalent to using a spherical wavefront in the resist. The diffracted images are shown in constant-intensity contours in the x - z space; x represents the lateral direction, and z represents the longitudinal direction. Figure 7.7 shows that the diffracted images in air and in the thick resist have the same lateral resolution because the NA in the resist remains unchanged. However, the DOF in the resist is improved because of a smaller θ in the resist.

When making a wafer exposure, the resist thickness must be kept within a fraction of the DOF derived from diffraction considerations using the NA, illumination condition, feature size, feature shape, and feature combinations. The remaining DOF, after deducting the resist thickness, is needed for focusing error, tilt, wafer and mask nonflatness, field curvature, scanning-induced errors, etc.

Figure 7.8 shows the situation when the resist thickness is 200 nm. Here, the diffracted image has two parts: one dictated by the wavelength in the coupling medium, and the other dictated by the wavelength in the resist. As the incident wave travels through the coupling medium, it is identical to a wave in a

homogenous medium, shown in Fig. 7.7, until it reaches the resist surface. Then, it splits into two waves: the transmitted wave and the reflected wave. The latter is a mirror image of the transmitted wave adjusted for the refractive index in the coupling medium. The former can be thought of as a part of the diffracted wave in the resist shown in Fig. 7.7, taken within the depth of 200 nm because we assume perfect antireflection coatings above and below the resist, as well as an absorption-free coupling medium and resist. Hence, a transmitted light with no change in intensity across the interface is plotted.

Multiple reflections between the resist-wafer interface and the resist-coupling-medium interface are very important for DOF. When they are present, one should not neglect them. Fortunately, these reflections are often reduced to a minimum to control the delineated image size, i.e., CD control. Figure 7.9 shows CD as a function of the resist thickness over a SiON-bottom antireflection coating at different k values over an oxide layer. When $k = 0.65$, the CD variation is less than 1.5%. The reflectivity swing corresponding to this CD variation is about 1%; this is typically what can be achieved at the manufacturing floor. Consequently, we may assume very good antireflection properties at both interfaces. Figure 7.8 is plotted with these assumptions. With all of the simplifications above, one can literally cut out the intensity contours in air

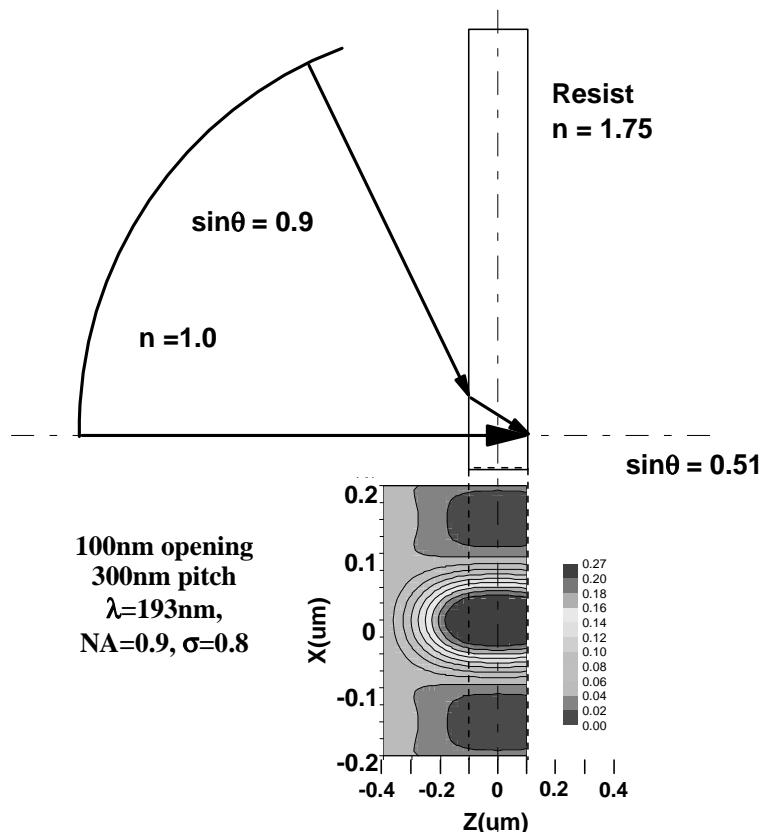


Figure 7.8 Same as Fig. 7.7, except for a 200-nm resist thickness.

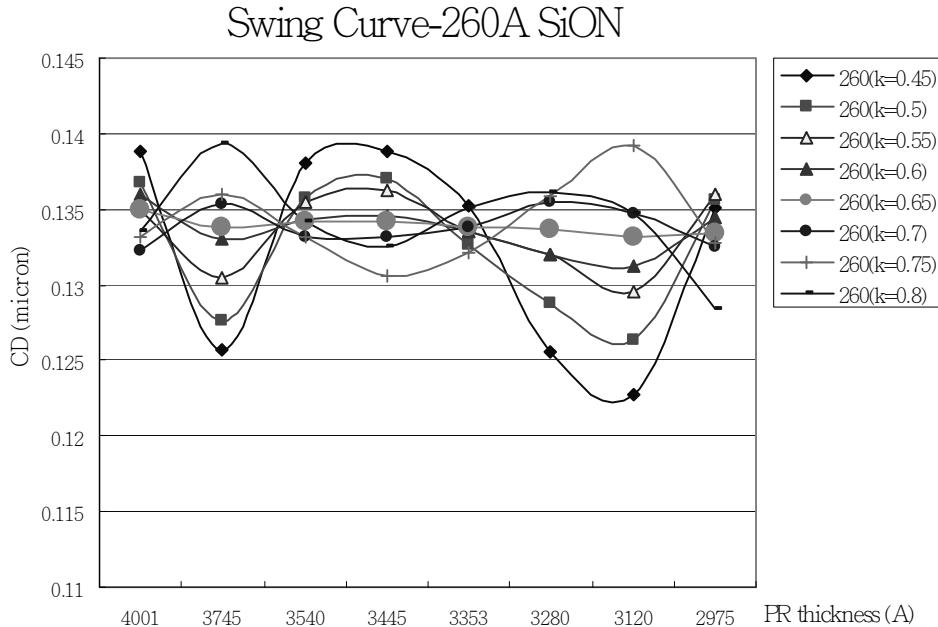


Figure 7.9 CD swing curve as a function of resist thickness on 26 nm of SiON.

between $z = -400$ nm and -100 nm in Fig. 7.7 and paste them to another cutout from the thick resist space between $z = -100$ nm and 100 nm to construct the diffraction pattern in Fig. 7.8, which illustrates the diffracted intensity distribution in two media. In our case, a computer program was written for diffracting in the composite media and pasting. Strictly speaking, spherical aberration can be corrected for only one of the two media. For our purpose of illustrating the diffracted image in two media, neglecting spherical aberration is sufficient.

7.3.2 Effects of wafer defocus movements

Because the resist thickness does not change, when the wafer moves in the longitudinal direction, only the resist surface moves. Therefore, the sensitivity of the wafer's longitudinal movement is mostly governed by the index of the coupling medium instead of the resist. This point is illustrated in Figs. 7.10 and 7.11, using the imaging example in Fig. 7.7. The indices used for the resist and for the coupling medium are 1.75 and 1.46, respectively. The former shows the constant intensity contours in air and in resist as the air-resist interface moves from -200 to 200 nm in 100 nm steps. $\Delta Z = 0$ is the focal plane. That is, the resist layer is perfectly centered when the air-resist interface is at -100 nm. The moving distance in 100-nm steps refers to movements in the resist imaging space. For an outside observer concerned with controlling the distance between the lens and the wafer, the steps should be adjusted for the difference in refractive index. That is, $100\text{ nm}/1.75$, resulting in 57-nm steps. Similarly, in the case of immersion, the 100-nm steps in the resist become $100 \times 1.46/1.75 = 83$

nm, as far as controlling the distance between the lens and the wafer is concerned. The use of the refractive indices to adjust the defocusing steps will become clear after understanding Eq. (7.10). When using simulation, one must be careful with the DOF given by the simulator. If it is given in the resist space, the above adjustments should be performed. If it is given in air, then the steps in a dry medium need no correction, but the step in a fluid-coupling medium should be multiplied by the refractive index of the fluid.

When there is low light absorption and high development contrast in the resist, the constant intensity contours are good indicators of the resist profile at different exposure dosages. From Fig. 7.10, one sees that the resist profile is symmetrical when it is centered at the focal plane. There is more exposure at the top of the resist and less at the bottom as the wafer moves away from the lens, and it is the opposite as it moves closer. The span in the resist profile is indeed stretched in the z -direction because of the higher index in the resist. However, the contours in the coupling medium dictate the starting point of the contours in the resist. The former are crucial for controlling the distance between the lens and the wafer. Figure 7.11 shares the same trend as Fig. 7.10. The only difference here is that the coupling medium now has an index of 1.46, thus extending the contours in the z -direction and producing more tolerance for the distance between the lens and the wafer. Note that at $\Delta Z = -83$ nm in the immersion case, the contours are still slightly better than without immersion at $\Delta Z = -57$ nm. It is even more obvious at the other ΔZ settings. This means immersion lithography tolerates larger longitudinal physical movements between the lens and the wafer than just $n_{CM}:1$.

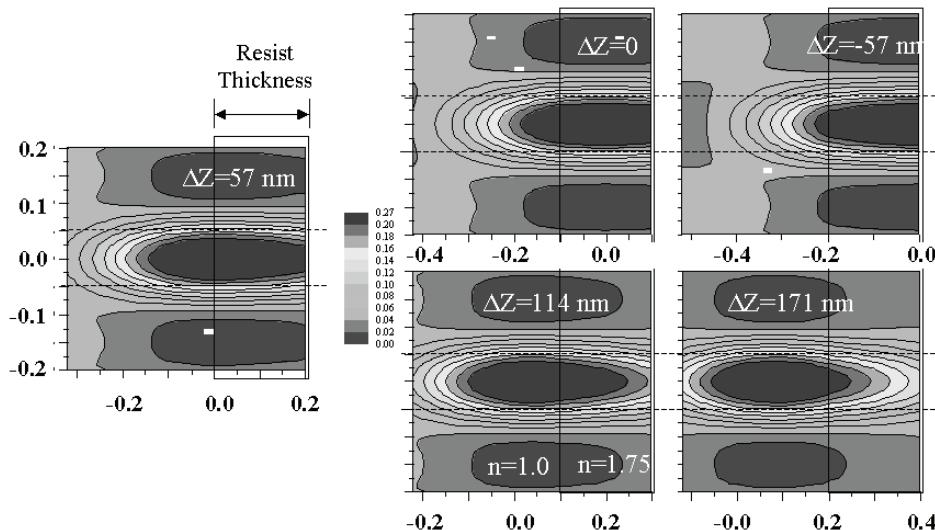


Figure 7.10 Resist layer in air at various ΔZ .

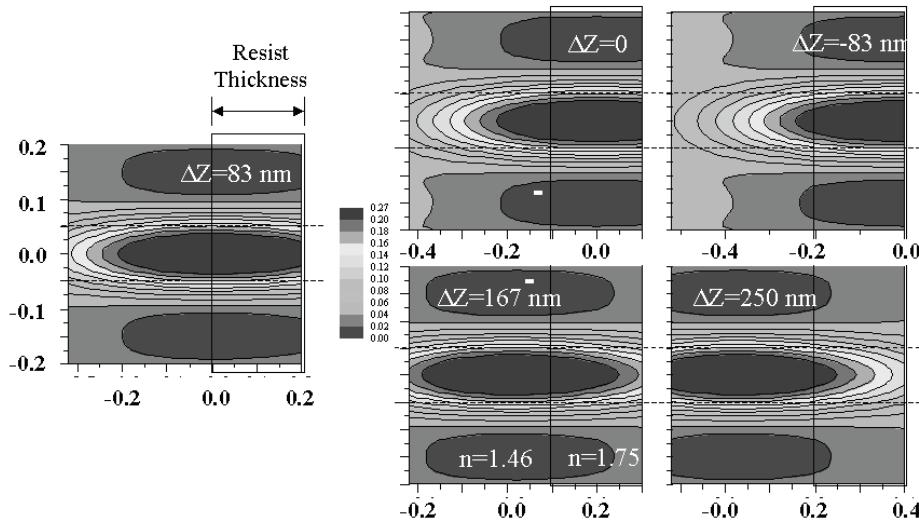


Figure 7.11 Resist layer in fluid at various ΔZ .

7.3.3 Diffraction DOF

The resolution and diffraction DOF in a single medium are represented by Eqs. (7.2) and (7.6). In the resist, the reductions of λ and $\sin\theta$ compensate for each other to keep resolution unchanged. The effect of $\sin^2(\theta/2)$ is stronger than the effect of wavelength reduction, resulting in a larger $\text{DOF}_{\text{diffrac}}$. Converting to dependency in NA, these equations become Eqs. (7.3) and (7.7). Equations (7.6) and (7.7) describe the situation of Fig. 7.7, when the entire diffraction DOF is contained within one single medium. They are different from Eq. (7.3) in terms of being independent of the medium. With these equations, Snell's law makes NA an invariant in any medium in the optical imaging train, and thus W stays constant. Just as $\text{DOF}_{\text{diffrac}}$ is not an invariant, neither is NHA.

Determining the correct k_3 in Eqs. (7.6) and (7.7) is very important. One takes the properly illuminated features of interest to either simulate or project the image through the lens and the coupling medium to the surface of the photoresist, then evaluates the range of defocus that keeps CD within specification for all features, and adjusts this range with the refractive index of the coupling medium. This sets k_3 to be used in these equations. One should bear in mind that at a high NA, light polarization makes the scaling equations less effective, as discussed in Sec. 7.4.

During wafer printing, the diffraction DOF must support the required DOF of the imaging system. The requirement contains focus error, wafer and mask flatness within the exposure field, wafer tilt with respect to the optical axis, scanning imperfections, image-plane deviation from the optical system, topography on the substrate, film nonuniformity, and resist thickness. Hence, for a given set of patterns on the mask, its diffraction DOF must be larger than the required DOF. These two types of DOF come from different sides. In a way, the

diffraction DOF is like a DOF reservoir for the required DOF to consume. Hence, a usable imaging system requires

$$DOF_{\text{diffrac}} \geq DOF_{\text{required}}. \quad (7.9)$$

To determine DOF_{diffrac} , one takes all of the representative feature types and sizes under the predetermined imaging condition, such as λ , NA, illumination setting, and lens aberrations, to evaluate the CD at a given range of exposure dosage and defocus. The E-D tree of each feature type and size is evaluated from the CD data. The common E-D window from the common E-D tree is used to set the DOF of a given exposure latitude (EL). Substituting into Eq. (7.6) or (7.7) yields k_3 .

The foregoing procedure takes care of evaluating the DOF in vacuum and has been documented in various publications.^{13,14,15} Since the refractive index of air is only slightly larger than that of vacuum, not much error is introduced. In a coupling medium having a refractive index much higher than that of vacuum, DOF_{diffrac} is increased by $n_{\text{coupling medium}}$. Similarly, in a thick-resist medium whose thickness is larger than DOF_{diffrac} , the DOF is increased by n_{resist} , as demonstrated in Fig. 7.7.

The diffraction DOF obeys the high-NA scaling law¹ as long as polarization effects are omitted. The polarization effects will be treated in Sec. 7.4.

7.3.4 Required DOF

To determine DOF_{required} , all of the contribution components are combined either linearly or quadratically, depending on whether a particular component is systematic or statistical.^{6,16,17,18} The following equation demonstrates our current understanding of DOF_{required} . While wavelength is reduced in the resist medium that covers the wafer substrate and wafer topography, thus increasing the focal range of the acceptable diffraction image, many other contributing components are based on the resist surface moving in the coupling medium. As illustrated in the figure, the DOF contribution from these other components should use the refractive index of the coupling medium.

$$\begin{aligned} \frac{1}{n_{CM}} DOF_{\text{required}} = & \frac{1}{n_{\text{resist}}} (THK_{\text{resist}} + Z_{\text{topo}} + \Delta Z_{IPD}) \\ & + \left[\frac{1}{n_{\text{resist}}^2} (\Delta Z_{\text{resist}}^2 + \Delta Z_{\text{topo}}^2 + \Delta Z_{FU}^2) + \frac{1}{n_{air}^2} \left(\Delta Z_{\text{focus}}^2 + \Delta Z_{SI}^2 + \frac{\Delta Z_{MF}^2 + \Delta Z_{\text{tilts at mask}}^2}{M^2} \right) \right]^{1/2}, \end{aligned} \quad (7.10)$$

where IPD denotes the image plane deviation, FU is film uniformity, CM is the coupling medium, SI is scanning imperfections, WL is wafer flatness, ML is mask flatness, and M is mask reduction ratio. All contributing components except resist

thickness are random, either within each exposure field or between exposure fields. By associating topography and film uniformity to n_{resist} , we made the assumption that the resist perfectly planarizes over the topography and the substrate film. We also assume that the image-plane deviation is completely contained in the resist thickness. ΔZ_{tilts} contains tilts at the wafer side as well as the mask side. Tilts at the mask side must be divided by M^2 . Film uniformity refers to local variations in the substrate films; they are mostly planarized by the resist. Therefore, their DOF contributions are governed by the resist index. Wafer flatness is spread out in a larger lateral range; therefore, the resist follows the wafer flatness conformably. As a result, its DOF contribution is governed by the index of the coupling medium.

The difference between a dry imaging system and an immersion system is in n_{CM} . Because of the larger n_{CM} of immersion systems, the DOF requirement is increased. Therefore, when the vacuum wavelength is identical to the immersed wavelength; a dry system requires less DOF. However, with the same light frequency, the immersed wavelength is shorter, resulting in a larger $DOF_{diffrac}$. It more than compensates for the loss in a higher required DOF.

7.3.5 Available DOF

DOF_{avail} for device fabrication is $DOF_{diffrac}$ minus the contribution from resist thickness. It is the room to contain the other defocus-contributing components. In a simple-minded way, DOF_{avail} is the DOF measured in the resist image at the desired resist thickness through an exposure-defocus matrix measurement. The measurement must be carefully performed to remove any measurement error and any contribution from wafer nonflatness, tilting error, etc. Analytically, the DOF_{avail} is

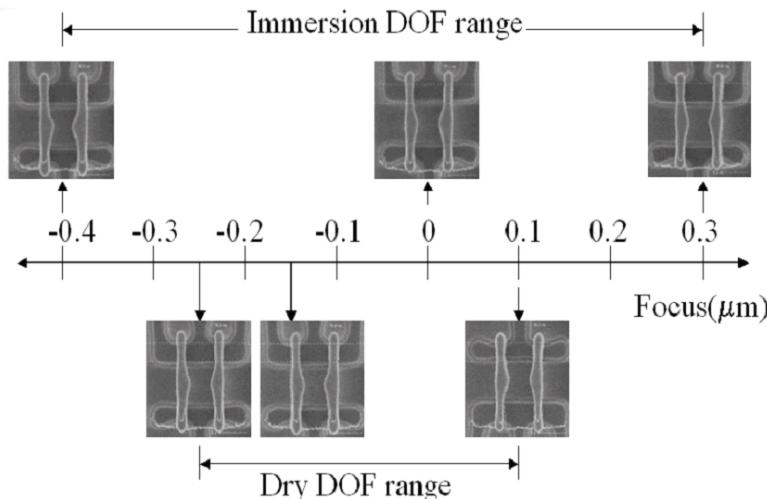
$$DOF_{avail} = k_3 \frac{\lambda_{coupling\ medium}}{\sin^2 \frac{\theta}{2}} - \frac{n_{CM}}{n_{resist}} THK_{resist}, \quad (7.11)$$

and

$$\begin{aligned} \frac{1}{n_{CM}} DOF_{avail} &= \frac{1}{n_{resist}} (Z_{topo} + \Delta Z_{IPD}) \\ &+ \left[\frac{1}{n_{resist}^2} (\Delta Z_{resist}^2 + \Delta Z_{topo}^2 + \Delta Z_{FU}^2) + \frac{1}{n_{air}^2} \left(\Delta Z_{focus}^2 + \Delta Z_{SI}^2 + \frac{\Delta Z_{MF}^2 + \Delta Z_{tilts\ at\ mask}^2}{M^2} \right) \right]^{\frac{1}{2}} \\ &+ \left[\frac{1}{n_{CM}^2} (\Delta Z_{tilts\ at\ wafer}^2 + \Delta Z_{WF}^2) \right]^{\frac{1}{2}}. \end{aligned} \quad (7.12)$$

Table 7.1 $\text{DOF}_{\text{diffrac}}$ and $\text{DOF}_{\text{avail}}$ of dry and immersion systems.

$\lambda(\text{nm})$	$\sin\theta$	k_1	k_3	$\text{DOF}_{\text{diffrac}}$	$\text{DOF}_{\text{avail}}$
193	0.9	0.466	0.35	240	126
132	0.616	0.466	0.367	451	289

**Figure 7.12** DOF of a polyimage on an active layer.

DOF_{avail} from an immersion system is larger than that from a dry system by more than a factor of the refractive index. Let us continue with the given example in Figs. 7.10 and 7.11. With a resist thickness of 200 nm, $n_{\text{resist}} = 1.75$ and $n_{CM} = 1.46$. DOF_{diffrac} and DOF_{avail} are tabulated in Table 7.1.

Here, DOF_{diffrac} is evaluated with n_{CM} and n_{resist} equal to 1 according to Eq. (7.3), and DOF_{avail} according to Eq. (7.11). The DOF_{avail} of this particular immersion system is almost three times that of the dry system. The slightly higher k_3 for the immersion system is due to less polarization-dependent stray light, which will be discussed in more detail in Sec. 7.4.

DOF improvements are confirmed experimentally. Figure 7.12 shows that the DOF of the etched polyimage exposed over the active layer is for a 90-nm SRAM circuit. The DOF determined by $\pm 10\%$ CD is 350 and 700 nm from dry and immersion imaging, respectively. The after-develop image has 314- and 622-nm DOFs.

Figure 7.13 shows the DOF of 130-nm contact holes determined with an E-D window. Axial illumination at $\sigma = 0.6, 0.75$ NA, and a 6% AttPSM mask were used. Individual and common E-D windows at 260-, 400-, and 430-nm pitches were plotted. The DOF improvement for each individual window is between 30% and 50%, and for the common window it is 51%.

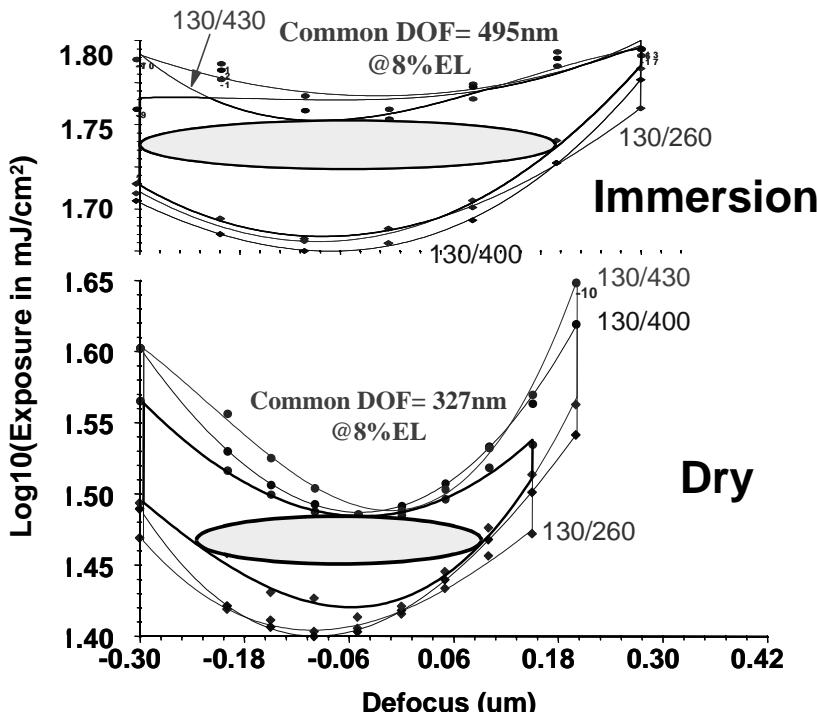


Figure 7.13 DOF of contact holes determined with an E-D window from immersion and dry exposures.

7.3.6 Preferred refractive index in the coupling medium

What is the optimum refractive index of the coupling medium? From the point of view of index matching to reduce reflection, it should be identical to the index of the resist, the lens material that the fluid is in contact with, or an optimum value between the two contacting materials. However, with the impressive progress made in antireflection coatings, index reflection reduction with the coupling medium is not that important. One should make the index as high as possible to increase the DOF. If such a medium is available, one should not hesitate to use an index higher than that of the resist. Granted, total internal reflection will prevent high spatial frequencies from entering the resist, but DOF is still increased as a function of its high index of refraction. The rejected frequencies are not available without the coupling medium anyway. Therefore, the refractive index in the coupling medium should be as high as possible.

7.3.7 Tradeoff between resolution and DOF_{diffrac}

The tradeoff between resolution and DOF_{diffrac} in an immersion system is similar to that of a wavelength reduction system. Equations (7.2) and (7.3) are used to demonstrate the tradeoff, neglecting polarization-dependent stray light for the time being. There are two extremes in the tradeoff: maximum resolution and maximum DOF_{diffrac}. To maximize resolution, the angle θ is maintained as high

as dry lithography can achieve. For example, $\sin\theta$ is kept at 0.9 in the presence of a high-index medium. This way, resolution is improved by $1/n_{CM}$, while $DOF_{diffrac}$ is reduced by the same factor. The situation is equivalent to simply reducing the vacuum wavelength. $DOF_{diffrac}$ reduction follows resolution improvement linearly instead of quadratically, as in the case of increasing NA. Another extreme is to keep the resolution constant by reducing θ to make $\sin\theta = (\sin\theta_o)/n_{CM}$. $DOF_{diffrac}$ is now improved by the following equation:

$$DOF_{diffrac} = k_3 \frac{\lambda}{\sin^2(\theta/2)} = k_3 \left(\frac{\lambda_o}{n} \right) \frac{1}{\sin^2 \left[\frac{1}{2} \sin^{-1} \left(\frac{1}{n} \sin \theta_o \right) \right]}. \quad (7.13)$$

The tradeoff is illustrated in Fig. 7.14. Here, Eqs. (7.2) and (7.3) are plotted in the $DOF_{diffrac}$ -versus-resolution space in the log scale, according to $k_1 = 0.315$ and $k_3 = 0.223$. λ and $\sin\theta$ are now parameters determining the resolution and attainable $DOF_{diffrac}$. This set of scaling factors is evaluated using the E-D forest methodology, 6% AttPSM, 8% EL, and $\pm 10\%$ CD tolerance. The common E-D window of three pitches corresponding to 1:1.5, 1:2, and 1:3 line-space ratios is used. In this figure, the three $DOF_{diffrac}$ curves corresponding to 193, 157, and 132 nm are plotted, showing that 193-nm immersion is indeed a generation better than 157-nm dry systems. At 193 nm and 0.7 NA, the resolution is 87 nm and $DOF_{diffrac}$ is 300 nm. A feasible range of improvement using water immersion is shown in the shaded area between the 193- and 132-nm curves. Preserving the resolution improves $DOF_{diffrac}$ to 480 nm with $\sin\theta = 0.48$. Preserving the 300-nm $DOF_{diffrac}$ by moving the immersed $\sin\theta$ to 0.59 produces a resolution of 70 nm. Using the same $\sin\theta$ at 0.7, but increasing the wavelength to 132 nm with immersion, produces a resolution of 59 nm while reducing $DOF_{diffrac}$ to 206 nm.

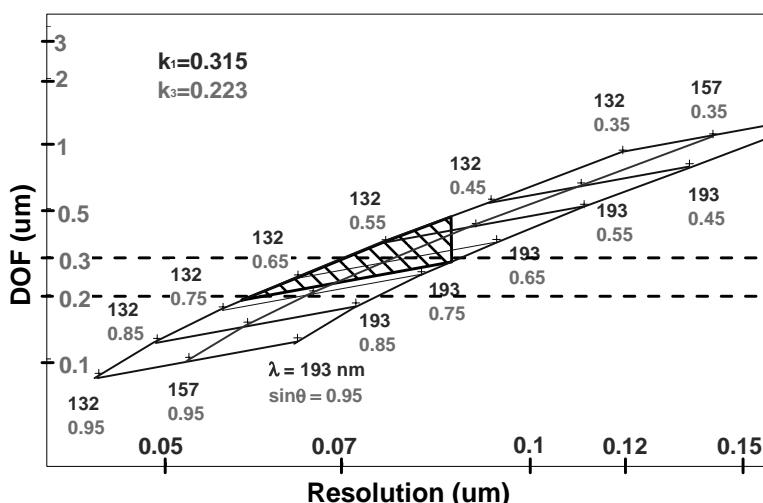


Figure 7.14 Tradeoff between resolution and $DOF_{diffrac}$ using annular illumination and AttPSM.

7.4 Polarization-Dependent Stray Light

Light, being an electromagnetic wave, has vectorial electric and magnetic fields. The orientation of these fields affects the transmission and reflection characteristics at the interfaces. Furthermore, the imaging process itself is also polarization dependent. The contrast of high spatial frequencies at the wrong polarization can be reduced. At a low NA, the polarization effects are not pronounced, and therefore have been ignored so far. Now, when NA is 0.6 or higher, especially when it is dramatically increased by immersion, the polarization effects can no longer be ignored.

In the last decade, vectorial illumination and imaging started to gather attention from lithographers. Flagello and Rosenbluth⁷ explored lithographic tolerances based on vector diffraction theory at high NA in the aerial image and in the resist. Recently Brunner et al.⁹ explained transverse magnetic (TM) imaging and film reflection at high angles. They also produced various experimental demonstrations. Smith and Cashmore¹⁹ showed polarization-dependent imaging in the resist and explored the impact of resist parameters.

This section analyzes two prominent polarization effects: (1) loss of contrast in TM polarization when spatial frequencies are recombined to form the image, and (2) stray light induced by specular reflection from the resist surface.

Both effects can be accounted for quantitatively in terms of stray light. These polarization-dependent stray lights (PDS)²⁰ are combined with the system stray light (SSL) normally existing in an optical imaging system into the intensity distribution of the optical image in order to generate E-D trees and windows that lead to the evaluation of DOF and EL of a set of given objects. We then compare the DOF loss due to PDS for line-space pairs, contact holes, disk illumination, annular illumination, and inclusion of more than the 1st-order spatial frequency in the object. A comparison is also made on 193-nm immersion and dry systems, as well as on 157-nm dry systems.

7.4.1 Imaging at different polarizations

The imaging process can be viewed as the transmitting and filtering of spatial frequencies. As light leaves the mask, it carries all of the spatial frequencies contained in the mask toward the imaging lens, which assumes the role of recombining these spatial frequencies at the image plane to reproduce the mask pattern. Because of finite NA, the lens also behaves like a low-pass filter, cutting off the higher spatial frequencies that exceed the limit imposed by the NA and transmitting the lower frequencies for recombination. Figure 7.15 shows the angular distribution of the spatial frequencies generated from the diffraction of the incident zero-frequency illumination beam by large and small openings in the mask. The lens pupil cuts off any spatial frequency outside the $\pm\theta$ range. The large opening diffracts into a smaller angular spectrum and is reproduced better than the smaller opening.

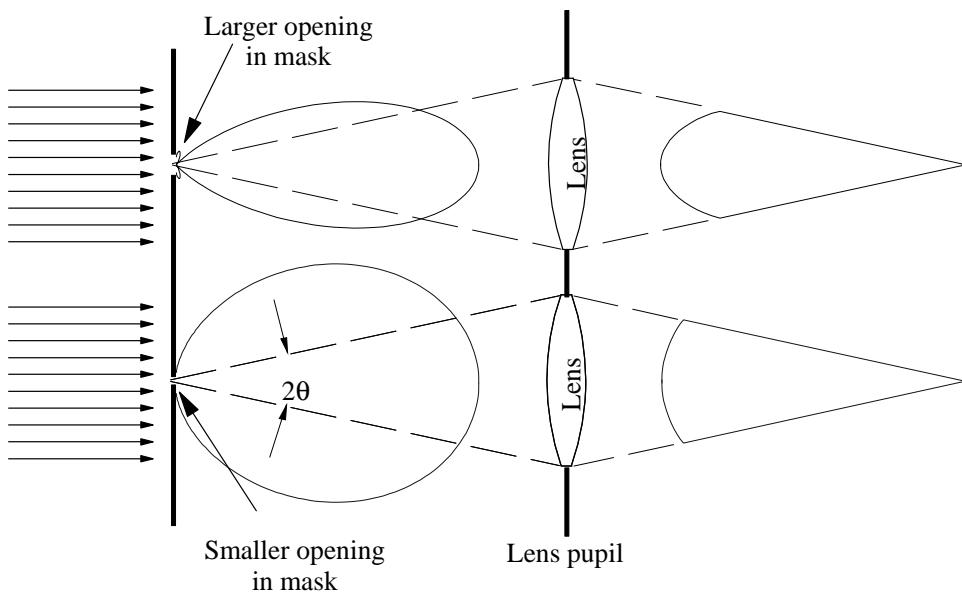


Figure 7.15 Diffraction from large and small openings. The lens cuts off the higher spatial frequency components, resulting in lost resolution. The diffraction patterns are normalized to the same intensity along the optical axis.

7.4.1.1 Recombination of spatial frequency vectors in the resist

The recombination process is polarization dependent. Let's examine a single pitch combining two spatial-frequency vectors in the transverse electric (TE) mode, making incident angles $\pm\alpha$ with respect to the optical axis of the lens, as shown in Fig. 7.16. The vectors represent the propagation direction of the positive and negative spatial frequencies. In the TE mode (i.e., when the electric field of the electromagnetic wave is perpendicular to the plane of incidence), the interference is straightforward. Let the electric field be E_1 and E_{-1} for each vector, and I be the intensity after the recombination:

$$E_1 = \hat{y}A_1 e^{ik(-x\sin\alpha+z\cos\alpha)} \quad (7.14)$$

and

$$E_{-1} = \hat{y}A_{-1} e^{ik(x\sin\alpha+z\cos\alpha)}, \quad (7.15)$$

where $k = 2\pi/\lambda$. The recombined intensity is

$$I = [E_1 + E_{-1}]^* [E_1 + E_{-1}] = A_1^2 + A_{-1}^2 + 2A_1 A_{-1} \cos(2kx\sin\alpha). \quad (7.16)$$

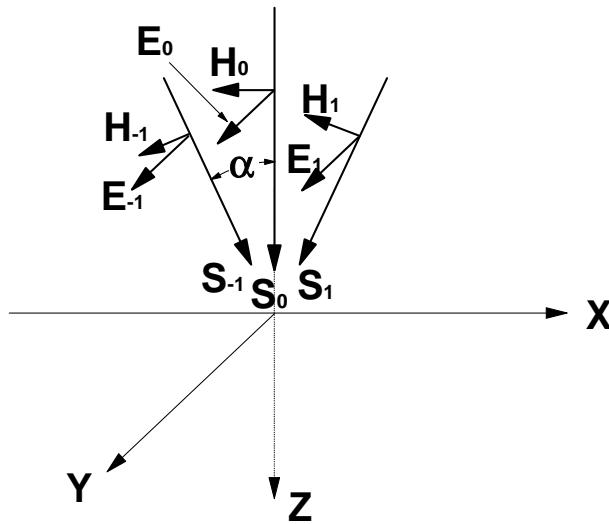


Figure 7.16 Imaging beams in TE mode.

Assuming $\alpha > 0$, when $2kx\sin\alpha = 0$, I becomes the maximum, and $2kx\sin\alpha = \pi$ becomes the minimum:

$$\text{Contrast} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2A_1 A_{-1}}{A_1^2 + A_{-1}^2}, \quad \alpha > 0. \quad (7.17)$$

Usually, $A_1 = A_{-1}$ and contrast = 1. Note that when $\alpha = 0$,

$$I_{\max} = I_{\min} = A_1^2 + A_{-1}^2 \quad (7.18)$$

and

$$\text{Contrast} = 0, \quad \alpha = 0. \quad (7.19)$$

Therefore, when two beams making finite angles with respect to the optical axis are symmetrical to it and form an image, it has a periodicity of $\lambda/2\sin\alpha$ and a contrast of 1, independent of the axial position, i.e., the DOF is infinite. The two-beam situation is exemplified by AltPSM or off-axis illumination. Note that we have used two coherent beams to illustrate imaging in the TE mode. When we are imaging through a lens using partially coherent illumination, the contrast is less than unity, and the DOF is finite. In the situation of off-axis illumination, one of the two interfering beams is actually the original 0th-order beam tilted by the off-axis light, and one of the 1st-order beams is thrown out of the coverage of the aperture angle θ . Hence, A_1 and A_{-1} are generally not identical to each other. In addition, when the tilting angle does not make the two beams symmetrical to the optical axis, many unwanted interference terms are produced, as given in detail in Sec. 6.3.2.1.3 and a prior article.²¹ Of course, one wants to optimize the

tilting angle, but only one pitch can be optimized out of all of the pitches in the circuit layout.

In many situations, the 0th-order beam is included in imaging, such as when the illumination is axial or when AltPSM is not used. Let's consider the case for three-beam imaging in the TE mode:

$$E_0 = \hat{y}A_0 e^{ikz}. \quad (7.20)$$

E_1 and E_{-1} were already given in Eqs. (7.14) and (7.15). The recombined intensity is

$$I = [E_0 + E_1 + E_{-1}]^* [E_0 + E_1 + E_{-1}]. \quad (7.21)$$

In the three-beam case, the possibility of $A_1 = A_{-1}$ is high. Assuming this condition leads to

$$I = A_0^2 + 2A_1^2 + 2A_1^2 \cos(2kx \sin \alpha) + 4A_0 A_1 \cos(kx \sin \alpha) \cos kz(\cos \alpha - 1). \quad (7.22)$$

The third term indicates a double-frequency component due to the interference of A_1 and A_{-1} . The fourth term carries the fundamental frequency dictated by the pitch of the object. The double-frequency component is usually filtered out by the lens aperture angle θ . Assuming $\alpha > 0$, when $kx \sin \alpha = 0$, I becomes the maximum, and $kx \sin \alpha = \pi$ becomes the minimum.

$$\text{Contrast} = \frac{4r \cos kz(\cos \alpha - 1)}{1 + 4r^2}, \quad \alpha > 0, \quad (7.23)$$

where

$$r \equiv \frac{A_1}{A_0}.$$

Note that even with coherent light, there is a z -dependence of the image contrast in the three-beam case. Also, since the electric-field vectors are combined as if they are scalars, the TE case is identical to the image formation by scalar prediction as long as the imaging involves one single medium.

It is more complicated in the TM mode when the magnetic-field vector of the imaging beams is perpendicular to the plane of incidence, as shown in Fig. 7.17. Let us first consider the two-beam case. Since the electric field is now an arbitrary vector in the x - z plane, it is decomposed into the x and z components for calculation:

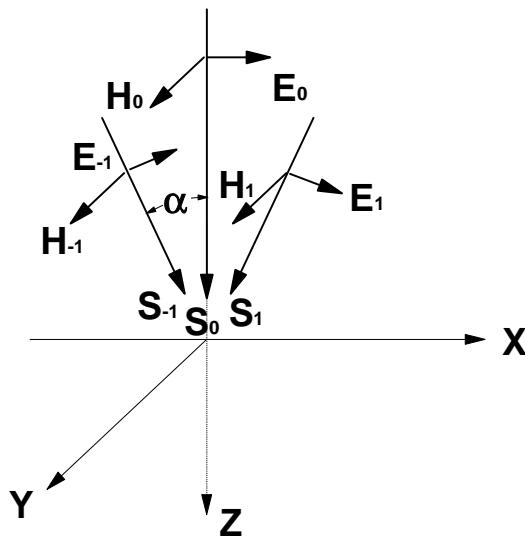


Figure 7.17 Imaging beams in TM mode.

$$E_l = \hat{E}_l A_l e^{ik(-x \sin \alpha + z \cos \alpha)}. \quad (7.24)$$

$$E_{lx} = |\hat{E}_l \cdot \hat{x}| A_l e^{ik(-x \sin \alpha + z \cos \alpha)} = A_l \cos \alpha e^{ik(-x \sin \alpha + z \cos \alpha)}. \quad (7.25)$$

Similarly,

$$E_{-lx} = |\hat{E}_{-l} \cdot \hat{x}| A_{-l} e^{ik(\sin \alpha + z \cos \alpha)} = A_{-l} \cos \alpha e^{ik(\sin \alpha + z \cos \alpha)}. \quad (7.26)$$

$$E_{lz} = |\tilde{E}_l \cdot \hat{z}| A_l e^{ik(-x \sin \alpha + z \cos \alpha)} = A_l \sin \alpha e^{ik(-x \sin \alpha + z \cos \alpha)}. \quad (7.27)$$

$$E_{-lz} = |\tilde{E}_{-l} \cdot \hat{z}| A_{-l} e^{ik(x \sin \alpha + z \cos \alpha)} = -A_{-l} \sin \alpha e^{ik(x \sin \alpha + z \cos \alpha)}. \quad (7.28)$$

The recombined intensity is

$$\begin{aligned} I &= E_x * E_x + E_z * E_z = [E_{lx} + E_{-lx}] * [E_{lx} + E_{-lx}] + [E_{lz} + E_{-lz}] * [E_{lz} + E_{-lz}] \\ &= A_l^2 + A_{-l}^2 + 2A_l A_{-l} \cos(2kx \sin \alpha) \cos 2\alpha. \end{aligned} \quad (7.29)$$

Substituting $2kx \sin \alpha = 0$ and $2kx \sin \alpha = \pi$ for I_{max} and I_{min} ,

$$Contrast = \frac{2A_l A_{-l} \cos 2\alpha}{A_l^2 + A_{-l}^2}, \quad \alpha > 0. \quad (7.30)$$

Equations (7.17) and (7.30) determine that the contrast in the TM mode is lower by the factor $\cos 2\alpha$ with respect to the TE or scalar case. That is,

$$\frac{Contrast_{TM}}{Contrast_{TE}} = \cos 2\alpha, \text{ 2 beams.} \quad (7.31)$$

For the three-beam case in the TM mode, we take the same E_1 and E_{-1} as in Eqs. (7.25) to (7.28) and add the 0th-order component E_0 :

$$E_0 = \hat{x}A_0 e^{ikz}. \quad (7.32)$$

With $A_1 = A_{-1}$,

$$\begin{aligned} I &= E_x^* E_x + E_z^* E_z = [E_{0x} + E_{1x} + E_{-1x}]^* [E_{0x} + E_{1x} + E_{-1x}] + [E_{1z} + E_{-1z}]^* [E_{1z} + E_{-1z}] \\ &= A_0^2 + 2A_1^2 + 2A_1^2 \cos(2kx \sin \alpha) \cos 2\alpha + 4A_0 A_1 \cos(kx \sin \alpha) \cos kz (\cos \alpha - 1) \cos \alpha. \end{aligned} \quad (7.33)$$

Assuming $\alpha > 0$, when $kx \sin \alpha = 0$, I becomes the maximum, and $kx \sin \alpha = \pi$ becomes the minimum.

$$Contrast = \frac{4r \cos kz (\cos \alpha - 1) \cos \alpha}{1 + 2r^2 (1 + \cos 2\alpha)}, \alpha > 0. \quad (7.34)$$

The comparison of Eqs. (7.23) and (7.34) to derive a contrast equation is less straightforward than the derivation of Eq. (7.31) in the TE case. Strictly speaking, the contrast in the TM case is

$$\frac{Contrast_{TM}}{Contrast_{TE}} = \frac{1 + 4r^2}{1 + 2r^2 (1 + \cos 2\alpha)} \cos \alpha, \text{ 3 beams.} \quad (7.35)$$

At a small α ,

$$\frac{Contrast_{TM}}{Contrast_{TE}} = \cos \alpha, \text{ 3 beams.} \quad (7.36)$$

The contrast in the TM case is now lower than that in the TE or scalar case by $\cos \alpha$ at a small α . For a large α , the contrast in the TM case is slightly larger than what Eq. (7.36) predicts. With unpolarized or circularly polarized light,

$$I = \frac{I_{TE} + I_{TM}}{2} \quad (7.37)$$

and

$$\text{Contrast} = \frac{A_1 A_{-1} (\cos 2\alpha + 1)}{2(A_1^2 + A_{-1}^2)}, \text{ 2 beams.} \quad (7.38)$$

$$\text{Contrast} = \frac{2r[\cos kz(\cos \alpha - 1)(\cos \alpha + 1)]}{1 + 4r^2}, \text{ 3 beams, small } \alpha \text{ only.} \quad (7.39)$$

Therefore, compared to the scalar results, the contrast differs by a factor of $(\cos \alpha + 1)/2$ for three-beam recombination and $(\cos 2\alpha + 1)/2$ for two beams. It means that at the largest possible α , i.e. $\pi/2$, the contrast drops to 0.5 for the former and 0 for the latter. Fortunately, recombination imaging takes place inside the resist layer, which has a refractive index larger than that of the coupling medium. The angle α is reduced in the resist according to Snell's law and can be much less than π . For example, when light is directly incident into the resist from air, $\sin \alpha_{\text{resist}} = (1 / n_{\text{resist}}) \sin \alpha_{\text{air}}$. In the case of immersion, $\sin \alpha_{\text{resist}} = (n_{\text{fluid}} / n_{\text{resist}}) \sin \alpha_{\text{fluid}}$. The light that does not contribute to contrast is harmful. It will be treated in Sec. 7.4.

Even though we have derived the $\cos 2\alpha$ and $\cos \alpha$ dependencies using coherent light, the relationship holds for partially coherent and incoherent light as well. These illumination conditions can be simulated with superposition of diffraction intensity from coherent point sources in the lens pupil.^{22,23}

7.4.1.2 Polarized refraction and reflection at the resist surface

The introduction of the resist surface complicates the situation. The impact of this extra surface in terms of scalar diffraction and on the DOF budget was given in Sec. 7.3. Here, we will examine the loss of the imaging light as a function of the polarization in the light beams. The Fresnel formulae²⁴ on reflection and refraction of polarized light are

$$T_{\text{TM}} = \frac{2n_1 \cos \alpha_1}{n_2 \cos \alpha_1 + n_1 \cos \alpha_2} B_{\text{TM}} \quad (7.40)$$

$$T_{\text{TE}} = \frac{2n_1 \cos \alpha_1}{n_1 \cos \alpha_1 + n_2 \cos \alpha_2} B_{\text{TE}} \quad (7.41)$$

$$R_{\times} = \frac{n_2 \cos \alpha_1 - n_1 \cos \alpha_2}{n_2 \cos \alpha_1 + n_1 \cos \alpha_2} B_{\text{TM}} \quad (7.42)$$

and

$$R_{\text{TE}} = \frac{n_1 \cos \alpha_1 - n_2 \cos \alpha_2}{n_1 \cos \alpha_1 + n_2 \cos \alpha_2} B_{\text{TE}}, \quad (7.43)$$

where

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2, \quad (7.44)$$

and T and R are amplitudes of the transmitted and reflected lights. In order to distinguish the amplitude of the electric vector in the two spaces, A_1 and A_{-1} in Eqs. (7.14) and (7.15) are renamed B_{11} and B_{-11} . Their counterparts in the transmitted space are B_{12} and B_{-12} , respectively. Hence, B_{12} and B_{-12} are related to B_{11} and B_{-11} through Eq. (7.41) for the TE mode and through Eq. (7.40) for the TM mode by setting $B_{11} = B_{TE}$, $B_{12} = T_{TE}$, etc.

The transmission in the TE and TM modes through an interface between $n_1 = 1$ and $n_2 = 1.75$ is plotted in Fig. 7.18. The differentiation in polarization is perceivable starting from $\sin\theta = 0.3$. However, it begins to make a difference from $\sin\theta = 0.6$. The loss of exposure energy is less important than the reflected light becoming stray light. At $\sin\theta = 0.8$, the total unpolarized transmission is about 90%, and reflection is about 10%. If completely converted to stray light on a wafer, 10% is significant, whereas a loss of exposure affects only throughput. From Eq. (7.42), $R_{TM} = 0$ when $n_2 \cos\alpha_1 - n_1 \cos\alpha_2 = 0$. This specific angle is called the Brewster angle.²⁵ The impact of the Brewster angle to optical imaging is not as important as the loss of transmitted light in the TE mode. According to Sec. 7.4.1.1, TE is the preferred polarization for reconstructing spatial frequencies. Unfortunately, this preferred component is more reduced than the TM component at high NA.

This problem can be alleviated with a top antireflection coating (TARC), which is applied to the resist surface. TARC couples more light into the resist, which reduces the reflected light that can contribute to stray light. In the case of immersion lithography, the difference between the refractive indices of the immersion fluid and the resist is lower than that between the air and the resist; the immersion fluid serves as an inherent TARC. With a fluid index of 1.46, the unpolarized transmitted light is more than 98%, leaving less than 2% to be reflected at $\sin\theta = 0.8$. Even with the unfavorable TE polarization, the

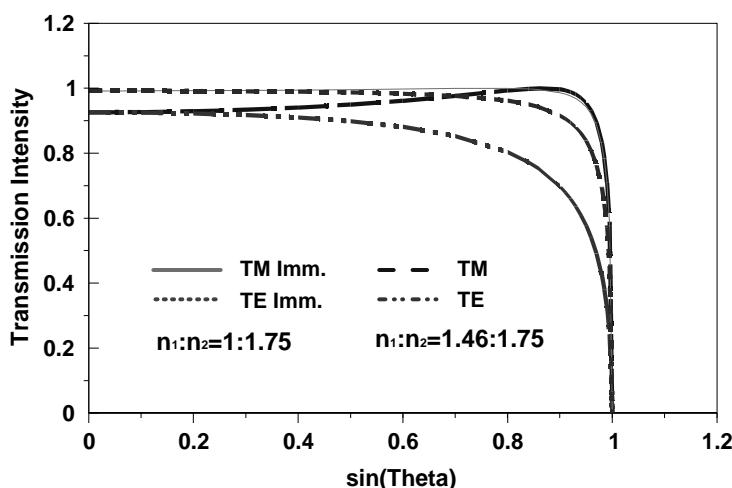


Figure 7.18 Polarization-dependent transmission at an interface. $n_1:n_2 = 1:1.75$ is used for the dry case, and $1.46:1.75$ is for the immersion case.

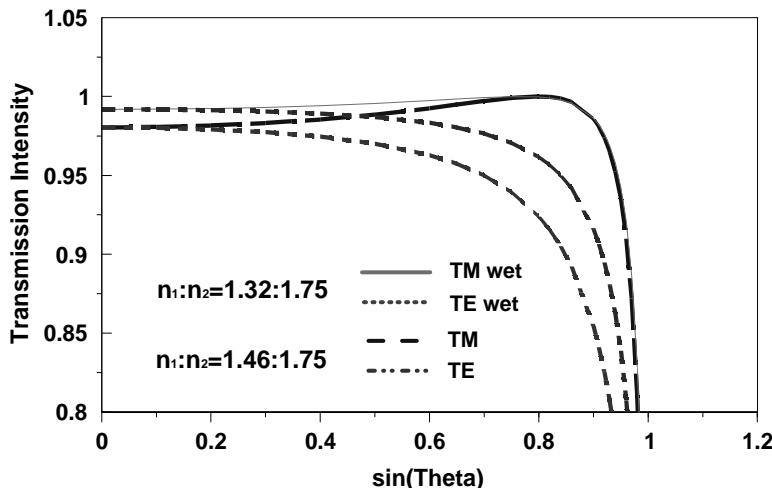


Figure 7.19 Polarization-dependent transmission at an interface. $n_1:n_2 = 1.46:1.75$ and $1.32:1.75$. The former is the situation with immersion lithography and the latter is the best index matching to air with TARC.

transmission and reflection are 96.2% and 3.8%, respectively. The situation is plotted in Fig. 7.18 to compare with the normal practice of using a wafer in air. The immersion case and the situation when using TARC are compared in Fig. 7.19. Here, the refractive index is taken to be the optimum for matching the resist to air, i.e., $\sqrt{1.75} = 1.32$. Immersion shows better reflection/transmission characteristics than TARCs.

7.4.1.3 Different effects of polarized illumination

In addition to the polarization dependence of spatial frequency recombination in the resist, pattern orientation in the mask also plays an important role in the final image on the wafer. Figure 7.20 shows the diffracted waves from a 3λ slit opening in the mask.²⁶ Both TE and TM illuminations are used. Note that TE here is defined as the electric-field vector parallel to the edge of the slit; ditto for TM. These types of illumination are illustrated in Fig. 7.21(c). In the previous sections, as long as the electric-field vector is parallel to the resist surface, it is TE, similar to Fig. 7.21(b). Unpolarized light regularly found in existing projection-imaging tools is depicted in Fig. 7.21(a). The electric-field vectors are randomly oriented. Back to Fig. 7.20, the diffracted light from the two illuminations is quite different. The TE diffracted wave satisfies the boundary condition that the electric field in a conducting screen is 0. That is, the electric field in the chrome part of the mask is 0. But, it has a structure in the transparent part of the mask plane. There are three peaks in the structure, separated by approximately one wavelength. The peak intensity in the mask plane is higher. There is a pronounced focus at about three wavelengths from the mask. The TM wave also satisfies the boundary condition that the magnetic field is unity in the

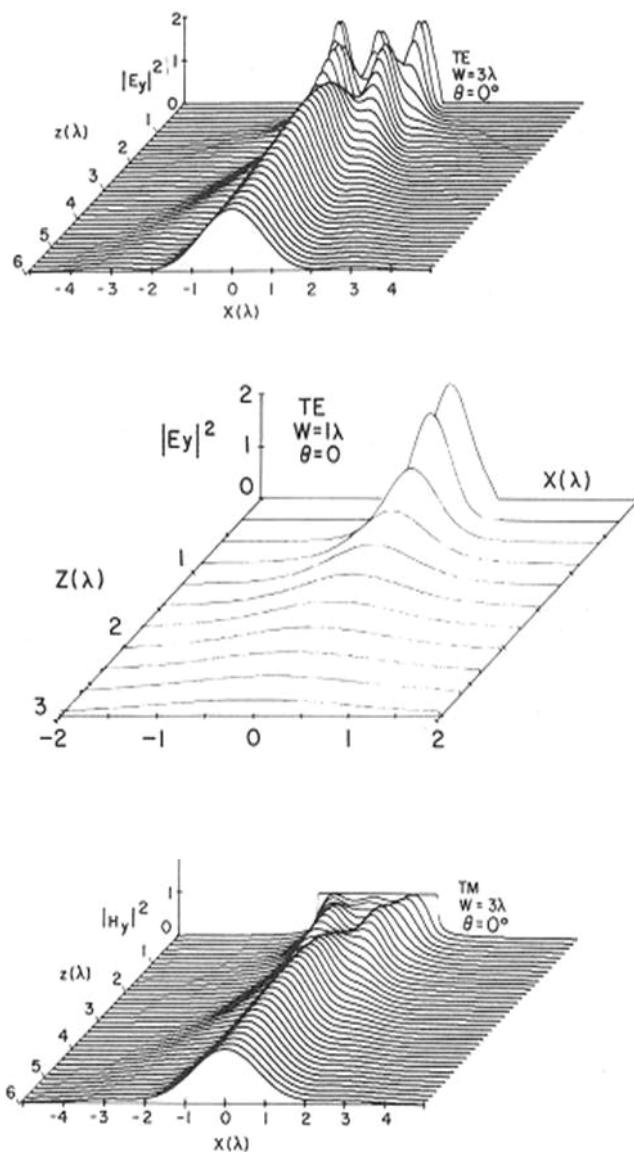


Figure 7.20 Polarization-dependent diffraction.

slit opening. It exhibits a weak three-peaked structure slightly away from the mask plane. The three peaks become two and finally one, just as in the case of TE illumination. When the slit width becomes 1λ , the peak in the TE slit is even higher. The peak is most pronounced at a slit width of about 0.7λ . When the slit is smaller, less light is transmitted through the TE slit. This is also predictable from the boundary condition. In fact, subwavelength transmission gratings can be used as a TM filter to reject TE light.

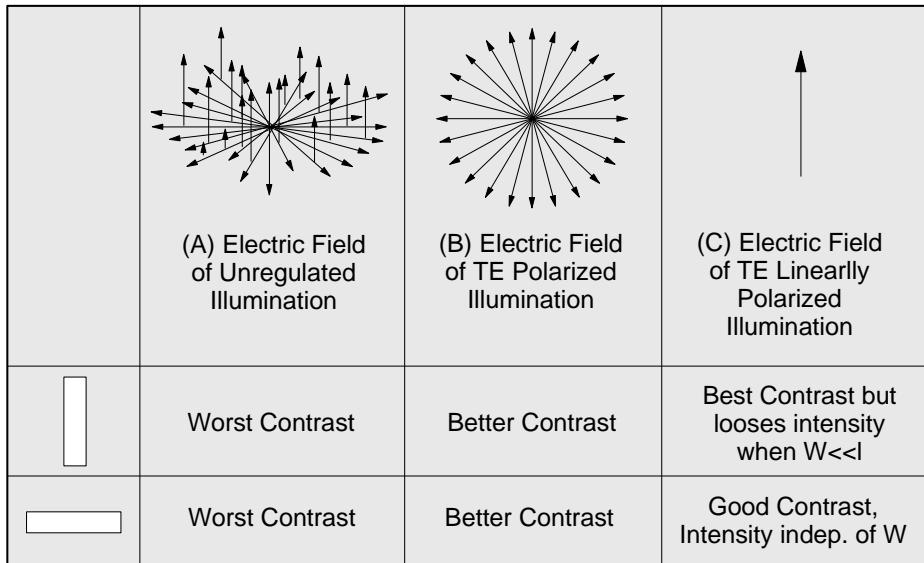


Figure 7.21 Effects of polarized illuminations to rectangular features.

With the three types of polarization conditions in Fig. 7.21, the situation in 7.21(a) is found in our daily imaging work. At a high NA, the light with an electric-field component perpendicular to the resist surface suffers from PDS; these effects are analyzed in Sec. 7.4.2 below. The imaging light combines without PDS in the case of B-type illumination. However, the electric-field vector can be randomly oriented with respect to the orientation of the line on the mask. To get the most of polarized illumination, the C-type illumination with vertically oriented patterns is the best. This puts a restriction on pattern orientation on the mask. Another resolution-enhancement technique (RET) calls for separating any arbitrary patterns into two perpendicular orientations and using dipole off-axis illumination. This complements the C-type illumination. The only tradeoff is double exposure, which reduces the throughput of an exposure tool, thus reducing productivity. It is well known that lining up the gates in one direction helps control the uniformity of gate CD. This also works in the direction of C-type illumination. Figure 7.22 illustrates a method to produce B-type illumination. Light can be manipulated to be azimuthal in the lens pupil to result in the elimination of the TM component while maintaining the symmetry in the orientation of the tangential electric-field vector. The orientation effect of B-type illumination is similar to annular illumination. Both are nondiscriminating to the orientation of the patterns on the mask, but the resolution-enhancement effect is not as strong as their discriminating counterpart. In the former case, the illumination is C-type; in the latter case, it is dipole illumination. Note that transmission and reflection at each optical surface between the light source and the resist have an effect on the polarization state of the actinic light. It is not straightforward to produce pure polarization states as B or C. However, residual unwanted polarization can be tolerated.

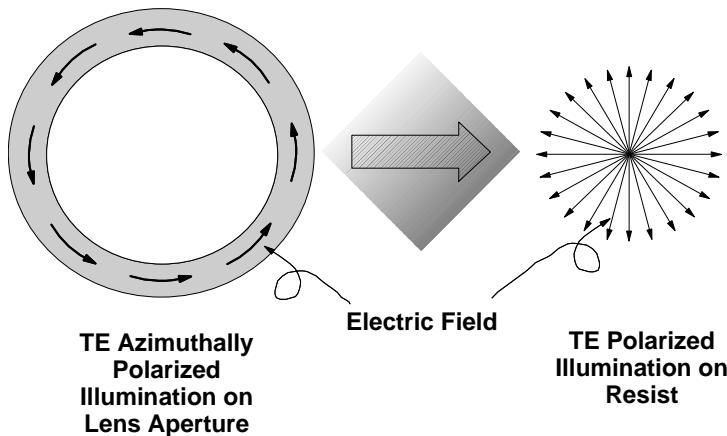


Figure 7.22 Azimuthally polarized illumination.

7.4.2 Stray light

After the major equations were introduced in Sec. 7.4.1 and the physics of stray light was explained, we now start to consider three types of stray light: SSL, PDS_{recom} (i.e., PDS from spatial frequency recombination in the resist), and $PDS_{reflect}$ (i.e., PDS from reflection off the resist and other surfaces on the wafer-film stack).

Figure 7.23 shows an axial beam illuminating the mask pattern, creating the spatial frequencies in different diffraction orders. The 0th-, 1st-, and 2nd-order beams are illustrated. This angular spectrum is strictly a function of the mask pattern and the illumination beam. It is independent of the imaging lens so far. In this illustration, the 2nd-order beams are filtered by the lens. Only the 0th- and the ±1st-order beams are incident on the resist surface. The transmitted light undergoes refractions following Snell's law. The intensity, resulting from the recombination of the electric-field vectors, exposes the resist. Specular reflection takes place at the resist surface, as shown. Though not plotted, specular reflection also takes place at all of the reflection surfaces on the wafer-film stack. Most likely, there is a bottom antireflection coating (BARC) underneath the resist. Hence, the two most significant reflections are those from the resist surface and from the BARC.

7.4.2.1 System stray light

Transmitted light that does not recombine becomes stray light. The lens itself induces SSL regardless of whether the imaging light is polarized. Basically, when a light beam strikes the lens surfaces, each lens surface contributes to stray light from residual reflections from the lens ARC, surface roughness, surface blemishes, and particulates on the lens surfaces, etc. This light beam can be the regular imaging light moving forward as well as the reflected light from the wafer surfaces, moving in the opposite direction. Strictly speaking, the SSL is a complicated function in the position affected by the lens configuration, including

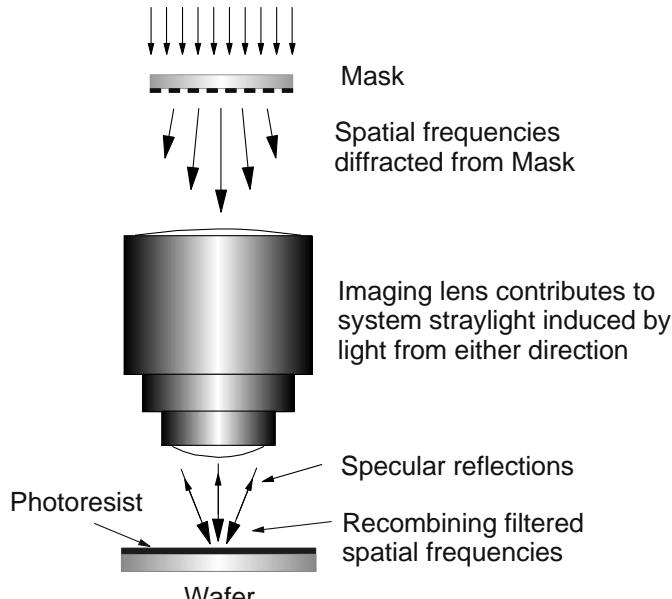


Figure 7.23 Imaging, reflecting, and stray light.

surface curvature, refractive indices, ARC characteristics, surface roughness, position of the lens elements, position of the apertures, reflectivity of walls and slopes, distribution of surface contamination, etc. To simplify the problem, the SSL is approximated by a uniform percentage of the incident light whose level is determined by the ratio of transparent and opaque areas in the mask.

The reflected light from the wafer surfaces also produces SSL when it strikes the lens surfaces. After leaving the lens, it is reflected by the mask absorber. More will be said in Sec. 7.4.2.3 on this type of reflection.

7.4.2.2 Stray light from recombination of spatial frequency vectors inside the resist

The normal lens imaging process reduces contrast for images containing higher spatial frequencies. This takes place when the mask-diffracted light passes through the lens. This diffraction contrast loss is taken care of with existing projection imaging simulations. After the lens, when spatial frequencies are recombining in the resist, the additional reduction of contrast in the recombination process shows up as stray light in its entirety; this is PDS_{recom} . As shown in Fig. 7.24, I_{min} is the background light in the resist, making up the stray light. Its portion with respect to the total light is normalized with $0.5(I_{max}+I_{min})$. So,

$$PDS_{recom} = \frac{2I_{min}}{I_{max} + I_{min}} = 1 - Contrast_{recom}. \quad (7.45)$$

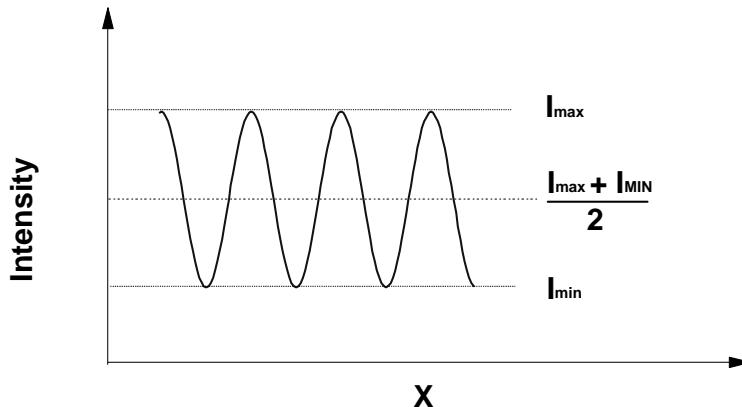


Figure 7.24 Contrast and stray-light-related parameters.

The contrast of linearly polarized light making an arbitrary angle ϕ with respect to the electric vector of the TE mode is a function of ϕ and the additional recombination contrast of the TE and TM modes:

$$\text{Contrast}_{\text{recom}}(\phi) = \sin^2 \phi \cdot \text{Contrast}_{\text{TM recom}} + \cos^2 \phi \cdot \text{Contrast}_{\text{TE recom}}. \quad (7.46)$$

$\text{Contrast}_{\text{TE recom}} = 1$ and $\text{Contrast}_{\text{TM recom}} = \cos \alpha_u$, where $\alpha_u = 2\alpha$ is for two-beam interference and $\alpha_u = \alpha$ is for three-beam interference. The angle ϕ is depicted in Fig. 7.25. Equation (7.45) becomes

$$PDS_{\text{recom}} = (1 - \cos \alpha_u) \sin^2 \phi. \quad (7.47)$$

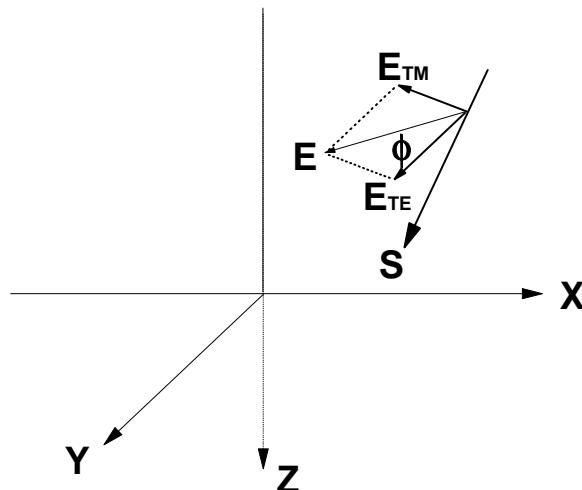


Figure 7.25 The electric field of an arbitrary linearly polarized beam, making an angle ϕ with respect to the direction of the electric field in the TE mode.

The most common illumination to date can be treated as having equal contributions from the TE and TM modes. That is,

$$\text{Contrast}_{CP \text{ recom}} = 0.5(\cos \alpha_u + 1), \quad (7.48)$$

where *CP* stands for circular polarization. The equation can be used for circular polarization, linear polarization with $\phi = \pi/4$, or completely random polarization.

Note that the contrast used to calculate stray light is strictly the contrast produced by the interference of two coherent imaging beams whose incident angles are dictated by the spatial frequency of the patterns in the mask, i.e., $\sin\alpha = \lambda/p$, where *p* is the periodicity of line-space pairs. Equations (7.38) and (7.39) are applied, using this angle. The aperture angle θ of the imaging lens should not be confused with α , where $\sin\theta$ sets the upper limit of α that is not blocked by the lens aperture. When $\sin\theta < \sin\alpha$, only the 0th-order beam passes through the lens; no image is formed. When $\sin\theta \geq \sin(n\alpha)$, the spatial frequencies $\lambda/p, 2\lambda/p, \dots, n\lambda/p$ recombine at the imaging plane to form the lens image. In lithography, because of economy, the imaging lens is extended to its full resolution potential. As a result, most imaging involves only the 1st-order spatial frequency. Occasionally, some larger pitches are blended into the circuit design, leading to 2nd-order spatial frequencies. In extreme cases pushing for superresolution, when $n\lambda/p > 1$, these frequencies turn into evanescent waves whose treatment is beyond the scope of this work.

When a two-beam resolution-enhancement technique (RET) is used, $\alpha = \sin^{-1}(\lambda/2p)$, instead of $\sin^{-1}(\lambda/p)$, as shown in Fig. 7.26. In the case of off-axis illumination, using one of the illumination beams as an example, it is tilted by an

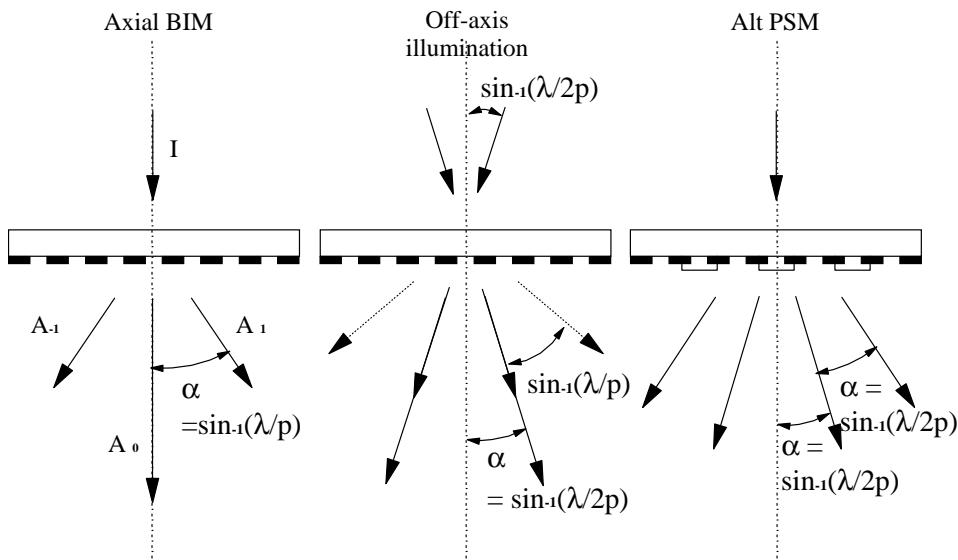


Figure 7.26 α in axially illuminated BIM, off-axis illumination, and AltPSM.

angle $\sin^{-1}(\lambda/2p)$; as a result, the 1st-, 0th-, and -1st-order beams now make angles $-\sin^{-1}(\lambda/2p)$, $\sin^{-1}(\lambda/2p)$, and $\sin^{-1}(3\lambda/2p)$ to the optical axis of the lens, respectively. The angle α for calculating the contrast is now $\sin^{-1}(\lambda/2p)$ instead of $\sin^{-1}(\lambda/p)$ in an axially illuminated binary intensity mask (BIM), equivalent to reducing the spatial frequency by half. In the case of AltPSM, because of the phase shift in every other object, the 0th-order beam becomes zero, and the two diffracted beams make angles $\pm\sin^{-1}(\lambda/2p)$ to the optical axis, also reducing the spatial frequency by half.

The higher harmonics are separated by $\sin^{-1}(\lambda/2p)$. Naturally, the contrast in the two- and three-beam situations is calculated according to Eqs. (7.31) and (7.36), respectively.

With the same object, the PDS_{recom} is similar in magnitude independent of the RET above because in the case of axially illuminated BIM, $\cos\alpha = \cos[\sin^{-1}(\lambda/p)]$, whereas in the other two cases, $\cos 2\alpha = \cos[2 \sin^{-1}(\lambda/2p)]$. Their magnitudes are almost identical in small α and become significantly different only when $\sin^{-1}(\alpha/2p) \geq \pi/8$, as shown in Fig. 7.27. Since the range of α in both cases is between 0 and $\pi/2$, the main difference between them is that the contrast in the three-beam case cannot become negative. The tradeoff of reducing the spatial frequency to resolve an unresolvable object is negative contrast in the TM mode. Negative contrast means the image is reversed. It does not start at $\alpha/p = 1$, as usually speculated, but at $\sin^{-1}(\alpha/2p) = \pi/4$. Therefore, two-beam interference has better contrast than three-beam interference, even beyond the resolution limit of the latter.

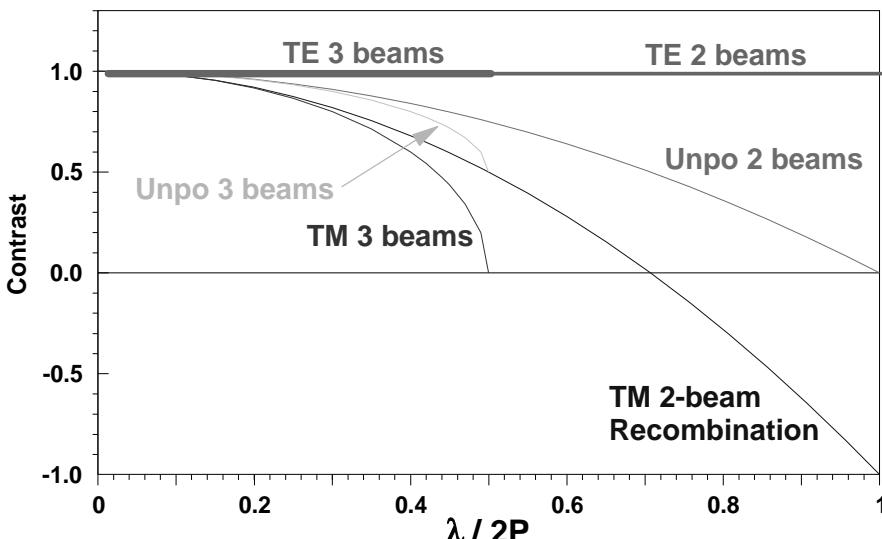


Figure 7.27 Contrast from a single object but with two- and three-beam illuminations.

7.4.2.3 Stray light from reflection off the resist surface

Only part of the specularly reflected light from the resist surface, as described in Sec. 7.4.1.2, contributes to $PDS_{reflect}$. After reflecting from the resist surface, light traces back to the mask as a conjugate of the image formed on the resist surface. Referring back to Sec. 5.3.3.4 and Fig. 5.15, any reflected light off the conjugate of the mask plane contributes to stray light to reduce the contrast of the resist image. The reflectivity of the absorber ranges between 10% and 20%. In view of the above argument, it is preferable to have a low-reflectivity mask absorber unless all reflections from the wafer are suppressed.

With an actual imaging lens consisting of more than 30 elements, thus about 70 glass-to-air interfaces, even the best ARC for each lens surface can add up to a few percent of total effective reflection. The reflected light, even under a perfect conjugated reflection relationship, still produces stray light because of an imperfect ARC at each lens surface and residual surface roughness.

7.4.2.4 Incorporating polarization effects to E-D windows

The E-D window methodology in Chapter 4 and Ref. 13 is used to determine the DOF and EL of a given set of design patterns. One takes all of the representative feature types and sizes under the predetermined imaging condition, such as λ , NA, illumination setting, and lens aberrations, to evaluate the CD at a given range of exposure dosage. The constant-CD contours in the E-D space, of the lower as well as the upper bounds of CD, are constructed. The DOF is the length in the defocus direction of the largest possible rectangle or ellipse within the contours, whereas the length in the exposure direction determines the EL. There is one set of contours for each critical dimension taken from each feature type and size. The common intersection of the upper contours and the lower contours forms the common E-D tree, and the corresponding largest rectangle or ellipse¹⁵ is the common E-D window. Rectangular E-D windows are the absolute bounds for EL and DOF, whereas elliptical windows are statistical in nature. They represent the DOF and EL corresponding at a CD distribution of a yet-to-be-determined number of standard deviation, $x\sigma$. There is no physically rigorous way to determine x . It is best determined statistically from manufacturing data.

Rectangular E-D windows are used consistently throughout this section. The rectangular shape intersects the E-D trees more precisely, especially in situations of a small aperture angle θ where the tree branches have a small curvature and are extended in the defocus direction. The working DOF is expected to be larger than those shown here by 50 to 150 nm.

The Signamizer software²⁷ is modified with the equations in this section to generate the results. To incorporate the SSL in the evaluation of E-D windows, a given level of constant intensity is superimposed on the aerial image based on the optical parameters, such as λ , NA, illumination setting, and lens aberrations. For example, a 5 to 10% constant intensity is added to the normalized intensity distribution of the aerial image of concern. This SSL level is typically found in a modern diffraction-limited imaging lens in a scanner. Though it is a complicated

spatial function of lens ARC, surface roughness, optical configurations, baffles, etc., SSL is usually treated as constant for a 1st-order estimate of stray light effects.

PDS is clearly dependent on the incident angle, even though we still distribute it evenly in the resist. The level of PDS is calculated as a function of the incident angle according to the formulae given in this section. For PDS from recombination of spatial frequency vectors, its dependence to α can be determined from Eqs. (7.17), (7.23), (7.30), (7.34), or (7.45).

The PDS from reflection off the resist surface is treated by multiplying the reflectivity given by Eqs. (7.42) and (7.43) with the reflectivity of the absorber and with the lens stray light, which is induced by imperfect lens ARC, surface roughness, optical configurations, baffles, etc., to an incident beam. Here, the incident beam is not from the illumination on the mask but from the reflection from the resist surface. Therefore, in the TM case,

$$PDS_{reflect} = \frac{R_{TM}}{B_{TM}}^2 (\rho R_{MA\ TM} + SSL), \quad (7.49)$$

where ρ is an adjustment factor that takes care of the partial contribution to stray light, which is discussed in Sec. 7.4.2.3. We have set it to 0.5 based on our experience. A rigorous setting awaits simulation with reflections from the multilayer stack on the wafer and the subsequent reflection from the mask absorber. For PDS due to reflection for the TE mode, simply substitute R_{TM} and B_{TM} with the corresponding B , R , and R_{MA} in Eq. (7.49). R_{MA} corresponds to reflection from the mask absorber. It is also polarization dependent. Hence, the total modification of the intensity distribution $I(x,z)$, calculated from a regular simulation program, is as follows:

$$I_{modified}(x,z) = [I(x,z) + I_{avg} (PDS_{recom} + PDS_{reflect}) + PD \cdot SSL]C, \quad (7.50)$$

where I_{avg} is $I(x,z)$ integrated over x at each z plane, PD is pattern density, and C is an adjustment constant to keep the total intensity integrated over the mask unchanged before and after inducing stray light. This way, the exposure level of different patterns is aligned to each other for evaluation of the common E-D window. The loss in the transmitted light due to reflection is also incorporated through C :

$$C = \frac{I_{avg}}{I_{avg} (1 + PDS_{in\ resist} + PDS_{reflection}) + PD \cdot SSL}. \quad (7.51)$$

Both PDSs are multiplied with I_{avg} because the light incident on the resist and transmitted into the resist has already passed the imaging lens and is diffracting

from there. The SSL is multiplied by PD because it is induced by light diffracted by the mask but not yet through the lens. The very action of the mask-diffracted light going through the lens induces the SSL.

7.4.2.5 Simulation results with PDS

Line-space pairs of line-to-space ratios 1:1, 1:2, and 1:3 on a BIM are used to demonstrate simulation with PDS. Figure 7.28 shows the DOF of these features at an 8% EL under 193-nm axial disk illumination at $\sigma = 0.82$. A 10% stray light from the optical system is included. The refractive index of the resist layer is assumed to be 1.75 and that of the immersion fluid is 1.46.

Let's first compare PDS and no-PDS (NPDS) effects within one feature type. Take the 1:1 pairs. At a very small aperture angle θ , there is no DOF because of a lack of resolution. As the resolution builds up at larger aperture angles, the NPDS curve shows an increase in DOF before it peaks out. With PDS, DOF starts at a larger θ because the DOF becomes available only after the contrast in the optical image overcomes the stray light. Note that the PDS stays constant as θ increases because it is a function of α determined by the spatial frequency component of the mask pattern, as discussed in Sec. 7.4.2.4. At a larger θ , the contrast increases while the PDS stays constant, thus accounting for the smaller difference between PDS and NPDS at larger aperture angles. Among the three features, the 1:3 pairs show the

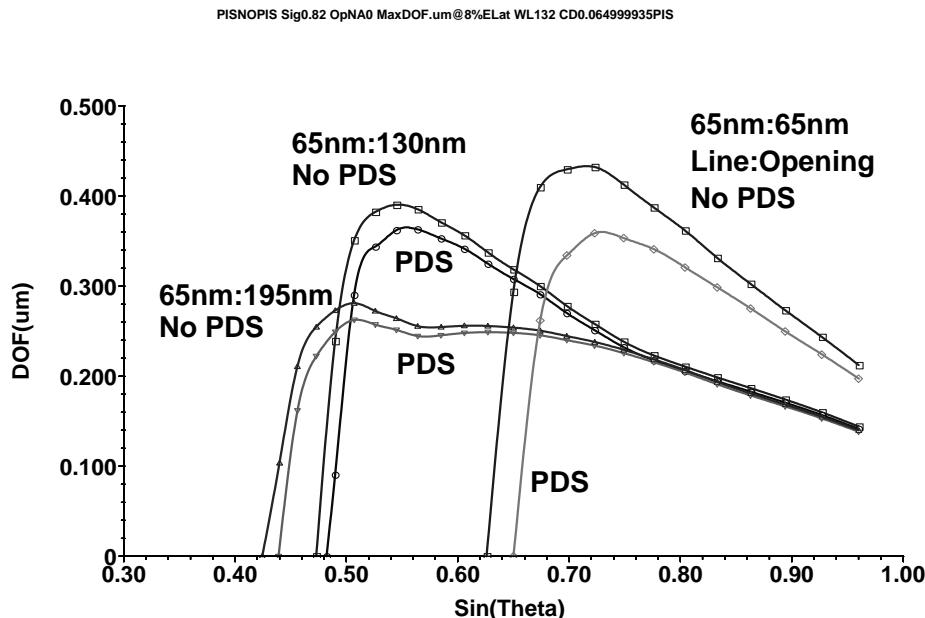


Figure 7.28 DOF as a function of $\sin\theta$, comparing PDS and NPDS with a 193-nm immersion in water using BIM, 65-nm lines, 1:1, 1:2, 1:3, $\sigma = 0.82$, 8% EL, $n_{\text{water}} = 1.46$, $n_{\text{resist}} = 1.75$, CD tolerance = $\pm 10\%$, and SSL = 10%.

Table 7.2 SSL, recombination PDS, reflection PDS, and reflectivity of the coupling-medium/resist interface according to the situation in Figs. 7.28 and 7.29. The second number in the last three columns includes both the first and the second harmonic.

System	L:O	PDS	PD _{SSL}	PDS _{recom}	PDS _{reflec}	Reflectivity
193-nm immersion	1:3	No	7.50%	—	—	—
		Yes	7.50%	4.70%	0.132%	0.88%
	1:2	No	6.67%	—	—	—
		Yes	6.60%	8.74%	0.170%	1.13%
	1:1	No	5.00%	—	—	—
		Yes	5.00%	23.43%	14.51%	100%
132-nm dry or 193-nm immersion	1:5	No	8.30%	—	—	—
132-nm dry	1:5	Yes	8.30%	0.94%, 3.86%	1.12%, 2.06%	7.46%, 15.8%
193-nm immersion	1:5	Yes	8.30%	2.03%	0.12%	0.825%

least difference between PDS and NPDS. The reason is that this group has the largest pitch. The contrast with a larger pitch is higher. In this particular situation, the recombination PDS for the 1:3, 1:2, and 1:1 features are 4.71%, 8.74%, and 23.43%, respectively; the reflection PDS is 0.13%, 0.17%, and 14.5%; and the reflectivity on the resist surface is 0.88%, 1.13%, and 100%. In the last case, the extremely high reflectivity is caused by the high incident angle in the coupling medium. The 130-nm pitch produces $\sin\alpha_u = 0.85$ in the resist at a refractive index of 1.75, but in the reflected beam it produces $\lambda_{CM}/p > 1$. Light is completely reflected. The PDS and reflectivity are summarized in Table 7.2.

The PDS stays constant with respect to θ only between each order of spatial frequency. That is, when $n\lambda/p \leq \sin\theta < (n + 1)\lambda/p$. Take the 1:5 pairs shown in Fig. 7.29 as an example. At a pitch of 390 nm and a fictitious dry wavelength of 132 nm, the 1st-order spatial frequency is $\sin\alpha = \lambda/p = 0.338$. It passes through a lens with $\sin\theta \geq 0.338$. As soon as $\sin\theta \geq 0.676$, the 2nd-order spatial frequency is also admitted, contributing more to stray light. Here, the DOF curve shows a drop to 246 nm from 253 nm at $\sin\theta = 0.68$, due to an increase in recombination PDS from 0.94% to 3.86%. The PDS, by reflection off the resist surface, also increases from 1.12% to 2.06%. Note that the change induced by the second harmonic is purely due to the increase of stray light from an additional order of light. The $\cos\alpha$ relationship in Eq. (7.36) is no longer exact, but we did not use a better formula in this simulation. A better formula should only slightly improve the simulation accuracy.

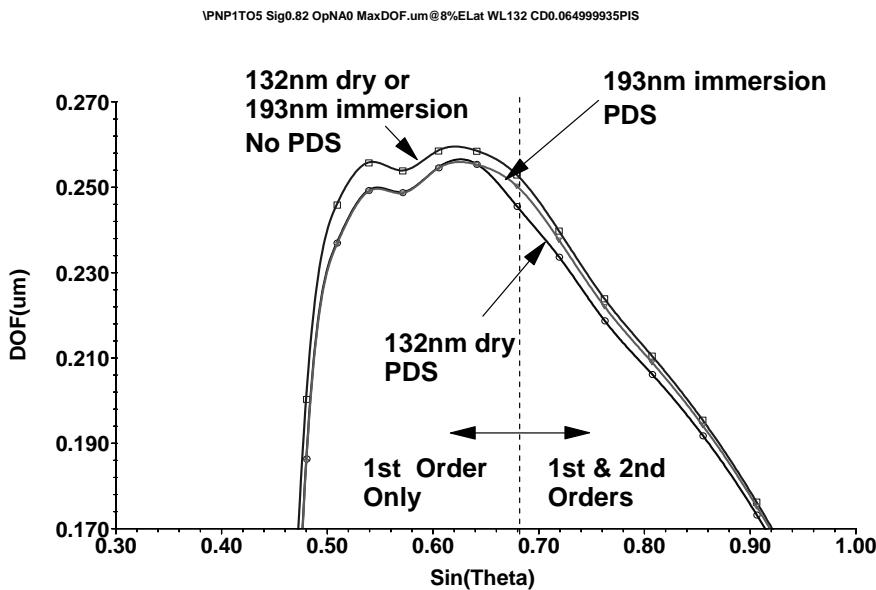


Figure 7.29 Same as Fig. 7.28, except the object consists of 65-nm lines at a 1:5 line/opening. Note the inclusion of the 2nd-order diffraction beam. The fictitious cases of a 132-nm dry system and a 193-nm immersion system, with PDS and NPDS, are compared.

Figure 7.29 also contains a curve from a 193-nm immersion system. There is no 2nd-order beam because at the mx mask side the wavelength is still 193 nm. Hence, $\sin\theta = 0.497/m$. The factor m is the reduction ratio of the imaging system. The wavelength only changes to 132 nm in the coupling medium. The second harmonic at $\sin\theta = 0.994/m$ is smaller than $\text{NA} = 0.92/m$ and is rejected by the imaging lens within the θ range of our plot. It is ironic that rejecting the higher-order harmonic produces a larger DOF; this is due to PDS.

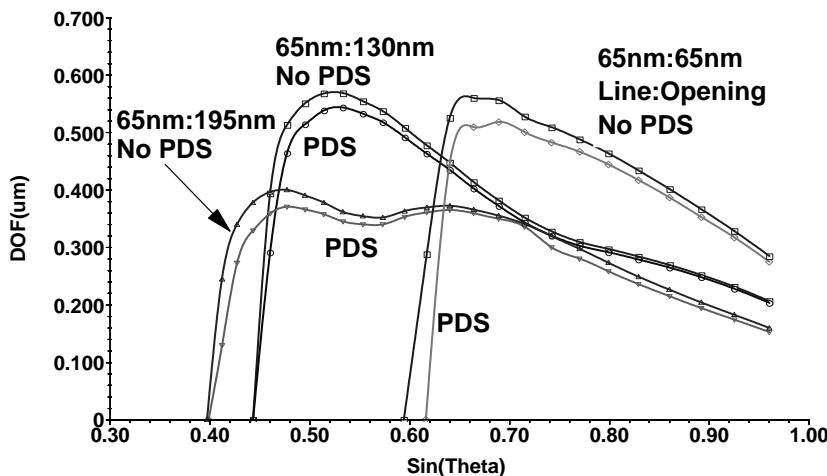
From Table 7.2, the recombination PDS is 2.03%. It is larger than that of the 132-nm dry case, which is expected. However, after the second harmonic is included in the image of the dry system at $\sin\theta \geq 0.676$, the recombination PDS of the dry system is now larger. The reflection PDS at 0.12% is expected to be much smaller than the counterparts of the dry system, even with only the first harmonic, because of better index matching, as was discussed in Sec. 7.4.1.2.

Let's turn our attention to a two-beam interference system produced by annular illumination. The features used in Fig. 7.30 are identical to those in Fig. 7.28, except that AttPSM with a 6% transmission and an annular illumination with $\sigma_{in} = 0.42$ and $\sigma_{out} = 0.84$ are used. The DOFs without PDS are much higher than the three-beam case, exemplified by BIM with disk illumination, as expected from this RET. The two-beam leverage discussed in Sec. 7.4.2.2 makes PDS_{recom} smaller than that in the three-beam case, especially for the 1:1 feature. The SSL, PDSs, and reflectivity of this case are summarized in Table 7.3.

Table 7.3 SSL, PDSs, and reflectivity used in Fig. 7.30.

System	L:O	PDS	PD·SSL	PDS _{recom}	PDS _{reflec}	Reflectivity
193-nm immersion	1:3	No	7.50%	—	—	—
		Yes	7.50%	4.49%	0.123%	0.82%
	1:2	No	6.67%	—	—	—
		Yes	6.67%	7.97%	0.124%	0.83%
	1:1	No	5.00%	—	—	—
		Yes	5.00%	17.94%	0.132%	0.88%

NOPISATT SigIn=.5Out=0.84 OpNA0 MaxDOF.um@8%ELat WL132 CD0.064999935PIS

**Figure 7.30** Same as Fig. 7.28, except for AttPSM and annular illumination $\sigma_{in} = 0.42$, $\sigma_{out} = 0.84$. The loss of DOF with PDS is less severe.

Does a dry system do better than an immersion system? Figure 7.31 shows the DOF of the 65-nm 1:1 feature versus $\sin\theta$ for three systems: 157-nm dry, 132-nm dry, and 193-nm immersion. The 193-nm dry system is not included because of a lack of resolution. In this figure, we also include the DOF simulated with neither SSL nor PDS. The systems with a complete absence of any stray light clearly outperform the others in the same wavelength group. The contributions from SSL and PDS with the 132-nm dry system are about the same, whereas despite smaller reflection PDS, the 193-nm immersion system shows less DOF than the 132-nm counterpart because PDS_{recom} is more significant here. When compared with the 157-nm dry systems, all 132-nm systems are substantially better, including the 193-nm immersion system.

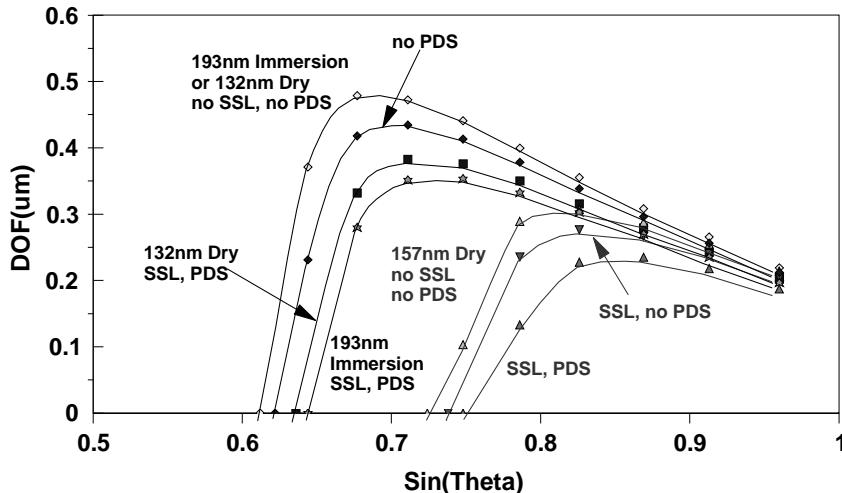


Figure 7.31 DOF of the 1:1 group in Fig. 7.28 for three systems: 157-nm dry, 132-nm dry, and 193-nm immersion. The curves with neither SSL nor PDS are included. The result from the 193-nm dry system is not included because there is no DOF.

For the 1:2 group shown in Fig. 7.32, the trend is similar but the margins are smaller. Here, the 193-nm dry system is included. In this figure, we plotted the DOF against NA instead of $\sin\theta$. The 193-nm immersion system now moves to the same NA regime as the 193-nm dry system. It means that the complexity of the 193-nm immersion lens is similar to that of the 193-nm dry system, but the DOF is between two to three times larger. The complexity of the 193-nm lenses is higher than that of the 157-nm system. However, because of reduced material quality, it is not obvious that there is any advantage with 157-nm lenses.

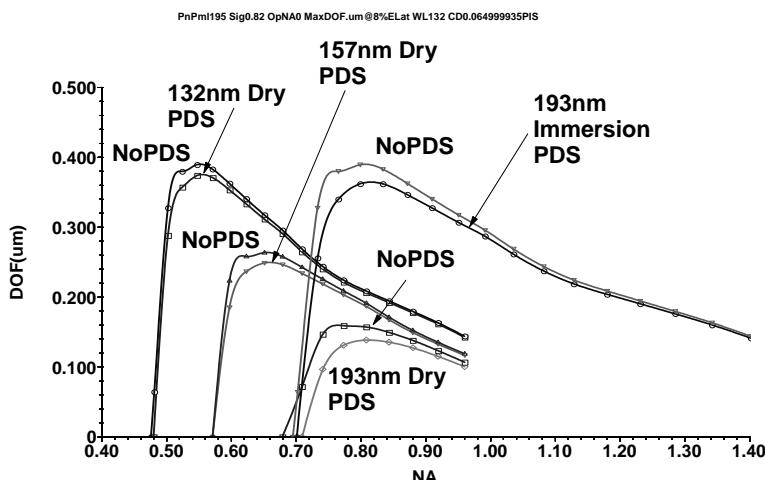


Figure 7.32 Same as Fig. 7.31, except the object is the 1:2 group. The 193-nm dry system is also in the comparison. Note that the horizontal axis is now in NA instead of $\sin\theta$.

Table 7.4 SSL, PDSs, and reflectivity used in Fig. 7.33.

System	H:S	PDS	PD·SSL	PDS_{recom}	PDS_{reflec}	Reflectivity
193-nm immersion	1:3	No	1.25%	—	—	—
		Yes	1.25%	4.71%	0.17%	0.88%
	1:2	No	1.67%	—	—	—
		Yes	1.67%	8.74%	0.13%	1.13%
	1:1	No	2.50%	—	—	—
		Yes	2.50%	23.43%	14.51%	100%

Figure 7.33 shows the DOF of 65-nm contact holes, with hole-to-space ratios of 1:1, 1:2, and 1:3, using AttPSM and disk illumination at $\sigma = 0.82$. This is now a 2D object pattern. The stray light levels are evaluated similar to the 1D case because the spatial frequencies in x and y are independent. However, the pattern density is now lower, reducing the relative impact of SSL. The SSLs in Table 7.4 are 25 to 50% lower than those in Table 7.2 and Table 7.3. The recombination PDSs are about three times smaller. Despite these smaller contributions, the DOF of contacts is smaller than that of line-space patterns due to smaller NPDS DOF to start with. The difference between PDS and NPDS is larger than that of line-space patterns because stray light is proportionally larger than image contrast.

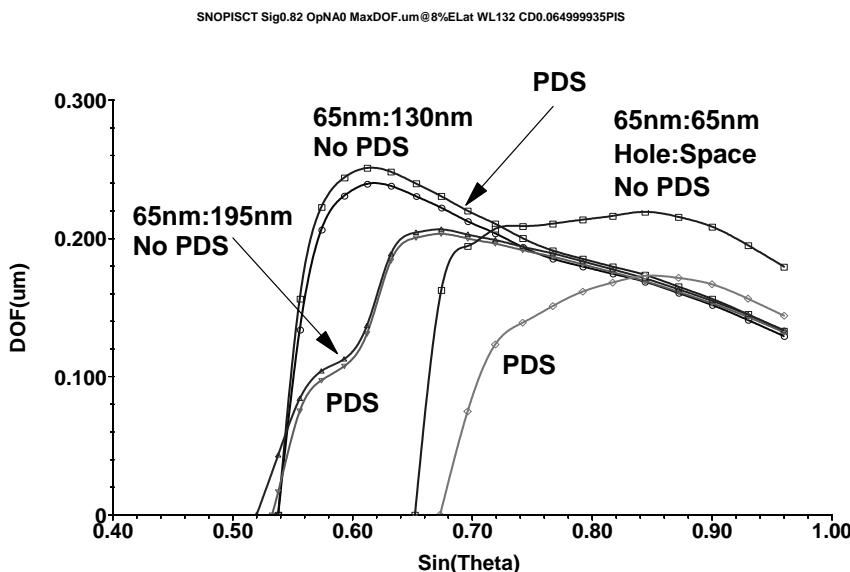
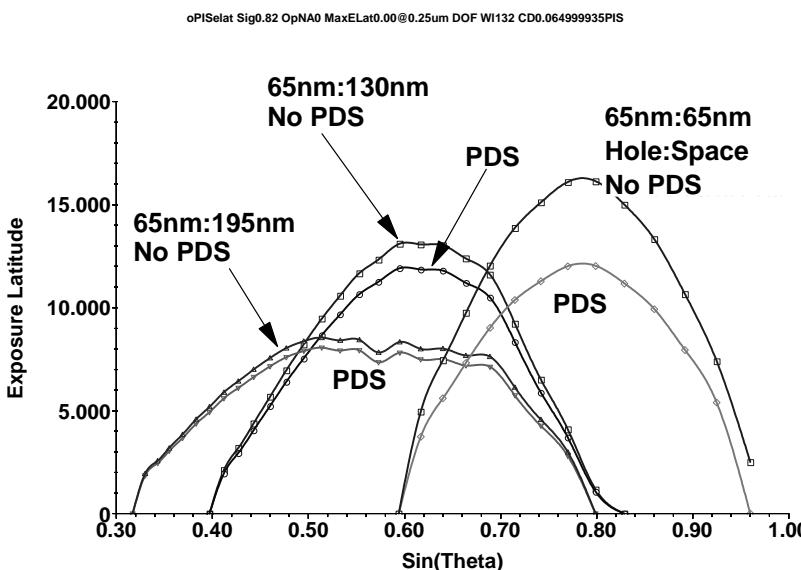


Figure 7.33 DOF of 65-nm holes 1:1, 1:2, 1:3 with PDS and NPDS, using 193-nm immersion in water and AttPSM, $\sigma = 0.84$, 8% EL, $n_{water} = 1.46$, $n_{resist} = 1.75$, CD tolerance = $\pm 10\%$, and SSL = 10%.

We have been characterizing PDS with DOF at a fixed EL. Now, the criterion is changed to EL with a fixed DOF. The situations in Figs. 7.28 and 7.33 are chosen for the comparison. The former examines 65-nm lines at 1:1, 1:2, and 1:3 line-to-space ratios. The latter covers 65-nm holes at hole-to-space ratios of 1:1, 1:2, and 1:3.

Figure 7.34 is the EL counterpart of Fig. 7.28. The trend of a smaller impact from PDS for the larger pitches is similar to that from the DOF criteria. One noted difference is that the difference between PDS and NPDS decreases monotonically with $\sin\theta$, according to the DOF criteria. In the EL case, the difference is zero at both ends of $\sin\theta$. It peaks at the optimum EL. That means the gain in EL from polarized illumination is most significant at the $\sin\theta$ for optimum EL. When one chooses to operate at the $\sin\theta$ for optimum DOF, the gain in EL is less. To explain this difference, the E-D windows used to determine the EL in Fig. 7.34 and the DOF in Fig. 7.28 are shown in Figs. 7.35 and 7.36, respectively. Actually, the E-D trees in these two figures are identical. Note that they stretch out in defocus at a low $\sin\theta$ then lose defocus, but gain EL as $\sin\theta$ becomes larger. Therefore, when E-D windows of the same EL are fitted within these trees, the NPDS tree reaches the acceptable EL first shows a large DOF, while the DOF from the PDS E-D tree is still zero. As $\sin\theta$ continue to increase, this difference becomes less significant; hence, the monotonic decrease in the difference in DOF between PDS and NPDS. On the other hand, the difference in EL in the E-D trees at low $\sin\theta$ is negligible, when the fixed DOF is set at a usable medium value. The EL peaks at intermediate $\sin\theta$ where the E-D windows are not stretching out near the extreme defocuses of the E-D trees.



Rectangular Window

Figure 7.34 Same as Fig. 7.28, except EL is plotted instead of DOF, which is set to 250 nm.

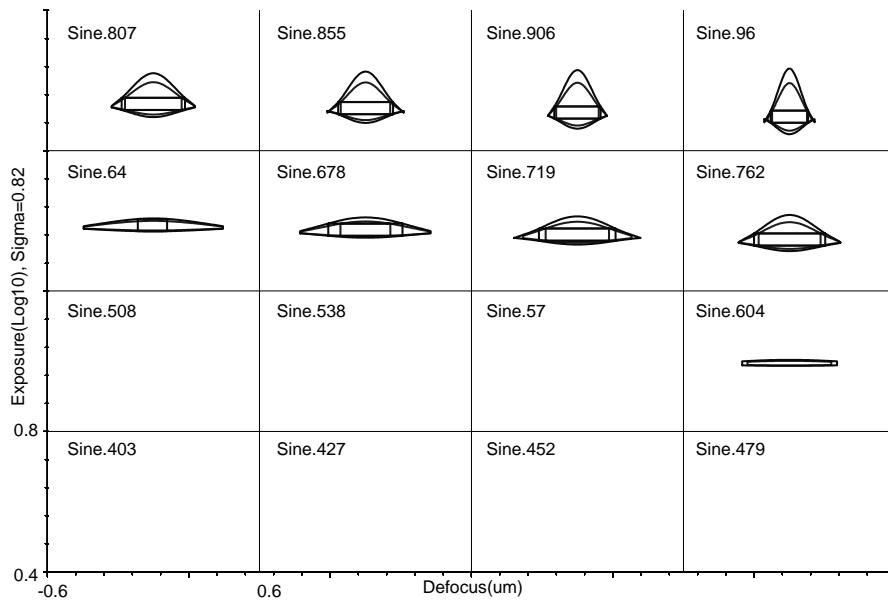


Figure 7.35 E-D windows to define the EL of the 1:1 curves in Fig. 7.34.

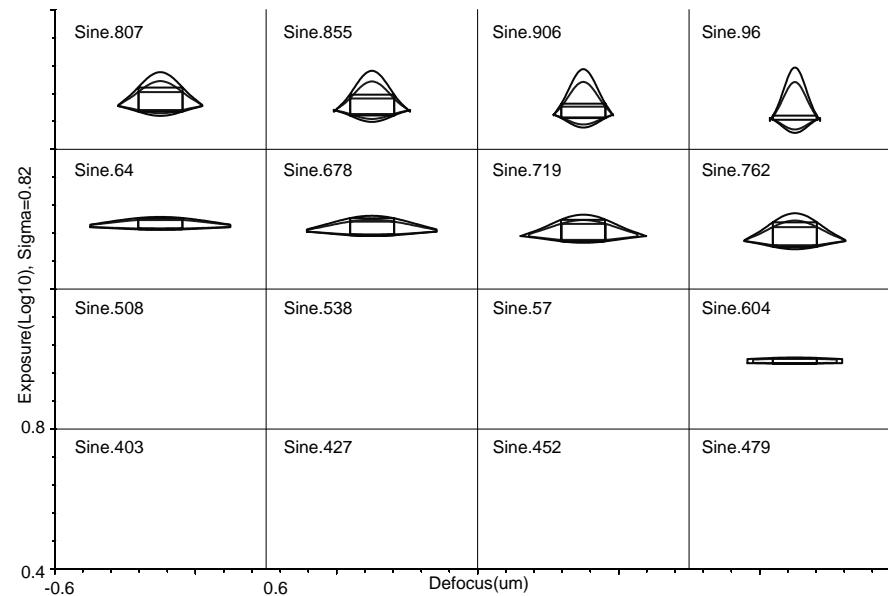


Figure 7.36 E-D windows to define the EL of the 1:1 curves in Fig. 7.28.

EL as a function of $\sin\theta$ for 65-nm holes using the same condition as Fig. 7.34 is plotted in Fig. 7.37. Here, the difference between PDS and NPDS for the tight pitch is more significant than in the case of DOF. For example, at $\sin\theta = 0.732$, the EL drops from 5.7% to 2.88%, which is almost half. The EL does not close out at a high $\sin\theta$ because of the smiling shape of the E-D trees, as shown in Fig. 7.38.

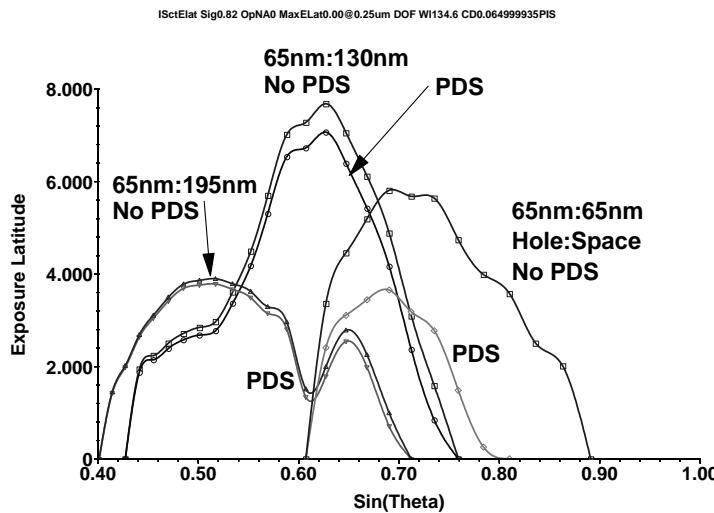


Figure 7.37 Same as Fig. 7.33, except EL is plotted instead of DOF, which is set to 250 nm.

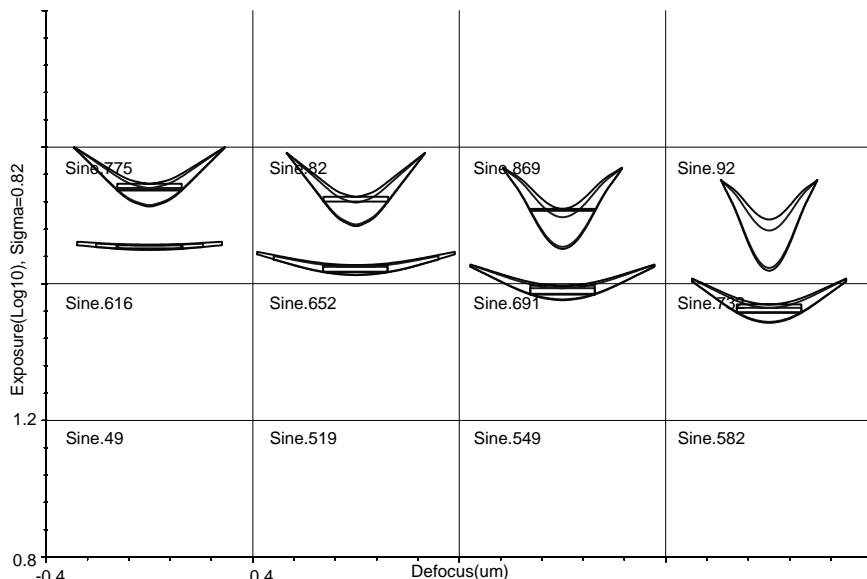


Figure 7.38 E-D windows to define the EL of the 1:1 curves in Fig. 7.34.

Polarized illumination helps DOF and EL. Its leverage is not uniform and is a strong function of imaging conditions. As shown in the previous figures, its leverage is in tight pitches. Intuitively, one tends to think that polarized illumination helps EL more than DOF. However, the actual situation is quite complicated. In evaluating these gains, one should bear in mind that DOF_{avail} is the relevant criteria instead of $DOF_{diffrac}$. Table 7.5 shows a comparison of these gains for the common window of 32-nm lines at 90- and 113-nm pitches using AttPSM and annular illumination.

Table 7.5 Gain in EL and DOF with polarized illumination. The simulation was based on the common window of 90- and 113-nm pitches for the 32-nm polyline.

AttPSM OAI	NA	Sigma	EL	DOF _{diffrac}	DOF _{avail}	Improv.
Unpolarized	1.35	0.96:0.48	5%	137	55	
	TE	1.3	0.96:0.48	5%	169	87
Unpolarized	1.25	0.96:0.48	2.43%	200	118	23/58%
	TE	1.25	0.96:0.48	3.57%	200	118
						47%

7.5 Immersion Systems and Components

This section will cover the hardware side of immersion systems. Starting from the configuration of an immersion system, the coverage includes the immersion fluid, immersion lens, bubbles, particulates, mask, and photoresist.

7.5.1 Configuration of an immersion system

The lens-based configuration (LBC), also known as the shower configuration^{28,29} and shown in Fig. 7.39, involves immersion fluid applied to and extracted from a small region covering the lens, and it is stationary with respect to the lens as the wafer is stepped or scanned. There are many motivations in adopting LBC:

- (1) The wafer stage is essentially identical to that of a dry system. It saves development cost and time.
- (2) It is possible to keep the alignment, focusing, and leveling setup unchanged.
- (3) The volume of fluid is small. Filling up the cavity can be very fast, maintaining wafer throughput.

Foreseeable problems include:

- (1) Near the wafer edge, the immersion region includes the wafer and areas outside the wafer. There is discontinuity between the wafer and the chuck as well as between the wafer chuck and the area outside the chuck. Maintaining the hydrodynamics and water cavity, and managing water extraction can be more difficult. Particles at the backside of the wafer tend to be washed to the surface.
- (2) The LBC head tends to leave trace amount of water or moisture behind, which is the root cause of water stains.
- (3) The photoresist will have inconsistent water-contacting history at different locations. As the wafer is stepped from field to field, the neighboring fields (or parts of them) are covered by water. This can happen to a field several times and not necessarily in sequence or the same number of times for each field. It depends on the stepping pattern and the location of the field.

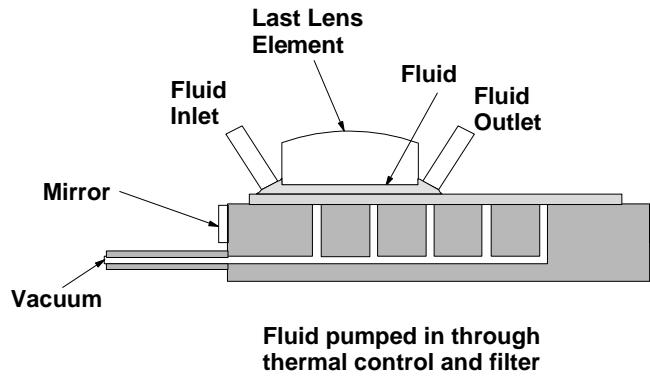


Figure 7.39 Lens-based configuration for immersion.

The wafer-based configuration (WBC), also known as the bath configuration,³⁰ involves a wafer completely immersed in a water-filled circulating tank during the wafer stage, as shown in Fig. 7.40. Water circulates under and above the stage continuously. It is filtered and temperature regulated as it moves in and out of the wafer surface area. The water can be completely drained for wafer loading and unloading. The cover prevents water from spilling over and foreign particles from falling in. The advantages of WBC are:

- (1) Exposure at the edge of the wafer is no different from that at the center.
- (2) Each field contacts the wafer for the same amount of time.
- (3) There is no concern about water stains left outside the LBC head.
- (4) Bubble generation from poor hydrodynamics near the edge of the wafer is not a problem.

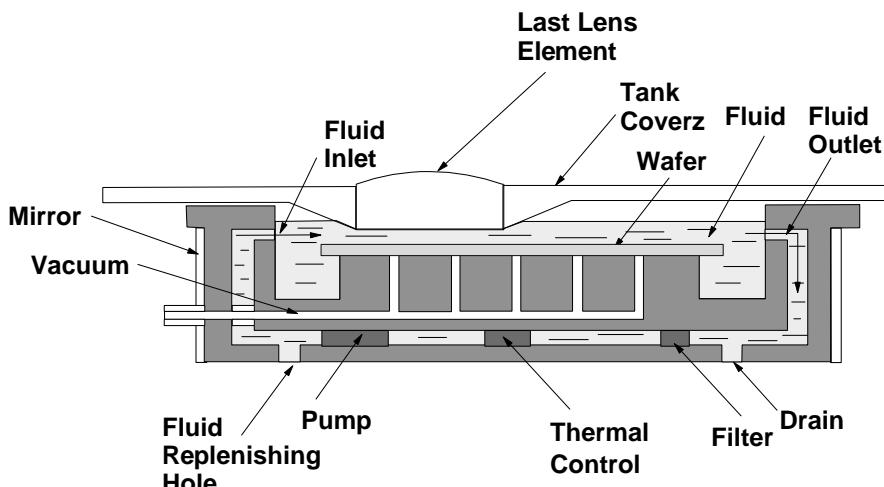


Figure 7.40 Wafer-based configuration for immersion.

Disadvantages include:

- (1) The pre-exposure and postexposure soaking times of each exposure field are different.
- (2) It takes more effort or more time to fill and drain the water.
- (3) Focusing, tilting, and alignment must be performed in the immersion mode if a twin stage is not used. However, from Eq. (7.10), $\text{DOF}_{\text{required}}$ is reduced if focusing and tilting are performed in the fluid, and if practical difficulties can be overcome.
- (4) Substantial redesign of the wafer stage is necessary.

7.5.2 The immersion medium

The optical requirements of the immersion fluid are a high refractive index and a high transmission. Its refractive index should be consistent and uniform, unaffected by thermal variations during exposure. Table 7.6 lists various immersion fluids⁵ for 193-nm and 157-nm light. The absorption coefficient of the perfluoropolyether material is too high. In order to limit light loss to 10% through the fluid, its thickness must be kept below 75 μm . The absorption coefficient is still too high at 193 nm. On the other hand, water appears to be the perfect medium for immersion at 193 nm. Not only is its absorption coefficient low, the refractive index is surprisingly high at 193 nm. In most other wavelengths, the refractive index of water is in the vicinity of 1.33.

The refractive index of water, as well as its thermal and spectral sensitivity at 193 nm, have been rigorously measured by Burnett and Kaplan.^{31,32} They used two independent methods to calibrate against each other. The first method fills a prism fluid cell with water. By carefully controlling the angle of the prism, measuring the deviation angle of the incident and refracted light, and using exactly parallel, very flat, and smooth fused-silica etalon plates as the windows for the fluid cell, the three optical constants can be determined to better than five digits. The second method measures the interference fringe positions in the transmittance spectrum from an etalon cell filled with water. It is also accurate to better than five digits. At 21.5° C and 193.39 nm, the refractive index and thermal and spectral sensitivity of water are 1.43662, $-1.00 \times 10^{-4}/^\circ\text{C}$, and $-0.002109 \text{ nm}^{-1}$, respectively. The immersed wavelength is 193.39 nm. It was used for all critical calculations. Some less critical calculations use the 1.46 value given in an earlier publication.⁵

Table 7.6 193-nm and 157-nm immersion fluids.

Fluid		λ_0 (nm)	<i>n</i>	a, cm^{-1} base 10	λ_1 (nm)
Perfluoro-polyether ^[5]	PFPE	157	1.37	6–19	115
		193	—	0.1–0.6	—
Water ^[32]	H ₂ O	193	1.43664	0.036	134

A fluid with higher index can support a higher NA. However, this is not without limit. Table 7.7 lists $\sin\theta$ in the lens, coupling, and resist media, as well as the attainable NA. Assuming that the last element in the lens is a flat piece of fused silica whose refractive index is 1.56, and 1.75 is to be the refractive index of the resist layer, $n_{fluid} = 1.44$ and $\sin\theta_{fluid} = 0.95$, the attainable NA is 1.37. Raising the index of the fluid to 1.56 leads to NA = 1.48. Further increasing the fluid index to 1.66 runs into the problem of a physically impossible $\sin\theta_{lens}$ of 1.01. The angle in the fluid must be cut back to $\sin\theta_{fluid} = 0.93$ to keep $\sin\theta_{lens} < 1$. The NA is now limited to 1.54. Any further increase in the fluid index cannot support a larger NA unless the last lens element is allowed to have a concave surface, as shown in Fig. 7.41. The curvature makes the thickness of the fluid uneven, resulting in a requirement of fluid transmission higher than that of water at 193 nm. This makes development of fluid with an index higher than 1.66 even more difficult.

In addition to the optical properties of the immersion fluid, many material characteristics of the fluid must be considered. First, the fluid cannot cause any physical or chemical change to the surface it contacts, namely, the resist surface, the lens, and its antireflection coating. Also, it cannot deposit contaminants to these surfaces. It is required to have a large dissolution capacity for gas, so that any bubble formed is readily dissolved. The viscosity should be low so that it can flow freely and will not slow down wafer scanning. The contact angle of the fluid

Table 7.7 Maximum attainable NA as a function of n_{fluid} .

	$n_{fused\ silica} = 1.56$		$n_{resist} = 1.75$	
n_{fluid}	$\sin\theta_{lens}$	$\sin\theta_{fluid}$	$\sin\theta_{resist}$	NA
1.44	0.88	0.95	0.78	1.37
1.56	0.95	0.95	0.85	1.48
1.66	0.99	0.93	0.88	1.54
1.66	1.01	0.95	0.90	n/a

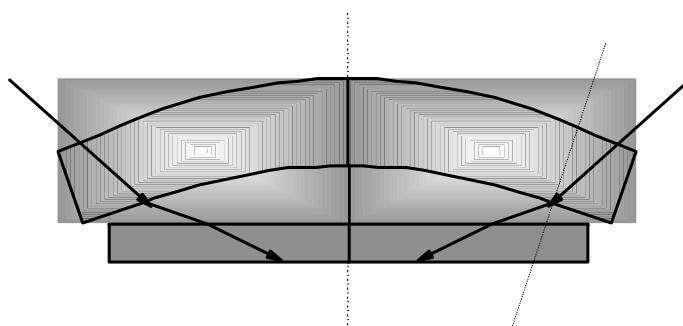


Figure 7.41 Curved lens/fluid interface to support hyper NA.

to the resist surface or the top-coating material on the resist, if used, should be carefully optimized. Too small an angle can lead to water residue at the area outside of the lens immersion head. With too large an angle, air bubbles may be generated during scanning. During development, a small contact angle is preferred for better wetting of the developer.

The thickness of the immersion medium should be determined by optical and hydrodynamic considerations. From the optical point of view, in addition to an acceptable transmission, the thickness bears a direct consequence of an aberration's sensitivity to a refractive index change. Suzuki³³ used the following formula to determine the relationship of the fluid thickness D to the angle θ of the image-forming beam in the fluid:

$$D \leq \frac{\cos \theta}{\Delta n} m\lambda , \quad (7.52)$$

where m is the fraction of the wavelength that the imaging wavefront cannot deviate from an ideal wavelength. Suzuki assumed that the fluid can be controlled to 0.01° C. Using Burnett's number of dn/dT , $1.00 \times 10^{-4}/^\circ$ C, $\Delta n = 1 \times 10^{-6}$. When $\sin \theta = 0.95$, $D \leq 3.1 \times 10^5 m\lambda$. Let $m = 0.02$ and $\lambda = 193$ nm, $D \leq 1.2 m\lambda$. D as a function of $\sin \theta$ is plotted in Fig. 7.42.

The fluid thickness also must satisfy hydrodynamic requirements. Flagello et al.²⁸ take the limits from the requirement of a laminar flow and an acceptable shear force on the projection lens to set the relationship between the fluid thickness and the scan speed using the density and viscosity of the fluid. They arrived at a fluid thickness between 1–2 mm to support a scan speed of almost 500 mm/sec for only a few mN disturbance forces.

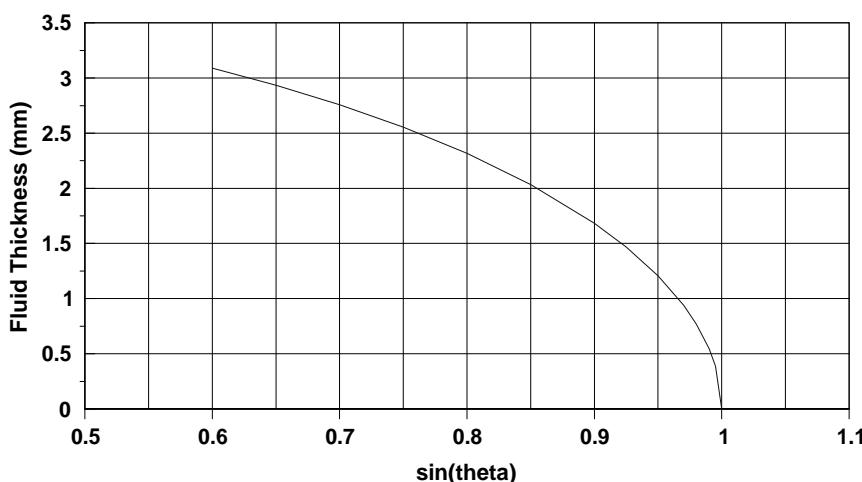


Figure 7.42 Upper limit of fluid thickness for wavefront error less than $\lambda/50$.

7.5.3 The immersion lens

The challenge in the immersion lens varies greatly depending on the NA required. When the NA of a dry system is to be maintained, as in the case of Fig. 7.4, it is not difficult to design and fabricate the immersion lens. The major change from dry to immersion is the reduction of the working distance to utilize the short optical distance required to minimize the effect of inhomogeneity. However, when $NA > 1$ is required, the bending angles in the lens are harder to manage.^{28,29} At $n_{CM} = 1.44$, the design limit appears to be 1.3~1.35 by some lens designers.ⁱⁱ The field size of the lens may must be reduced to keep the size of the projection optics within physical and economic viability. It is well known that field-size reduction lowers productivity of the exposure tool substantially. Economic considerations will dictate the desirability of these choices, as well as the adoption of a different lithographic system.

One must distinguish the use of immersion lithography in reaching the DOF and resolution extremes. For the former, it extends the life of an existing imaging system for near-term new technology nodes, such as the 65-nm node, and broadens the manufacturing margin of these systems. For the latter, there is no dry system that can rival these applications. The limitation of immersion lithography rests where all dry systems cease to function.

7.5.4 Bubbles in the immersion medium

The effect of bubbles in the immersion medium is mostly in redirecting the imaging light in the form of scattering, resulting in reduction of the image contrast. The scattering effect is a strong function of the size and quantity of the bubbles, as well as the distance from the resist surface.

Marston³⁴ investigated light scattering caused by bubbles in seawater. He analyzed the cases of simple bubbles and those having a coating of foreign material at the surface, using Mie theory as well as physical and geometrical optics approximations. The part of his investigation relevant to immersion lithography includes total reflection from bubbles, angle-averaged scattering at the critical angle, enhanced backscattering, and optical radiation pressure on bubbles. Interested readers may find much more work in the references of his article. Marston's explanation on the lifetime of the simple and the coated bubbles are applicable to immersion lithography. Simple bubbles are readily dissolved in the suspending liquid. The extra coating is the major cause of undissolved bubbles. We must make sure that there is no detrimental impurity in the water that preserves the bubbles. Fortunately, Marston showed that the magnitude of scattering from either type of bubbles is similar.

ⁱⁱ Private communications.

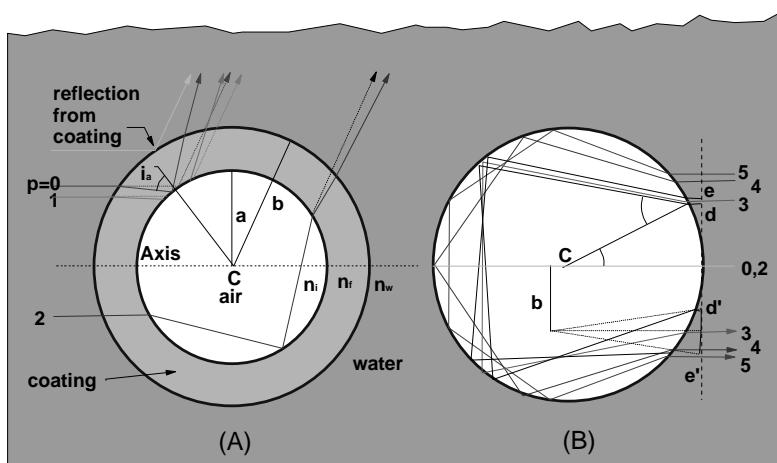


Figure 7.43 Scattering of bubbles reported by Marston.³⁴ (a) Reflection and refraction of rays by a simple and a coated bubble. The rays from the former type are drawn as a dashed line. (b) Backward-directed rays for a bubble in water. (© 1989 IEEE.)

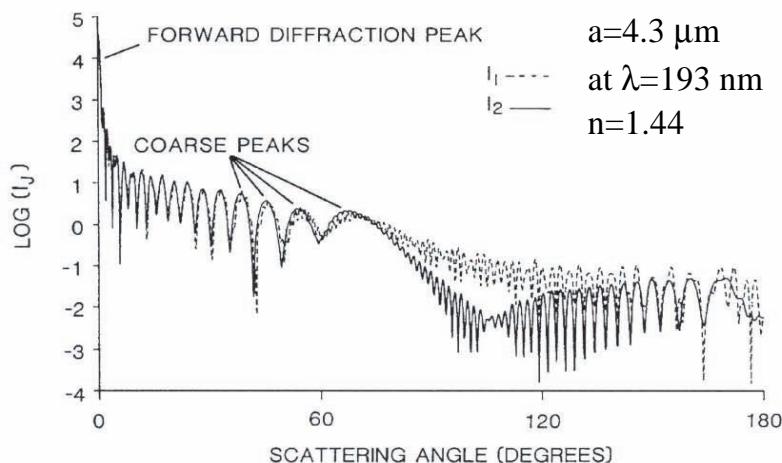


Figure 7.44 Normalized scattering irradiance evaluated from Mie theory as a function of the scattering angle for bubbles with radius $a = 100\lambda_{\text{water}}/\pi = 4.3 \mu\text{m}$.³⁴ (© 1989 IEEE.)

The situations investigated by Marston are shown in Fig. 7.43. The normalized scattered irradiance I_j for a bubble of radius $a = 4.3 \mu\text{m}$ is plotted in Fig. 7.44 as a function of scattering angle, i.e., the angle the scattered light makes with the optical axis. The physical irradiance of the scattered and incident light, i_j and i_{inc} , are related to I_j as follows:

$$i_j / i_{\text{inc}} = I_j a^2 / 4R^2, \quad (7.53)$$

where

$$I_j = |S_j|^2 (\lambda/\pi a)^2. \quad (7.54)$$

S_j is the complex scattering amplitudes from the Mie series. TE polarization is denoted by $j = 1$, and TM is denoted by $j = 2$. R is the distance from the bubble to the observation point. R must be at the far field for the results to be exact. Marston does not specify the originating point of R because he is mostly concerned with the far field. Let's assume that the originating point of R is the center of the bubble at a reasonably far field. Take $R = 0.5$ mm, which is halfway in a proposed immersion fluid thickness of 1 mm. $i_j / i_{inc} = I_j \times 1.8 \times 10^{-5}$. With the 4.3- μm bubble, the scattering irradiance is quite small, except for a narrow range of small scattering angles. But when the bubble is closer to the resist surface, i_j / i_{inc} can be much larger. Hence, the distance of the bubble from the resist surface is a very important factor.

Even though I_j is plotted for only one bubble size, we have used it to gain insight into the significance of bubbles of this size range in optical imaging. It is also useful to check the result of our work using the Mie series.³⁵ Figure 7.45 shows the scattering spectra of bubbles with sizes of 100 nm, 1 μm , and 10 μm . The Marston result with a 4.3- μm bubble fits in the ranges of our 1- μm and 10- μm spectra. The shape of the Marston spectrum resembles that of our 10- μm case. Note that we used $n_{water} = 1.44$ at 193 nm, while Marston used $n_{water} = 1.33$ at the visible wavelengths.

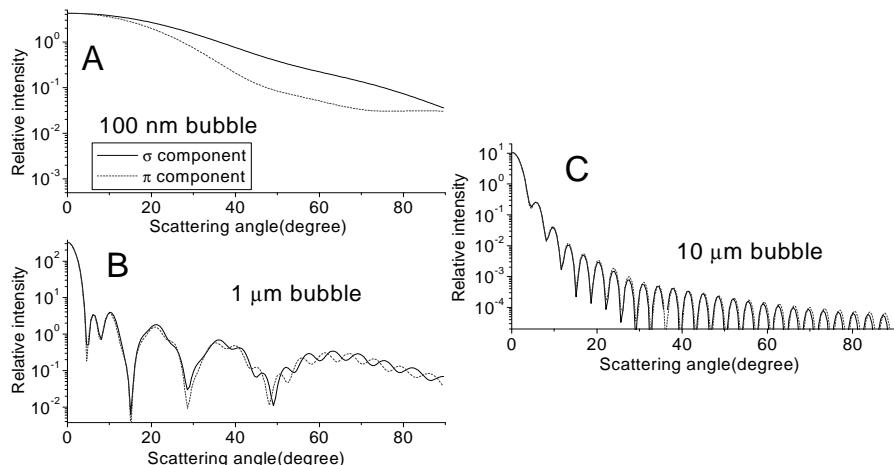


Figure 7.45 Scattering spectra for bubble sizes (a) 100 nm, (b) 1 μm , and (c) 10 μm .

Gau et al.³⁵ use a statistical approach and the scattering cross sections evaluated with Mie theory to investigate the impact of bubbles on imaging. When that happens, the energy redirected by a group of bubbles is

$$P = \left(\frac{Na^3}{V} \right)^{\frac{2}{3}} \frac{\sigma_s}{\sigma_g}, \quad (7.55)$$

where N is the number of bubbles, V is the volume of liquid where the bubbles are equally distributed, a is the radius of the bubble, and σ_s/σ_g is the normalized scattering cross section. Taking all of the possible combinations of each diffraction order, this energy becomes stray light to reduce the contrast of the optical image.

Bubbles do deteriorate the optical image. It is best not to allow any bubble in the immersion fluid during exposure, which is not difficult. There are three possible causes for bubbles in the fluid:

- (1) the release of dissolved gas in the fluid,
- (2) trapped air due to turbulent flow, or
- (3) outgassing of the resist.

The release of dissolved gas is mostly due to the saturation of gas already dissolved in the fluid. When temperature or pressure changes the saturation point, gas is released. Since degassing tools are commercially available, there is no reason not to pretreat it before dispensing it to the coupling space. Doing so also facilitates the dissolution of any bubbles created by a turbulent flow or outgassing. The turbulent flow should be avoided by hydrodynamic design of the immersion enclosure. Bubbles from resist outgassing are also quickly dissolved in the immersion fluid.³³ Even if they do not dissolve, it is dubious that resist outgassing spoils the recorded image. Most resists outgas in a dry system. When these bubbles are released into the immersion fluid, the size and density near the resist/coupling-medium interface in the medium must be similar to those near the interface inside the resist, maintaining the same level of perturbation to the image. As they move away from the resist surface, the size may increase due to a combination of bubbles. It may also decrease due to dissolution into the fluid. When the net effect is a decrease in bubble size, they are less of a problem when compared to bubbles near the interface. Even if the net effect is an increase in bubble size, the density of the bubbles will be reduced accordingly, and the scattering effect decreases when the bubbles are farther away from the interface, according to Eq. (7.53). Now, the size and density of bubbles in an acceptable resist for dry exposure systems must not affect image recording in a detectable way. Otherwise, we would have discovered it with the innumerable exposures taken in worldwide manufacturing. Following the foregoing argument, the

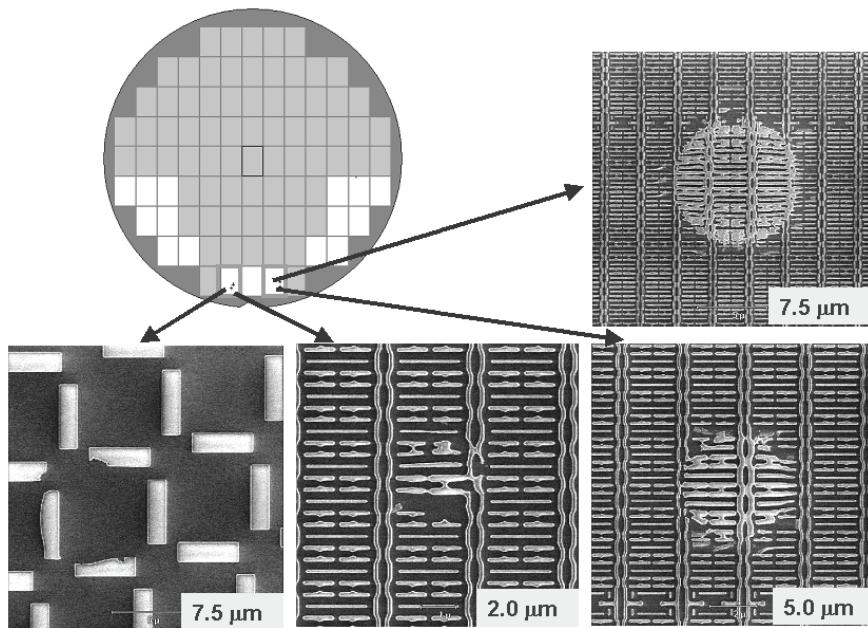


Figure 7.46 Resist image affected by bubbles.

perturbation of the bubbles inside the resist and in the fluid is of similar significance. Hence, outgassing from these resists does not affect the immersion image in a detectable way.

Using degassed water eliminates most concerns about bubbles. However, there is a possibility that a coating of foreign material at the surface of the bubbles prevents their dissolution, as described by Marston. This remains to be seen with quality-controlled water immersing the actual lens and resist surfaces. Should that happen, certain materials must be excluded or included in the immersion system to prevent the formation of the coating.

Concluding from the discussions above, the bubbles that can affect imaging are larger than a few micrometers. They are most likely generated by lack of laminar flow or trapping of air during high-speed scanning. Figure 7.46 shows the resist image affected by air bubbles in the fluid during exposure. The size of these bubbles ranges from 2 μm to over 7 μm .

7.5.5 The mask

To the first order, the mask for immersion lithography is the least impacted. The illumination wavelength on the mask is unchanged. All of the materials used in a dry system remain unchanged. These are the blank, absorber, and pellicles, as well as the fabrication process. Is there a need to immerse the mask? In a 4X reduction system, the NA at the mask side is only 0.325, even when the NA at the wafer side is 1.3. There is no need to immerse the mask to increase NA beyond unity.

7.5.6 Subwavelength 3D masks

Despite the need to change the general configuration and materials of the mask, immersion lithography enables a given light frequency to extend its resolution limit to the point that even 4X enlargement of a feature size still makes the mask pattern subwavelength. For example, 32 nm at the wafer side becomes 128 nm at the mask side. It is only 0.66λ . Subresolution assist features can be as small as 0.15λ on the mask. There are other subwavelength features on the mask, such as jigs and jugs induced by optical proximity correction (OPC). In this case, the 3D nature of mask patterns and polarization effects must be taken into account. There is an anticipated dependence of mask diffraction in feature orientation, absorber thickness, and profile. OPC must be able to take this effect into consideration. Otherwise, other measures must be taken. In addition, there are difficulties in making masks containing very small features. From the wafer-imaging side, as NA becomes very large, the size of the lens, its material, and fabrication costs can be unbearable.

These can be alleviated by reducing the field size of the lens from the well-established standard of $26 \times 33 \text{ mm}^2$ and increasing the reduction ratio from 4X to 8X. Lens design and fabrication become more forgiving. Mask making also becomes easier, having to cover much larger features with less of a feature count. CD control and cycle time are automatically improved. Therefore, field-size reduction and reduction increases complement each other and can be cost effective. They have attracted attention from potential beneficiaries.

The only concern with field-size reduction is the loss of productivity on the exposure tools due to a smaller field size. The time spent in stepping through smaller fields increases, thus dropping the wafer throughput of exposure tools. A previous work³⁶ on productivity loss from 4X to 5X masks showed that the difference in cost between 4X and 5X is about 10%. from the cost difference between 4X and 8X would be significantly more.

Note that in the entire semiconductor manufacturing process consisting of 30~40 masking layers, if only one layer has a tool that requires a smaller field size, all other tools must reset their field size, unless extremely good field stitching can be implemented. Cost savings on a tool may apply on one layer, but productivity loss affects all other layers. Perfect stitching is inherent in the scanning direction on scanners. To increase the field size this way, one must use larger masks, such as 9-in. masks instead of 6-in. masks. Similar to the change in wafer size, the entire industry must support this move.

Hence, productivity loss is by far the dominating factor. Mask magnification should not be increased unless prohibited by the law of physics. Before taking the drastic measure of increasing the reduction ratio and shrinking the field, the limitation from the law of physics can be delayed by using a solid immersion mask,³⁷ as will be discussed in Sec. 7.8.2.

7.5.7 The photoresist

As with any type of lithography, photoresist is an indispensable component in immersion lithography. Without it, the mask image cannot be successfully replicated and transferred. Fortunately, no serious problems have been encountered to expose and develop resists in an immersion environment.³⁸ Some immersed resists suffer from pattern collapse during development, but many resists are not affected.

Besides the optical requirement of a high refractive index, low optical absorption, and minimal outgassing, the resist must withstand a moment of contact with water. The longest possible time of contact is the time to expose a wafer in WBC. At 100 wafers per hour, the water contact time per wafer is on the order of one minute. With LBC, the contact time is reduced by about 1/50, but some areas may be revisited, depending on the step-and-scan scheme. During contact with water, materials such as the photoacid, photoacid generator, or resin can leech into the water. Water also soaks into the resist. The latter can affect resist development characteristics, while the former can become a source of contamination of the lens or a skin coating for bubbles, preventing its rapid dissolution. If these material exchanges are exposure dependent, then exposure and development uniformity may suffer.

In addition to water contact time, the water contact angle of the resist is also important.³⁹ Instead of the usual way of measuring the contact angle, it is measured with a water droplet on an inclined surface to simulate the situation of water being dragged by the immersion head over the wafer during scanning and stepping in a LBC system. Figure 7.47 shows the water droplet on an inclined wafer surface, the receding and the advancing contact angles, as well as the static contact angle. The advancing angle indicates the wetting property of the advancing immersion head as well as the receding immersion head with the receding angle. The receding contact angle influences water stains, and the advancing contact angle affects bubble formation.

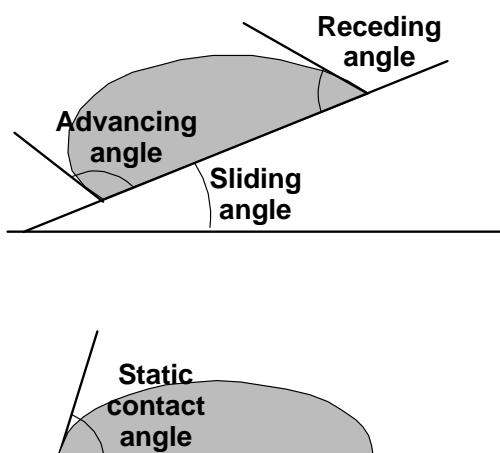


Figure 7.47 Receding, advancing, contact, and sliding angles.

In order to control the contact angles more effectively and isolate the resist material from the immersion fluid, a topcoat is often used. This topcoat works like a TARC in processing. For process simplicity, the topcoat is preferably removable during the development process. There are topcoats that need a separate solvent removal step. Initially, the latter is more effective. However, because of the need for simplicity, developer-removable topcoats have rapidly progressed. Ultimately, a single resist layer without a topcoat is the most preferable process for cost. The need of a topcoat is not obvious. Some resists work perfectly without a topcoat. Rather than specifying the leaching specification too conservatively (thus leading to a limited choice of resist materials and necessitating topcoats), a better prevention of lens contamination is an effective periodic cleaning.

The refractive index of the resist determines the angle of the spatial frequency recombination, which, in turn, affects PDS. Hence, a high refractive index is always preferred for photoresists, unless there are tradeoffs in light absorption, imaging performance, or other resist-processing characteristics.

7.6 Impact on Technology

Having taken in all of the considerations of the immersion-lithography system, the key question is what technology nodes can be used to manufacture semiconductor products? Here, we evaluate $\text{DOF}_{\text{diffrac}}$ and $\text{DOF}_{\text{avail}}$ by simulation for the poly-, contact, and metal layers of the 65-, 45-, and 32-nm nodes, using 193-nm and 157-nm dry and immersion-lithography systems. We also briefly studied the imaging feasibility of 157-nm immersion for the 22-nm node.

7.6.1 Simulation for an immersion system

The simulation is based on the Signamizer scheme⁴⁰ incorporating k_3 high-NA scaling,¹ PDS, and SSL. $\text{DOF}_{\text{diffrac}}$ and $\text{DOF}_{\text{avail}}$ are evaluated. The technology requirement is based on a minimum half pitch, CD, and the other relevant pitches. No isolated features are considered because the use of dummy features is assumed. We have systematically set the polyphoto CD to the technology node, i.e., a 65-nm CD for the 65-nm node, etc. The CD tolerance is simplistically set at a $\pm 10\%$ CD. For the polylayer, the half pitches are taken to be the CD of the last node, i.e., 65 nm for the 45-nm node, etc. For the contact layer, the CD is 1.5 times half pitch, rounded to a close integer. Pitches 2X, 2.5X, 3X, 4X, and 5X CD are included for the common E-D window. The metal CD is taken to be about 1.4 times half pitch. The ratios of line openings to opaque spaces are 1:1, 1:1.5, 1:2, 1.5:1, and 2:1. The resist thickness is 200, 150, and 100 nm for the polylayer of the 65-, 45-, and 32-nm nodes, respectively. It becomes 250, 175, and 100 nm for contact and metal. The resist must thin down to maintain a 3:1 height-to-width ratio and consume less of the required DOF. SSL is set at 10% for the 65 nm node, then 5% for the other two nodes. It is on the pessimistic side; however, the stray-light level is adjusted for pattern density. For example, SSL is multiplied by 0.5 for equal lines/spaces and by 0.25 for equally spaced contact

holes with 1:1 hole-to-space ratio. The half pitches, CD, and features assumed for the three technology nodes for the poly-, contact, and metal layers are shown in Table 7.8.

The results are plotted and tabulated in the following sections. They are taken from DOF- $\sin\theta$ curves based on the Signamizer scheme. DOF-NA curves can be readily obtained by multiplying $\sin\theta$ with the refractive index of the coupling medium. Sample plots on the 45-nm node using 193-nm immersion and the corresponding E-D windows at the $\sin\theta$ values are shown in Figs. 7.48 and 7.49, respectively. All of these calculations use 134.6 nm as the immersed wavelength and $n = 1.4366$ as the refractive index. The immersion wavelength is based on a 193.39-nm vacuum wavelength and the five-digit refractive index. Here, the 130-nm minimum pitch requires the largest θ to resolve. Nevertheless, the 195-nm pitch gates the DOF. The individual biases on the features optimize the common E-D window. They are very well optimized between $\sin\theta = 0.71$ and $\sin\theta = 0.82$. The common DOF coinciding with the 195-nm curve is the evidence.

Both DOF_{diffrac} and DOF_{avail} are listed in the tables that follow. Even though it is preferable to compare DOF_{avail}, providing DOF_{diffrac} makes it more straightforward to compute DOF_{avail} at a new resist thickness. However, only DOF_{avail} is plotted. We expect the exposure tools to support a DOF_{avail} of 250 nm for the 65-nm node, then 150 and 100 nm for the 45- and 32-nm nodes, respectively, based on discussions with exposure-tool vendors. Historically, these numbers have been 400 nm for the 130-nm node and 350 nm for the 90-nm node.

Table 7.8 Poly-, contact, and metal half-pitch layers, CD, and features assumed for the 65-, 45-, and 32-nm nodes.

Polynode	½ Pitch 1	Line ½ pitch next mode	Pitch 1 2.0	Pitch 2 2.5	Pitch 3 3.0		
65	90	65	180.0	225.0	270.0		
45	65	45	130.0	162.5	195.0		
32	45	32	90.0	112.5	135.0		
Contact node	½ Pitch	Hole 1.0	Pitch 1 2.0	Pitch 2 2.5	Pitch 3 3.0	Pitch 4 4.0	Pitch 5 5.0
65	100	100	200.0	250.0	300.0	400.0	500.0
45	70	70	140.0	175.0	210.0	280.0	350.0
32	50	50	100.0	125.0	150.0	200.0	250.0
Metal node	½ Pitch last node	Line 1.0	Pitch 1 2 (1:1)	Pitch 2 2.5 (1:1.5)	Pitch 3 3 (1:2)	Pitch 4 2.5 (1.5:1)	Pitch 5 3 (2:1)
65	90	90	180.0	225.0	270.0	225.0	270.0
45	65	65	130.0	162.5	195.0	162.5	195.0
32	45	45	90.0	112.5	135.0	112.5	135.0

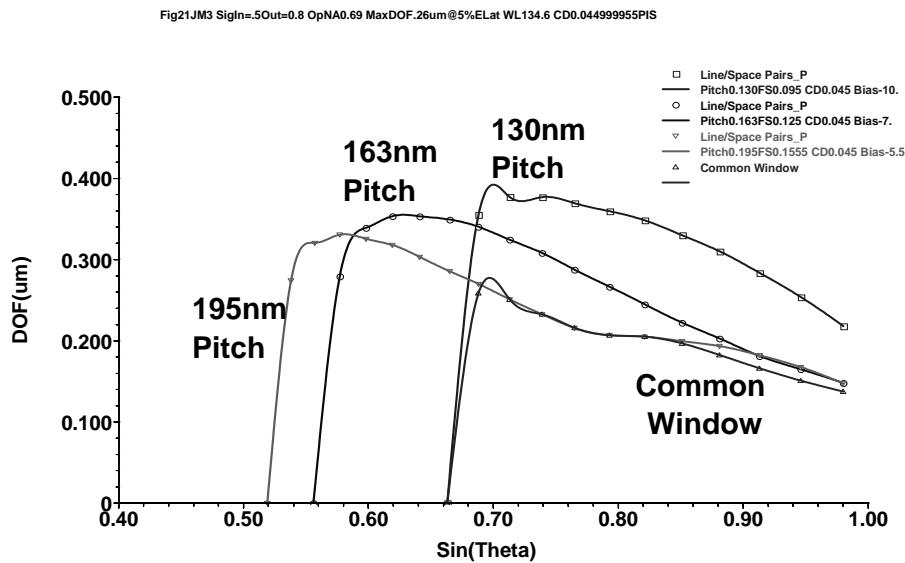


Figure 7.48 45-nm node DOF versus $\sin\theta$ with 193-nm immersion, $\lambda = 134.6$ nm, 6% AttPSM, $\sigma_{in} = 0.42$, $\sigma_{out} = 0.84$, CD 45 nm $\pm 10\%$, pitches 130, 163, 195 nm, 8% EL, and 10% SSL. Biases are -10 , -6 , and -4.5 nm.

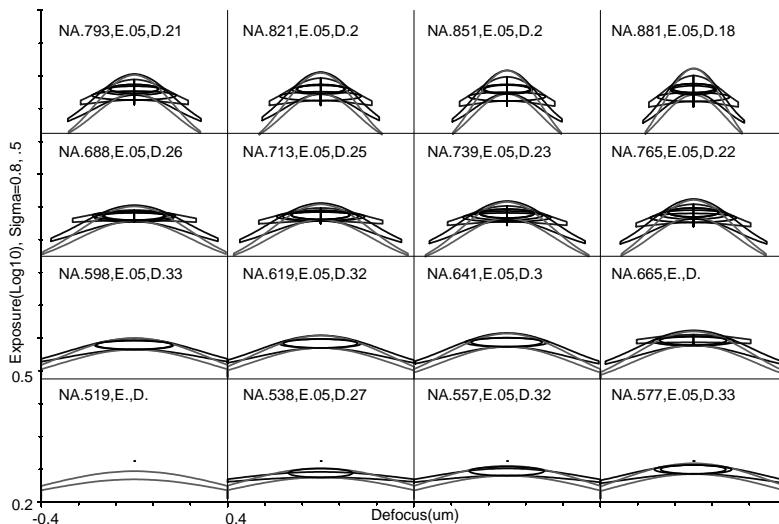


Figure 7.49 E-D windows leading to Fig. 7.48.

7.6.2 Polylayer

The simulation results on the polylayer for the three technology nodes are shown in Fig. 7.50, whose parameters are listed in Table 7.9. At the 65-nm node, the $DOF_{diffrac}$ is 270 nm and the DOF_{avail} is 156 nm for 193-nm dry systems using annular off-axis illumination and 6% AttPSM. The required $\sin\theta$ or NA is 0.888.

The k_1 value based on the half pitch is 0.414. Changing to a 157-nm dry system relaxes the NA to 0.739, while it increases $\text{DOF}_{\text{avail}}$ to 277 nm. The 193-nm immersion system provides a 315-nm $\text{DOF}_{\text{avail}}$ at NA = 0.925. Switching to a lower NA of 0.839, a 286-nm $\text{DOF}_{\text{avail}}$ is obtained. Note that, unlike the dry systems, the NA is no longer identical to $\sin\theta$ in the immersion case. Only the 157-nm dry and 193-nm immersion systems support the 250-nm $\text{DOF}_{\text{avail}}$ requirement. The NA = 0.839 requirement is important because this type of immersion tool is quickly available. Even at the nonoptimum lower NA, its $\text{DOF}_{\text{avail}}$ is still higher than either the 193-nm or the 157-nm dry system.

In 2008, the 65-nm node is being manufactured in almost all semiconductor fabs. Due to cost and timing issues, dry exposure tools are preferred. No immersion tool is needed. It was accomplished with a combination of stronger RET, relaxed design rules, raising the minimum pitch, forbidding the pitches that gate DOF, and better $\text{DOF}_{\text{required}}$ from the exposure tools. Some of the anticipated performance-enhancing techniques for the 45- and 32-nm nodes could also have been applied to the 65-nm node.

Turning to the 45-nm node, there is no chance to use a 193-nm dry system with annular illumination and a 6% AttPSM, regardless of the NA. The 157-nm system provides a 132-nm $\text{DOF}_{\text{avail}}$, and the 193-nm immersion system provides a 155-nm $\text{DOF}_{\text{avail}}$, which is acceptable. However, if necessary, there are several ways to produce a larger margin:

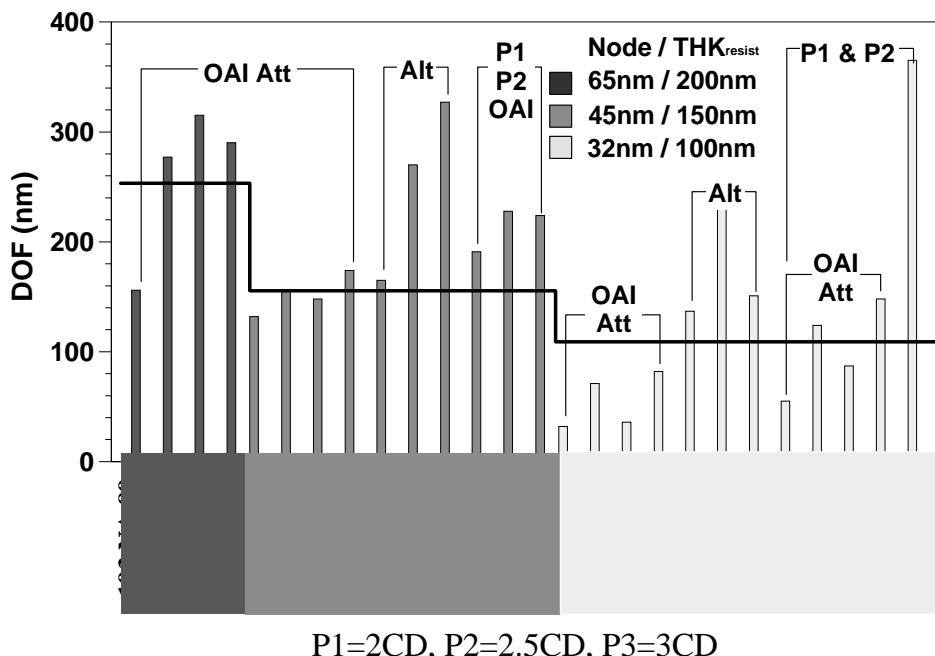


Figure 7.50 $\text{DOF}_{\text{avail}}$ of the polylayer.

- (1) Tighten the DOF control of the exposure tool.
- (2) Use a stronger RET, such as AltPSM or quadrupole illumination. The former can produce 165-, 270-, and 327-nm $\text{DOF}_{\text{avail}}$ for 193-nm dry, 157-nm dry, and 193-nm immersion systems, respectively.
- (3) Remove the most restrictive pitch. Removing the 195-nm pitch produces a $\text{DOF}_{\text{avail}}$ of 191 nm for the 157-nm dry system and 228 nm for the 193-nm immersion system with annular illumination and 6% AttPSM. Note that in this situation, $\text{DOF}_{\text{avail}}$ with BIM is only 4 nm smaller than that of AttPSM.
- (4) Use the TE polarization. A $\text{DOF}_{\text{avail}}$ of 148 nm can be obtained with 157-nm dry systems and 174 nm with 193-nm immersion.

At the 32-nm node, the 157-nm dry system, using 6% AttPSM and annular illumination, no longer supports a common window. 193-nm immersion produces 32 nm of $\text{DOF}_{\text{avail}}$, and 157-nm immersion produces 71 nm of $\text{DOF}_{\text{avail}}$. Resorting to the TE mode yields 36 and 82 nm, respectively. The relief is

Table 7.9 DOF of the polylayer at the three technology nodes.

		$n_{\text{assist}} = 1.75$				$n_{193\text{nmCM}} = 1.4366$				$n_{157\text{nmCM}} = 1.37$			
65 Poly		90	CD=65nm, P1=180nm, P2=225nm, P3=270nm	THK _{resist} (nm) = 200	10%SSL, 8%Elat								
Line No.	λ (nm)	HP(k_t)	sin θ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET		Feature Bias (nm)			
65P 1	193	0.414	0.888	0.888	0.84:0.42	270	156	AttPSM OAI		B-26,-23,-20			
65P 2	157	0.424	0.739	0.739	0.84:0.42	391	277	AttPSM OAI		B-25,-22,-18.5			
65P 3	134.6	0.431	0.644	0.925	0.84:0.42	479	315	AttPSM OAI		B-25,-22,-18.5			
65P 4	134.6	0.390	0.584	0.839	0.84:0.42	450	286	AttPSM OAI		B-25,-22,-20			
45 Poly		65	CD=45nm, P1=130nm, P2=163nm, P3=195nm	THK _{resist} (nm) = 150	5%SSL, 6%Elat								
Line No.	λ (nm)	HP(k_t)	sin θ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET		Feature Bias (nm)			
45P 1	157	0.325	0.784	0.784	0.8:0.4	218	132	AttPSM OAI		B-9,-6.5,-5			
45P 2	134.6	0.325	0.674	0.968	0.8:0.4	278	155	AttPSM OAI		B-8,-5,-3.8			
45P 3	157	0.325	0.784	0.784	0.8:0.4	234	148	AttPSM OAI TE		B-9,-8,-5			
45P 4	134.6	0.325	0.674	0.968	0.8:0.4	297	174	AttPSM OAI TE		B-11.5,-9.5,-8			
45P 5	193	0.280	0.831	0.831	0.3	251	165	AltPSM		-7,-5,-4			
45P 6	157	0.283	0.684	0.684	0.3	356	270	AltPSM		-5,-4,-6			
45P 7	134.6	0.275	0.569	0.817	0.3	450	327	AltPSM		0.2,-5			
Pitches 130, 163 nm													
45P 8	157	0.337	0.814	0.814	0.8:0.4	277	191	AttPSM OAI		B-12.5,-10			
45P 9	134.6	0.340	0.704	1.011	0.8:0.4	351	228	AttPSM OAI		B-12.8,-10			
45P 10	134.6	0.339	0.701	1.007	0.8:0.4	347	224	BIM OAI		B-12.8,-10.2			
32 Poly		45	CD=32nm, P1=90nm, P2=113nm, P3=135nm	THK _{resist} (nm) = 100	5%SSL, 5%Elat								
Line No.	λ (nm)	HP(k_t)	sin θ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET		Feature Bias (nm)			
32P 1	134.6	0.296	0.885	1.271	0.96:0.48	114	32	AttPSM OAI		B-5,-2.5,-1			
32P 2	111	0.310	0.765	1.117	0.96:0.48	149	71	AttPSM OAI		B-10,-8,-6.5			
32P 3	134.6	0.294	0.879	1.263	0.96:0.48	118	36	AttPSM OAI TE		B-5,-2.5,-1			
32P 4	111	0.308	0.759	1.040	0.96:0.48	160	82	AttPSM OAI TE		B-10,-8,-7			
32P 5	134.6	0.265	0.793	1.139	0.3	219	137	Alt PSM		B-8,-2,-6			
32P 6	111	0.269	0.663	0.908	0.3	308	230	Alt PSM		B-10,-4,-9			
32P 7	134.6	0.265	0.793	1.139	0.3	233	151	Alt PSM TE		B-7,-4,-9			
Pitches: 90, 113 nm													
32P 8	134.6	0.314	0.940	1.350	0.96:0.48	137	55	AttPSM OAI		B-8.8,-6			
32P 9	111	0.321	0.793	1.086	0.96:0.48	202	124	AttPSM OAI		B-10,-7.7			
32P 10	134.6	0.303	0.905	1.300	0.96:0.48	169	87	AttPSM OAI TE		B-9,-6.5			
32P 11	111	0.313	0.771	1.056	0.96:0.48	226	148	AttPSM OAI TE		B-10,-7.7			
32P 12	134.6	0.265	0.793	1.139	0.3	447	365	Alt PSM		B-17,-7			

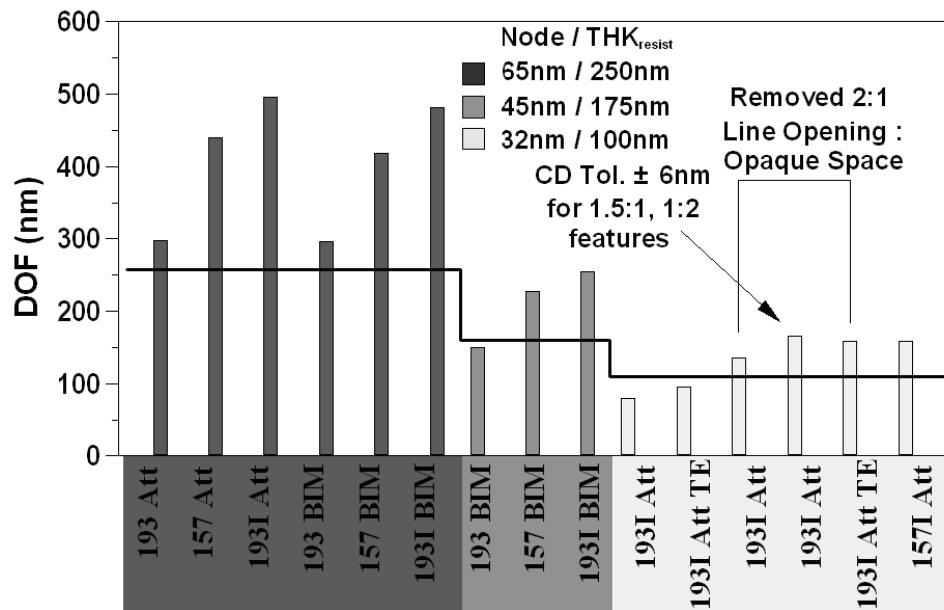


Figure 7.51 DOF_{avail} of the contact layer.

insignificant. DOF_{avail} of the previous four situations is insufficient to cover the 100-nm tool DOF control. Switching to a strong RET such as AltPSM suffices. Staying with 6% AttPSM/annular illumination and restricting the pitch to 90 and 113 nm produces a sufficient DOF_{avail} for 157-nm immersion but not for 193-nm immersion. Tightening of DOF_{required} must be pursued.

7.6.3 Contact layer

Figure 7.51 and Table 7.10 show the DOF of the contact layer at the three nodes. At the 65-nm node, the 272- and 286-nm DOF_{avail} for 157-nm dry and 193-nm immersion systems are acceptable. The 193-nm dry system is not. However, 65-nm manufacturing of the contact layer was still managed with dry systems by similar measures used for improving the DOF of the polylayer.

At the 45-nm node, DOF_{avail} for 157-nm dry and 193-nm immersion systems is insufficient, if staying with 6% AttPSM. However, the DOF_{avail} from the latter system is much closer to the tool requirement. There is no obvious reason that the focus control of the exposure tool cannot be tightened slightly to meet the requirement. Alternately, one may resort to restricted pitch, reduced resist thickness, a slight change in the design rule, or a stronger RET to bridge the gap. For example, keeping the pitch range from 140 to 210 nm facilitates a DOF_{avail} of 172 nm, as shown in the table. Changing the resist thickness to 150 nm results in DOF_{avail} = 163 nm. Relaxing the CD tolerance to ± 6 nm between the 280- and 350-nm pitches extends DOF_{avail} to 166 nm.

Table 7.10 DOF of the contact layer at three technology nodes.

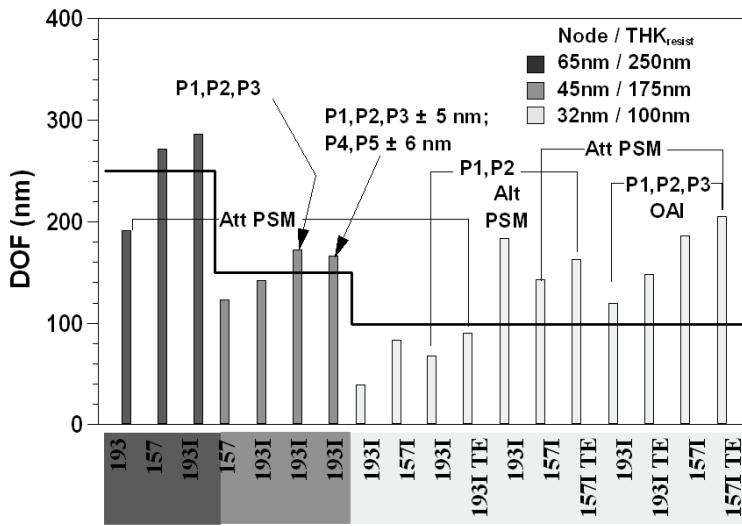
		$n_{\text{resist}} = 1.75$				$n_{\text{193nmCM}} = 1.4366$				$n_{\text{157nmCM}} = 1.37$			
65C T		100	CD=100 nm, Pitch=200, 250, 300, 400, 500 nm				$\text{THK}_{\text{resist}} (\text{nm}) = 250$				10%SSL, 8%Elat		
Line No.	λ (nm)	HP(k_1)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
65C 1	193	0.389	0.713	0.713	0.78	334	191	AttPSM	0,5,7,9,5,11				
65C 2	157	0.385	0.604	0.604	0.78	415	272	AttPSM	0,4,6,5,9,9,5				
65C 3	134.6	0.383	0.516	0.741	0.7	491	286	AttPSM	0,3,4,5,7,7				
45C T		70	CD=50 nm, Pitches=140, 175, 210, 280, 350 nm				$\text{THK}_{\text{resist}} (\text{nm}) = 175$				5%SSL, 6%Elat		
Line No.	λ (nm)	IIP(k_1)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
45C 1	157	0.339	0.760	0.760	0.78	223	123	AttPSM	0,4,5,7,8				
45C 2	134.6	0.328	0.631	0.906	0.78	286	142	AttPSM	0,4,5,5,7,5,8				
45C 3	134.6	0.328	0.631	0.906	0.78	310	166	AttPSM, $\pm 6\text{nm}$	0,4,5,5,4,7,8,8,3				
Pitches: 140, 175, 210 nm													
45C 4	134.6	0.330	0.635	0.912	0.78	316	172	AttPSM	0,4,5,5,4				
32C T		50	CD=50 nm, Pitches=100, 125, 150, 200, 250 nm				$\text{THK}_{\text{resist}} (\text{nm}) = 100$				5%SSL, 5%Elat		
Line No.	λ (nm)	HP(k_1)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
32C 1	134.6	0.319	0.858	1.233	0.86	121	39	AttPSM	0,3,8,4,5,6,5,7				
32C 2	111	0.316	0.702	0.962	0.86	161	83	AttPSM	0,3,8,4,5,6,5,7				
Pitches: 100, 125 nm													
32C 3	134.6	0.321	0.863	1.240	0.86	150	68	AttPSM	0,3,9				
32C 4	134.6	0.309	0.833	1.197	0.86	172	90	AttPSM TE	0,3,9				
32C 5	134.6	0.329	0.885	1.271	0.3	266	184	Alt PSM	10,1				
32C 6	111	0.323	0.716	0.981	0.86	221	143	AttPSM	0,3,8				
32C 7	111	0.310	0.688	0.943	0.86	241	163	AttPSM TE	0,3,9				
Pitches: 100, 125, 150 nm													
32C 8	134.6	0.296	0.798	1.146	0.88:0.44	202	120	AttPSM OAI	-4,-1,2,0,4				
32C 9	134.6	0.289	0.779	1.119	0.88:0.44	230	148	AttPSM OAI TE	-4,-1,6,0,1				
32C 10	111	0.303	0.672	0.921	0.88:0.44	264	186	AttPSM OAI	-4,-1,0,3				
32C 11	111	0.298	0.661	0.906	0.92:0.46	283	205	AttPSM OAI TE	-4,-1,0,3				

For the 32-nm node, $\text{DOF}_{\text{avail}}$ with 193-nm or 157-nm immersion using 6% AttPSM is again insufficient. To stay with 193-nm immersion, two approaches can be taken:

- (1) Use AltPSM and the pack-and-seal technique,³⁰ and remove the 150-, 200-, and 250-nm pitches.
- (2) Use annular illumination on 6% AttPSM and remove the 200- and 250-nm pitches. If one wants to use 157-nm immersion instead, he must remove three pitches with disk illumination or two pitches with annular illumination.

7.6.4 Metal layer

The simulation results for the metal layer are shown in Fig. 7.52, and its parameters are listed Table 7.11. Note that there are three pitches, but five types of features, because the line-opening:opaque-space ratios are 1:1, 1:1.5, 1:2, 1.5:1, and 2:1. At the 65-nm node, the $\text{DOF}_{\text{avail}}$ of 298, 440, and 496 nm are abundant. There is even room to replace AttPSM with BIM. At the 45-nm node, the $\text{DOF}_{\text{avail}}$ for 193-nm dry, 157-nm dry, and 193-nm immersion systems are also usable with BIM.



P1=2CD, P2=2.5CD, P3=3CD, P4=4CD, P5=5CD

Figure 7.52 DOF_{avail} of the metal layer. Off-axis illumination is used throughout.

Table 7.11 DOF of the metal layer at three technology nodes.

		$n_{\text{resist}} = 1.75$				$n_{193\text{nmCM}} = 1.4366$				$n_{157\text{nmCM}} = 1.37$			
		65 M1	90	CD=90, P1=180, P2=225, P3=270 nm				THK _{resist} (nm) = 250	10%SSL, 8%Elat				
Line No.	λ (nm)	HP(k ₁)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
65M 1	193	0.332	0.712	0.712	0.76:0.38	441	298	AttPSM OAI	5,0,5,19,27				
65M 2	157	0.328	0.573	0.573	0.76:0.38	583	440	AttPSM OAI	5,0,5,19,27				
65M 3	134.6	0.330	0.493	0.708	0.76:0.38	701	496	AttPSM OAI	5,0,5,19,27				
65M 4	193	0.332	0.712	0.712	0.76:0.38	439	296	BIM OAI	5,0,5,19,27				
65M 5	157	0.333	0.581	0.581	0.76:0.38	562	419	BIM OAI	5,0,5,19,27				
65M 6	134.6	0.330	0.493	0.708	0.76:0.38	687	482	BIM OAI	5,0,5,19,27				
		45 M1	65	CD=65, P1=130, P2=162.5, P3=195 nm				THK _{resist} (nm) = 175	5%SSL, 6%Elat				
Line No.	λ (nm)	HP(k ₁)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
45M 1	193	0.299	0.888	0.888	0.84:0.42	250	150	BIM OAI	0,5,8,11,17				
45M 2	157	0.319	0.771	0.771	0.84:0.42	327	227	BIM OAI	0,5,8,10,14				
45M 3	134.6	0.321	0.665	0.955	0.84:0.42	399	255	BIM OAI	0,5,8,10,14				
		32 M1	45	CD=45, P1=90, P2=112.5, P3=135 nm				THK _{resist} (nm) = 100	5%SSL, 6%Elat				
Line No.	λ (nm)	HP(k ₁)	sinθ	NA	Sigma	DOF _{diff}	DOF _{avail}	RET	Feature Bias (nm)				
32M 1	134.6	0.292	0.874	1.256	0.92:0.46	161	79	AttPSM OAI	0,4,6,8,5,13,1				
32M 2	134.6	0.292	0.874	1.256	0.92:0.46	177	95	AttPSM OAI TE	0,4,6,7,5,11,5				
32M 3	111	0.289	0.714	0.978	0.92:0.46	237	159	AttPSM OAI	0,4,6,7,5,13				
Pitches: 90, 112.5, 135(1.5:l only) nm													
32M 4	134.6	0.296	0.884	1.270	0.92:0.46	217	135	AttPSM OAI	0,4,6,5,9,2				
32M 5	134.6	0.287	0.857	1.231	0.92:0.46	240	158	Same as above but	0,4,5,8,5,11,5				
32M 6	134.6	0.297	0.888	1.276	0.92:0.46	248	166	AttPSM OAI, ≈6nm	0,4,7,5,9,5				

At the 32-nm node, the DOF_{avail} for 193-nm immersion using 6% AttPSM and annular illumination is insufficient. Adding polarized illumination brings it close to the tool focus control limit, whereas 157-nm immersion has ample DOF_{avail} without resorting to TE illumination. AltPSM is not that useful for the metal layer because of the requirement to establish a common E-D window for

lines and spaces at different pitches. To stay with 193-nm immersion, the 2:1 feature at the 135-nm pitch can be dropped to yield a $\text{DOF}_{\text{avail}}$ of 135 nm. It can be extended to 158 nm with polarized illumination. Further relaxing the CD control to ± 6 nm for the 1:2 feature at the 135-nm pitch and the 1.5:1 feature at the 112.5-nm pitch can extend the $\text{DOF}_{\text{avail}}$ to 166 nm without resorting to polarized illumination.

7.6.5 Recommendation for the three technology nodes

For the 65-nm node, either the 157-nm dry system or the 193-nm immersion system is acceptable. The latter has a larger margin of $\text{DOF}_{\text{avail}}$. The 193-nm dry system requires more enhancements, such as tightening the tool focus control, stronger RET, or dropping some pitches. Even though the 157-nm dry system can support the DOF requirement, the technology is not yet ready because of the lack of photoresist, soft pellicle, and CaF_2 in quality and quantity. The consideration on the contact layer is similar. For the metal layer, $\text{DOF}_{\text{avail}}$ for all three systems is acceptable.

For the 45-nm node at the polylevel, the 193-nm immersion system and the 157-nm dry system are incidentally above and below the cutoff $\text{DOF}_{\text{avail}}$ requirement of 150 nm, regardless of whether polarized illumination is used. Further tightening the focus control of the exposure tool can change this, of course. Going to AltPSM or annular illumination with a designated pitch range moves both systems over the cutoff point. AltPSM can even move the 193-nm dry system over the cutoff point. The selection of AltPSM or BIM with annular illumination must take both performance and cost into consideration. At the contact layer, 193-nm immersion can be made to work more easily. At the metal layer, $\text{DOF}_{\text{avail}}$ for all three systems are acceptable.

For the 32-nm node, from pure simulation results, 157-nm immersion provides more $\text{DOF}_{\text{avail}}$ than 193-nm immersion. However, if the practical difficulties of 157-nm immersion lithography still cannot be overcome by the time of the 32-nm node, there are combinations of RET and design-rule changes that make it possible to use 193-nm immersion lithography.

Polarized illumination can improve the DOF of a given setup. However, its improvement is incremental. The imaging performance of BIM is not much worse than AttPSM. However, it may be adopted in many cases to save mask cost.

The optimum NA requirement of the three technology nodes with 193-nm water immersion is summarized in Table 7.12. The NA to support tool focus control is listed under each layer in each node. The poly-NA is taken from Table 7.9, lines 65P4, 45P2, and 32P5; the contact NA is taken from Table 7.10, lines 65C3, 45C4, and 32C8; the metal NA is taken from Table 7.11, lines 65M1, 65M6, 45M1, 45M3, and 32M6. The “Tool” column provides the suggested NA for the tool. For the 65-nm node, the required tool NA is taken to be a two-digit round off of the highest NA that processing needs, namely 0.84. For the 45-nm node, tools of 0.97 NA suffice. However, that is too small an increment for the

Table 7.12 NA required for 193-nm water immersion at the three technology nodes.

Node	System	Poly	Contact	Metal	Tool
65 nm	dry	—	—	0.712 (65M1)	—
	immersion	0.839 (65P4)	0.741 (65C3)	0.708 (65M6)	0.84
45 nm	dry	—	—	0.888 (45M1)	—
	immersion	0.968 (45P2)	0.912 (45C4)	0.955 (45M3)	1.1
32 nm	immersion	1.14 (32P5)	1.15 (32C8)	1.28 (32M6)	1.3

65-nm tools. Making the maximum NA 1.1 on the 45-nm tools comfortably covers the need of all of the critical layers, plus some subnodes, before reaching the 32-nm node. So the 45-nm tool can turn out to be quite a cross-generation workhorse. At the 32-nm node, the maximum tool NA is again picked to handle all of the 32-nm critical layers.

7.7 Practicing Immersion Lithography

When water-immersion 193-nm lithography was proposed, the industry had already invested in 157-nm scanner development and the infrastructure to support it. There was a strong desire to remain on the 157-nm course, until it was realized that producing a large quantity of high-quality CaF₂ is extremely difficult. Overcoming the high absorption in a 157-nm resist had a similar problem. Developing a durable soft pellicle was also very difficult. The industry switched to water-immersion 193-nm lithography. Exposure tools have become available. Fervent efforts in resist development quickly followed. The mask infrastructure did not need much change. The burden has switched to semiconductor manufacturers to make immersion lithography a viable high-volume manufacturing technology.

7.7.1 Printing results

193-nm water immersion has been applied to logic devices and circuits to explore its potential and shake out its problems. Figure 7.53 shows polysilicon images of 90-nm-node static random access memory (SRAM) overlaid on the active layer.⁴¹ The former was exposed on a 0.75-NA 193-nm immersion scanner at ASML, and the latter was exposed on an equivalent dry system at TSMC. It demonstrated full-field imaging, a usable resist system, an acceptable overlay, and a large

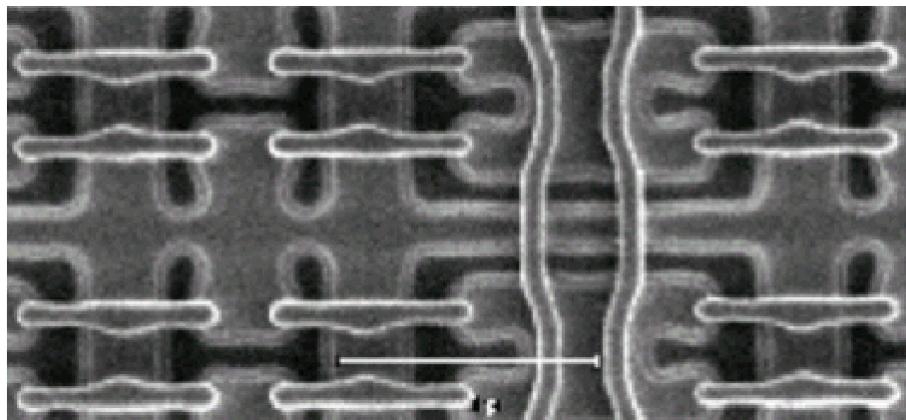


Figure 7.53 Immersion chip using 90-nm-node SRAM, and the polylayer overlays on the active layer. The former was exposed with a 0.75-NA 193-nm immersion scanner.

DOF. Turning to a 0.85-NA 193-nm immersion scanner, 55-nm node SRAM chips were delineated. Figure 7.54 shows the metal layer image of this chip at different field locations. Figure 7.55 compares the number of good dies to bad ones in a dry immersion split at the contact layer.⁴² The ratio is 72:70 versus 62:80, despite the immersion-induced defects. There could have been 25 more good dies, attaining 97:45, if a nonphoto misoperation could be discounted.

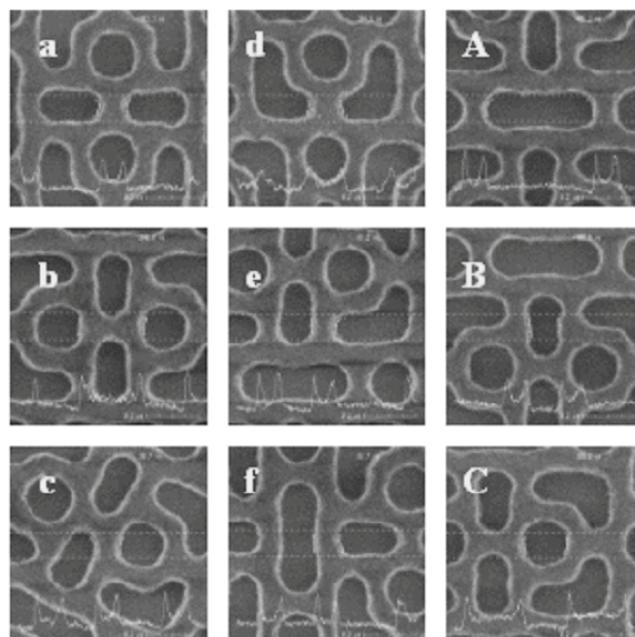


Figure 7.54 0.4- μm^2 55-nm-node SRAM metal layer exposed with a 0.85-NA 193-nm immersion scanner.

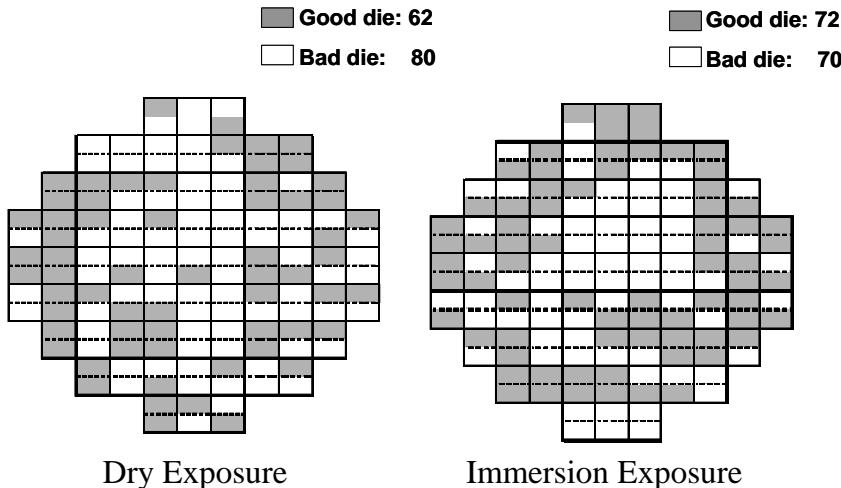


Figure 7.55 65-nm SRAM immersion yield from an R&D lot. One specific test chip was split for dry and immersion exposures at the contact layer. Yield would be higher if the center zone of the immersion wafer were not damaged by a nonphoto operation.

7.7.2 Defect reduction

Immersion lithography has come a long way. The wafer throughput and overlay performance has been steadily improved. There is no longer concern about these issues for 45-nm manufacturing, except for the ever-persisting demands of productivity and packing density. Perhaps the most imminent challenge is to reduce the defect level to a single digit as dry lithography. By press time of this book, many semiconductor factories can control defects to an acceptable level. The major differentiation lies in whether a topcoat above the resist is needed. The company that does not need a topcoat enjoys a cost savings.

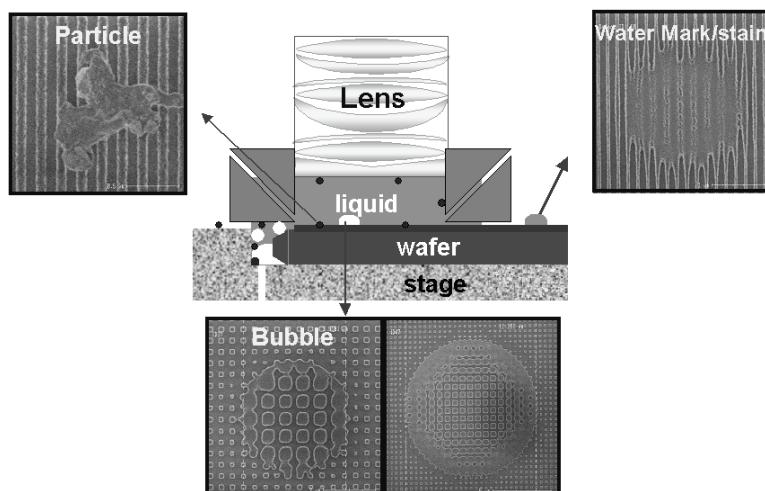


Figure 7.56 Major defects in immersion lithography: particles, water stains, and bubbles.

Figure 7.56 shows three major defect types from immersion exposure, namely water stains, bubbles, and particles. Water stains are due to residual wetting of the resist surface after the immersion head passes through. It can be controlled by preventing residual wetting, by material treatment, and/or by special processing. Bubbles in the water during exposure modify the resist image as if a magnifying glass is inserted during exposure. Bubbles are mostly induced by disruption of hydrodynamics during stepping and scanning. Particles can be in the water supply or carried from any surface that water contacts and carried to the wafer. One primary suspect for particle entry is under the wafer near its edge.

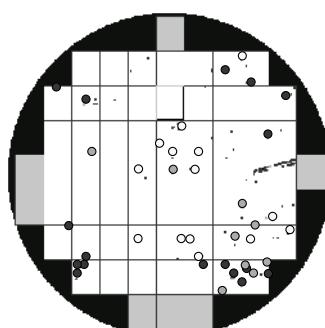
Figure 7.57 shows an effective postexposure soaking method to remove water stains produced in a given range of conditions. New immersion hood designs have significantly reduced the number of bubbles.

Figure 7.58 shows the result of a multiple-exposure study on a plain wafer, aiming at differentiating static and dynamic defects. Out of the 85 defects found on the wafer, only about one-quarter survived after five miniexposures. The other defects that were at different locations during each exposure were voted away. The experiment was conducted with a topcoat on the photoresist and with a postexposure bake to simplify the type of generated defects. We can attribute most of the dynamically located defects to bubbles and the static ones to particles.

7.7.3 Monitoring the immersion hood and special routing

Well-designed immersion systems, together with hydrophilic materials, can prevent bubbles. A postexposure soak or carefully selected resist/topcoat can prevent watermarks. Even just using a topcoat can greatly reduce the number of harmful particles after development. Unfortunately, none of the above methods can prevent the printing of particles deposited before exposure. A unique special routing technique⁴³ can reduce the number of particle-printing defects to a single digit per wafer. This special routing methodology is the follow-up of a special technique to monitor the immersion hood.

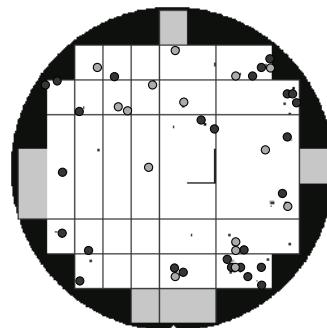
No post-exposure soak



Total defect count 170

100 water-stain defects ○

With post-exposure soak



Total defect count 70

Zero water-stain defect

Figure 7.57 Postexposure soak reduces water-stain defects.

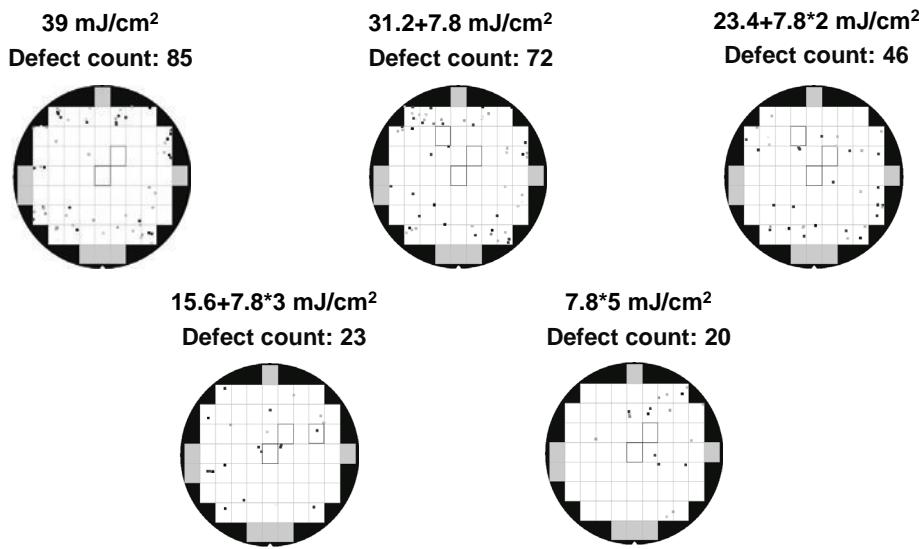


Figure 7.58 Immersion defect study using multiple exposures.

The immersion hood (IH) is an important component of immersion scanners for supplying, confining, and draining the water during exposure. It is circular in shape and placed under the bottom lens element. Its edge is outside the exposure field; that is, the area of coverage of the immersion fluid is larger than the exposure field. To confine the immersion fluid, there is an air knife surrounding the IH with an inward-oriented force in the radial direction, preventing water leakage to the outside of the immersed area during wafer movement. Figure 7.59 conceptually depicts the IH configuration, showing the direction of the air and water flows, the positions of the exposure slit, and the exposure field.

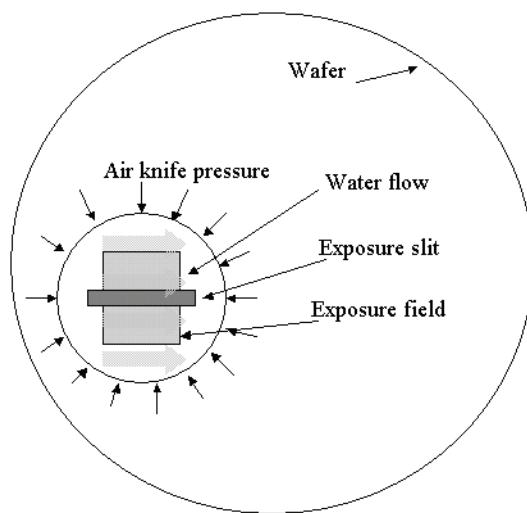


Figure 7.59 Conceptual configuration of the immersion hood.

Since particle printing is an important cause of defects in immersion lithography, it is necessary to identify where the particles come from and how they are printed.

The first possibility is that the particles suspended in water near the resist surface are printed. By considering the 500-mm/sec wafer scan speed and the 1.5-liter/min water flow, the particles in water are relatively static during exposure while the wafer passes by rapidly underneath. The resist image would consist of stripes if particles in suspension were printed. Stripe-like defects have never been observed.

The second possibility is depicted in Fig. 7.60. During exposure, it is possible that particles are transported through water from the IH to attach to the wafer surface, and those inside the exposure slot are printed. There is no direct evidence for or against the existence of this type of defects. Nevertheless, there is no way to prevent them from printing, except to entirely eliminate particles in the water.

The third possibility is that particles left from the previous exposure fields are printed. They may get there by the water droplet leaking out of the IH during the wafer scan. There are indeed water leaks during the wafer scan. Water leaks take place if the IH is not designed properly, the scan speed is too high, or the wafer surface is too hydrophilic. Figure 7.61(a) shows the wafer defect distribution accumulating the defect map of 20 bare-silicon test wafers. Straight line and arc trajectories are clearly observed. Water can only leak along the tangential directions at the edge of the IH along its path of travel, since the water at the IH edge is constrained only in the radial direction. In addition to water leakage in tangential directions at the edge of the IH, the IH also leaks due to the inertial force during stage acceleration, especially at the rear edge while the stage starts to move, leading to arc-shaped defect distribution. Hence, the defect map shown in Fig. 7.61(a) consists of straight lines and arcs from vertical and horizontal movements, respectively. The vertical direction is the wafer-scanning direction.

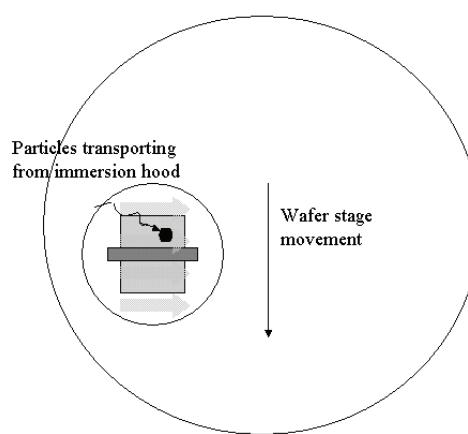


Figure 7.60 Particles transported through water from the immersion hood to the wafer during exposure.

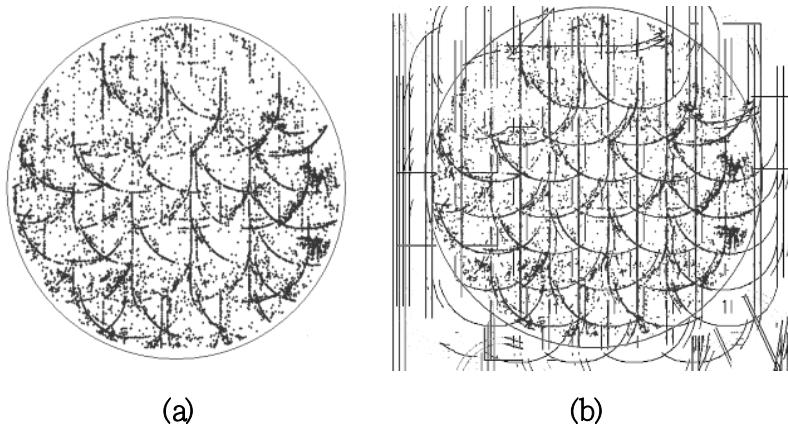


Figure 7.61 (a) Accumulated defect distribution from 20 bare-silicon wafers. (b) Model fitting after considering wafer-stage trajectories during exposure.

Figure 7.61(b) shows the result of fitting the defect distribution to a special model used by the authors of Ref. 43 to describe the phenomenon of the third possibility. The good fitting shown affirms this possibility. If an air curtain containing circularly laid out air holes was used, Fig. 7.61 would contain radial streaks of defect due to low air-pressure regions between the holes around the edge of the IH, enhancing leakage during stage acceleration. The defect maps from Fig. 7.61 were taken from an IH using the air knife instead of the air curtain. The former supports a continuous radial airflow and does not produce radial streaks, thus the absence of such in Fig. 7.61.

Accumulating defect distributions from many wafers and fitting them to the predetermined model has become a powerful tool for defect analysis. From that, a special routing technique was developed. To explain special routing, let us first look at Fig. 7.62, which illustrates where particles are residually left on the wafer after scanning. We classify them as effective and ineffective residuals according to whether they are printed. The residuals that are left on the fields

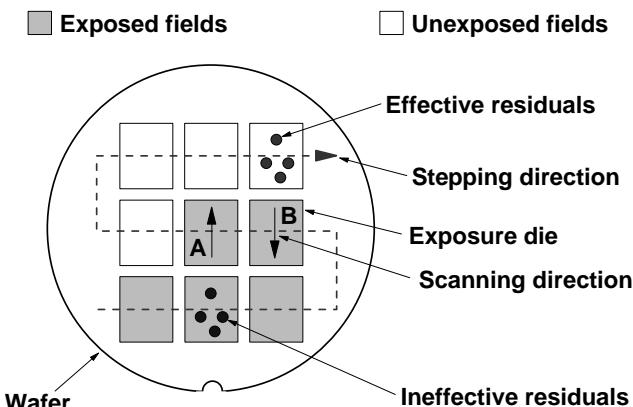


Figure 7.62 Effective residuals and ineffective residuals. The effective residuals are left by field scan B, while ineffective residuals are left by field scan A.

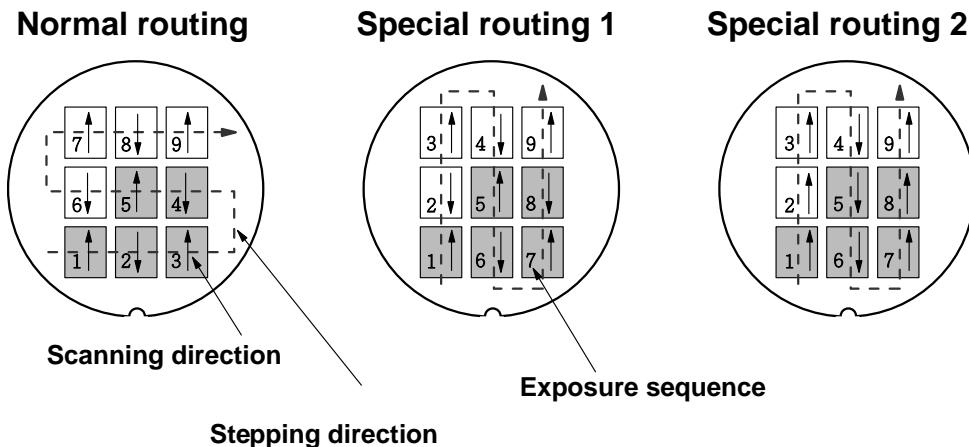


Figure 7.63 Normal routing and two defect-reducing special routings. The normal routing is used for dry-scanner production runs, while special routings 1 and 2 are designated for immersion defect reduction.

already exposed do not contribute to printed defects. They are ineffective residuals as indicated in the figure. The effective residuals in the figure are particles deposited in the unexposed fields. They are printed subsequently if they remain on the wafer without being stirred up by the moving water. Since the hood is larger than the exposure field, it is possible that residuals are left from more than one neighboring field.

Special routings can overcome this problem. Figure 7.63 shows the normal exposure routing and two special routings. In the normal routing, the fields are exposed row by row, and the scan directions alternate upward and downward, so that the time used to move the reticle back can also be used to expose the wafer for better productivity. The scan directions are perpendicular to the field-stepping directions. Our special routing exposes the fields column-wise with scan directions parallel to the field-stepping directions. Special routing 1 has opposite scan directions alternately assigned in a column, while the scan directions are all parallel to the field-stepping directions for special routing 2.

Figure 7.64 depicts the effective and ineffective residuals generated from normal routing. Right after scanning field 1, residuals are left outside the fields. They are ineffective ones. The wafer is now stepped to field 2, then scanned in the opposite direction. The residuals are left at the area for future exposures. Hence, they are effective residuals. After all nine fields are scanned, there are three clusters of effective residuals.

Figure 7.65 depicts the effective and ineffective residuals generated from special routing 1. The locations of the effective residual cluster are different as shown. There are also three clusters of effective residuals after scanning all nine fields. The nine fields are used for clarity of illustration. In reality, the number of fields on a 300-mm wafer is on the order of a hundred. Let us examine whether special routing is less affected by effective residuals (even though Figs. 7.64 and

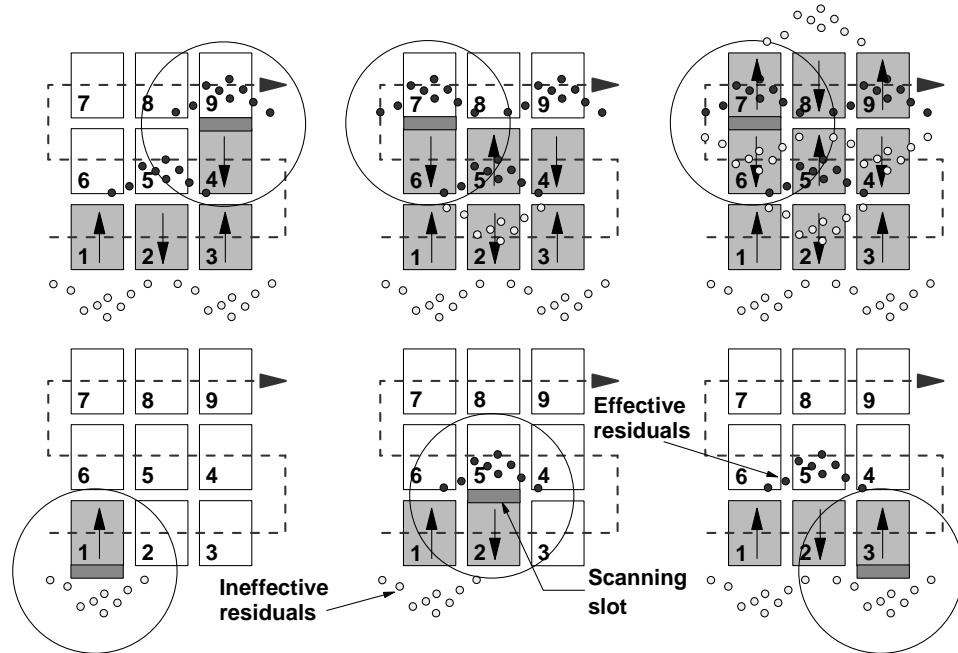


Figure 7.64 Effective and ineffective residuals generated from normal routing.

7.65 show that there is an identical number of clusters). If we add two more rows above fields 7, 8, and 9 in the case of normal routing, i.e., fields 10 to 15, the distribution of effective and ineffective clusters will be identical to that of fields 4 to 9. There will be three additional clusters. Similarly, we can add two more rows above fields 3, 4, and 9 for special routing 1. Fields 3a, 3b, 3c, and 3d that will be scanned between fields 3 and 4. Then, two additional fields will be scanned after field 9. There will still be three additional effective clusters for the two additional rows. The defect-reduction capability of special routing 1 is not obvious. So far, we have only considered the arc trajectory of the defect map. Adding consideration of the tangential trajectory may differentiate the defect characteristic of the two routing schemes.

Figure 7.66 shows the effective and ineffective residuals generated from special routing 2. All residuals are ineffective. We expect this special routing to significantly reduce defects. Comparing two wafer lots of 25 wafers each, the wafers using special routing 2 average 4.8 defects per wafer against 19.7 defects per wafer.

Even though special routing confirms the impact of water leakage at the IH, it reduces the wafer throughput. Special routing should not be used in high-volume production.

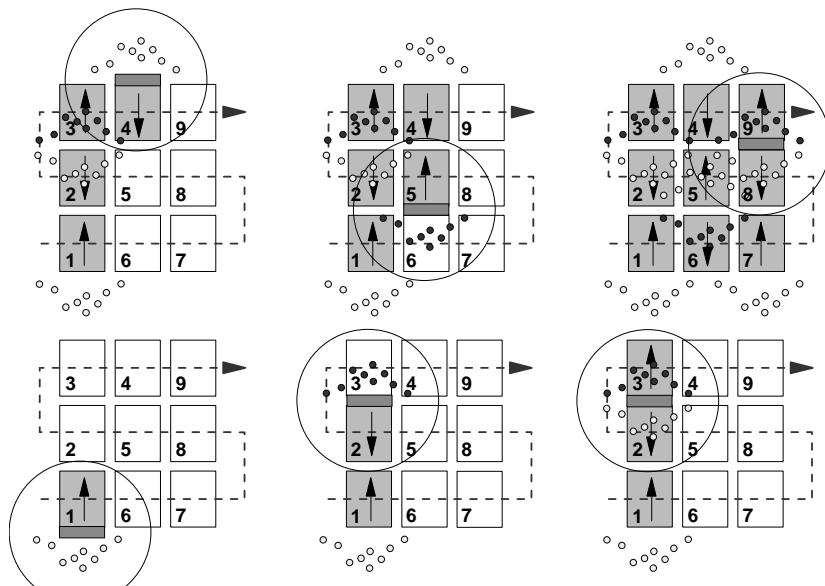


Figure 7.65 Effective and ineffective residuals generated from special routing 1.

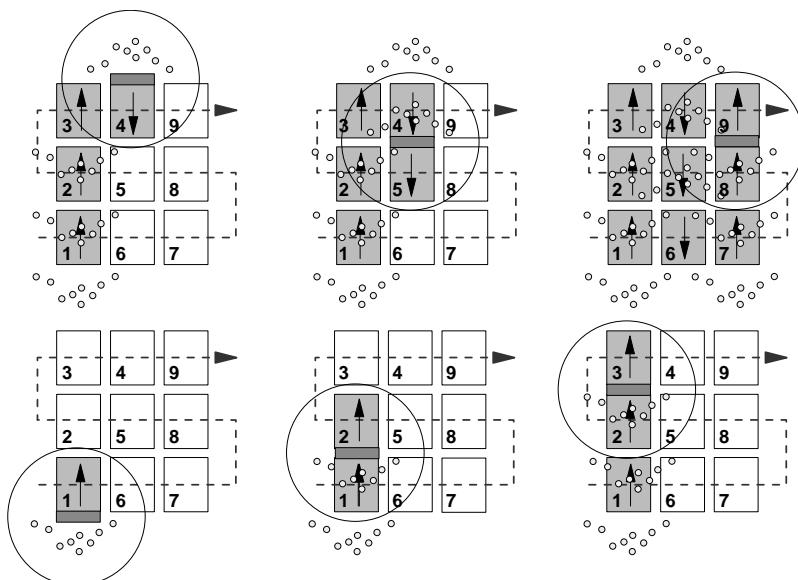


Figure 7.66 Effective and ineffective residuals generated from special routing 2.

7.7.4 Other defect-reduction schemes

The following schemes are feasible for high-volume production.

7.7.4.1 Wafer and equipment cleanliness

Flowing water can stir up particles otherwise dormant in the exposure process. It can carry particles generated by or attached to any surface along its path. Water

itself can contain particles from its source. All causes must be painstakingly eliminated. The supplied water must be carefully filtered. All parts in contact with water must be thoroughly cleaned during the assembly process. After the exposure tool is put to service, these surfaces must be periodically cleaned. All semiconductor-processing equipment with water passage requires periodic cleaning to remove particles and other contaminants. Immersion lithographic equipment is no exception. The wafer edge is susceptible to many nonlithography processes. It must be taken care of. The wafer backside is usually not as well kept as the front side, and thus it is particle prone.

7.7.4.2 Wafer seal ring

There is an easy way to isolate backside particles from reaching the front side. A thin, soft material can cover the gap between the wafer edge and the wafer chuck, as shown in Fig. 7.67. The soft material with a thickness of less than 0.1 μm is attached to a thicker frame for support and transportation. The wafer chuck is recessed to allow the seal ring to lay flat at the wafer edge. A vacuum through the gap between the wafer edge and the frame tightens the seal ring against the contacted surfaces. When the LBC immersion head is at the wafer edge, water does not leak through it, bubble-generation turbulence is eliminated, and particles from the back of the wafer are sealed to prevent them from being stirred up. Figure 7.68 shows the loading and unloading of the wafer with the seal ring and its installed frame. A frame raiser lifts the seal ring and the frame above the wafer loading/unloading position so that normal wafer loading/unloading is unperturbed. At the raised position, the frame and seal ring can be conveniently replaced.

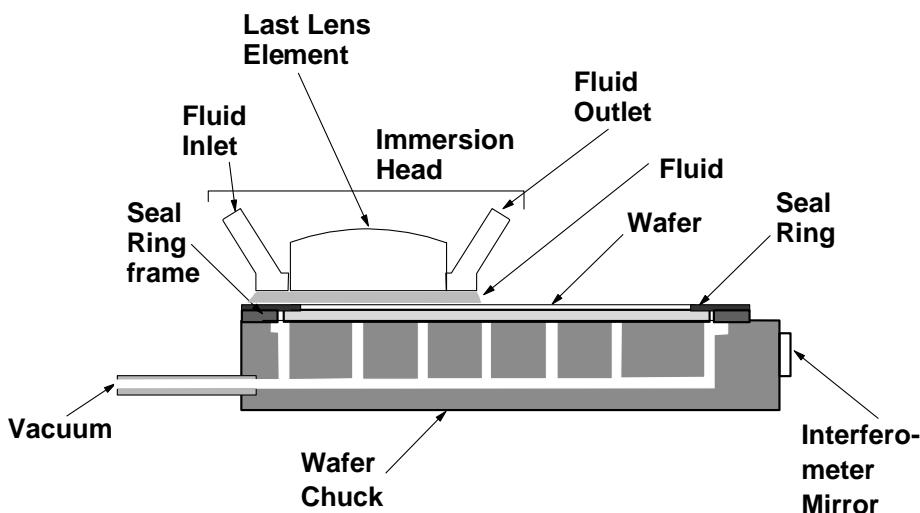


Figure 7.67 Wafer-chuck-supporting edge-seal ring on a seal-ring frame placed at the edge of the wafer. The frame and wafer are coplanar.

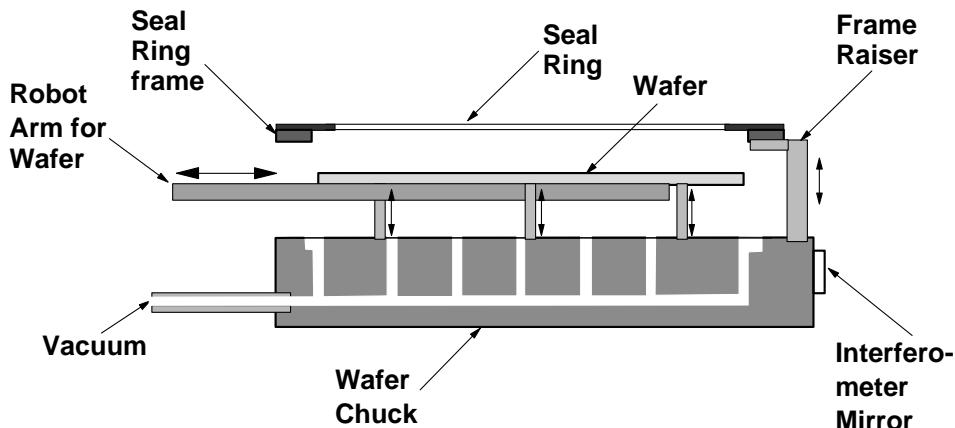


Figure 7.68 Wafer-edge seal ring and seal-ring frame raised above the wafer, which is lifted by pins to allow a robot arm to load/unload the wafer to/from the wafer chuck.

7.7.5 Results

A defect monitor chart for an ASML 1700i immersion scanner at TSMC is shown in Fig. 7.69. Two resist systems were monitored. The defect characteristic is quite different with these two resist systems. PR type 2 is clearly much more preferable than PR type 1 in terms of defect count per wafer. Figure 7.70 is a fall-on particle trend chart for the polylayer of the 45-nm logic node. Except for some excursions, the defect level is typically in the single digits and is often zero.

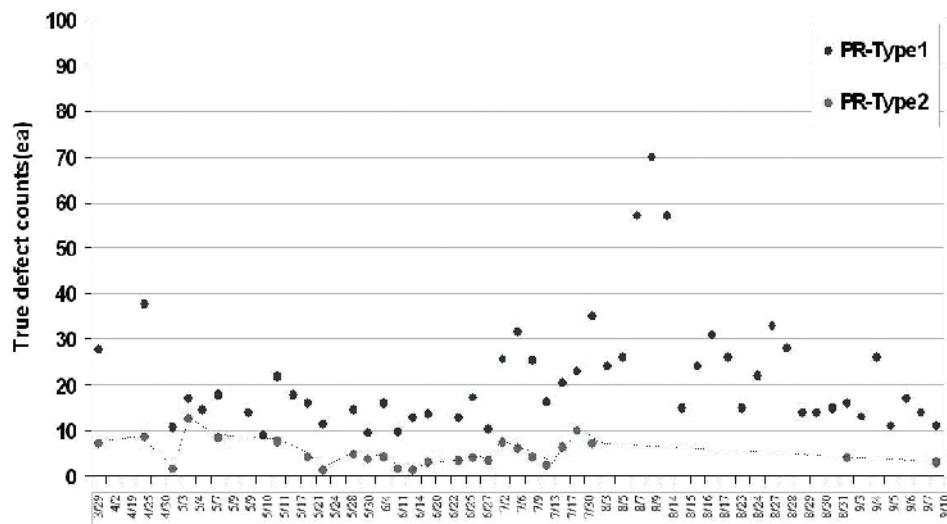
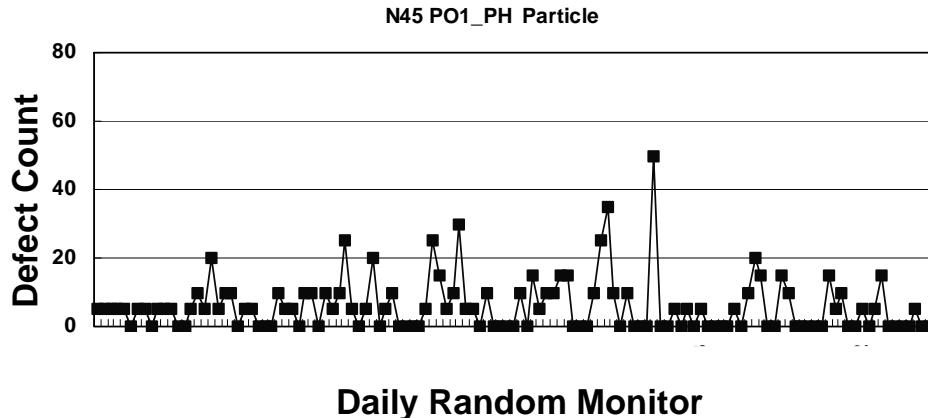


Figure 7.69 Defect monitor chart for ASML1700i.



Daily Random Monitor

Figure 7.70 Fall-on particle trend chart for the polylayer.

The yield and the distribution of the quiescent supply current (IDQ) of the 65-nm product integrated with immersion-imaged active, poly-, contact, and metal-1 layers, successfully achieved the dry-based production level, using the ASML XT-1400E and 1400I scanners with 0.93-NA dry and immersion lenses, respectively. The comparison of IDQ distribution is shown in Fig. 7.71. Even though immersion lithography is not needed for mass production of 65-nm products. It is a unique node that both dry and immersion scanners are available for comparing yield and electrical performance.

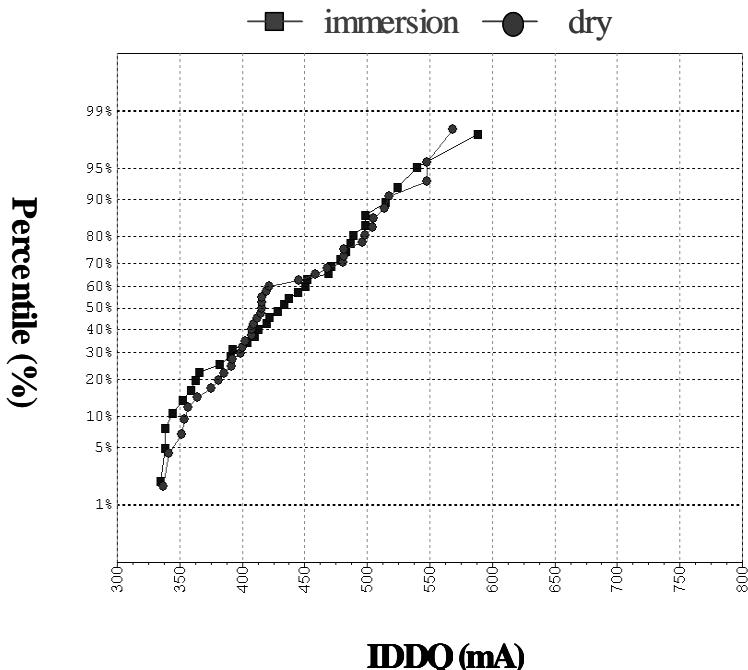


Figure 7.71 IDQ distribution comparing products using all dry layers to those built with active, poly-, contact, and metal-1 layers with immersion lithography.

7.8 Extension of Immersion Lithography

A few more techniques can be used on 193-nm water-immersion lithography to extend beyond 45-nm half pitch, namely by using higher-index materials, solid-immersion masks, polarized illumination, and double exposures.

7.8.1 High-index materials

Air with a refractive index close to 1 has been limiting the spatial frequency coupling from the lens to the resist. Water with a refractive index of 1.44 at 193 nm is much better. The theoretical coupling limit sets the maximum NA of water-immersion 193-nm systems to 1.44. However, using a realistic upper limit of $\sin\theta = 0.95$ for the imaging lens, the realistic NA limit is 1.368. To break through this limit, a higher-index coupling fluid must be used. Figure 7.72 depicts the possibilities. Staying at $\sin\theta = 0.95$, a fluid with $n = 1.56$ moves the limit to NA = 1.482. Because the refractive index of fused quartz is about 1.56 and that of CaF₂ is lower, an immersion fluid with an index higher than 1.56 cannot support a higher NA with a flat lens bottom. For example, at $n = 1.66$ the maximum realistic NA is still 1.482, except that due to a smaller angle in the fluid, DOF is larger. Staying with $n < 1.56$ in the lens material, the only way to increase NA beyond 1.482 is to bend the last lens surface, as shown in Fig. 7.41. Having a nonflat bottom lens surface requires extremely low absorption in the coupling medium. Otherwise, light would be absorbed unevenly in the coupling medium, leading to imaging difficulties. Hydrodynamics also suffers. In order to keep the bottom lens surface flat, higher-index optical materials are needed. Though still not ready for practical applications, 193-nm immersion fluids with an index ranging from 1.5 to 1.8 and lens materials with an index ranging from 1.6 to 2.0 were presented at the Second International Symposium on Immersion Lithography.^{44,45,46}

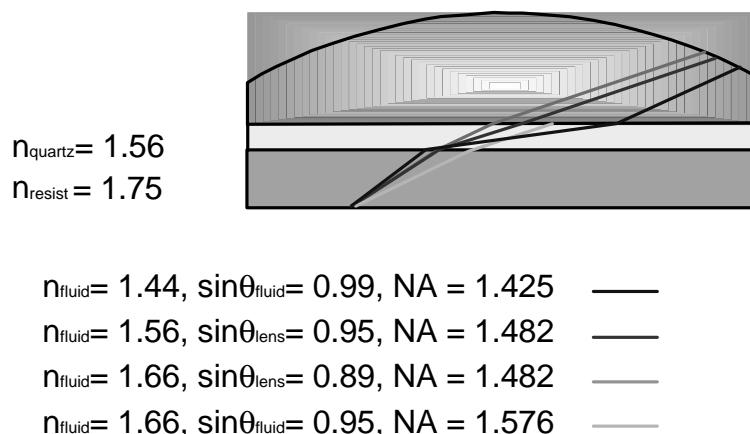


Figure 7.72 Impact of the fluid index.

7.8.2 Solid-immersion mask

CD at 32 nm is subresolution on the mask. Even at 4X magnification, it is only 0.663λ for the 193-nm ArF wavelength. When subresolution-assist features are used, they can be as small as 0.166λ on the mask. There are also subresolution jigs and jugs induced by optical proximity correction (OPC). In addition, the absorber thickness is not negligible. A 70-nm thickness is 0.36λ . These 3D subwavelength features react to the illumination nonlinearly and can lead to imaging difficulties, not to mention challenges in making masks that contain very small features.

Instead of using a larger magnification at the mask, a solid immersion mask can be used to delay the onset of the 3D subwavelength effects. The solid immersion mask is depicted in Fig. 7.73. A high-index transparent material planarizes over the 3D absorber structure on the mask. This way, the illuminating wavelength is reduced according to the refractive index of the planarizer. Reduction up to 70~80% is possible. Even though the wavelength is restored after light leaves the planarizer, the subwavelength diffraction zone is already taken care of.

In addition to having a high refractive index, the solid immersion material must be transparent to the imaging light. Its optical property should be consistent after repeated exposures. The material must be coated by using a planarizing method. If coated otherwise, such as using a conformal coating method, it must be polished for planarization. The coating process should not trap any particulates or bubbles to imbed defects on the mask. It should be strippable without damaging the mask. A resist resin would be a good starting candidate. Figure 7.74 shows improvement in the measured mask error enhancement factor (MEEF) of an uncoated mask, which was consecutively coated with 200 nm and 500 nm of a high-index material. The CD is 65 nm over a pitch range from 180 nm to 1 μ m. Here, MEEF is improved from 3.2 to 2.5, and DOF is improved from 220 to 270 nm for the 180-nm-pitch polytest patterns using the 500-nm-thick planarizer. The 200-nm-thick planarizer helps little.

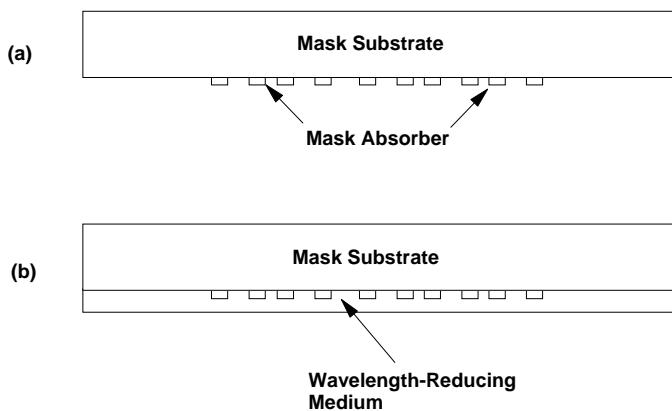


Figure 7.73 (a) A typical binary intensity mask. (b) A mask immersed with a wavelength-reducing solid medium.

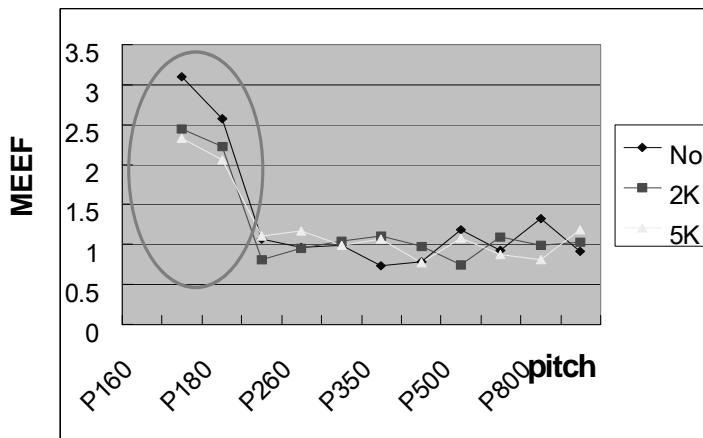


Figure 7.74 MEEF from an uncoated mask, a mask coated with 200 nm of high-index material, and a mask coated with 500 nm of high-index material.

7.8.3 Polarized illumination

Polarization and its impact on optical imaging in high-NA dry and immersion systems have been discussed in detail in Sec. 7.4. Here, we capture the impact for extending immersion lithography. A light wave propagating with an arbitrary orientation of the electric field vector toward the wafer can be split into two components: TE, with the electric field vector parallel to the resist surface, and TM, with the magnetic field vector parallel to the same. In terms of a combination of spatial frequencies, the TE mode is most efficient. All TE spatial frequencies have their electric field vectors collinear to each other. They superimpose completely, leaving no uncombined stray light. With the TM mode, the combination efficiency of the electric field vectors is a function of the incident angle, less efficient with larger angles. The uncombined components travel in opposite directions and become stray light to reduce the image contrast. When the illumination is restricted to only the TE mode, the image contrast is preserved, even at large incident angles.

In a projection imaging system, the illuminating light is split into 0, ± 1 , and ± 2 orders by the mask. They are scattered by surface roughness and multireflected by an imperfect antireflection coating at the lens surfaces. They are also reflected by the mask absorber. All turn into system stray light, which is polarization dependent, but the effect is not strong because of the averaging effect of many different reflecting surfaces, as well as incident and reflecting angles. The $\pm 2^{\text{nd}}$ -order beams blocked by the lens aperture may also turn into stray light because of incomplete blocking. Now the 0th- and $\pm 1^{\text{st}}$ -order beams that transmit and reflect at the resist/wafer and resist/coupling-medium interfaces are strongly polarization dependent. The TM-transmitted beams into the resist do not completely recombine. The reflected TE and TM beams are scattered and multireflected by the lens surface as well as reflected by the mask absorber, just like the incident beams. They induce another level of stray light. The resultant contrast loss leads to a reduction in the EL and DOF.

Table 7.13 Polylayer EL and DOF gain with polarization, a 32-nm line, and 90-nm and 113-nm pitches.

AttPSM OAI	NA	Sigma	ELAT	DOF _{diffrac}	DOF _{avail}	Improv.
Unpolarized	1.35	0.96:0.48	5%	137	55	—
TE	1.3	0.96:0.48	5%	169	87	23/58%
Unpolarized	1.25	0.96:0.48	2.43%	200	118	—
TE	1.25	0.96:0.48	3.57%	200	118	47%

Given an 8% EL, the DOF of 65-nm equal lines and spaces is plotted in Fig. 7.31 for 193-nm water immersion, 193-nm dry, and 132-nm dry systems.¹⁹ The wavelength of 132 nm is used because of the earlier reported value of the refractive index of water^{4,5} at 1.46. Since adjusting the wavelength by 1.4% would not produce any perceivable difference in the figure, 132 nm is not updated. The curve with the largest DOF is based on either a perfect 193-nm immersion system or a 132-nm dry system. As soon as SSL is considered, the DOF drops to the NPDS curve. DOF drops further with PDS included. The loss is more significant for the 193-nm immersion system than the 132-nm dry system. The worst-performing group is related to 157-nm systems with the two stages of DOF reduction due to SSL and SSL+PDS. A typical way to interpret the curves is that 193-nm immersion systems are always better than 157-nm dry systems—stray light for stray light. Another way to look at them is to consider the room for improvement when the illumination is polarized and SSL is reduced. Table 7.13 lists the gain in DOF_{diffrac} and DOF_{avail}¹¹ with TE polarized illumination at a 5% EL for the common E-D window of 32-nm lines, 90-nm and 113-nm pitches, using AttPSM and OAI. The gain in EL with DOF_{diffrac} kept constant at 200 nm is also shown. The gains are 23% in DOF_{diffrac}, 58% in DOF_{avail}, and 47% in EL.

7.8.4 Double exposures and pitch splitting

To further the limit, pattern pitch is a very important item to control. After the minimum pitch is reached, the limit can be extended by splitting the mask pattern into larger pitches to delineate them separately without cross-exposure interference onto the same areas on the wafer. The split-pitch technique is illustrated in Fig. 7.75. The contact-hole pattern in the final image is split into two sets of contact holes whose size and pitch are sustainable with the imaging technology of the previous node. The size is shrunken after double exposures combine the images of the larger holes. Avoiding unacceptable interference of the two exposures is key to the success of the split-pitch technique. Even though the interference of the two exposures is incoherent, it should still be avoided when resolution is pushed very hard. A straightforward but expensive way to do so is to develop and etch the pattern from the first exposure before following with the second one. A more economical way is to render the resist photoinsensitive, such as with crosslinking, then overcoat with another layer of resist for the

second exposure. The most desirable way to avoid this unacceptable interference is to develop a resist with sufficient contrast to support double exposure.

Figure 7.76 illustrates the working principle of double imaging. Figure 7.76(a) uses coherent illumination on an object containing two adjacent openings. The diffracted electric field from each opening is plotted in black. The intensity of the combined electric field is plotted in dots. The peaks and valley of the intensity are almost washed out between the two features, making them difficult to resolve. If the features are separated further, thus decreasing the resolution requirement, the electric field between them will be lower, and the valley of the resultant image intensity will have a higher contrast. To improve the resolution, we can expose the two openings one at a time on the same resist material. This double-exposure situation is depicted in Fig. 7.76(b). The combined image has a correlation between the two openings only through the intensity recorded in the resist. It indeed has more contrast than that shown in Fig. 7.76(a). Note that the small intensity at the midpoint between the two features is not zero. The resist is partially exposed away from the line edges at their dark side. For a single exposure, this residual exposure is below the threshold for development. With two exposures, the partial exposures accumulate and can become larger than the threshold. The contrast of the resist must be higher than that used for single exposure to further suppress the threshold. Alternately, to make the accumulated exposure zero for ultimate resolution, the resist image must be transferred one at a time, e.g., etched into the substrate one at a time. This is a double-patterning process.

The resist images of single and double exposures and double patterning are shown in Fig. 7.77. In the case of single exposure, the line ends join at the after-develop image (ADI), even though etching separates the line ends as shown in the after-etch image (AEI). In the case of double exposure, the four line segments in the middle consists of only two long line segments. A second exposure having a horizontal transparent segment in the middle cuts the two lines into four segments; they are well separated. The AEI is a faithful reproduction of the ADI, but the line ends at these cuts are sharpened due to residual resist exposure from

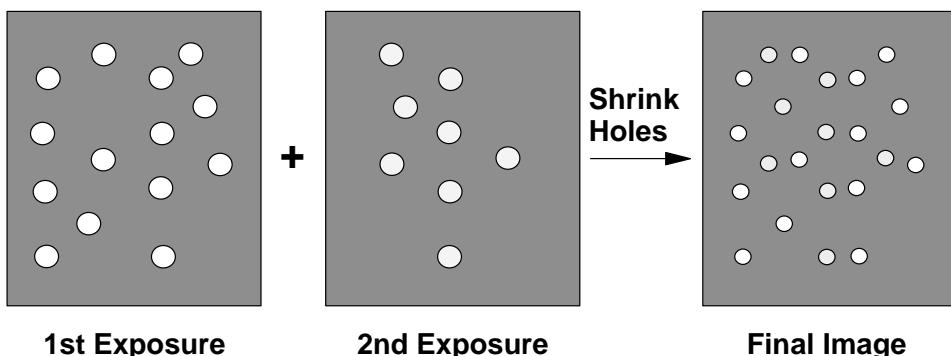


Figure 7.75 Pitch splitting.

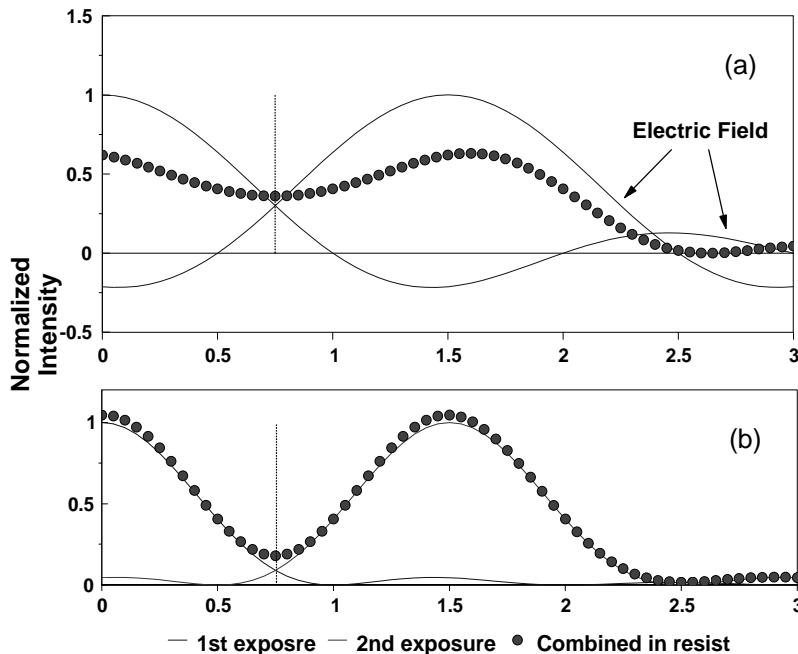


Figure 7.76 Combined intensity in a resist. (a) Coherent illumination from a single exposure. (b) An incoherent combination of light with a double exposure.

the first exposure. In the case of double patterning, the AEI of the first exposure is etched out. Then, the resist is stripped, and a new resist layer is recoated. The same horizontal cutting mask is used to delineate the cut all of the way to the AEI. There is no evidence of any interaction of the two patterns in the two layers.

7.8.5 Pack-unpack

The pack-unpack technique^{30,47} uses a second resist exposure to handle the cross-interference issue. This double-exposure technique removes the less favorable pitch. In the case of contact holes, either with AltPSM or OAI, the less-favorable pitches are the large ones all of the way to the isolated holes. They suffer from a small DOF. Hence, a mask with padded holes filling the gaps left by the desired holes is first exposed, followed by the exposure of a second mask containing unpacking features to fill the unwanted holes or reopen the desired holes. The resist image from the first exposure is covered with a second resist layer, as shown in Fig. 7.77. This double-exposure technique produces a larger DOF. The alignment of the second exposure is less critical than many other double-exposure techniques because its function is only to remove the unwanted holes. Figure 7.78 also shows the SEM image of a working second resist system. The alcohol-based resist system reported in Ref. 47 does not affect the first resist image. It images well as seen in the two SEM images at two magnifications.

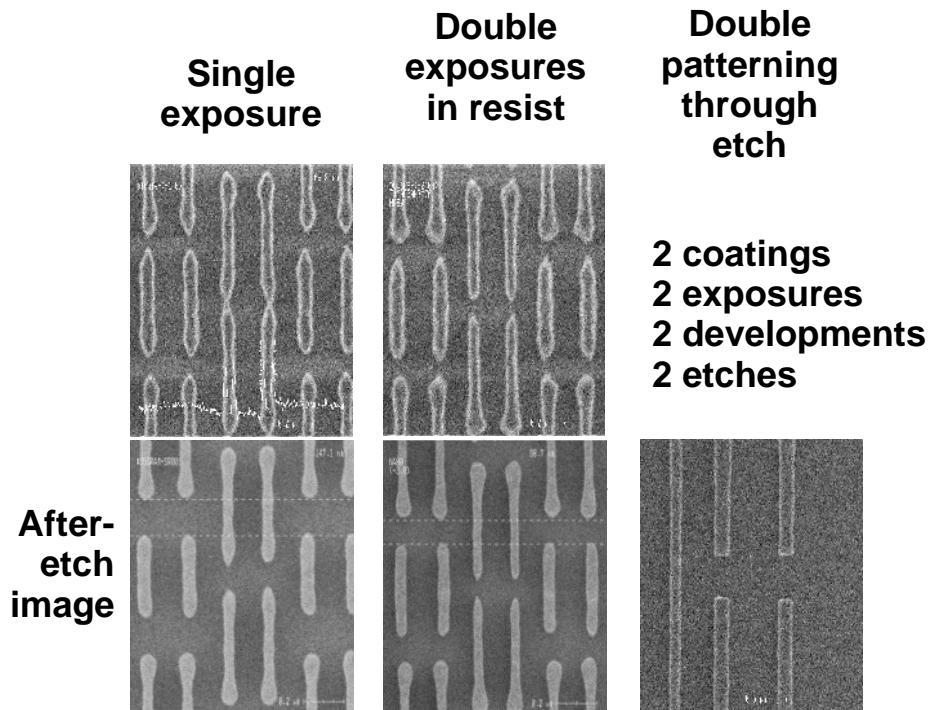


Figure 7.77 Comparison of single and double exposures with double patterning.

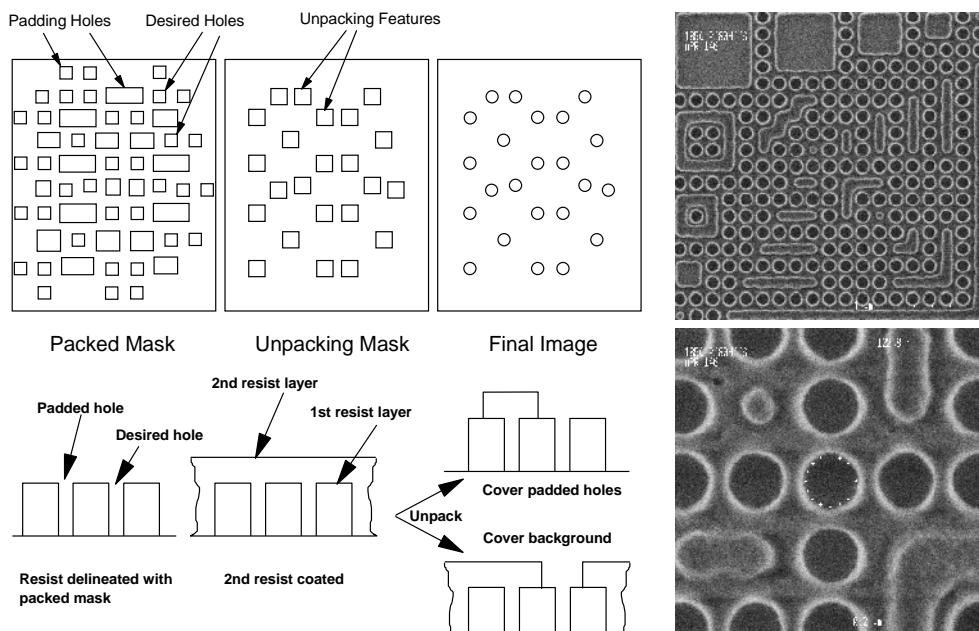


Figure 7.78 Pack-unpack double exposures.

7.8.6 Overcoming the throughput penalty with double imaging

All double-imaging techniques have a potential to lose at least 40% of the throughput at the exposure tool, maybe less if the tool is specifically optimized for double exposure. It is likely that this will happen, unless the other lithographic systems can outperform in imaging and cost.

All double-imaging techniques suffer from the potential loss of exposure-tool throughput by at least 40%. Moreover, the overlay error between the two exposures consumes the overlay error budget. There are proposals to overcome these problems. To reduce the overlay error, the wafer can remain on the wafer stage while the mask is changed. Because of 4X magnification at the mask, the alignment of the mask is more accurate than that of the wafer. However, mask change is time consuming. So one mask change serves the exposure of two wafers without compromise of alignment accuracy. This method is called ABBA, meaning mask A is replicated to wafer 1, then mask B to wafers 2 and 3 and so forth. The highest exposure-tool productivity is achieved with the AAAA...BBBB exposure scheme. Yet, it has the largest overlay error.

To reduce the reticle alignment time for the ABBA scheme, one can implement a long reticles stage that is capable of holding two masks simultaneously and has a scanning distance slightly more than the length of two masks.⁴⁸ The two masks are prealigned to each other on the common stage. Therefore, the wafer needs to be aligned only once to just one of the masks. Between exposures A and B, the reticle stage is simply shifted from one predetermined position to the other. The masks do not have to be unloaded and loaded. The load/unload position of the reticle stage should be at the middle of the stage so that each reticle can complete one scan cycle back and forth without incurring extra time for travel.

To completely remove the double-imaging throughput penalty and overlay-error component, two illuminators are used to cover the two masks on the long reticle stage. A long beamsplitter having a small cross section slightly larger than the width of the scanning slot combines the two beams from the two masks into the imaging lens,⁴⁹ as shown in Fig. 7.79. This way, only the number of condensers and the power of the light source need to be doubled. The increase in cost is much less than using two scanners to maintain the single-exposure throughput.

This technique applies to all double-imaging schemes, with the exception of double patterning. As discussed with Fig. 7.76, because of the accumulation of residual resist exposure, the contrast of the resist must be higher than that used for a single exposure. If this high contrast is developed, the double-mask technique can make the cost of double patterning identical to single patterning. The impact will be tremendous.

7.9 Conclusion on Immersion Lithography

If all of the extension techniques are successful, immersion lithography will be extended, as shown in Fig. 7.80. High-index materials, polarized illumination,

and solid immersion on the mask supports up to a 1.55 NA. Split pitch has the potential to reduce k_1 to 0.15 from 0.3, suggesting a half pitch of 19 nm, an unbelievable $\lambda/10$ dimension. Without high-index materials, the limit is 21 nm. In comparison, EUV lithography with a split pitch from 0.4 k_1 and a 0.3 NA can resolve a 9-nm half pitch, about two generations further from 193-nm immersion.

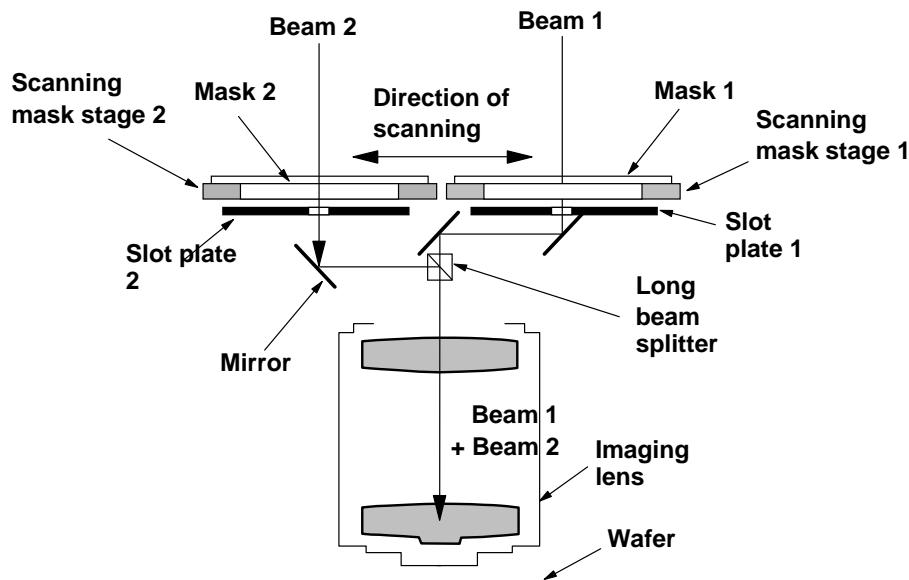


Figure 7.79 Double mask, long beam splitter, imaging lens, and wafer assembly.

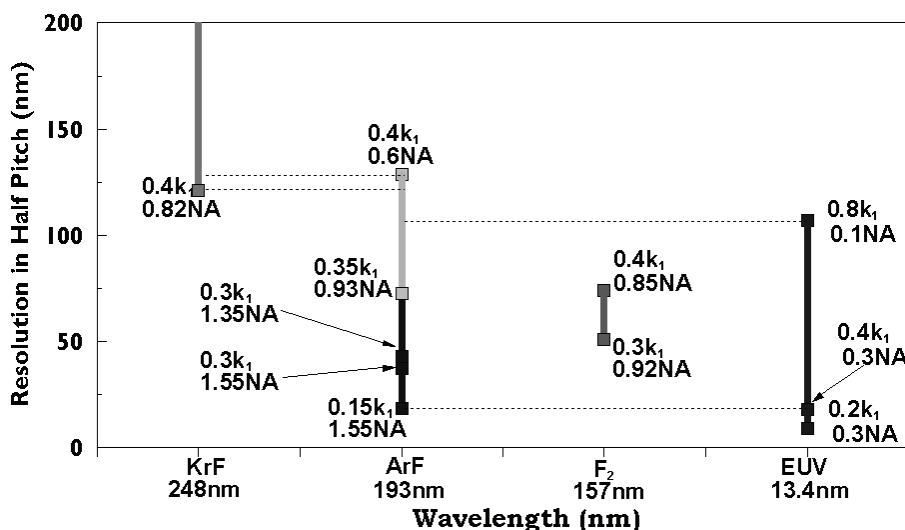


Figure 7.80 Final view of optical lithography

Whether immersion lithography will get there depends not only on the success of new materials and new processes, but also on whether other competing technologies are more economically viable and easier to practice. Another factor to consider is the extendibility of the potential successor. If completely new approaches and infrastructures are needed to move to another technology for one or two generations, there will be insurmountable inertia to stay on what is well understood. Innovation and resources will be directed toward that goal.

The next chapter explores potential successors to optical lithography and examines the cost and infrastructure needed for them.

References

- 1 B. J. Lin, "The k_3 coefficient in nonparaxial λ/NA scaling equations for resolution, depth-of-focus, and immersion lithography," *J. Microlith. Microfab. Microsyst.* **1**, pp. 7–12 (2002).
- 2 H. Kawata, J. M. Carter, A. Yen, and H. I. Smith, "Optical projection lithography using lenses with numerical apertures greater than unity," *Microelectron. Eng.* **9**, pp. 31–36 (1989).
- 3 J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid immersion deep-ultraviolet interferometric lithography," *J. Vac. Sci. Technol. B* **17**(6), pp. 3306–3309 (1999).
- 4 M. Switkes and M. Rothschild, "Immersion lithography at 157 nm," *J. Vac. Sci. Technol. B* **19**, pp. 2353–2356 (2001).
- 5 M. Switkes and M. Rothschild, "Resolution enhancement of 157 nm lithography by liquid immersion," *J. Microlith. Microfab. Microsyst.* **1**, pp. 225–228 (2002).
- 6 B. J. Lin, "The future of subhalf-micrometer optical lithography," *Microcircuit Eng.* **6**, pp. 31–51 (Sept. 1987).
- 7 D. G. Flagello and A. E. Rosenbluth, "Lithographic tolerances based on vector diffraction theory," *J. Vac. Sci. Technol. B* **10**, pp. 2997–3003 (1992).
- 8 D. G. Flagello, T. Milster, A. E. Rosenbluth, "Theory of high-NA imaging in homogeneous thin films," *J. Opt. Soc. Am. A* **13**, pp. 53–64 (1996).
- 9 T. A. Brunner, N. Seong, W. D. Hinsberg, J. A. Hoffnagle, F. A. Houle, M. I. Sanchez, "High numerical aperture lithographic imagery at the Brewster angle," *J. Microlith. Microfab. Microsyst.* **1**, pp. 188–196 (2002).
- 10 B. J. Lin, "Immersion lithography and its impact on semiconductor manufacturing," *J. Microlith. Microfab. Microsyst.* **3**, pp. 377–395 (2004).
- 11 B. J. Lin, "Depth of focus in multilayered media—a long-neglected phenomenon aroused by immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**, pp. 21–27 (2004).
- 12 V. N. Mahajan, "Calculation of primary aberrations: refracting systems," Chapter 5, Sec. 12, p. 318 in *Optical Imaging and Aberrations, Part 1, Ray Geometrical Optics*, SPIE Press, Bellingham, Washington (1998).

- 13 B. J. Lin, "Exposure-defocus forest," *Jap. J. Appl. Phys.* **33**, p. 6756 (1994).
- 14 B. J. Lin, "Partially coherent imaging in two dimensions and the theoretical limits of projection printing in microfabrication," *IEEE Trans. Electron. Dev.* **ED-27**, pp. 931–938 (1980).
- 15 C. A. Mack, "Focus effects in submicron optical lithography, part 4: metrics for depth of focus," *Proc. SPIE* **2440**, pp. 458–471 (1995).
- 16 D. G. Flagello, H. van der Lann, J. van Schoot, I. Bouchoms, and B. Geh, "Understanding systematic and random CD variations using predictive modeling techniques," *Proc. SPIE* **3679**, p. 162 (1999).
- 17 S. Inoue, M. Itoh, M. Asano, and K. Okumura, "Desirable reticle flatness form focus deviation standpoint in optical lithography," *J. Microlith. Microfab. Microsyst.* **1**(3), p. 307 (2002).
- 18 S. Sethi, A. Flores, P. McHale, R. Booth, S. Graca, S. Frezon, and C. Fruga, "Improving the backend focus budget for 0.5 μ m lithography," *Proc. SPIE* **2440**, p. 633 (1995).
- 19 B. W. Smith and J. Cashmore, "Challenges in high NA, polarization, and photoresists," *Proc. SPIE* **4691**, p. 11 (2002).
- 20 B. J. Lin, "Simulation of optical projection with polarization-dependent stray light to explore the difference between dry and immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**, pp. 9–20 (2004).
- 21 B. J. Lin, "Off-axis illumination—working principles and comparison with alternating phase-shifting masks," *Proc. SPIE* **1927**, p. 89 (1993).
- 22 D. Gabor, "Light and Information," in *Astronomical Optics and Related Subjects*, Z. Kopal, Ed., North-Holland Publishing Co., p. 17 (1956).
- 23 D. Gabor, "Optical transmission," in *Information Theory*, C. Cherry, Ed., Butterworths, London, p. 26 (1956).
- 24 M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 6th Edition, Cambridge Univ. Press, p. 40 (1998).
- 25 M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 6th Edition, Cambridge Univ. Press, p. 43 (1998).
- 26 B. J. Lin, "Electromagnetic near-field diffraction of a medium slit," *J. Opt. Soc. Am.* **62**, pp. 977–981 (1972).
- 27 B. J. Lin, "Signamization," *Proc. SPIE* **2726**, p. 71 (1996).
- 28 D. Flagello, J. Mulkens, B. Streefkerk, and P. Graeupner, "The benefits and limitations of immersion lithography," *J. Microlith. Microfab. Microsyst.* **1**, pp. 104–114 (2004).
- 29 S. Owa and H. Nagasaka, "Advantage and feasibility of immersion lithography," *J. Microlith. Microfab. Microsyst.* **1**, pp. 97–103 (2004).

- 30 B. J. Lin, "Semiconductor foundry, lithography, and partners," *Proc. SPIE* **4688**, p. 11 (2002).
- 31 J. H. Burnett and S. G. Kaplan, "Measurement of the refractive index and thermo-optic coefficient of water near 193 nm," *J. Microlith. Microfab. Microsyst.* **1**, pp. 68–72 (2004).
- 32 R. Gupta, J. H. Burnett, U. Griesmann, and M. Walhout, "Absolute refractive indices and thermal coefficients of fused silica and calcium fluoride near 193 nm," *Appl. Optic.* **37**, pp. 5964–5968 (1998).
- 33 A. Suzuki, "Immersion lithography update," *International Sematech 2nd Immersion Workshop* (July 2003).
- 34 P. L. Marston, "Light scattering from bubbles in water," *Oceans 89 Proc.* **4**, pp. 1186–1193 (1989).
- 35 T. S. Gau, C. K. Chen, and B. J. Lin, "The image characterization of bubbles in water for 193-nm immersion lithography," *J. Microlith. Microfab. Microsyst.* **1**, pp. 61–67 (2004).
- 36 B. J. Lin, "4X / 5X mask considerations for the future," *ASML 157nm Users' Forum* (Sept. 2000).
- 37 B. J. Lin, H. T. Lin, and H. C. Hsieh, "A device and method for providing wavelength reduction with a photomask," Patent Pending.
- 38 T. Hirayama, "Resist and cover material investigation for immersion lithography," *International Sematech 2nd Immersion Workshop*, Almaden, California (July 2003).
- 39 A. Otoguro, J. Santillan, T. Itani, K. Fujii, A. Kagayama, T. Nakano, N. Nakayama, H. Tamatani, and S. Fukuda, "Development of high refractive index fluids for 193nm immersion lithography," *International Sematech 2nd International Immersion Symposium*, Bruges, Belgium (Sept. 2005).
- 40 B. J. Lin, "Signamization," *Proc. SPIE* **2726**, pp. 71–81 (1996).
- 41 J. H. Chen, L. J. Chen, T. Y. Fang, T. C. Fu, L. H. Shiu, Y. T. Huang, N. Chen, D. C. Oweyang, M. C. Wu, S. C. Wang, C. H. Lin, C. K. Chen, W. M. Chen, T. S. Gau, B. J. Lin, R. Moerman, W. Gehoel-van Ansem, E. van der Heijden, F. de Johng, D. Oorschot, H. Boom, M. Hoogendorp, C. Wagner, and B. Koek, "Characterization of ArF immersion process for production," *Proc. SPIE* **5754**, pp. 13–22 (2005).
- 42 C. H. Lin, S. C. Wang, K. S. Chen, C. Y. Chang, T. C. Wu, M. C. Wu, M. T. Lee, J. H. Chen, S. W. Chang, Y. S. Yen, Y. H. Chang, T. C. Fu, T. S. Gau, and B. J. Lin, "193-nm immersion lithography for 65-nm and below," *International Sematech 2nd International Symposium on Immersion Lithography*, Bruges, Belgium (Sept. 2005).
- 43 F.-J. Liang, H. Chang, L.-H. Shiu, C.-K. Chen, L.-J. Chen, T.-S. Gau, and B. J. Lin, "Immersion defect reduction (1)—Analysis of water leaks in an immersion scanner," *Proc. SPIE* **6520**, 65204U (2007).

- 44 S. Peng et al., "New developments in second generation 193nm immersion fluids for lithography with 1.5 numerical aperture," *International Sematech 2nd International Symposium on Immersion Lithography*, Bruges, Belgium (Sept. 2005).
- 45 Y. Wang et al., "Material design for highly transparent fluids of the next generation ArF immersion lithography," *International Sematech 2nd International Symposium on Immersion Lithography*, Bruges, Belgium (Sept. 2005).
- 46 Y. Inui et al., "Fluoride single crystals grown by the CZ method," *International Sematech 2nd International Symposium on Immersion Lithography*, Bruges, Belgium (Sept. 2005).
- 47 D. C. Owe-Yang, S. S. Yu, H. Chen, C. Y. Chang, B. C. Ho, C. H. Lin, and B. J. Lin, "Double exposure for the contact layer of the 65-nm node," *Proc. SPIE* **5753**, pp. 171–180 (2005).
- 48 B. J. Lin, "Two-mask step-and-scan aligner," U.S. Patent 6,777,143 (2005).
- 49 B. J. Lin, "Mask superposition for multiple exposures," U.S. Patent 7,394,080 (2008).

Chapter 8

Outlook and Successors to Optical Lithography

Optical lithography has been the workhorse of semiconductor fabrication since the inception of integrated circuits. The lensless proximity printing system gradually gave way to projection-printing systems, and one-to-one replication systems became reduction systems. It took this latest form from the 0.15-NA 436-nm g-line lens, featuring resolution over 2 μm with a 0.8 k_1 factor, all of the way through raising the NA until the lens became too expensive to build at that time, reducing wavelength to reposition the NA for the next round of increases, and lowering k_1 whenever the pace of NA and wavelength changes are behind the circuit shrinking roadmap, as shown in Fig. 8.1.

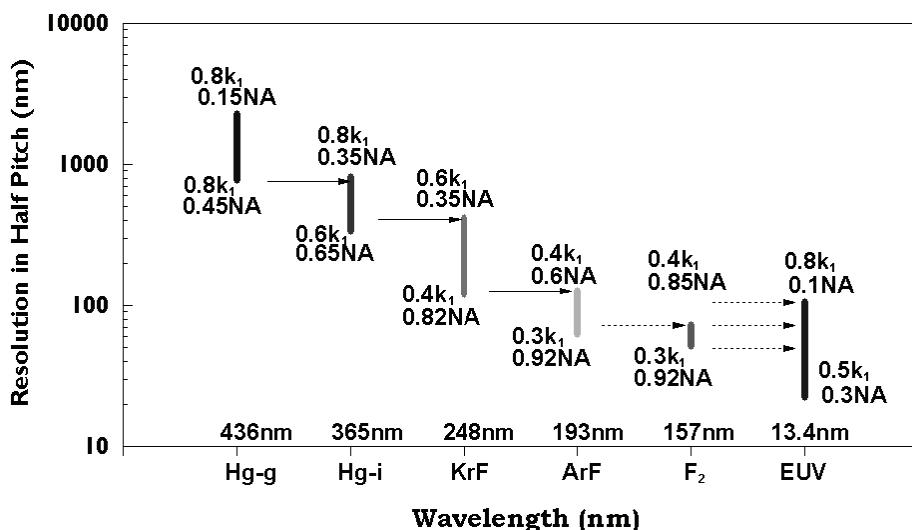


Figure 8.1 Evolution of projection optical lithography from a 0.15-NA g-line.

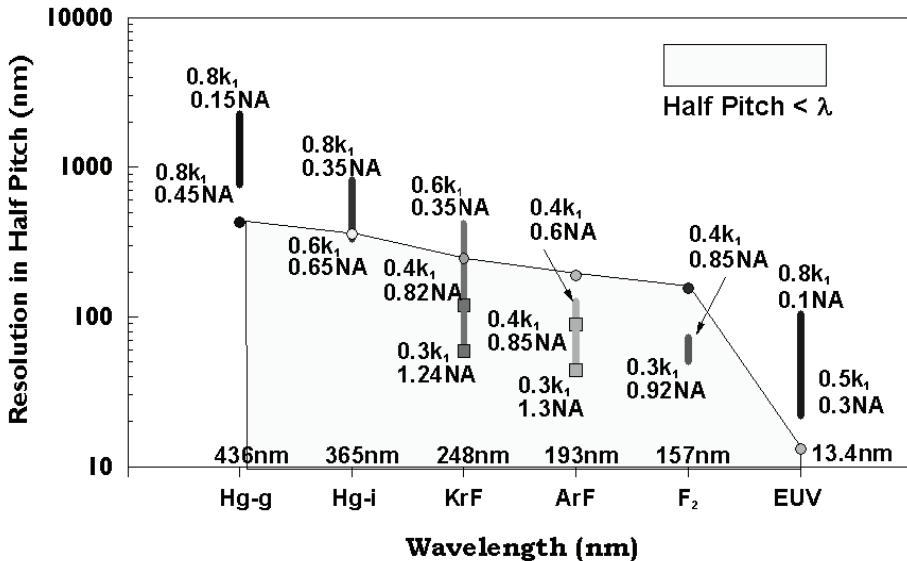


Figure 8.2 Evolution of optical lithography, including immersion lithography.

To date, NA and k_1 have been extended in full at 193 nm. Using 157-nm or 13.4-nm wavelengths is not easy. As a result, 193-nm water-immersion lithography extends lithography and changed its evolution path, as shown in Fig. 8.2. With the potential NA increased to the 1.3 regime, and k_1 lowering to 0.3, 193-nm water-immersion exposure systems already supercede what 157-nm dry systems can do. The vertical position of the circular dots indicates the actinic wavelength to compare with the delineated half pitch. Starting from 248 nm, the half-pitch resolution has been pushed below the actinic wavelength, more so when NA is further increased and k_1 further lowered. Only when 13.4 nm is made practical, where a high NA and a low k_1 are extremely difficult, the resolution will then become larger than the wavelength. Incidentally, 13.4 nm is an order of magnitude smaller than the ArF wavelength in water, which has a refractive index of 1.44.

8.1 Outlook of Optical Lithography

The tradeoff between resolution and DOF, as a function of λ and NA, is governed by Eqs. (4.3) and (4.4). Constant-NA lines at varying λ and constant- λ lines at varying NA are plotted logarithmically in the resolution-DOF space in Fig. 8.3, with resolution as the x axis and DOF as the y axis. Logarithmic scale was chosen to better cover the orders of magnitude in resolution and DOF. Initially, when the paraxial scaling Eqs. (4.1) and (4.2) were used, the constant-NA and constant- λ lines were straight lines in the logarithmic scale. The rhombic shape of the line clusters looks like a galaxy; together they include every possible

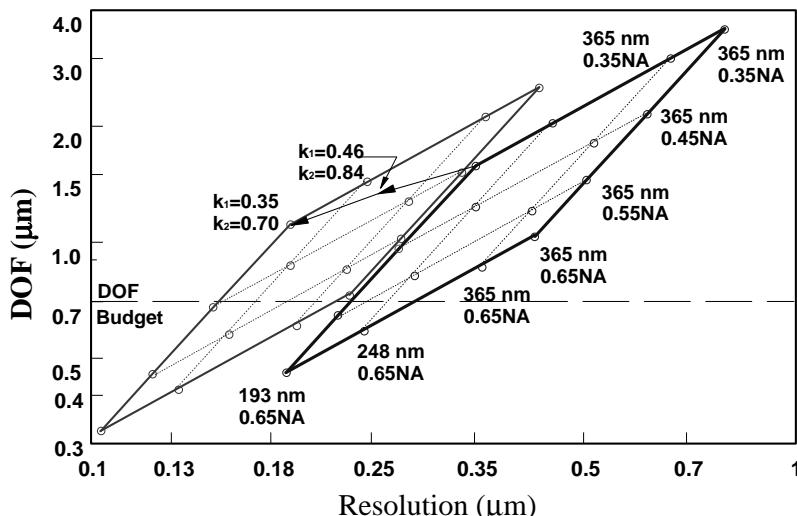


Figure 8.3 Optical lithography galaxy in the k_2 era. moving forward by k_1 reduction.

combination of NA and λ in optical lithography, hence the name optical lithography galaxy. At a high NA, the straight lines are bent toward a lower DOF. The scaling equations work strictly for transverse electric (TE) polarization only. The stray light induced by the transverse magnetic (TM) component affects the DOF (as a function of NA) as a function of the refractive indices in the coupling medium and in the resist, as was discussed in Sec. 7.4. The galaxies represent imaging only at the most favorable TE polarization.

8.1.1 Optical lithography galaxy for logic gates

The optical lithography galaxies for a logic gate are shown in Fig. 8.4. Logic gate patterns are typified with unequal lines and spaces within the pitch. With a positive resist, the opaque lines in the mask are smaller than the openings, thus smaller than half of the given pitch. The smallest pitch has a ratio of 1:1.5, which is resist line/resist space. The galaxies of three imaging schemes are plotted: namely, attenuated phase-shifting mask (AttPSM) and extreme ultraviolet (EUV) with 1:1.5, 1:2, and 1:3 line-to-space ratios, as well as the alternating phase-shifting mask (AltPSM) imaging scheme covering only the 1:1.5 and 1:2 pitches. The DOF is defined as the vertical diameter of the largest common elliptical E-D window of all of the pitches in each imaging scheme, using 8% EL, 5% stray light, unpolarized light, $n = 1.75$ for the resist, and 10% mask reflectivity. After picking a wavelength and a feature size, the NA, illumination, and bias of each feature are optimized for the largest DOF using the Signamizer¹ modified for polarized illumination, stray light, mask reflectivity, and resists.² Here, 193 nm and 65 ± 6.5 nm are the wavelength and feature size, respectively, for a 6% AttPSM. The optimized NA is 0.787, $\sigma_{out} = 0.76$, $\sigma_{in} = 0.38$. The bias of the 1:1.5 feature is -11 nm; 1:2 feature, -6 nm; 1:3 feature, -2 nm; resulting in a

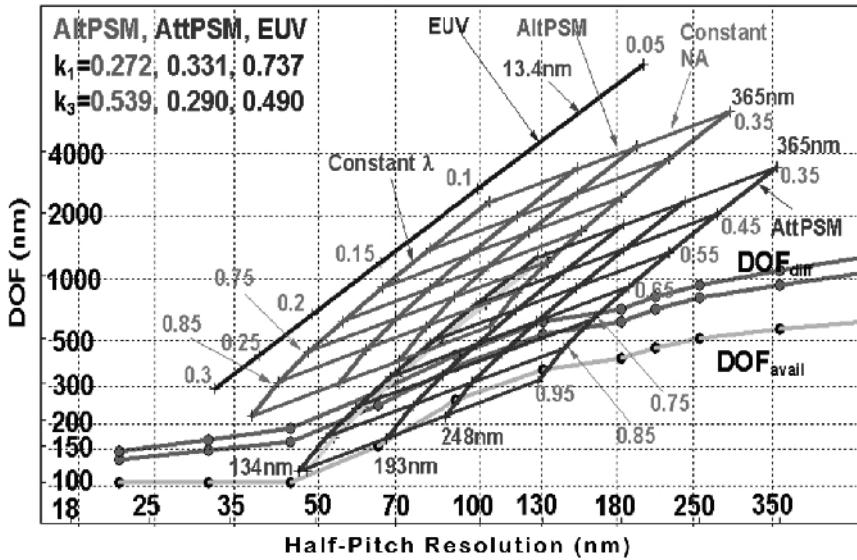


Figure 8.4 Lithography galaxies based on AltPSM 1:1.5 to 1:2, AttPSM 1:1.5 to 1:3, and EUV 1:1.5 to 1:3.

292-nm DOF. The resolution and DOF scaling coefficients k_1 and k_3 are 0.331 and 0.290, calculated from Eqs. (4.3) and (4.4). All combinations of λ and NA are constructed from this set of k_1 and k_3 for each imaging scheme. Note that there is another 134-nm constant- λ line slightly below the 134-nm line in the galaxy. The latter is plotted as if the 134-nm wavelength was a wavelength in vacuum, using straightforward scaling with k_1 and k_3 . With the former 134-nm line, the short wavelength is achieved by water immersion of a 193-nm vacuum wavelength, and its DOF performance is slightly worse than that of the equivalent vacuum wavelength, as was discussed in Sec. 7.4.2.5. The effect was illustrated in Fig. 7.33.

For EUV at 13.4 nm, the critical dimension (CD) used to evaluate k_1 and k_3 is 32 ± 3.2 nm. The feature bias is -1 , -0.5 , and 0 nm for the 1:1.5, 1:2, and 1:3 ratios, respectively. Since DOF drops insignificantly from NA = 0.15 to 0.25, we chose a high k_1 at 0.737 for better CD control. The corresponding k_3 is 0.490. The binary intensity mask (BIM) and axial illumination at $\sigma = 0.82$ are used. Except

Table 8.1 Parameters of the three imaging schemes for the polylayer.

	CD and tol (nm)	Exposure latitude	L:S	Bias	Illum.	Pol.	Stray light	Mask reflec.	n_{resist}	$n/\sin\theta/\text{DOF}(\text{nm})$
AttPSM dry	65 ± 6.5	8%	1:1.5, 1:2, 1:3	-10, -6, -2	$\sigma_{\text{out}}=0.76$ $\sigma_{\text{in}}=0.38$	unpol.	5%	10%	1.75	1/0.787/292
AttPSM immersion	65 ± 6.5	8%	1:1.5, 1:2, 1:3	-9.4, -4.3 -1	$\sigma_{\text{out}}=0.76$ $\sigma_{\text{in}}=0.38$	unpol.	5%	10%	1.75	1.44/0.566/445
AltPSM	32 ± 3.2	5%	1:1.5, 1:2	-17, -7	$\sigma=0.3$	TE	5%	10%	1.75	1.44/0.911/246
EUV	32 ± 3.2	8%	1:1.5, 1:2, 1:3	-1, -0.5, 0	$\sigma=0.82$	unpol.	15%	1%	1.1	1/0.247/424
EUV	22 ± 2.2	8%	1:1.5, 1:2, 1:3	-2, -1, 0	$\sigma=0.82$	unpol.	15%	1%	1.1	1/0.203/358

for mask reflectivity at 1%, stray light at 15%, and a resist refractive index at 1.1, all other parameters used for the Signamizer simulation are identical to those of AttPSM. Here, mask reflectivity refers to the reflectivity of the absorber instead of the substrate. The parameter impacts the contrast of the image—the lower the reflectivity, the better. Since a single wavelength has been chosen for EUV, the galaxy collapses to a constant- λ line.

For AltPSM, 193-nm immersion lithography imaging is used. This exemplifies a low- k_1 scheme pushing optical lithography to its very limit. Only two features are used for the common window. The EL is reduced to 5%. TE polarization is used. The CD used to evaluate k_1 and k_3 is 32 ± 3.2 nm. The feature bias is -17 and -7 nm for the 1:1.5 and 1:2 features, respectively. The refractive index of the immersion fluid is 1.44. Stray light level is suppressed to 5%. The axial illumination σ is 0.3. $k_1 = 0.272$. $k_3 = 0.538$. The parameters used for these three imaging schemes are summarized in Table 8.1.

In the galaxy plot, two DOF lines are given. $\text{DOF}_{\text{avail}}$ refers to the available DOF. A detailed discussion on $\text{DOF}_{\text{avail}}$ was given in Sec. 7.3.5. $\text{DOF}_{\text{diffrac}}$ refers to the DOF evaluated from diffraction. The points in the galaxy must be above $\text{DOF}_{\text{diffrac}}$ to meet the DOF requirement. Also, from the same section, $\text{DOF}_{\text{diffrac}}$ is different for dry and immersion systems because of different optical thicknesses from the resist of an identical physical thickness. Therefore, two $\text{DOF}_{\text{diffrac}}$ curves are shown. The one above is for immersion systems. These DOF criteria are not constant. They decrease as a result of a reduced resist thickness and an improvement in the exposure tool technology, which is based on discussions in Sec. 8.1. The AttPSM galaxy bends at a high NA farther down than what is shown.

Figure 8.5 further explores the extension of the strong resolution-enhancement technique (RET) and EUV. Pitch splitting (see Sec. 7.8.4) is applied to AltPSM, enabling $k_1 = 0.271/2$ and k_3 remaining at 0.538. Further drops in k_1 for EUV are also shown in Fig. 8.5. The scaling coefficients are evaluated using $\text{CD} = 22 \pm 2.2$ nm. The bias to the 1:1.5 and 1:2 features is -2 and -1 nm, respectively. The DOF is 358 nm at $\text{NA} = 0.3$, leading to $k_1 = 0.417$ and $k_3 = 0.278$; with pitch splitting, $k_1 = 0.209$. These galaxies indicate that, from the perspective of imaging and DOF, there still is room for photon lithography to move through many generations below 65 nm.

8.1.2 Optical lithography galaxy for contact holes

The optimum printing conditions are dependent on the feature type. For contact holes, we used 100 ± 10 nm for 193-nm dry systems. All parameters are shown in Table 8.2. Most of them are similar to those in Table 8.1, except that axial illumination is used. The stray light for contact holes is reduced because of more dark areas on the mask compared with the polylayer.

The galaxies for contact holes are shown in Fig. 8.6. They operate at higher k_1 , and, hence, are shifted down and to the right with respect to the galaxies for the polylayer.

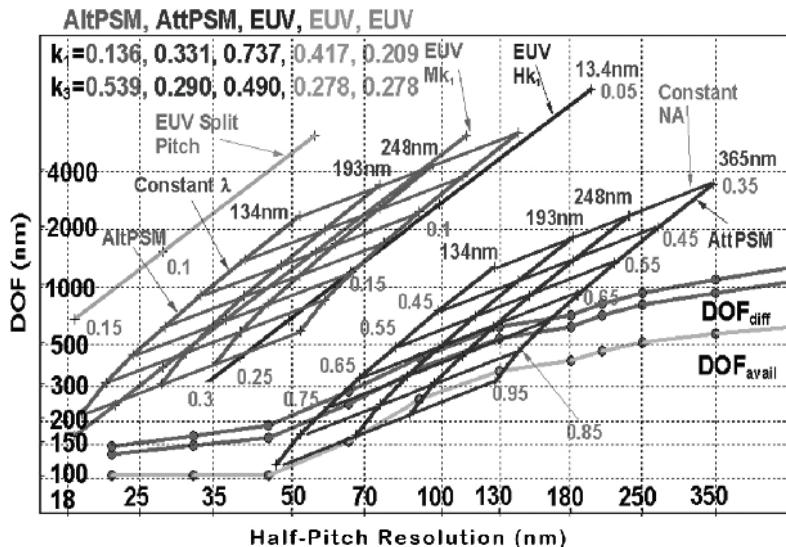


Figure 8.5 Lithography galaxies based on AltPSM 1:1.5 to 1:2, AttPSM 1:1.5 to 1:2 with pitch splitting, and EUV 1:1.5 to 1:3 at two k_1 values and pitch splitting.

Table 8.2 Parameters of the three imaging schemes for the contact layer.

	CD and tol (nm)	Exposure latitude	Hole/Sep.	Bias	Illum.	Pol.	Stray light	Mask reflec.	n _{resist}	n/sinθ/DOF(nm)
AttPSM dry	100±10	8%	1:1,1:1.5, 1:2,1:3,1:4	0.6, 7,10,11	σ=0.78	unpol.	3%	10%	1.75	1/0.635/ 393
AttPSM immersio	70±7	8%	1:1,1:1.5, 1:2,1:3,1:4	0,4,5,6,7,7	σ=0.78	unpol.	3%	10%	1.75	1.44/0.658/ 268
AltPSM	50±5	5%	1:1,1:1.5	0,-5.5	σ=0.3	TE	3%	10%	1.75	1.44/0.879/ 329
EUV	40±4	8%	1:1,1:1.5, 1:2,1:3	0,2,2,2	σ=0.78	unpol.	10%	1%	1.1	1/0.24/ 530

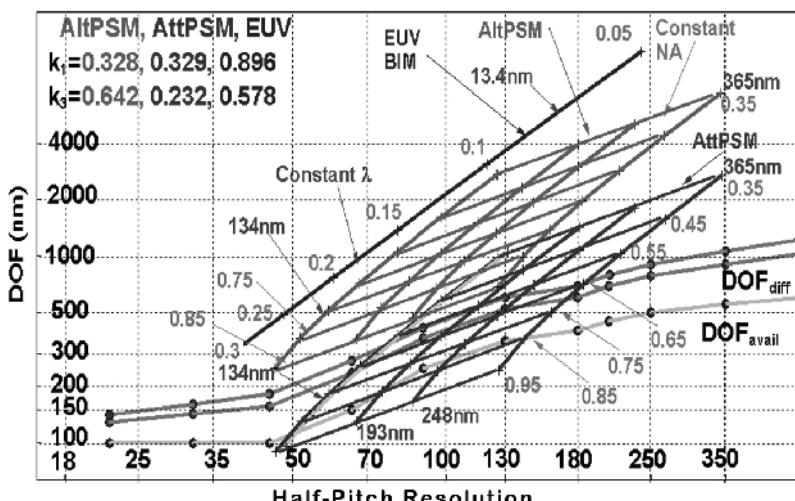


Figure 8.6 Lithography galaxies for contact holes based on AltPSM 1:1 to 1:1.5, AttPSM 1:1 to 1:5, and EUV 1:1 to 1:4.

Table 8.3 Parameters of the three imaging schemes for the metal layer.

	CD and tol (nm)	Exposure latitude	L.S	Bias	Illum.	Pol.	Stray light	Mask reflec.	n_{resist}	$n/\sin\theta/\text{DOF}(\text{nm})$
AttPSM dry	70±7	8%	1:1;1:1.5;1:2; 1.5:1;2:1	0,4,5,7, 13,13	$\sigma_{\text{out}}=0.8$ $\sigma_{\text{in}}=0.4$	unpol.	5%	10%	1.75	1/0.881/ 266
AttPSM immersion	50±5	8%	1:1;1:1.5;1:2; 1.5:1;2:1	0,3,5, 8,12	$\sigma_{\text{out}}=0.76$ $\sigma_{\text{in}}=0.38$	unpol.	5%	10%	1.75	1.44/0.866/ 181
AltPSM	50±5	5%	1:1;1:1.5;1:2; 1.5:1;2:1	0,3,5,5,2	$\sigma=0.3$	TE	5%	10%	1.75	1.44/0.79/ 311
EUV	30±3	8%	1:1;1:1.5;1:2; 1.5:1;2:1	0,0,0,2,0	$\sigma=0.82$	unpol.	15%	1%	1.1	1/0.196/ 241

8.1.3 Optical lithography galaxy for equal lines and spaces

Patterns of the metal layer are characterized with five line-space ratios, namely 2:1, 1.5:1, 1:1, 1:1.5, 1:2, and three pitches. Their galaxy parameters are shown in Table 8.3. The galaxies are shown in Fig. 8.7. A medium k_1 is used for EUV, since metal imaging is less demanding. Its location is shared with that of AltPSM with pitch splitting, which is not explicitly plotted. EUV with a pitch splitting extends over 193-nm immersion.

8.2 EUV Lithography

8.2.1 Introduction

Optical lithography in the regime of $k_1 = 0.3$ is very complicated. The image contrast is low, leading to a significant mask error enhancement factor (MEEF). For contact hole patterning, the MEEF can be more than 4. Line ends can have MEEF as high as 10. In addition, the shape of patterns at low k_1 is rounded out and requires a significant amount of optical proximity correction (OPC) to keep

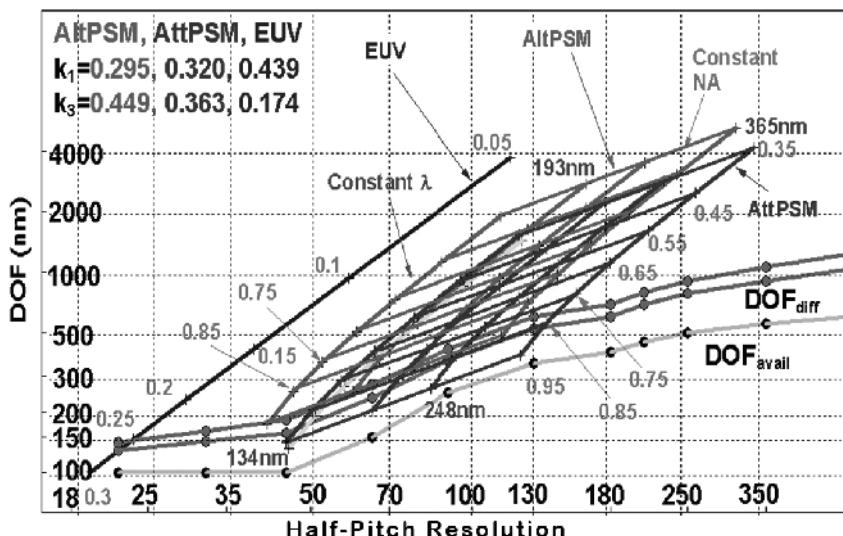


Figure 8.7 Lithography galaxies for the metal layer based on AltPSM 2:1 to 1:2, AttPSM 2:1 to 1:2, and EUV 2:1 to 1:2.

them usable. Extreme-ultraviolet lithography^{3,4} (EUVL) uses a 13.4-nm wavelength. Recently, it has migrated to 13.5 nm. Either wavelength offers an order of magnitude reduction in wavelength from the water-immersed ArF wavelength. It presents an opportunity to bring k_1 back to above 0.5. No wonder EUVL has attracted large research and development efforts worldwide. The size of the EUVL development effort has dwarfed two other well-known but unsuccessful lithography development efforts, namely 157-nm lithography and x-ray proximity printing. Take the 32-nm half pitch. At the water-immersed ArF wavelength of 134 nm and $\sin\theta = 0.95$, k_1 is 0.227. Turning to EUV at $\sin\theta = 0.25$, k_1 becomes 0.597. Initially, soft x ray⁵ was used to identify this regime of lithography. Later, the name EUV was adopted to indicate that this lithography is an extension of UV lithography, not x-ray lithography.

With such a dramatic drop in wavelength, the imaging system is substantially different from existing systems. First, any substance heavily absorbs EUV light, even gases; therefore, the optical path must be in vacuum. Second, following its heavy absorption characteristics, there is no transmitting material; EUV imaging depends on reflection. However, the EUV reflectivity is only on the order of a few percent. Reflectivity is built up with a multilayer stack consisting of 40- to 50-layer pairs of Mo and Si, to achieve 65~70% reflectivity, enabling an all-reflective optical system that includes the illumination optics, the mask, and the imaging optics. Figure 8.8 shows an EUV imaging system developed at Sandia

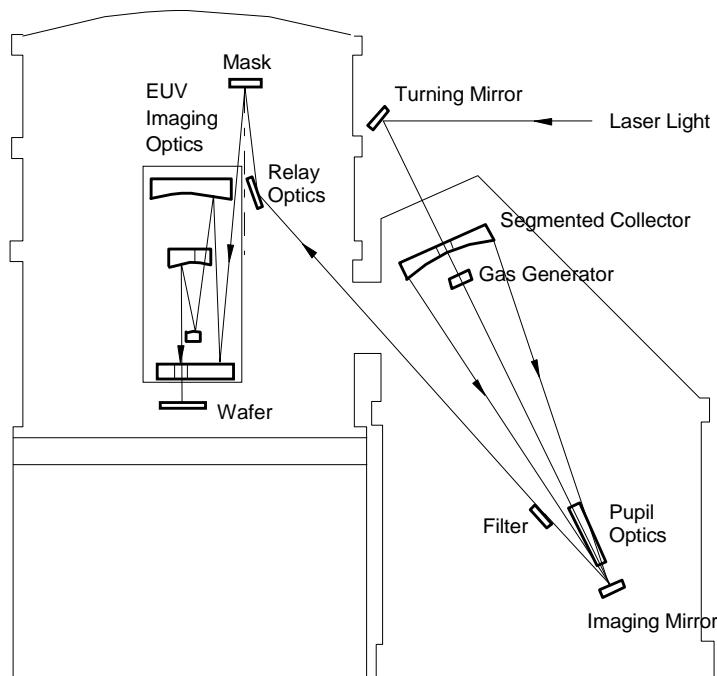


Figure 8.8 EUVL system developed at the Sandia National Laboratories and the Lawrence Livermore National Laboratory.

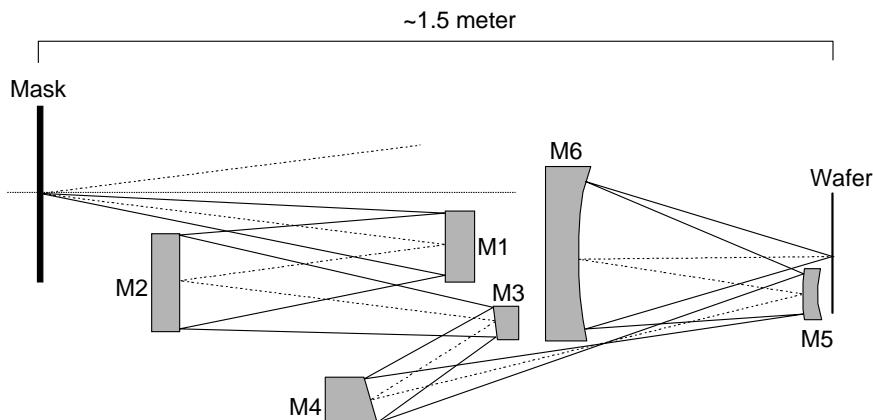


Figure 8.9 EUVL projection optics developed by Zeiss for the ASML α tool.

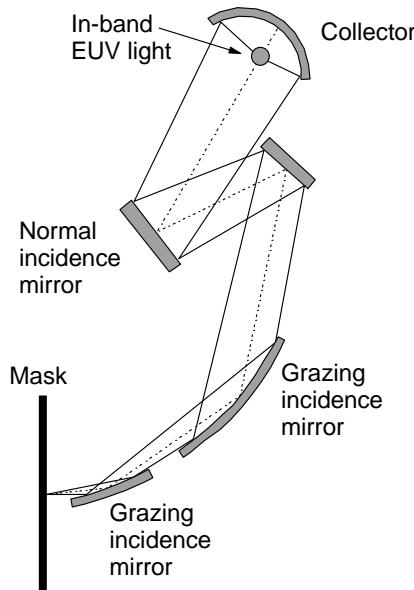


Figure 8.10 Example of an EUVL illuminator for the ASML α tool.

National Laboratories and Lawrence Livermore National Laboratory.^{6,7} A YAG laser beam hits a Xe-cluster target to generate the 13.4-nm light. The light is collected to go through the pupil optics, imaging optics, filter, and relay optics to illuminate the mask. Due to the lack of a beamsplitter, the incident and reflection beams are oblique, making an angle in the order of 6 deg with respect to the direction perpendicular to the mask. The mask pattern undergoes four reflections in the imaging lens. The imaging light passes through a blocking slit to expose the resist on the wafer. The NA of this system is 0.1. Zeiss⁸ has proposed a 0.25-NA system consisting of six reflecting surfaces, as shown in Fig. 8.9. Together with the two mirrors in the illuminator shown in Fig. 8.10, reflective mask, and a light-collecting mirror, there are 10 normal-incidence reflecting surfaces.

Similar to an optical scanner, the imaging field takes the form of a slit that has a length of 26 mm. The entire $26 \times 33\text{-mm}^2$ field is filled by scanning. Unlike the straight-slot field of an optical scanner, the slit is curved.

8.2.2 EUV sources

An EUV scanner needs to be economical for manufacturing a semiconductor. It is anticipated that the scanner itself will cost more than a 193-nm immersion scanner. Therefore, it is very important to keep the throughput and running cost of EUV scanners comparable to those of conventional scanners, if not less. Hence, the EUV source and resist sensitivity are required to sustain a 100-wafers-per-hour (wph) throughput, if not more. This source power requirement is dictated by the power P required to expose the resist, the reflectivity of the mirrors, and the number of reflecting surfaces, etc.

Let us assume that $P \text{ mJ/cm}^2$ is needed to expose the resist, 700 cm^2 is needed to expose a 300-mm wafer, and 65% of the time each wafer spends in the scanner is for exposure, then $30P \text{ mW}$ of in-band EUV power is needed per wafer to sustain the 100-wph throughput. To reach $30P \text{ mW}$, the EUV in-band light generated at the source, spreading out to 2π steradian (sr), is $26.7P \text{ W}$. It first goes through a membrane filter that absorbs 50% of the light output. The collector mirror sustains a solid angle that is 70% of 2π . Its reflectivity is assumed to be 68%, identical to the other normal-incidence multilayer-coated mirrors. There are two of such mirrors for the condenser and six for the projection optics, assuming the configuration of ASML's α tool.⁸ Mask reflectivity is also assumed to be 68%. The reflectivity of the grazing incidence mirrors⁹ is taken to be 80%. Though not specified in Ref. 9, following Ref. 10, the integrator in the illuminator most likely consists of two multifaceted mirrors at a 60% reflectivity. We also incorporate a 10% loss for path length attenuation and another 10% loss for bandwidth mismatch in the illuminator as well as in the projection optics, as estimated in Ref. 10. A typical assumption for EUV resist sensitivity is 5 mJ/cm^2 , leading to EUV output of 134 W, according to the foregoing calculations. It is very close to the 115 W given by Meiling et al.¹⁰ The EUV power level at each component is shown in Fig. 8.11. The power levels are marked in Fig. 8.12 at key components in a composite optical layout from Figs. 8.9 and 8.10.

Table 8.4 lists the power incident on the light collector, light integrator, mask, mirrors, etc., to support a throughput of 100 300-mm wafers per hour, using the parameters for deriving the $26.7P \text{ W}$ requirement for an in-band EUV $2\pi \text{ sr}$ output at the source. The occurrences of $> 20\text{-W}$ incident power are shaded to indicate caution needed for special heat handling.

Table 8.5 lists the power absorbed into the EUV components. Absorbed power above 10 W is shaded. Resist sensitivity up to 50 mJ/cm^2 is used because 30 and 50 mJ/cm^2 are used for line-space and contact-hole delineation, respectively, for 193 nm. These low sensitivities are required to ensure acceptable resist linewidth roughness.

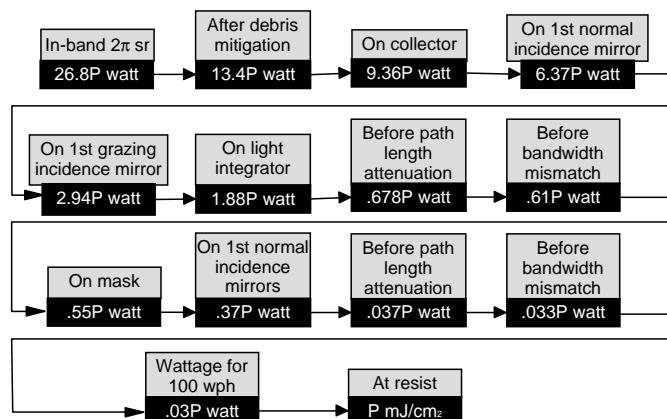


Figure 8.11 Power level at each component.

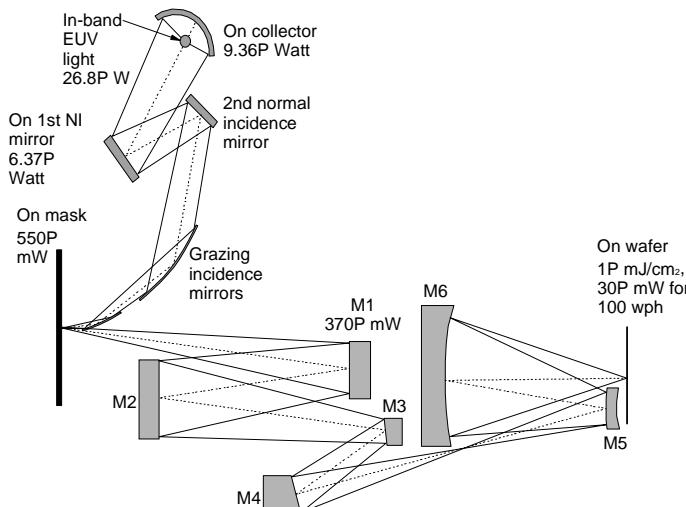


Figure 8.12 EUV power indicated on the optical layout combining the illuminator of Fig. 8.10 and the projection optics of Fig. 8.9.

Table 8.4 Power incident on EUV components for 100 wph at various resist sensitivities, according to 26.74P W.

resist dosage (mJ/cm ²)	1	2	5	10	20	30	50
wattage for 100 wph	0.030	0.060	0.15	0.30	0.60	0.90	1.50
into 1st NI mirror (W)	0.12	0.24	0.60	1.20	2.39	3.59	5.98
into mask (W)	0.18	0.35	0.88	1.76	3.52	5.27	8.79
into light integrator (W)	0.75	1.51	3.77	7.53	15.1	22.6	37.7
into 1st GI mirror (W)	0.59	1.18	2.94	5.89	11.8	17.7	29.4
into 1st NI mirror (W)	2.04	4.07	10.2	20.4	40.7	61.1	102
into collector (W)	3.00	5.99	15.0	30.0	59.9	89.9	150
into debris mitigator	13.4	26.7	66.9	134	267	401	669

Table 8.5 Power absorbed into EUV components at various resist sensitivities to support 100 wph.

resist dosage (mJ/cm ²)	1	2	5	10	20	30	50
wattage for 100 wph	0.030	0.060	0.15	0.30	0.60	0.90	1.50
before BW mismatch (W)	0.033	0.066	0.17	0.33	0.66	1.00	1.66
before propag. atten. (W)	0.037	0.074	0.18	0.37	0.74	1.11	1.85
before 6 NI mirrors (W)	0.37	0.75	1.87	3.73	7.47	11.2	18.7
on mask (W)	0.55	1.10	2.75	5.49	11.0	16.5	27.5
before BW mismatch (W)	0.61	1.22	3.05	6.10	12.2	18.3	30.5
before propagation (W)	0.68	1.36	3.39	6.78	13.6	20.3	33.9
before light integrator (W)	1.88	3.77	9.42	18.8	37.7	56.5	94.2
before 2 GI mirrors (W)	2.94	5.89	14.7	29.4	58.9	88.3	147
before 2 NI mirrors (W)	6.36	12.7	31.8	63.6	127	191	318
on collector (W)	9.4	18.7	46.8	93.6	187	281	468
spread into 2pi sr (W)	13.4	26.7	66.9	134	267	401	669
in-band 2pi sr (W) (before debris mitigation)	26.7	53.5	134	267	535	802	1,337

EUV light is generated primarily by discharge-produced plasma (DPP)^{11,12} or laser-produced plasma (LPP).¹³ The schematics of a DPP EUV source from Ref. 12 are shown in Fig. 8.13. The main features of these DPP sources are the anode and the cathode. Electric energy is stored between them. When a discharge is triggered, it produces the plasma to emit EUV light.

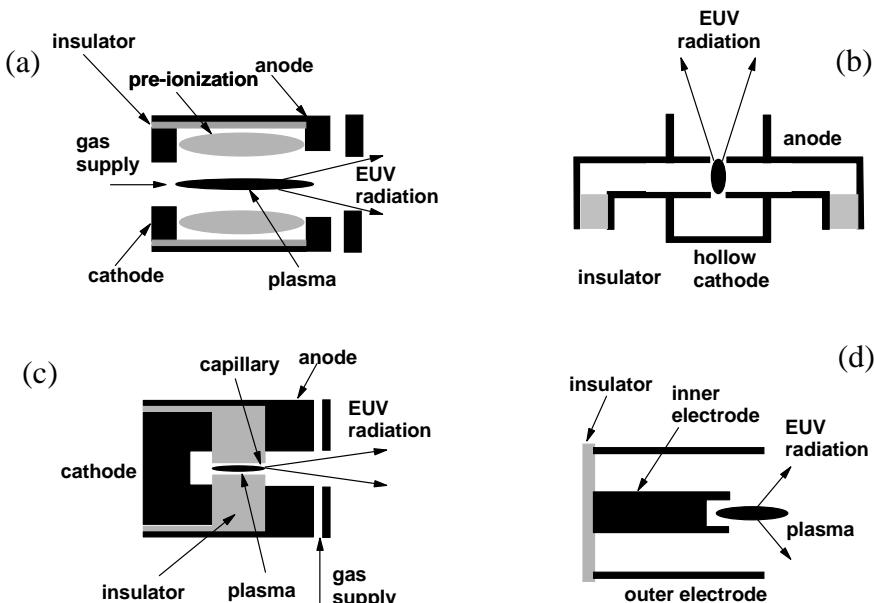


Figure 8.13 A schematic of a DPP EUV source: (a) Z-pinch with preionization, (b) hollow-cathode plasma-ignited Z-pinch, (c) gas-filled capillary, and (d) the plasma focus.

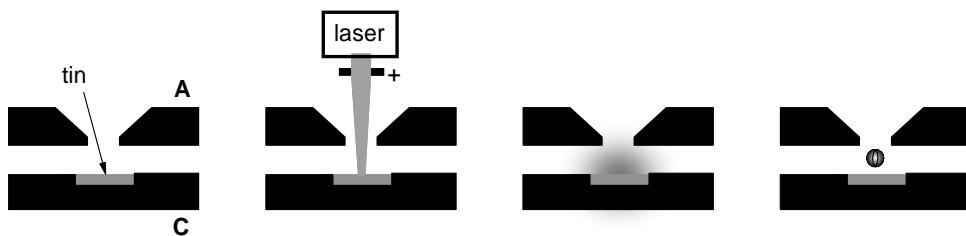


Figure 8.14 Schematic of laser-initiated DPP EUV source: A = anode, C = cathode.

Figure 8.14 shows a special case of a DPP source according to Ref. 13. The laser beam evaporates the tin target into a conducting vapor cloud that short-circuits the anode and cathode to create a discharge to energize the tin for EUV output. LPP sources use a pulsed laser to generate hot and dense plasma to produce EUV light. No electric energy is stored to energize the plasma. The laser beam supplies the energy. Therefore, much more power is needed for the LPP source than is needed for the laser-ignited DPP source.

Figure 8.15 uses the Cymer LPP source in Ref. 14 as an example. Here, a Nd:YLF seed laser is tuned to a 351-nm wavelength to be amplified by two XeCl amplifier lasers that operate at 400 W each and are expected to ultimately operate at 3.5 kW each at 16 kHz. Light from these lasers hits Li droplets to produce high-density plasma to emit EUV light. A 5-sr mirror collects the EUV light. Newer Cymer LPP sources use CO₂ lasers to hit Sn droplets to produce the plasma for EUV light.

Generally, DPP sources are more efficient because they use direct conversion from electric power. DPP sources are also more compact. LPP sources need a conversion from electric power to laser output, then from laser power to EUV power, making them less efficient. On the other hand, DPP sources tend to generate more debris than do LPP sources. The debris damages the collector mirror and shortens its lifetime.

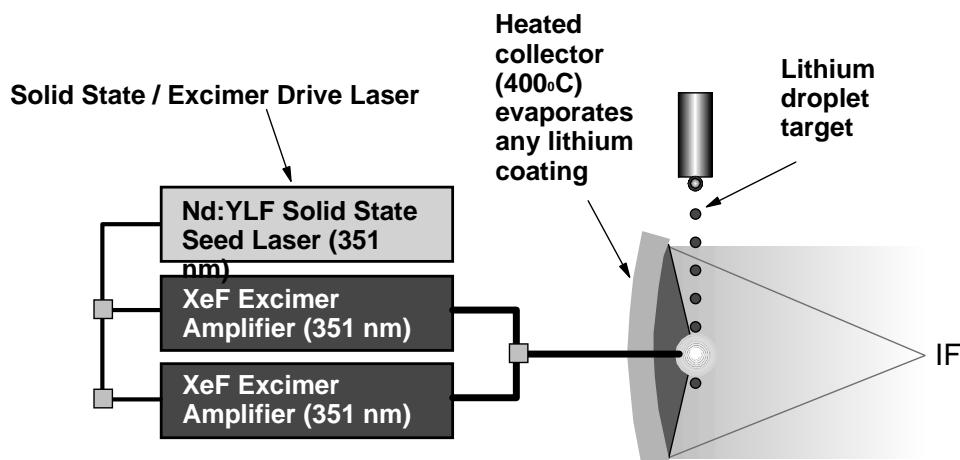


Figure 8.15 Example of LPP EUV source—Cymer EUV source.

Xe and Sn have been used as the target. The latter has a higher conversion efficiency yet tends to produce more debris, which contaminates the collector mirror. The best reported lifetime is several hundred hours, several orders of magnitude from the 10,000 hours required. The conversion efficiency of Xe DPP sources is between 0.5 and 1%; that for Sn DPP sources is between 1 and 2%. Most of the unconverted power turns into heat and must be removed. It is customary to consider removing 20 to 50 kW from the DPP source, as LPP needs a factor of 2 to 5 more.

Only the power radiating in the 2% bandwidth is usable. The light outside this band is harmful. Due to the peaking of the spectral reflectivity of the multilayer reflector, reflectivity drops sharply outside the usable bandwidth, and the unreflected light turns into heat, further complicating the heat management of EUV optics.

8.2.2.1 Wall-power requirement of EUV systems

Power consumption of exposure tools used to be a negligible issue. However, with an exotic light source, two-stage conversion efficiency, and a surge in raw-energy costs, power consumption warrants careful examination. The power required of a Xe DPP source with 500 W of in-band power is 700 kW or higher; for an LPP source, 1.25 GW or higher is required. The vacuum system and high-speed stage are estimated to require 200 kW, totaling 1.45 GW for an EUV scanner. We calculated the raw input power required to run 59 100-layers-per-hour tools to sustain a 130,000 wafers-per-month capacity on 20 critical layers per wafer. The throughput is adjusted for typical field-size mix, availability, and utilization. The 150 MW required to run a 12-in. fab for such a capacity is used as the reference for the power needed for the exposure tools for the critical layers. The power for immersion scanners supporting double patterning assumes a 200-layers-per-hour single-exposure throughput. These immersion tools consume 115 kW to run the light source and 50 kW to run the exposure unit. The power to run the tracks is not included. To run 59 immersion scanners this way requires 6.49% of the total power for the fab. To run 59 EUV scanners with an LPP source adds 57% of the power required to run the entire fab.

8.2.3 EUV masks

The EUV mask is a reflective mask. Due to the lack of a beamsplitter, the illumination must make an angle with respect to the optical axis of the mask. This angle is typically 6 deg to keep the optical components out of each other's way, yet introduce a minimum off-axis effect. Similar to the mirrors in the EUV optics, reflectivity is built up in a multilayer stack, with each layer pair contributing 1.5~2% of the total reflectivity of the stack. The EUV mask shape and size can be made similar to those of UV masks and can physically fit into conventional mask writers and inspection and repair tools. Not much change is expected for the EUV mask writer. However, mask inspection requires actinic light. New mask inspection tools need to be developed. Mask repair is expected

to be nontrivial, especially for the multilayer stack. Hence, the EUV mask repair yield is expected to be lower than that of conventional masks.

8.2.3.1 Configuration of EUV masks

An EUV mask consists of a substrate, multilayer stack, capping layer, buffer layer, and absorber layer, as shown in Fig. 8.16. Due to the reflective optics, the mask substrate need not be transparent to the actinic light. The requirements are low thermal expansion, rigidity, stability, ease to smoothen and flatten, and good adhesion for the multilayer stack. Similar to all reflectors in the EUV optical train, EUV masks achieve their reflectivity with a multilayer thin-film stack consisting of 40- or 50-layer pairs of Mo and Si^{14,15} with a combined thickness of 6.94 nm per layer pair. The Mo thickness is about 40% of the thickness of the layer pair. To protect the multilayer stack, 2 to 3 nm of Ru¹⁴ and 11 nm of Si¹⁵ have been used as the capping layer. Buffer layers used include 50-nm SiO₂ and 10- to 50-nm CrN.¹⁵ The absorber layer is generally 70 to 100 nm in thickness, and it includes Al, Cr,¹⁶ TaN,¹⁶ TaBN,¹⁵ and W.

8.2.3.2 Random phase shifting

The reflectivity of TaBN¹⁷ is 2%. Assuming a blank reflectivity of 65%, a 2% absorber reflectivity produces a 94% contrast that is equivalent to a 3% absorber transmission in a transparent mask. Assuming the same reflection-induced phase shift at the absorber and the reflector surfaces, the absorber must be embedded and flushed with the reflecting top surface or at $n\lambda/2$ above or below the reflecting top surface, to prevent phase shifting from taking place. The amount of phase shift is

$$\phi = 4\pi t/\lambda, \quad (8.1)$$

where t is the distance between the top of the absorber and the top of the reflecting multilayer, as shown in Fig. 8.17. A thickness of $\lambda/4 = 3.35$ nm induces a π phase shift due to the doubling of the optical path in the reflective mask. To keep the phase shifting error within 6 deg, as required in typical phase-

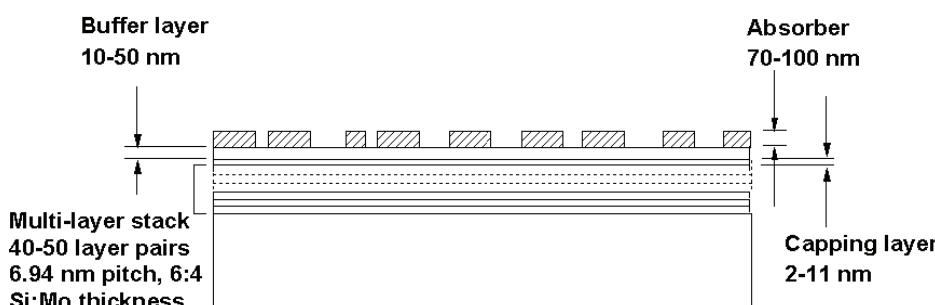


Figure 8.16 Configuration of an EUV mask.

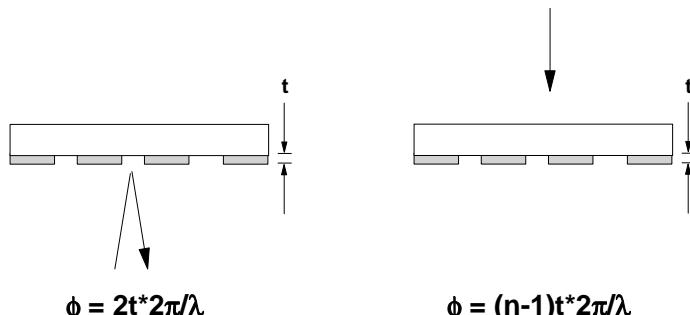


Figure 8.17 Thickness comparison of reflective and transmissive phase-shifting masks.

shifting mask (PSM) technology, the thickness of the absorber must be controlled within 0.11 nm. Hence, PSM technology is very difficult with EUV. It is very important to keep the absorber reflectivity low to make the mask a BIM.

If the thickness of the absorber or of a phase-shifting layer is not controlled to the required specification, random phase shifting takes place, reducing the common E-D window, as shown in Fig. 8.18(a). Here, the DOF based on four absorber thicknesses producing phase shifts of 0, 45, 90, and 180 deg is evaluated using the E-D window methodology. The absorber has a reflectivity of 2%. The feature is 22-nm opaque lines at a 55-nm pitch. On-axis illumination at $\sigma = 0.66$, $\lambda = 13.5$ nm, based on an 8% EL and a $\pm 10\%$ CD tolerance are used. The stray-light level is assumed to be 15%. The DOF-versus- $\sin\theta$ curves in Fig. 8.18(b) from the four absorber thicknesses are very similar, having the peak DOF above 300 nm. However, the common E-D window supports a DOF of only about 200 nm.

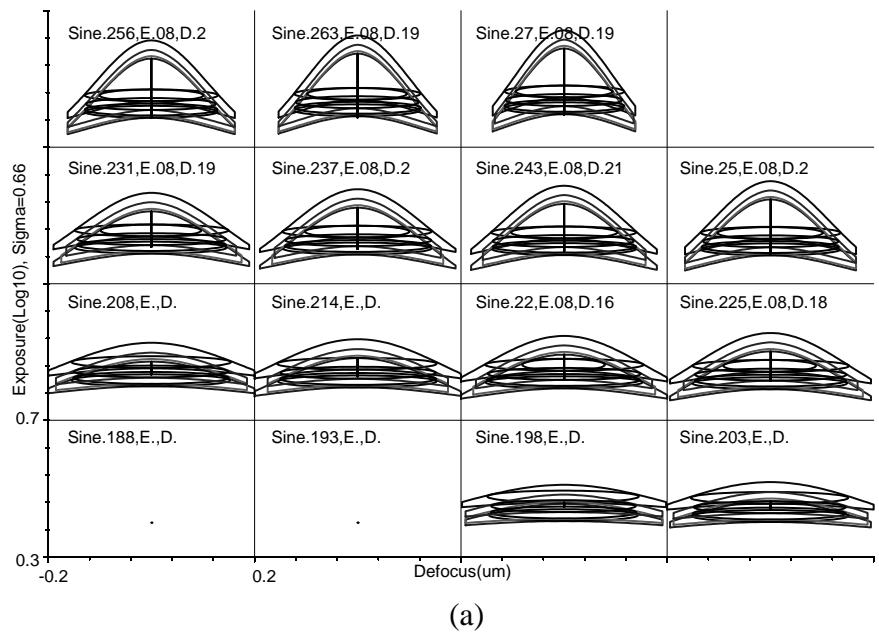
8.2.3.3 Effects of oblique incidence

8.2.3.3.1 Overlay mismatch

The mask pattern must be adjusted for the obliquely incident illumination beam, which makes an angle on the order of 6 deg. The pattern must be expanded in the x -direction by 0.55%, i.e., $1/\cos(6 \text{ deg})$, to overlay the image for mix-and-match. For a rigorous comparison of the wafer image and the image in the mask database, this difference must be kept in mind. This cosine effect in pattern asymmetry is manageable.

8.2.3.3.2 Pattern shadowing

A more serious effect comes from the shadow cast by the 6-deg beam. For absorber thicknesses between 70 and 100 nm, the shadow zone where the edge image is blurred is between 7.3 and 10.5 nm, correspondingly. The shadow has the effect of lowering the contrast of the edges perpendicular to the incident beam. This effect has been simulated.^{18, 19} The MEEF in the features perpendicular to the incident beam is indeed worse than those in the other direction. Figure 8.19 shows the shadow of the edges due to the 6-deg incident beam.



VILST2P3 Sig0.66 OpNA0.24 MaxDOF.21um@8%ELat WL13.5 CD0.021999978PIS

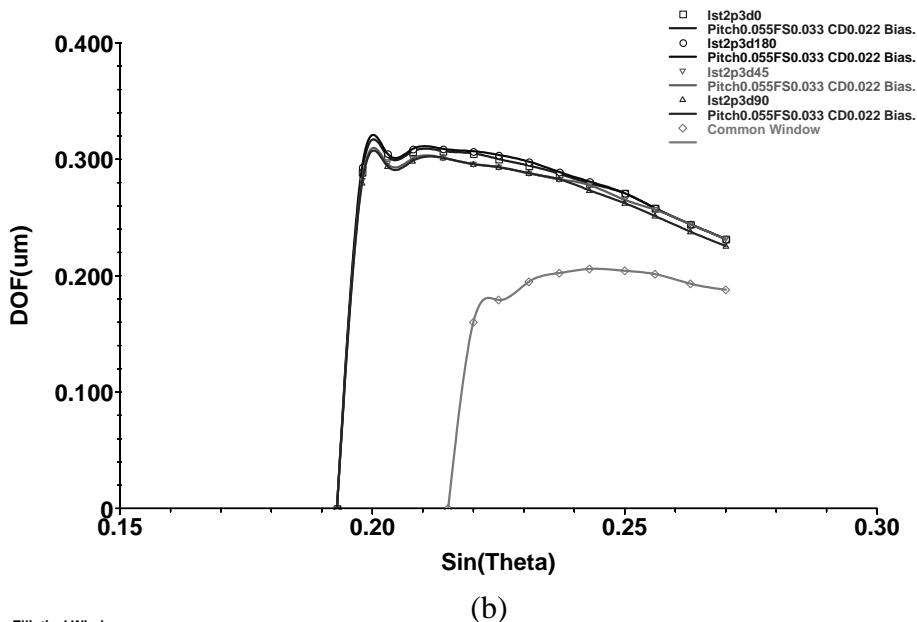


Figure 8.18 (a) Individual and common E-D windows. The four individual windows are evaluated with phases of 0, 45, 90, and 180 deg, respectively, for a 22-nm opaque line at a 55-nm pitch, based on $\sigma = 0.66$, $CD_{tol} = \pm 10\%$, $EL = 8\%$, $\lambda = 13.5$ nm, and 15% stray light. The absorber has a reflectivity of 2%. (b) DOF from the individual and common E-D windows in (a).

8.2.3.3.3 Overlay error from mask flatness

Another serious effect comes from the use of a reflective mask. When there is a slight tilt of the mask—caused by mask flatness, the wedge effect, or errors in orienting the reflecting surface of the mask—the tilt introduces a lateral image displacement. Even worse, the tilt angle is first multiplied by two because the mask tilt affects both the incident angle and the reflection angle, which is then multiplied by the mask reduction factor m , as illustrated in Fig. 8.20.

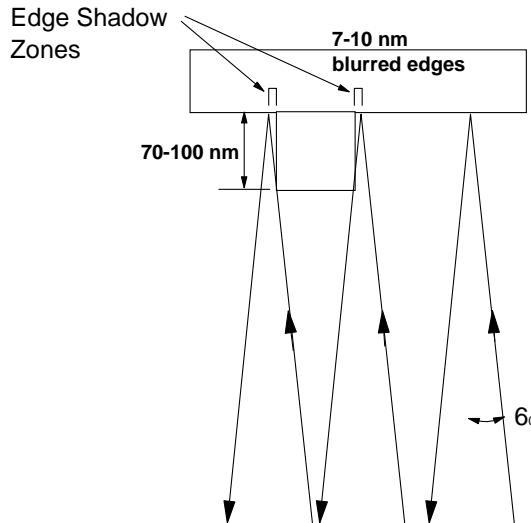


Figure 8.19 Shadow of edges due to off-axis beam effects.

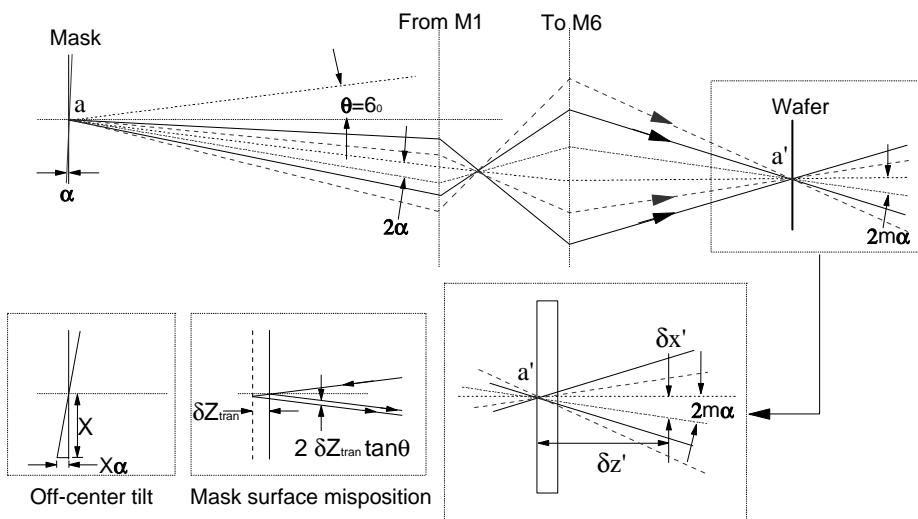


Figure 8.20 Lateral and longitudinal position errors due to mask rotation and translation.

Here, we took the rays from the M1 and M6 masks to the wafer in Fig. 8.9 and omitted the other rays. We then tilted the mask by angle α and the three reflected rays from the mask by 2α . Assuming a conjugate relationship between points a and a' , all rays regardless of the angle of tilt converge at a' . The final angle that the tilted principle ray makes is $2m\alpha$. As a result, the longitudinal deviation from the focal plane $\delta z'$ is related to the lateral image displacement $\delta x'_{rot}$ by

$$\delta x'_{rot} = 2m \sin \alpha \delta z', \quad (8.2)$$

where $\delta z'$ equals the DOF of the EUV images. We just analyzed the situation of mask tilt at the rotation center. Now, let us look at a point located laterally from the rotation center by X . In addition to $\delta x'_{rot}$, a longitudinal dislocation of $\delta z_{rot} = X \sin \alpha$ is introduced from the mask focal plane. The corresponding longitudinal displacement in the image plane is

$$\delta x'_{rot} = \frac{X}{m^2} \sin \alpha. \quad (8.3)$$

Another possible method for inducing a longitudinal dislocation from the focal plane involves the mask being translated in the z -direction δz_{tran} , producing

$$\delta x'_{tran} = \frac{\delta z_{tran}}{m^2} \quad (8.4)$$

in the image plane. In addition, because of the finite incident angle θ , a lateral image shift due to mask translation $\delta x'_{tran}$, is produced.²⁰

$$\delta x'_{tran} = \frac{2}{m} \delta z_{tran} \tan \theta. \quad (8.5)$$

The mask roughness components are shown in Fig. 8.21. Pure translational error is exemplified by δz_{tran1} ; pure rotational error is exemplified by δz_{rot1} , which is defined by α_1 and X_1 .

The diffraction DOF in an EUV system should be larger than the following components:

$$DOF_{diffrac} \geq DOF_{budget} = \delta x'_{scanner} + \frac{resist\ thickness}{n} + \delta x'_{tran} + \delta x'_{rot}, \quad (8.6)$$

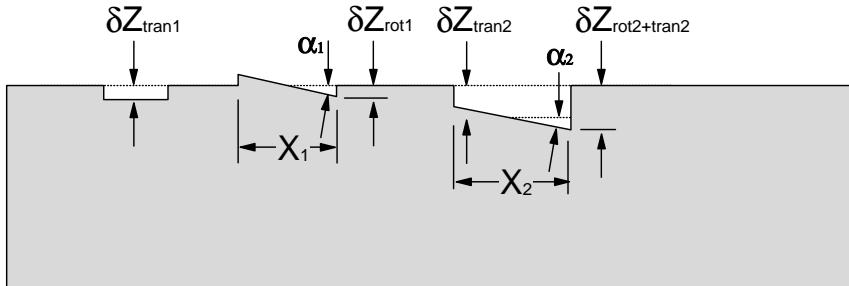


Figure 8.21 Mask roughness components.

where $\delta z_{scanner}$ contains the contribution for scanner focusing errors, lens image plane deviation, scanning errors, wafer chucking tilt, wafer topography, etc., all of which are similar to what is expected from a 193-nm scanner. n is the refractive index of the resist.

$$\delta x'_{tran} + \delta x'_{rot} \leq DOF_{diffrac} - \delta x'_{scanner} - \frac{\text{resist thickness}}{n} + \delta z . \quad (8.7)$$

At a 32-nm half pitch, according to Fig. 8.4, using $k_1 = 0.525$ and $k_3 = 0.348$, $DOF_{diffrac}$ is 348 nm. If $\delta z'_{scanner} = 100$ nm, resist thickness = 70 nm, and $n = 1.1$ at the EUV wavelength, then

$$\delta z'_{tran} + \delta z'_{rot} < 171 \text{ nm} . \quad (8.8)$$

The overlay limit is used to break down the specification of each component as follows. The total lateral image shift from Eqs. (8.2) and (8.5) is

$$\begin{aligned} \delta x'_{total} &= 2m \sin \alpha (\delta z'_{tran} + \delta z'_{rot}) + \frac{2}{m} \delta z_{tran} \tan \theta \\ &= \frac{2}{m} [(\sin \alpha + \tan \theta) \delta z_{tran} + X \sin^2 \alpha] \leq \text{overlay tolerance} \equiv \Delta x' . \end{aligned} \quad (8.9)$$

Solve $\sin \alpha$ in Eq. (8.9) by making the left side of the equation equal to $\Delta x'$:

$$\sin \alpha = \frac{-\delta z_{tran} \pm \sqrt{\delta z_{tran}^2 - 2X \delta z_{tran} \tan \theta + 2mX \Delta x'}}{X} . \quad (8.10)$$

Usually one-third of the minimum feature is used as the overlay tolerance, and one-third of this overlay tolerance can be allocated to $\Delta x'$, assuming that the other traditional overlay components can be squeezed to accommodate this extra reflective-mask-induced component. For a 32-nm half pitch, the minimum feature after lithography is generally 22 nm. $\Delta x'$ becomes 2.44 nm. When $\alpha = 0$,

$\delta z'_{rot} = 0$, $\delta z_{tran} = m^2 \delta z'_{tran} \leq 46.5$ nm, meaning the mask flatness must be better than 46.5 nm. The best flatness of current commercially available masks is one order of magnitude larger. This restriction is more severe than that from Eq. (8.8). With $\theta = 6$ deg, substituting $\delta z_{tran} = 46.5$ nm and $X = 1$ mm, $\sin\alpha$ becomes 4.65×10^{-5} . Substituting into Eq. (8.3) yields $\delta z'_{rot} = 2.91$ nm, which is safely within the limit posed by Eq. (8.8). Figure 8.22 plots $\sin\alpha$ as a function of X at $\delta z_{tran} = 10$ to 46.51 nm.

8.2.3.4 EUV mask fabrication

The patterning of EUV masks should be similar to UV mask patterning, in terms of mask writing. The mask writer probably needs the least modification. One concern is that because of the stringent requirement of reflective mask roughness and flatness, the chucking in the mask writer and in the exposure tool should be as close to each other as possible.

Any mask must be defect free. Otherwise, the defect will be replicated in every field on the wafer. The defect-free requirement starts from the mask blank, and applies to the multilayer and all of the subsequent layers on top of it. Fortunately, the multilayer deposition process is forgiving for small particles remaining on the surface of the substrate. The protrusion is smoothed layer by layer, even though perturbation of the reflectivity near the particles still occurs.²¹ This only works for defects smaller than 50 nm in size and near the bottom of the stack. Small particles that are farther away from the mask substrate are less forgiving.

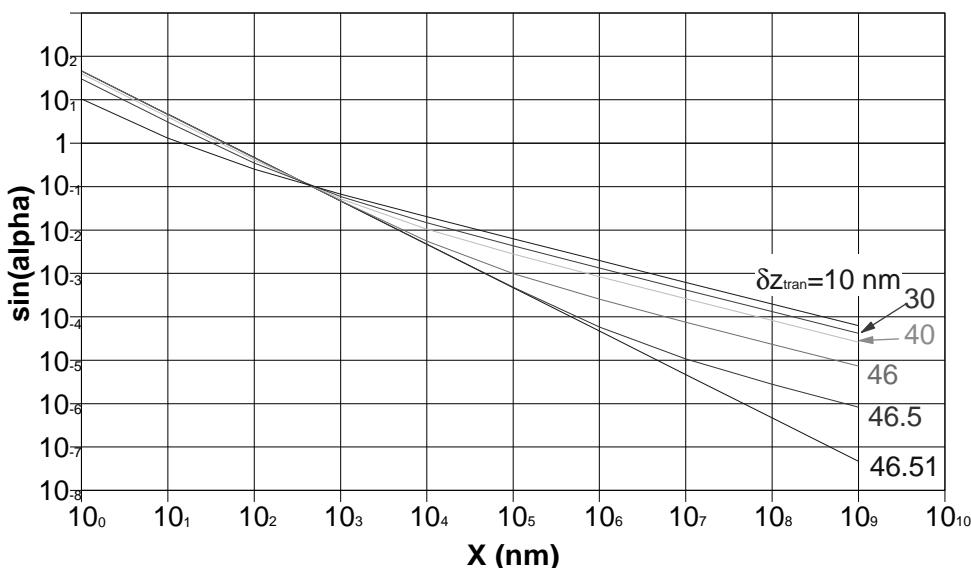


Figure 8.22 Plot of $\sin\alpha$ as a function of X at different values of δz_{tran} , for $\Delta x' = 2.44$ nm.

To delineate the absorber, one must protect the multilayer from being damaged by etching and any related processing, hence the need of a buffer layer. The capping layer is meant to protect the multilayer from environmental corrosion. Etch protection is generally left to the buffer layer. The buffer layer also protects the multilayer during the repair process. Mechanical removal¹⁶ of the absorber defect with atomic-force-microscope-type equipment tends to leave debris that must be carefully removed. The mechanical removal has been shown to stop within the buffer layer while not scratching the capping layer and the multilayer. E-beam repair is new but holds promise.²²

The mask must be inspected after patterning to locate the repair sites. Inspection at the actinic wavelength calls for a completely new inspection technology. The general thinking is to inspect at a convenient UV wavelength, e.g., 257 nm in continuous wave (CW). To facilitate such off-wavelength inspection, it is necessary that there is sufficient contrast between the absorber and the reflecting layer stack at the inspection wavelength and that the defects visible in EUV imaging are detectable. To date, many EUV printable defects have escaped inspection at off-wavelengths. Actinic inspection may be inevitable. Bright EUV source and reflective inspection optics are required.

8.2.3.5 Absence of mask pellicle

All optical masks enjoy the protection of a pellicle, as we discussed in Sec. 5.3.2. Any particle accidentally falling on the mask is kept at a safe distance from focus. Even for thicknesses of a few micrometers, there is no transparent material at the EUV wavelength. It is not feasible to protect the mask during exposure as can be done for regular UV pellicles. The proposal for EUVL is to keep the mask under an opaque cover during storage and transportation. It is without protection only during exposure in the scanner. The EUV optics are designed with the mask facing down at the exposure position to minimize the chance of particles falling on it. However, electrostatic chucking may unexpectedly attract particles.

8.2.4 EUV projection optics

Typical layouts of EUV optical components are shown in Figs. 8.8 to 8.10. Smoothness of these components is critical to reduce stray light. Note that reflection doubles the optical path difference from longitudinal surface changes in the optical element. Reflective optical elements are much more sensitive than the refractive elements illustrated in Fig. 8.17. To maintain the same level of stray light, reflective surface roughness and surface configuration must be four times better than the refractive ones. The requirement of $\lambda/50$ in 193-nm refractive optical elements translates to $13.5/200 = 0.067$ nm, which is a fraction of the dimension of an atom. For the moment, the NA in the EUV system is only 0.25. A mid-spatial-frequency roughness of 0.14 nm can keep the system flare to 10%.^{10,23} This precision must be maintained during fabrication, mounting, and assembly in the lens and scanner factories, as well as during the alignment and

realignment in vacuum on the wafer-manufacturing floor. Contamination buildup at the optical surfaces can also increase roughness above specification.

Contamination of the optical surface can deteriorate the ultrahigh smoothness and precision of the optical elements. Oxidation is irreversible. The capping layer on the multilayer structure is designed to prevent oxidation. Carbon buildup can presumably be cleaned. Assuming that the number of cleaning cycles is infinite, the cleaning time and frequency should be kept at an acceptable level that does not disrupt production or reduce the cost of ownership. A current 10-hour cleaning frequency may or may not be acceptable depending on the time required, complexity, risk, and cost. The specification for a mirror lifetime is 30,000 hours or about 3.5 years. One thousand hours or 1.4 months have been demonstrated on a synchrotron beam line.²⁴ There is hope that switching to pulse radiation will extend the life even further by optimizing the pulse frequency and duration to the adsorption and desorption characteristics of the molecules and the multilayer. However, at the same average power, pulse radiation delivers a higher peak power, and the dependence of adsorption and desorption can be nonlinear for instantaneous power. A demonstration of a mirror lifetime at realistic operating conditions remains to be seen.

Reflective optics has the inherent advantage of a polychromatic response, thus a broad bandwidth. Instead of the fraction-of-a-picometer bandwidth required by 193-nm lithography, the bandwidth requirement of EUVL is 2%. Most multilayer coatings meet this requirement,¹⁰ as exemplified by Fig. 8.23. A key requirement is to center all of the coatings to 13.5 nm. A spectral mismatch can waste valuable EUV power. Power loss due to multilayer mismatching is about 10%.⁹

Heat on the optical elements should be carefully managed. Figure 8.11 and Tables 8.1 and 8.2 give an idea of the heat load to bear by the optical elements. The heat must be quickly and uniformly removed to maintain the subnanometer accuracy of the optical surfaces. From the point of heat management, it is desirable to use a fast resist.

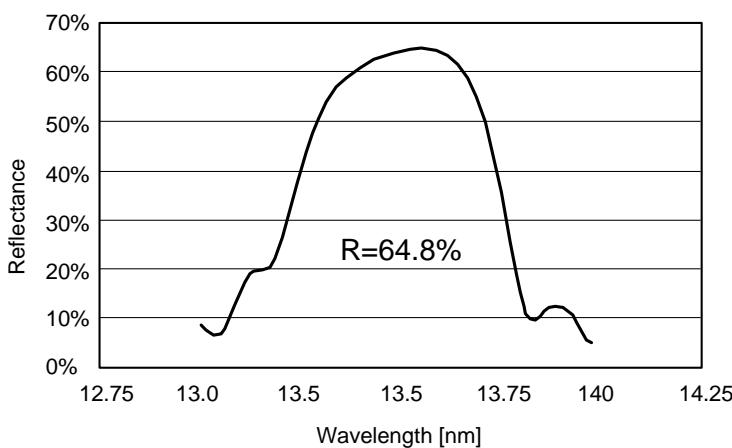


Figure 8.23 Spectral reflectivity of an EUV multilayer coating.

8.2.5 Wall-power consumption

A fast resist is also desirable for keeping the electric power in check for better power management. A typical power requirement²⁵ of EUV exposure tools using a DPP source is on the order of 150 kW, assuming a 5-mJ/cm² resist sensitivity. The same requirement using an LPP source is between 300 and 650 kW.

Recent data puts LPP source power requirementsⁱ at 450 kW for a 10-mJ/cm² resist. Together with 190 kW for scanner operation, an LPP-based EUV scanner requires 640 kW compared to 165 kW for an ArF immersion scanner. With lower resist sensitivity for realistic imaging, less optimistic collection efficiency and source conversion efficiency, as well as two more mirrors in the optical train, each EUV scanner may consume²⁶ 1.5 to 13 MW.

8.2.6 EUV resists

Recent publications on EUV resists have been rare. In the entire 2005 SPIE Emerging Lithographic Technologies Conference, out of the 85 papers submitted for EUV lithography, only two were on resists. Even in the adjacent conference on Advances in Resist Technology and Processing, only seven out of 148 papers were on this topic. Figure 8.24 shows the distribution of the papers in sources, optics/coating, systems/performance, masks, metrology, and resists. The number of papers on EUV sources is highest, followed by equal distribution for other disciplines. EUV development research appears to be undertaken mainly by physicists, especially high-energy physicists.

The resist issues for EUV change rapidly with time. When soft x-ray-projection lithography was proposed, the main concern about resists was their high absorption. The absorption coefficient α was estimated²⁷ to be larger than 3 μm^{-1} . EUV light penetrates to 200 nm, but the effective imaging thickness is about 60 nm. During that time, 0.25 μm was considered a formidable resolution for optical lithography.²⁸ Only this author predicted a 0.13- μm resolution.²⁹ EUV

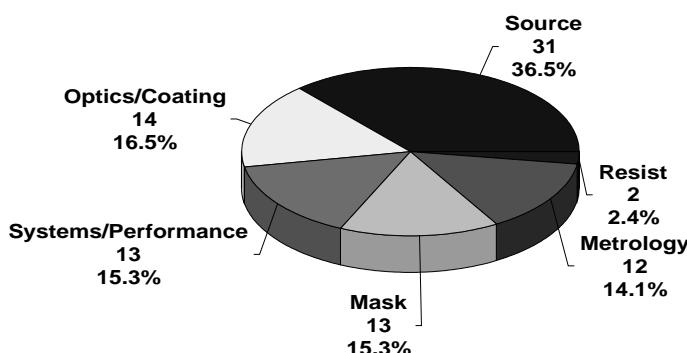


Figure 8.24 EUV papers submitted for the 2005 SPIE Emerging Lithographic Technologies Conference.

ⁱTentative specification from potential EUV scanner supplier.

lithography was proposed in view of the pessimistic outlook on optical lithography. Therefore, resist absorption was clearly an issue. Bilayer and trilayer resist systems were considered necessary to support EUV.

However, optical lithography marched on. The opportunity for a successor will barely arrive at a 32-nm half pitch, projected to take place at least 20 years after the initial EUV proposal. At 20~30-nm feature sizes, resist thickness is no longer limited by absorption. Due to the capillary force and other factors during resist development/rinse as we discussed in Sec. 6.2.4, resist images with larger than a 3:1 height-to-width aspect ratio tend to collapse. The resist thickness is limited to less than 100 nm for a 30-nm feature size.

On the other hand, the throughput of optical scanners is approaching 150 wph, while the field size remains $26 \times 33 \text{ mm}^2$ at a 4X reduction. Any successor will have to measure up to the cost of ownership of a state-of-the-art scanner. With the cost of equipment projected to be up to two times the price of an optical scanner and throughput at 100 wph, an EUV scanner is already inferior in terms of cost and productivity. Hence, high resist sensitivity is the only hope for making EUV economical. If resist sensitivity remains at 30 to 50 mJ/cm² for LER control, then the source power requirement will be extremely difficult to meet, as can be seen in Table 8.4. Even if the high-energy physicists succeeded in producing an extremely intense EUV source, the incident power on the optical elements and the resultant heat will be very difficult to manage. For throughput and heat management, it is desirable to have resist sensitivity between 1 and 2 mJ/cm². Of course, EUV shot noise at such resist sensitivity must be considered, in addition to LER.

At least four resist companies have tried to develop an EUV resist with a low line edge roughness (LER) and a high sensitivity. No less than 48 samples have been evaluated.³⁰ Sensitivity versus linewidth roughness (LWR) is shown in Fig. 8.25. Two resists reported³¹ in 2005 with a 18.3-mJ/cm² sensitivity and a 4.1-nm

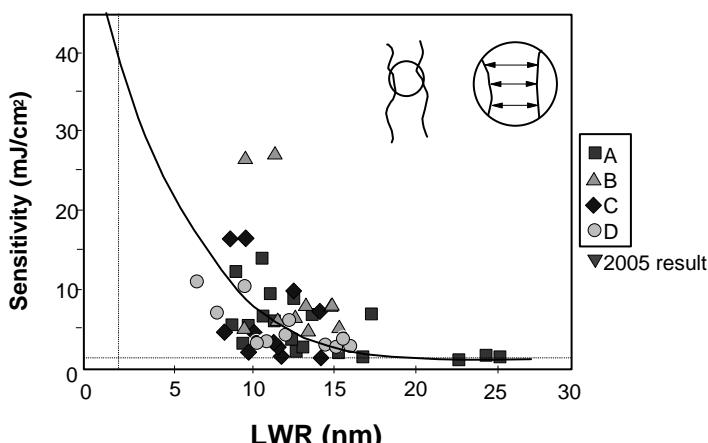


Figure 8.25 Tradeoff between photospeed and LWR for resists from vendors A to D from 2004 results. Two 2005 data points have been added. The asymptotic curve is added.

LER and a 35-mJ/cm^2 sensitivity and a 5-nm LER were added to the chart. Note that the LER values were converted to LWR by multiplying by $\sqrt{2}$. The required 1-mJ/cm^2 sensitivity and the 2-nm LWR for the 32-nm node is located at the intersection of the horizontal and the vertical dotted lines. This author took the liberty to draw an asymptotic curve on the chart, showing how resist sensitivity statistically trades off with LWR. A breakaway from the curve is needed to meet the sensitivity-LWR requirement.

8.2.7 EUV OPC

At a relatively high k_1 , OPC for EUVL has been expected to be extremely simple. However, two new phenomena pose unconventional challenges to OPC for EUVL. Stray light in the EUV image is not only an order of magnitude higher than that of conventional optics, it is field dependent, i.e., the level of stray light varies at different intrafield locations. Hence, the contrast and exposure level are field dependent; they must be compensated for according to the location on the mask.

The second phenomenon is also field dependent and originates from pattern shadowing, discussed above in Sec. 8.2.3.3.2. Here, shadowing is shown in 3D, as reported by Lorusso.³² The 6-deg beam is incident at the radial direction with respect to the illuminated curved slit. Hence, the shadowing effect changes as a function of the orientation of the feature as well as due to its position in the slit. Figure 8.26 shows a horizontal and a vertical line at three locations of the slit and the corresponding shadows. Note that the line feature is assumed infinitely long. Therefore, the shadow is drawn for its edges but not for its ends. Figure 8.27 is a top view of the same line features but oriented in four different directions. For the 45 and 135-deg lines, the shadow is not symmetrical with respect to the center of the slit because of the relative orientation of the illuminating beam and because the line is different at the ends of the slit.

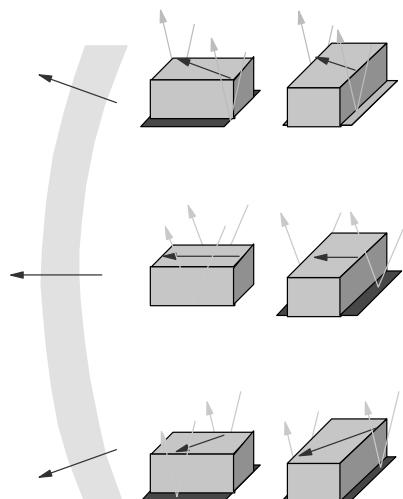


Figure 8.26 Shadowing due to the orientation in illumination. (Image courtesy of Lorusso and Kim.³²)

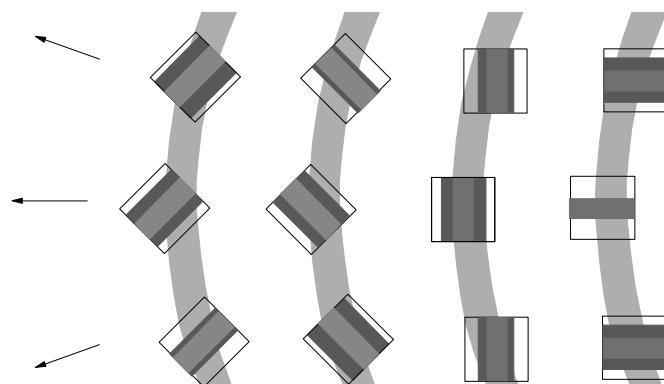


Figure 8.27 Shadowing of edges as a function of the position at the illuminated slit. (Image courtesy Lorusso and Kim.³²)

As a result of the positional dependence of the shadows, the delineated linewidth varies according to its position in the scanning slit. This must be compensated for with field-dependent OPC.

8.2.8 Summary of EUVL

EUVL offers an opportunity to extend optical lithography by two generations. It has the potential to gain DOF and make OPC simpler. Tremendous progress has been made in EUV optics, multilayer coating, sources, and masks. As a lucratively funded project with a high commercial return, much knowledge in physics, materials, processes, and metrology has been gained. The related technology has been established, and many highly talented scientists have been gainfully employed. Volumes of reports have been produced. Science has advanced.

The need for investment has not ceased. To spend the funds wisely, further research to increase source power must concentrate on conversion efficiency. Mitigation of debris and extending the life of the collector are important. By far, the highest leveraged development should be aimed at the resist. Ideally, 1~2-mJ/cm² resists with less than 2 nm of LER should be developed while keeping in mind the problem of shot noise. Otherwise, too much heat must be managed. High wall power makes the technology not only costly but also environmentally unfriendly.

What other challenges are there? A short list would include off-axis optics, adjustable illumination, adjustable NA, stray light, atomic-level configuration and smoothness of optical components, their shortened lifetime from debris, contamination and high-energy radiation, as well as mask defect, inspection and repair, mask cost, absence of mask pellicles, mask flatness, random phase shifting, field-dependent OPC, system uptime, throughput, footprint, and cost.

8.3 Massively Parallel E-beam Maskless Imaging

Optical lithography, with all foreseeable extensions, can reach a half pitch of 19 nm. However, starting at 32 nm, one should start comparing it with potential successors to phase in the most economically viable technology. EUV lithography is popularly considered the only successor. However, Sec. 8.2 showed quite a few uncertainties. Massively parallel e-beam (MEB) maskless lithography (ML2) makes high-throughput lithography possible with e-beam that has a higher resolution and DOF potential than 193 or 13.5 nm. The maskless direct writing (DW) scheme eliminates costly and time-consuming mask making. In addition, the critical-dimension uniformity (CDU) budget allocated for the mask can now be removed.

8.3.1 Introduction to e-beam imaging

E-beam DW emerged in the 1970s,³³ due to concerns about optical lithography stopping at a 1-μm half pitch. However, it has been handicapped by throughput, despite tremendous efforts to increase it. Unfortunately, the throughput requirement has increased much faster.³⁴ E-beam lithography found an application in mask writing instead.^{35,36,37} The fast-moving electron quickly outpaced optical pattern generators that used mechanical means to move a single microscopic objective to form the pattern. Only later did laser writing with multiple heads³⁸ regain the upper hand in writing speed against e-beam mask writers. They gain speed by using multiple optical beams (16, 32, etc.). Recently, a new optical mask-writing schemeⁱⁱ offers potential for an even higher throughput. A bank of 1 million electronically addressable mirrors is used. To speed up throughput of e-beam writers to support wafer writing, parallel writing using multiple electron beams is the only way to go.

E-beam has virtually no limit in wavelength-dependent resolution. An equivalent wavelength of electrons with particle energy E_0 in eV is given by³⁹

$$\lambda_{electron} = 1.226E_0^{-1/2} \text{ nm.} \quad (8.11)$$

The relationship of photon energy in eV to wavelength is also given in Ref. 39:

$$\lambda_{photon} = 1239.9 E_0^{-1} \text{ nm.} \quad (8.12)$$

Tables 8.6 and 8.7 show the wavelengths in photons and electrons applicable for lithography. The two tables show that even the longest wavelength in electrons is two to three orders of magnitude shorter than the shortest photon wavelength.

Due to the short wavelength, electro-optical systems need only a very small NA to satisfy the resolution requirement. Take a 2-keV e-beam whose

ⁱⁱ Micronic mask writer.

wavelength is 27.4 pm. If k_1 is 0.8, then the NA required to resolve a 16-nm half pitch is 0.00137. The resolution limit of e-beam lithography is due to space-charge effects, the imperfection of the electron optics, and scattering in the recording medium.

- *The space-charge effect:* Electrons, being charged particles, repel against each other, exhibiting a space-charge effect that enlarges a focused spot where the electrons cross over each other. A higher beam current enhances the space-charge effect. For high-throughput applications, the best way to treat space-charge effect is to minimize the number of beam crossovers or simply eliminate them.
- *Imperfection of the electron optics:* One of the most significant imperfections in electron optics, such as an uneven focusing field at different radii from the optical axis, causes spherical aberration. Off-circular apertures cause astigmatism. The finite aperture shape induces diffraction. Beam energy spread causes chromatic aberration.
- *Scattering:* When electrons enter the resist layer and the substrate, they scatter in the forward and backward directions. Backscattering dominates high-energy beams, and forward scattering dominates low-energy beams. Scattering enlarges the beam spot, as well as increases beam-to-beam interaction to produce e-beam proximity effects. Actually, the discovery and correction of e-beam proximity effects⁴⁰ predated OPC. For energy below 5 keV, the e-beam minimally penetrates the resist, and backscattering from the substrate is negligible. Most electrons are put to use in exposing the resist. Above 50 keV, most of the electrons penetrate deeply into the substrate, and backscattering takes place inside the substrate. Proximity effects in the resist are also not as severe. At approximately 25 keV, both forward scattering and backscattering must be taken into account.

Table 8.6 Photon wavelengths and equivalent energy in eV.

	g-line	i-line	KrF	ArF	F2	EUV	x ray
(nm)	436	365	248	193	157	13.4	1
eV	2.84	3.40	5.00	6.42	7.90	91.8	1240

Table 8.7 Electron wavelengths as a function of E_0 .

	300 eV	1 keV	2 keV	5 keV	10 keV	25 keV	50 keV	100 keV
(nm)	0.0708	0.0388	0.0274	0.0173	0.0123	0.0078	0.0055	0.0039

8.3.2 MEB ML2 systems proposed

MEB ML2 systems are polarized into two extremes in acceleration voltage: low-voltage (LV) systems using up to 5 kV and high-voltage (HV) systems using 50 to 100 kV. Both systems have low proximity effects, as we discussed in Sec. 8.3.1. Because the power incident on the resist is proportional to the product of the acceleration voltage and the beam current, LV MEB systems usually aim at high-throughput applications. On the other hand, HV MEB penetrates the entire resist layer. Most resists for 248- or 193-nm exposures can be exposed with HV MEB. E-beam mask blank resists are also for HV MEB exposure.

8.3.2.1 MAPPER MEB ML2 system

The LV MEB ML2 system⁴¹ proposed by MAPPER is shown in Fig. 8.28. Output from a single electron source floods the collimator lens to send the beam to an aperture plate consisting of 13,000 distributed apertures, as shown in Fig. 8.29. After going through the aperture array and the condenser array, the unwanted beams are blanked off by a blunker array that gets instruction from the optical fibers, which illuminate photosensitive circuits on the blunker array to generate an electric signal for blanking. The beam deflector array deflects all unblanked beams to scan in sync. The beam stop array stops the beams that are blanked off, and the projection lens focuses the unblanked beams to the wafer.

The writing scheme is shown in Fig. 8.29. For a 45-nm half-pitch system, the 13,000 beams are distributed 150 μm apart in a $26 \times 10 \text{ mm}^2$ area. The separation ensures negligible beam-to-beam interaction. Only the beam blur recorded in the resist due to forward scattering calls for proximity correction. The beams are staggered row by row at a distance of 2 μm . The deflection range of each beam is slightly larger than 2 μm to allow a common stitching area between beams, so that the 150- μm spaces in each row are filled with patterns after 75 rows of beams pass through. Each beam is blanked off for the unexposed areas. The beam size and locations are drawn closer to scale in Fig. 8.30, which shows the overlapping of the scanned beams and the length covered by each beam. Due to the staggering of the beams, the full pattern starts after approximately 10 mm from the start of the scan. Fortunately, the eight fields in an entire row on the 300-mm wafer can be scanned without interruption. Therefore, the 10-mm lead is needed only once per row, instead of for each field as in a masked scan-and-repeat system. The 35-nm spot is made possible to delineate smaller features by use of the 2.25-nm grid and edge gradient.

Recently, the independent writing of 99 beams was demonstrated.⁴² Figure 8.31 shows the result of fully exercising the aperture array, the blunker array, the deflector array, and the electrostatics optics from the assembly of these MEMS-fabricated components. Since mechanical stage scanning is not yet available, the beams were deflected electronically in both directions. Gratings at four orientations were written to the hydrogen silsesquioxane (HSQ) resist within the 2- μm square zone covered by each beam. Figure 8.32 shows how these grating are written. The vertical gratings were written with the beams skipping every

other column; the horizontal gratings were written with the beams skipping every other row. For 45 and 135-deg orientations, the beams were written in every row but were skipped at every other spot on the row with staggering from row to row. In Fig. 8.33, dense contact and iso contacts, dense vertical trenches and horizontal trenches, iso vertical trenches and horizontal trenches in 40-nm of polymethylmethacrylate (PMMA) are shown. The images are repeated in 42 beams. Each pattern in Fig. 8.33 represents the image from one beam. The CD mean and 3σ of each feature are shown in Table 8.8. The CD distribution of the patterns in each beam is tighter than it is between beams. Note that the dosage used is different for the patterns. Proximity correction was not used in the experiment.

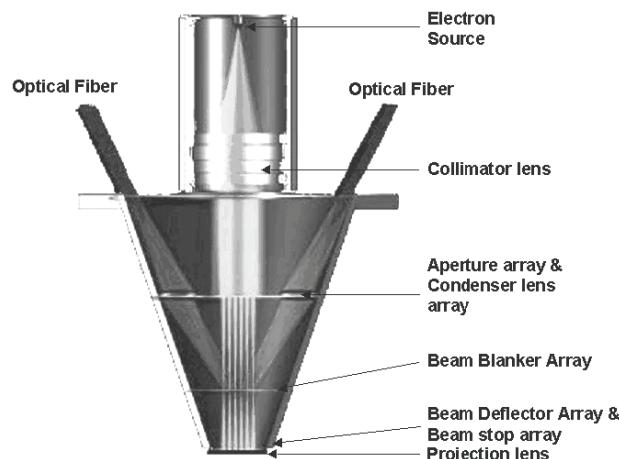


Figure 8.28 A LV MEB ML2 system proposed by MAPPER. (Image courtesy of B. J. Kampherbeek et al.)

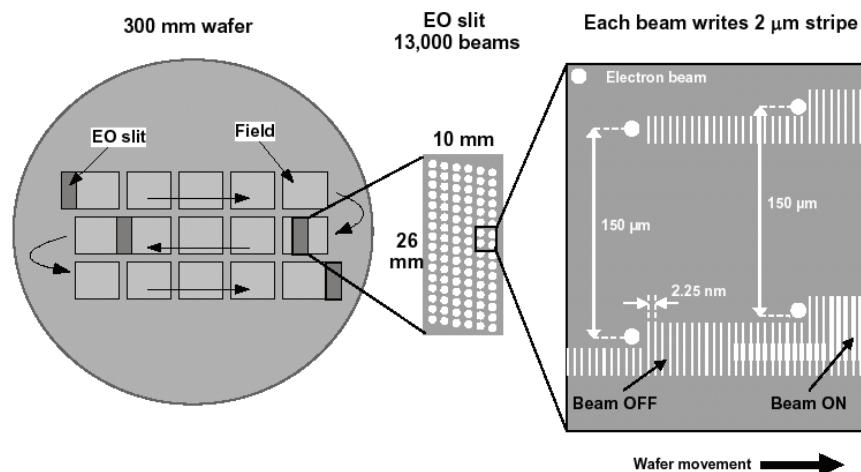


Figure 8.29 The aperture array and writing scheme of the MAPPER LV MEB ML2 system. (Image courtesy of B. J. Kampherbeek.)

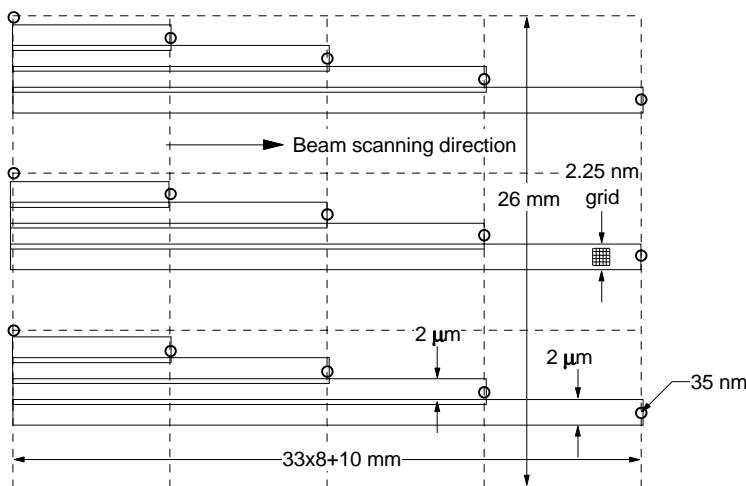


Figure 8.30 MAPPER beam locations and coverage.

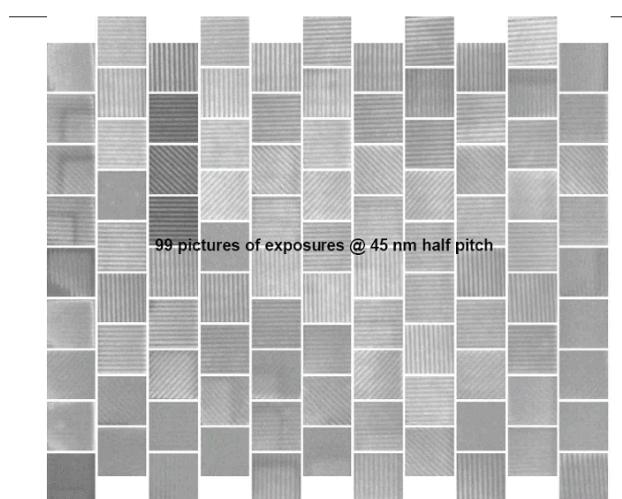


Figure 8.31 MAPPER 45-nm half-pitch e-beam results. (Image courtesy of B. J. Kampherbeek.)

Table 8.8 Exposure dosage, CD mean, and 3σ of 42 beams from six representative patterns.

Pattern	CD mean (nm)	CD 3σ (nm)	Dosage ($\mu\text{C}/\text{cm}^2$)
Dense hole	55.2	2.5	125.7
Iso hole	47.7	4.9	145.0
Dense vertical trench	43.0	5.1	78.6
Dense horizontal trench	42.3	5.6	78.6
Iso vertical trench	50.8	4.2	125.0
Iso horizontal trench	50.8	3.4	125.0

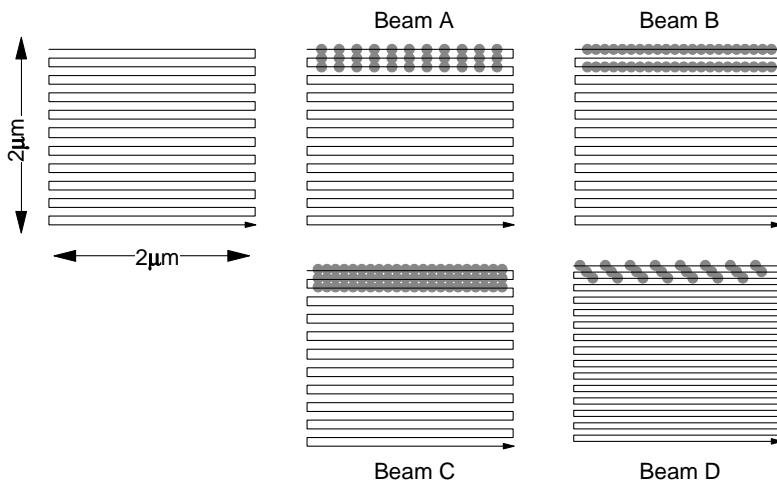


Figure 8.32 The writing scheme of the four patterns in Fig. 8.31.

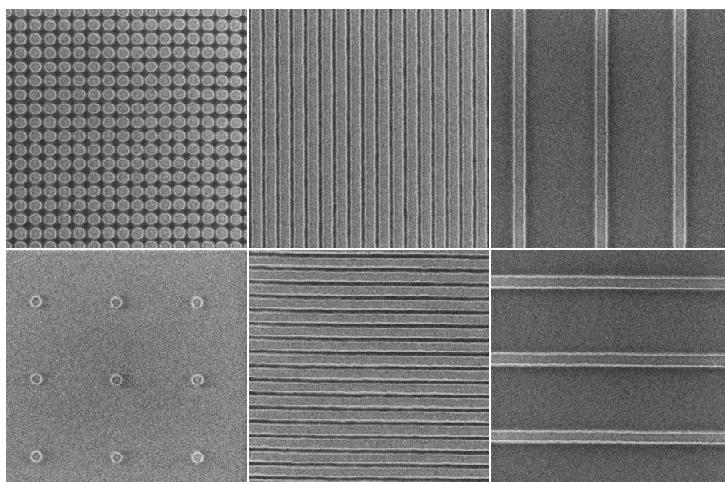


Figure 8.33 Six types of pattern in 40-nm PMMA using 42 beams. 3σ CD ranges from 2.5 to 5.6 nm, with nominal CD of 45 nm. (Courtesy of B. J. Kampherbeek et al.)

8.3.2.2 LV system with individual sources

For simplicity and reliability, many MEB ML2 systems use a single source to flood an array of apertures. However, the limitation of closely locating the MEB can sacrifice better e-beam placement for data reduction or space-charge reduction. Multiple sources can help, in principle, to increase the current for exposure. In an MEB ML2 system with individual sources,⁴³ the output of a field-emitter array (FEA) goes into a microchannel amplifier (MCA) array, which runs at a saturation mode so that the output current stays constant regardless of the source fluctuations. Each MCA can cover multiple sources to

further ensure stability at the output of the MCA. The electrons then go to an integrated multiple-electron-beam lens array (EBLA).

Not much has been reported on the EBLA, except that the array covers a $26 \times 33\text{-mm}^2$ field and that 13,965 channels will be used. The acceleration voltage is 5 kV. At 1.2 nA output per channel from the EBLA, $20 \mu\text{C}/\text{cm}^2$ at five 300-mm wafers per hour for a 45-nm spot size is expected at the wafer.

8.3.2.3 HV MEB ML2 system

A 100-keV HV MEB ML2 system^{44,45} uses a single source and a single electron column to image a multi-aperture array. As shown in Fig. 8.34, one electron gun illuminates the aperture plate system through an electrostatic condenser. The apertures are $5 \times 5 \mu\text{m}^2$ in size, separated by 35 μm from each other and distributed in a $20 \times 20 \text{ mm}^2$ area. A two-stage reduction lens, each demagnifying the beams by approximately 14X, is used. The first stage consists of an electrostatic lens and the second is magnetic. There is one beam crossover in each stage. There is a total demagnification of 200X, making the incident beams on the wafer $25 \times 25 \text{ nm}^2$ in size; the array is distributed in a $100 \times 100\text{-}\mu\text{m}^2$ area. The beams are now 175 nm apart.

The aperture plate system is illustrated in Fig. 8.35. The electron beam from the single source has 5-keV energy when it reaches the cover plate, which blocks off the electron in the unwanted areas. They then pass through a blanking plate consisting of apertures and deflector electrodes. CMOS circuits are built on another area of the blanking plate that is not shown in the figure. They send a few volts to the deflection electrodes to blank off the unwanted beams. Data can be sent via optical cables to switch the CMOS circuits, just as in the MAPPER system. The final plate in the aperture plate system is the aperture plate that is

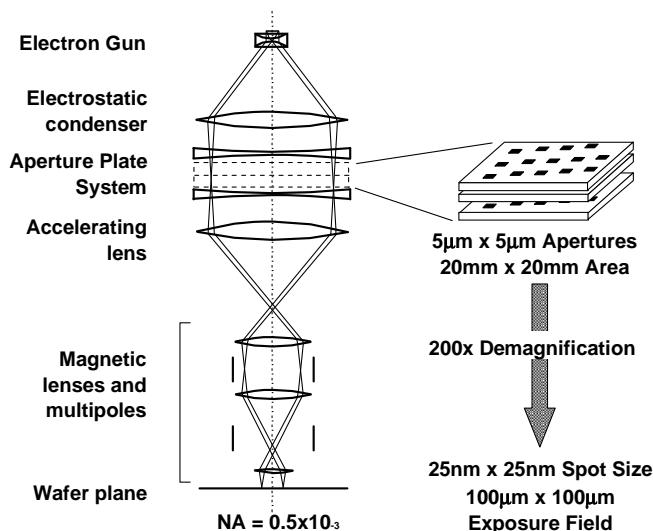


Figure 8.34 The schematics of the PML2 electron optical column. (Image courtesy of C. Klein et al.)

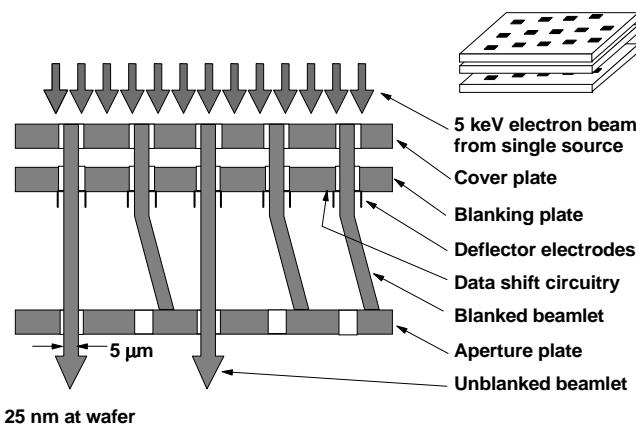


Figure 8.35 Aperture plate system principles. (Image courtesy of C. Klein et al.)

mounted 15 mm below the blanking plate. The plate stops the blanked beams deflected by the blanking plate. It also defines the final aperture of the beams that are unblanked. A proof-of-concept system is designed to supply 0.5 μ A to provide 18 μ C/cm². An early production tool is planned for 1.9 million apertures and a 3.5- μ A beam current to support 2 wph at the same dosage of the proof-of-concept tool.

Recent imaging results⁴⁶ from a 15-keV system are shown in Fig. 8.36. Figure 8.36(a) contains 16-, 22-, 32-, and 45-nm half pitches, a 1:1 line-space, and isolated resist lines in the HSQ negative resist of 50-nm thickness. Figure 8.36(b) shows linewidth as a function of exposure time, demonstrating the resolution potential of the 15-keV system and good linewidth control. The resolution of a 50-keV system is expected to be higher. Dynamic demonstration with programmable blanking is expected.

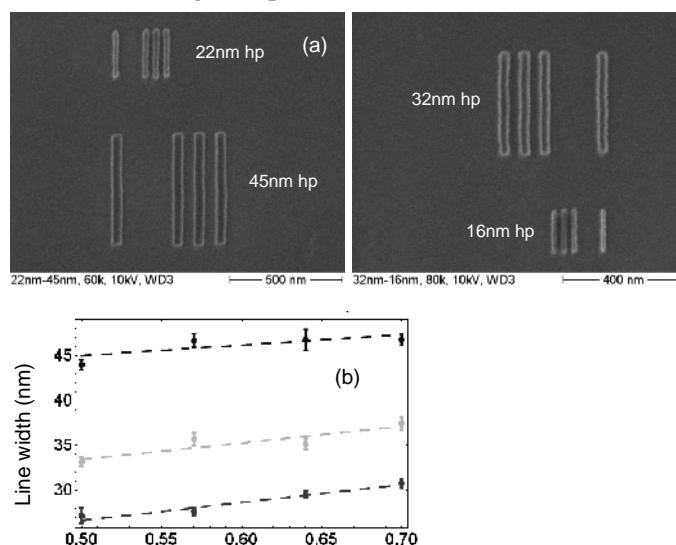


Figure 8.36 IMS PML2 electron optics results. (Image courtesy of C. Klein et al.)

8.3.2.4 MBS MEB ML2 system

The MBS system⁴⁷ utilizes multiple micro-e-beam columns to provide parallelism. Because the microcolumns are only a few millimeters in diameter, as many as 88 of them can be mounted on top of a 300-mm wafer, providing 88X improvement in writing coverage per unit time. As shown in Fig. 8.37, the beams go through the electron optics without crossover, which limits the beam current to prevent space-charge-induced blurring. The leftmost column depicts the unblanked beam focused to the center of the aperture. The middle column shows the beam being blanked off, and the rightmost column shows the unblanked beam deflected to its widest position of about 50 μm off axis. The beam shaping and blanking optics enclosed within the slashed boundary can be changed according to the writing strategy. Figure 8.38 shows an assembly of two microcolumns.

Multibeam systems (MBS) use third-order beam shaping to increase the beam current on the wafer without causing aberration. As depicted in Fig. 8.39, the shape of the beam is optimized at the beam-defining aperture to support a square image with a large current on the order of 1000 A/cm^2 . This way, the throughput can be increased by another factor of between 25 and 50 to couple with the 88X gained from clustering the microcolumns. Using a vector scan further gains another factor of between 2 and 5. The working principle of 3rd-order beam shaping or patterned beam-defining aperture (PBDA) is shown in Fig. 8.40. The electrons in the inner and outer rays constructively contribute at the depicted locations of the square image.

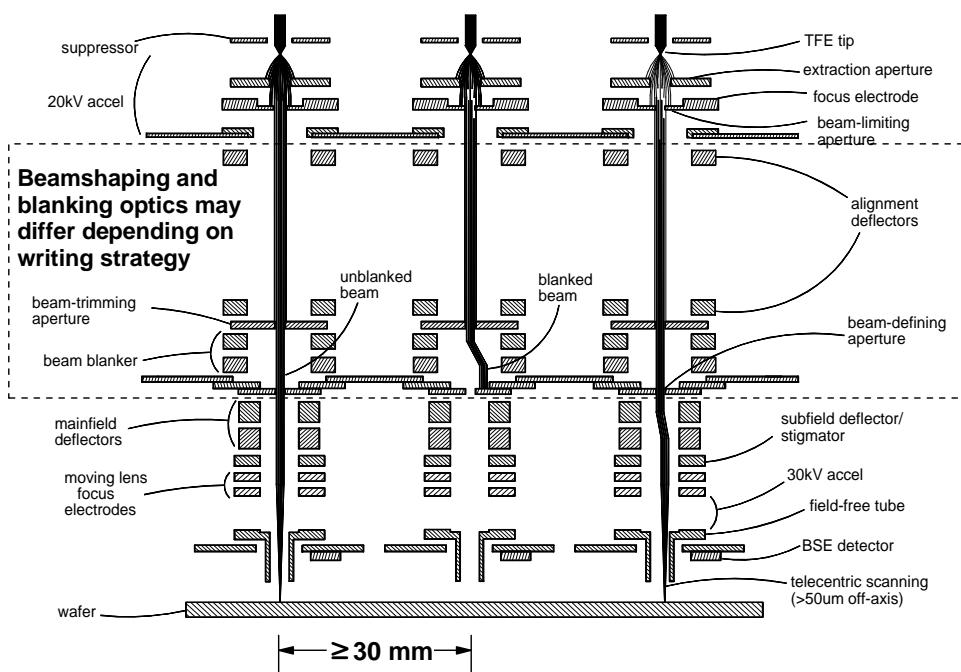


Figure 8.37 One-block EBDW columns from MBS. (Image courtesy of N. W. Parker.)

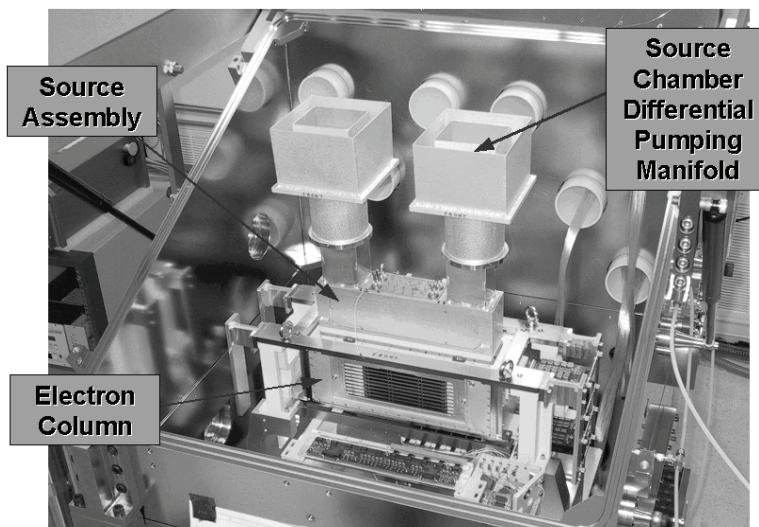


Figure 8.38 Microcolumns from MBS. (Image courtesy of N. W. Parker.)

8.3.2.5 Reflected e-beam lithography

The reflected e-beam lithography (REBL) system⁴⁸ shown in Fig. 8.41 takes advantage of a magnetic prism, which bends the beam by 90 deg. The e-beam is accelerated to an intermediate voltage to enter the magnetic prism. After bending, it is collimated and decelerated to the order of 2 volts (V) before reaching the dynamic pattern generator (DPG), which is basically a CMOS circuit with its last metal layer facing the magnetic prism. Each metal pad is programmed to switch between 0 and 2 V. The pads at 0 V absorb the incoming electrons, while those at

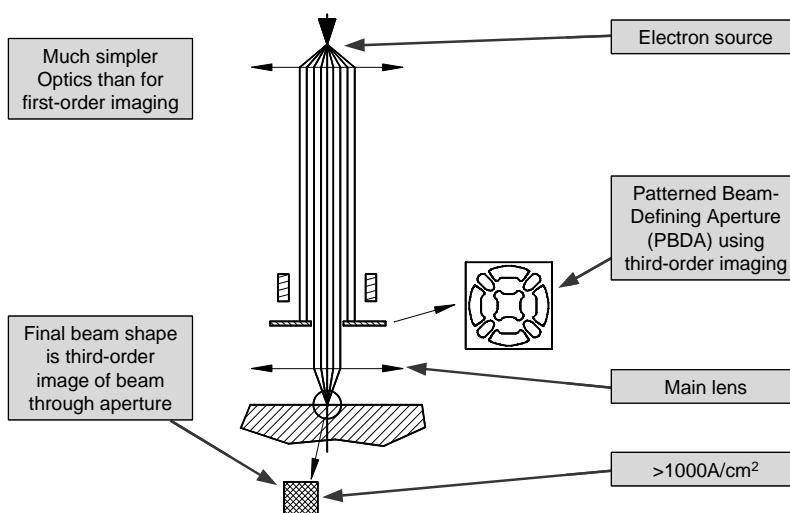


Figure 8.39 Third-order beam shaping. (Image courtesy of N. W. Parker.)

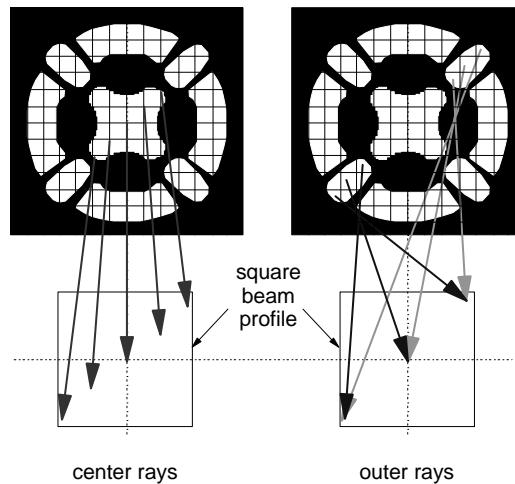


Figure 8.40 Working principle of patterned beam-defining aperture. (Image courtesy of N. W. Parker.)

2 V reflect them. The reflected beam is accelerated back to the intermediate voltage to be bent by the magnetic prism toward the wafer. The beam is now accelerated to the final 50 keV and demagnified to the desired spot size. As many as 2000×2000 pixels can be fabricated on the DPG. REBL is a virtual e-beam projection-printing system, as is Scalpel,⁴⁹ PREVAIL,⁵⁰ and EPL.⁵¹ The main difference is the elimination of a difficult e-beam transmissive mask. The magnetic prism facilitates an e-beam reflective mask that can be programmed at electronic speed.

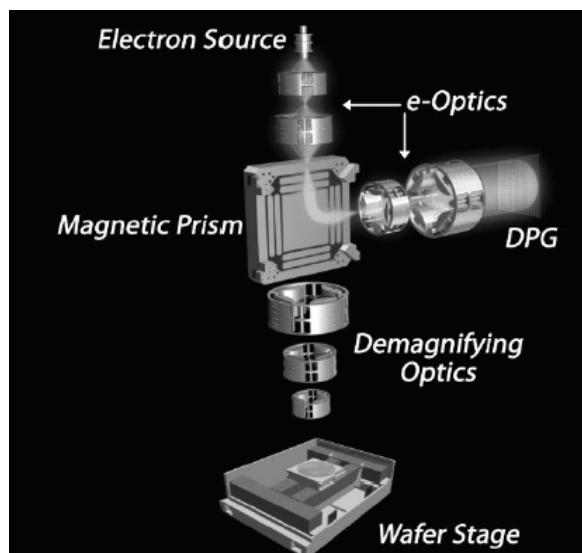


Figure 8.41 The working principle of reflected e-beam lithography (REBL). (Image courtesy of KLA Tencor.)

8.3.3 Comparison of the different systems

There are two major categories of MEB ML2 systems: HV and LV. HV systems employ acceleration voltages of 50 kV or higher, whereas LV systems use 5 kV or lower. High-energy electrons more deeply penetrate the resist and the substrate. Most scatterings take place inside the wafer instead of in the resist. Therefore, an e-beam proximity effect is small. Due to good penetration through the resist, HV systems can use resist thicknesses similar to those in optical lithography. An LV system can only penetrate a few tens of nanometer in the resist. Bilayer resist systems should be used when a thicker resist image is needed for pattern transfer after the lithography step.

However, at a high current, the power, thus the heat, incident to the resist can be substantial. Take $30 \mu\text{C}/\text{cm}^2$ as the resist sensitivity, at ten 300-mm wafers per hour, in which 65% of the time is for exposure to cover a 700-cm^2 area, and $90 \mu\text{A}$ is needed on the wafer. A 100-kV system imposes 9 watts (W) on the wafer; a 5-kV system only imposes 0.45 W.

Even so, wafer/resist heating for 5 keV e-beam systems at 10 wph is better than ArF immersion systems at 200 wph. The 50 keV e-beam system heats the resist even less. However, most of its power goes into the substrate and may create a problem there. Table 8.9 shows the incident power on the wafer, in the resist and through the resist, from 5-keV and 50-keV e-beam exposure, as well as from ArF exposure. The throughput is 10 wph for the e-beam systems and 200 wph for the ArF system. Using a resist thickness of 50 nm, the power in the resist of the 5-keV system is estimated to be 10% of the power on the wafer, resulting in $4 \text{ mW}/\text{cm}^2$ in the resist and $35.5 \text{ mW}/\text{cm}^2$ in the wafer. The power in the resist is 80% of that of the ArF system, whereas the power through the resist into the wafer is only 36% of that of the ArF system. With the 50-keV e-beam, we estimate that 0.1% of the incident power is absorbed by the resist. Hence, the power in the resist and through the resist is 8% and 400% of those of ArF exposures, respectively, amounting to $0.4 \text{ mW}/\text{cm}^2$ and $394.3 \text{ mW}/\text{cm}^2$.

In terms of alignment, HV systems may use their own beam to interrogate the alignment marks or employ conventional optical means. LV systems, with their limited ability to penetrate the resist, will most likely depend on conventional optical means. The MAPPER system seems to adopt readily to optical scanners with twin stages, off-axis alignment, wafer mapping, etc.

Table 8.9 Power incident on the wafer, in the resist, and through the resist.

Voltage on wafer kV	Resist sensitivity $\mu\text{C}/\text{cm}^2$	Thruput (wph)	Exposure time/field (sec)	Pattern density	Energy on wafer (mJ/cm ²)	Power	Power	Power
						on wafer	in resist	thru resist (mW/cm ²)
5	$60 \mu\text{C}/\text{cm}^2$	10	3.80	50%	150	39.5	4.0	35.5
50	$60 \mu\text{C}/\text{cm}^2$	10	3.80	50%	1500	394.7	0.4	394.3
ArF	$30 \text{ mJ}/\text{cm}^2$	200	0.14	50%	15	104.2	5.0	99.2

The systems can be further categorized in terms of the number of beams, sources, columns, and even wafers. Single-beam systems are definitely not of interest here because of low wafer throughput. Therefore, all systems considered here use multiple apertures and blanking. Multiple-source systems have the flexibility in beam placement for better writing schemes and reduction of the space-charge effect between beams. There is also the potential to get more current on the resist. Electrons do not must be wasted by being blocked at the beam-defining areas. However, source uniformity, stability, and lifetime are of concern. MCA seems to be a good scheme to overcome some of the problems. Using a single source to flood expose a beam-defining plate is more popular among the systems proposed because there is a lot of know-how in single-source technology, and not much breakthrough is required to satisfy the ever-increasing throughput need. For some systems with a single source, such as the MAPPER system, high source brightness helps to increase throughput.

Similarly, a single column takes advantage of well-established e-beam imaging systems, but it cannot cover the full $26 \times 33 \text{ mm}^2$ field. The beams tend to distribute equally in x and y , covering a fraction of a millimeter in either direction. In the scan direction, the number of beams covering the same feature can be redundantly large, even though they can be used to fine tune the dimension of the features. Multiple columns have more flexibility in layout, if one is not too concerned about wasting the circular output of a single source. The combination of multiple columns and multiple sources is the least wasteful. The multiple-column systems usually distribute the columns to cover the entire 26-mm field width. Multiple conventional columns, instead of MEMS-fabricated columns, are less compact and less flexible. The number of these columns also tends to be limited.

Some systems with small electron optics may use the cluster concept to apply parallelism to wafers. Figure 8.42 depicts this concept by showing the footprint of a ten-exposure head cluster in a compact layout compared with the footprint of an ArF immersion scanner and an EUV scanner. The light source unit of the ArF is placed at the subfloor, and the data-processing equipment for the ML2 system is also placed in the subfloor. Besides compact writing heads, the wafer stage also needs to be small to enable clustering. Fortunately, the size of the wafer stage is strongly dependent on the required scanning speed. For a scanner operating at a wafer throughput in excess of 100 wph, the stage speed can approach 650 mm/sec, and the reticle stage at 4X needs a top speed of 2600 mm/sec. High acceleration and deceleration are needed to support the high top speed, necessitating a large stage-travel distance. With reticle scanning, the scan direction must be reversed to cover the reticle, whereas in a maskless system, there is no need to reverse the scan direction after scanning each field. Figure 8.43 shows the scanning direction of MEB ML2 scanners and ArF immersion scanners. There are many more reversals in direction with the ArF scanner. For the MEB ML2 system, the small number of scan reversals can be completely removed if parallelism is applied to the exposure fields. For example,

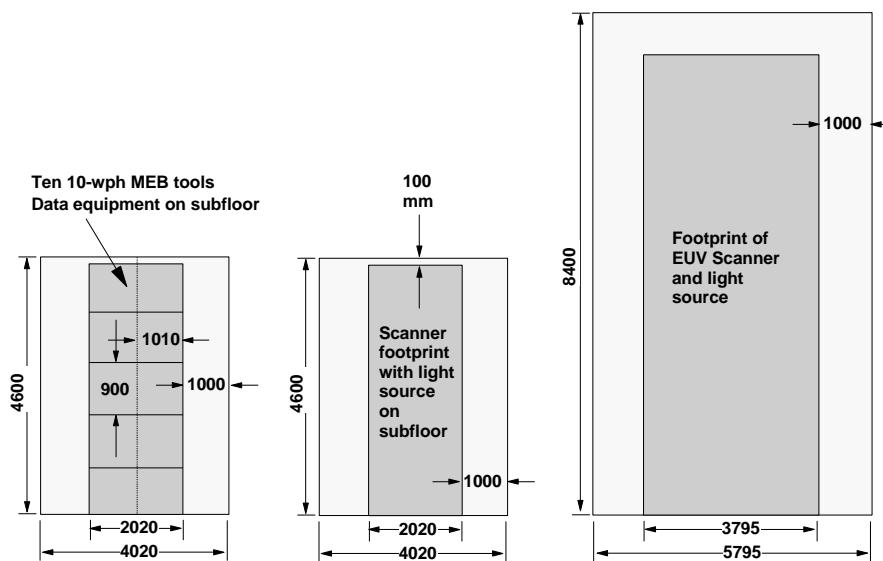


Figure 8.42 The footprint of an MEB ML2 cluster compared with an ArF immersion scanner and an EUV scanner.

if all 11 fields in a row can be covered by 11 exposure heads and scanned simultaneously, there is no need to reverse the scan direction at all. The cluster shares one single high-speed track, and the data-processing equipment further enhances the potential to lower the cost of MEB ML2. For the MBS system, adding parallelism can increase the throughput by another factor of 10. However, the microcolumns, even though quite economical, are more expensive than the MEMS optics.

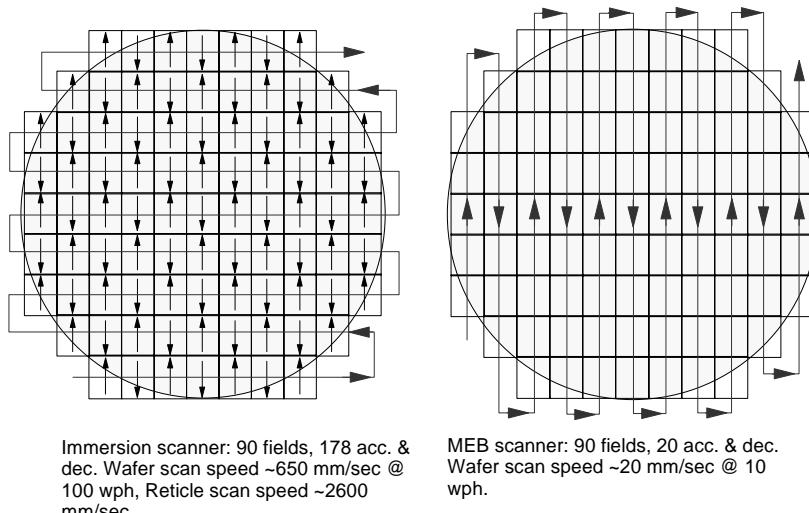


Figure 8.43 The stepping sequence of MEB scanners and ArF immersion scanners.

All MEB ML2 systems must be concerned with stitching. High-accuracy stitching is needed between beams and between subfields. The latter is required for systems that cannot cover the whole field width. Beam position stability and control are very critical. E-beam systems tend to be contaminated while writing on resist. Contaminated e-beam systems induce charging and make the beams behave erratically. LV systems with electron optics very close to the wafer are more susceptible to contamination.

Controlling the massive number of beams or pixels so that the CD and overlay are within specification is the least-known part of the process. Data flexibility enables the correction of systematic deviations. Unpredictable deviations may come from contamination and electric or magnetic noise; they must be eliminated or isolated.

To ensure that the electron optics are working properly and that there is no miswriting, quick inspection of the electron optics or the wafer image is necessary, especially when the technology is new. As confidence is gained, the frequency of inspection can be reduced. Experience in mask writing indicates that miswriting rarely occurs. Most mask defects are induced by processes such as developing, etching, cleaning, and repair.

8.3.4 Data volume and the rate of DW systems

The data rate of a high-throughput MEB ML2 system is another concern. MEB ML2 needs a high data rate and expensive data-handling hardware. Take a 300-mm wafer filled with 90 $26 \times 33\text{ mm}^2$ fields for the system described in Ref. 43 to produce 10 wph. With a $2.25 \times 2.25\text{-mm}^2$ pixel size and a 10% overscan, there are 190 T-bits or 21.2 T-bytes. To achieve a throughput of 10 wph with a nonexposure time of 0.2 sec per field, the exposure time per field is 3.8 sec, leading to a data rate of 3.8 Gbps (gigabits per second). The existing data rate is 10 Gbps for fiber transmission, slightly higher than 3 Gbps for field-programmable gate array (FPGA), but it can be raised to over 10 Gbps with an application-specific integrated circuit (ASIC). Perhaps the bottleneck is in the Si-based CMOS sensor for the blanker array, which has the data rate of only 1 Gbps. Higher-rate Si-based CMOS blanking, GaAs-based blanking, or a data-compression scheme must be implemented. Much of the cost of an optical scanner is now transferred to data-processing equipment. Fortunately, the cost reduction of data-processing equipment predictably follows Moore's law.

8.3.5 Power consumption of MEB ML2

For MEB ML2, 120 kW is estimated for the cluster of 10 sources and exposure units. The data path to support this cluster of 10 without sharing is estimated to require 250 kW. In it, 15 kW is attributed to the dynamic random access memory (DRAM) to store two sets of 23 terabytes of writing data compressed by 10X. In addition there are 1.65 kW for 13,000 optical transceivers and 1.1 kW for 1,100 FPGA, each supporting 12 optical transceivers. To be conservative, 7.25 kW is added to the total wall power for the data path to support each e-beam column.

With 10 e-beam columns, the data path uses 250 kW of wall power. In addition, each source and column requires 12 kW of wall power. Hence, the 10-column cluster needs 370 kW of wall power in total. Running 59 100-layer/h clusters requires 14.6% of the fab power. Sharing the data path by 5X reduces the wall power to 6.8% of fab power, similar to the power required of the immersion scanners. For the power consumption of ArF and EUV scanners, revisit Sec. 8.2.2.1 above.

8.3.6 Extendibility of MEB ML2 systems

The throughput of MEB ML2 can be severely impacted at each node advance. Inherently, the data volume per field increases twofold. The writing speed must be doubled to maintain the same throughput. When the resist sensitivity reaches the shot-noise limit, further reducing the feature size for the next generation requires reducing the resist sensitivity to maintain the same number of electrons for exposure; throughput is further reduced. The resist thickness must be reduced for each generation because e-beam scattering is the limiting factor of e-beam resolution. There are not many known methods to reduce e-beam scattering except by reducing the resist thickness, atomic number of the resist, or optimizing the acceleration voltage for scattering. Out of the three possibilities, only thickness reduction is effective. Therefore, at each node advance, the resist thickness must be reduced. Increasing the etch resistance for pattern transfer is required.

In comparison, optical systems need to increase the NA for higher resolution. Increasing NA inevitably reduces DOF, which also calls for a smaller resist thickness. Increasing NA in EUV may require adding more mirrors in the imaging optics, thus inducing more power loss, leading to throughput loss.th At the shot-noise limit, photon-based systems also suffer from needing to reduce the resist sensitivity for the next generation, which also decreases throughput. The resist height-to-width limit imposed by the resist collapse also calls for a reduction in the resist thickness in photon-based systems.

8.3.7 Comparison of 4X mask writing to 1X wafer writing

There have been concerns about the feasibility of writing 1X patterns on wafers, while 4X mask writing is reaching its limit. Actually, the difficulty is of the same order, if not easier for 1X wafer writing. Take a 32-nm line. To control the line to $\pm 10\%$ requires a CD tolerance of ± 3.2 nm. After applying a 60% CD allowance from the total CD budget, the CD tolerance on the 4X mask is ± 7.7 nm. With MEEF = 2 for line-space, $CD_{mask\ tol} = \pm 3.9$ nm. With MEEF = 4 for contact holes, $CD_{mask\ tol} = \pm 1.93$ nm; MEEF = 10 for line ends, $CD_{mask\ tol} = \pm 0.77$ nm. There are jigs and jugs to pattern on the 4X mask, whereas simple patterns can be directly written on the wafer, even after e-beam proximity correction.

8.3.8 Applications for MEB ML2

We now discuss foreseeable applications for MEB ML2. Bear in mind that e-beam proximity correction is much simpler than OPC. Cycle-time reduction using MEB ML2 is appreciable.

- *The main workhorse for a 32-nm half pitch and below*—From the preceding arguments, it is feasible to make MEB ML2's cost, space, and power comparable to current workhorses, such as ArF immersion scanners or EUVL. When the foregoing concerns are overcome, MEB ML2 will be attractive for applications at any number of wafers per mask, regardless of logic, memory, or flash manufacturers. The savings in mask cost and cycle time, as well as the elimination of the nuisance of mask pellicles, mask contamination under high-energy exposures, and electrostatic discharge, are added bonuses.
- *Next-node technology development*—Presently, technology development is gated by the availability of the exposure tool. With MEB ML2 tools tuned for next-generation lower-throughput exposure, technology development can start much earlier.
- *Emergency fix at production lines*—Firefighters on production lines generally look at OPC for quick fixes. However, it still takes time to make a mask with the problem-fixing OPC. It will be wise to pass the problematic lot with DW exposure so that sufficient time can be taken for a thorough fix of the root cause. In this case, there are tradeoffs between MEB ML2 and optical DW. The former is quicker in proximity correction turnaround, but special e-beam processes must be predeveloped and ready to go. Actually, the latter also needs special process development, albeit a little closer to the base-line processes.
- *Prototyping*—Prototyping with ML2 has always been the primary application endorsed by ML2 promoters because of the gain in time, flexibility, and savings in mask cost. Here, we consider it just one of the applications.
- *Mask writing*—Mask writing has been expensive because of the small market size for mask making. By taking away some mask-making opportunities, MEB ML2 propels mask making into more expensive regimes. The only way to remedy the situation is also to make the ML2 machine a mask writer to spread the development cost of either system. A side benefit is the capability to debug and fine tune a mask writer with wafers first. Mask blanks are just too expensive for such purposes.
- *Exposing critical layers for small lots*—Many masks were made to expose less than 100 wafers. In this case, ML2 can be much more economical. It also facilitates a quicker start for wafer exposures.

Most of the above applications are more suitable with HV systems because they can share the same resist with the optical system. On the other hand, the

potentially higher throughput of LV MEB ML2 systems make them a formidable cost saver for mainstream manufacturing of semiconductors.

8.3.9 Summary of MEB ML2

MEB ML2 shows some promise in being a successor to optical lithography. It has inherent high-resolution potential and a large depth of focus. With the advent of MEMS technology, massively parallel MEB systems are becoming more feasible. Many schemes can spread out the beams to reduce space-charge effects. Multiple sources add flexibility in placement and potentially more current to the wafer. However, e-beam systems are known to be more difficult to control. Contamination and charging always must be carefully prevented and closely watched. Stitching and overlay are not straightforward. E-beam source brightness is still wanting for high-volume throughput. MEMS-fabricated systems must be made to work. Vacuum systems are larger and more costly than atmospheric ones. It is more time consuming and expensive to maintain and repair vacuum exposure systems. Furthermore, the loss of momentum to optical lithography makes it difficult to acquire funding to develop robust, manufacture-worthy MEB ML2 systems.

8.4 Outlook of lithography

As discussed in the last chapter, optical lithography is marching toward the end of its legendary longevity in being a semiconductor-manufacturing workhorse. After using the 193-nm wavelength and water immersion, polarized illumination, and improved masks, the only way to stretch optical lithography farther is to split the pitch by double patterning to reduce k_1 to 0.15 from 0.3, suggesting a half pitch of 21.5 nm. However, starting from a 32-nm half pitch, there are opportunities for other technologies to compete with 193-nm immersion.

We now compare the challenges of optical lithography as it faces its end and its possible contenders, which consist of MEB ML2 systems, EUV, and nanoimprint.

8.4.1 Double patterning

Double patterning uses two exposures and two subsequent processing steps consisting of resist coatings, baking, developing, cleanings, etchings, etc., to delineate the pattern of a semiconductor layer. The cost more than doubles because the second pattern-transfer cost is now attributed to the lithography cost.

The pattern must be split into two to make two masks. Therefore, the pattern must be split friendly. Otherwise, more than two masks must be used. Figure 8.44 shows simple contact-hole and line-space patterns where the distances a , b , and c cannot be resolved using a single mask. One of the two split masks still has a forbidden distance. Splitting the exposure also requires tightening of the overlay between the two exposures.

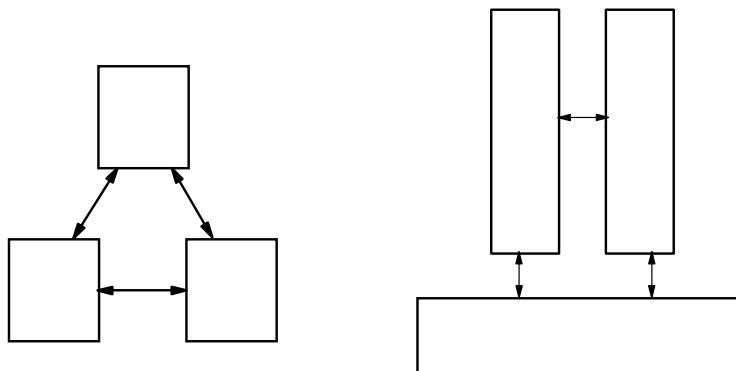


Figure 8.44 Patterns that need to be split into more than two masks.

In order to reduce cost, 1coat-2expo-1etch (1C2P1E) and 2coat-2expo-1etch (2C2P1E) processes are being developed. The former is the least expensive, but it relies on a hypercontrast resist to suppress the cross contamination of residual exposures. Such a resist must be developed. The latter method develops the first resist image, freezes it, and then coats it with the second resist. Many variations of the 2C2P1E scheme are being intensively developed. Some of the schemes may not realize substantial savings because the exotic materials and required treatments tend to be expensive.

In short, double patterning is not only costly, but also requires more design-rule restrictions to make the circuit layout split and alignment friendly. Cost reduction from the processing side requires further development.

8.4.2 EUV lithography

EUV lithography promises single-exposure, simple OPC and has commanded worldwide development efforts at an unprecedented running cost in addition to hundreds of million dollars of sunk cost.

Will it be cost effective? The system is inevitably huge due to the need for high-speed reticle and wafer stages in vacuum. No doubt, the production-worthy hardware will be extremely expensive. The key to suppression of wafer exposure cost is raised throughput. With resist sensitivity staying at a reasonable level in the order of 30 mJ/cm^2 to reduce line-edge roughness and shot-noise effects, the only avenue for high throughput is a high-source power, which also tends to be expensive and environmentally unfriendly, if it can even be made to work.

EUV resists may not provide cost relief. Even if the material production costs were contained, the EUV exposure tools that resist companies will need to purchase inevitably adds to the cost of EUV resists. The high cost of a defect-free EUV mask substrate is well known. The cost of a mask inspection may become a significant factor because actinic light is needed. The cost of yield loss and efforts to prevent it, due to exposure without a pellicle and radiation-induced mask contamination, has yet to be determined from field exposures at manufacture-worthy throughput.

In short, even if every technical problem in EUV lithography was solved, the high exposure cost and high raw-energy consumption may prove detrimental.

8.4.3 MEB ML2

E-beam maskless lithography also has decades of history. It has not been the industry's workhorse for wafer imaging, except for making personalized interconnects in the earlier days at IBM. With circuit complexity ever increasing, the improvement of throughput, even though impressive by itself, could not catch up. E-beam directwrite did become a workhorse for mask making.

To increase the e-beam throughput to a cost-effective level, a high degree of parallelism must be used. It can be in the number of minicolumns, the number of MEMS-fabricated microlenses, pixels in a programmable mask, or numbers of MEMS-fabricated apertures. The exploration of parallelism need not stop here. Some of the above can be favorably combined. Wafers can be exposed simultaneously in a cluster. Without clustering, throughputs on the order of 10 to 40 wph have been claimed. Some are dependent on pattern density due to space-charge-limited current. Some are not.

Even though e-beams have extremely high resolution and depth of focus, scattering in a resist spreads a well-focused beam and fundamentally limits resolution. The only way to reduce scattering is to use a smaller resist thickness. It is foreseeable that resist thickness on the order of 50 nm or less must be used. A multilayer resist system or highly selective hard mask must be used.

The data rate of MEB ML2 is quite demanding. Heavy-duty data processing equipment is needed. Much of the cost in optics for an optical scanner is transferred to data-processing equipment. Fortunately, the cost reduction of data-processing equipment predictably follows Moore's law.

8.4.4 Nanoimprint lithography

Nanoimprint, the lensless replication technique, seems to have unlimited resolution capability and the molding tool costs mere pennies. Nanoimprint has taken contact printing to a higher level. Instead of letting the mask contact the photoresist while still occupying its own space, the imaging medium is mingled with the molding template, i.e., the mask in nanoimprint. Just like contact printing, nanoimprint is susceptible to defects. Mingling the medium with the mold does not offer promise of fewer defects.

The mold has a much shorter lifespan than the optical mask, but child molds can be replicated effortlessly using the same molding technique. Grandchildren can also be produced. Children and grandchildren must be stored and managed.

The template polarity is changed from mother to children, and again for grandchildren. The preferred stud polarity is reversed to the less-preferred pit polarity. Figure 8.45 depicts the three mold generations. The preferred stud form is chosen for the mother mold. As a result, the grandchild mold has the same polarity as the mother mold.

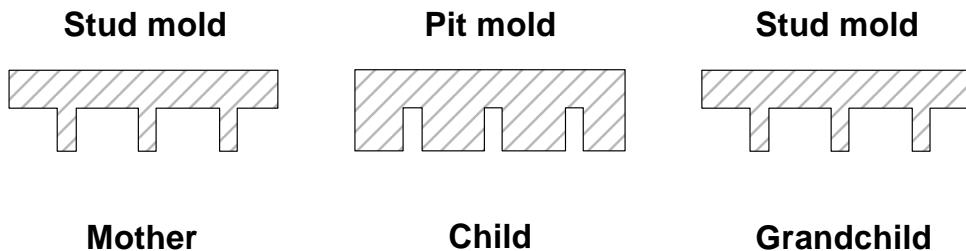


Figure 8.45 Mother, child, and grandchild molds.

Molding is a paced process. Pressing the template into the fluid takes time in order to overcome the resistance from the viscosity of the fluid and to avoid creating bubble-producing turbulence. Waiting for bubbles to dissolve before molding takes time. Waiting for the fluid to solidify also takes time. And finally, pulling out the mold is a slow process. Figure 8.46 depicts schematically the four causes of time consumption in nanoimprint technology.

Producing a uniform residual film under the bottom of the template does not take time, but a rather stringent specification for the flatness and roughness of the mold and the wafer as well as the forbiddance of nanoparticles under the template. In addition to the above intratemplate unevenness of the residual film, the thickness of the film between moldings is also difficult to control. Figure 8.47 shows the situation of the intra- and interfield residual film.

8.5 Conclusions

EUVL systems have the largest momentum and highest expectation to succeed. However, the list of challenges is still long. The higher the development cost, the more financial burden the industry will bear, regardless of its success or failure.

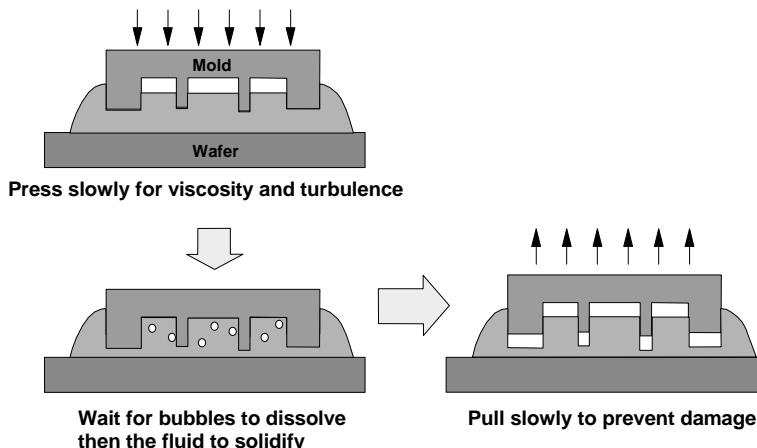


Figure 8.46 Nanoimprint time-consuming causes.

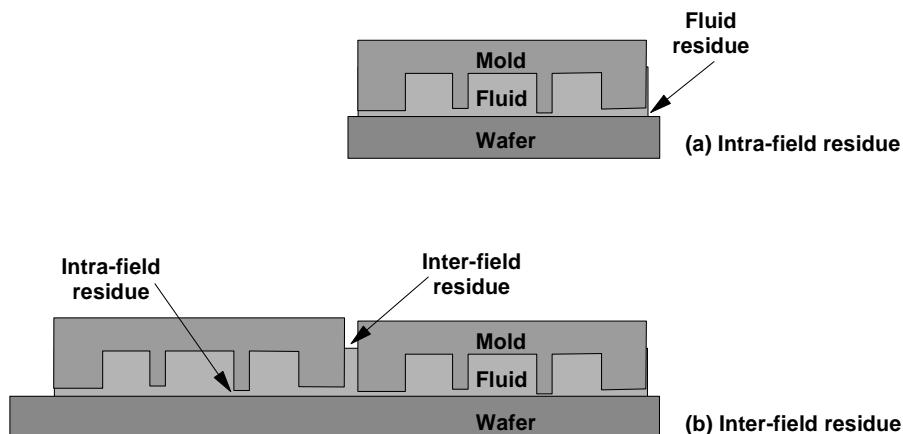


Figure 8.47 Intra- and interfield residue control in nanoimprint.

MEB ML2 systems have the potential to be cost effective. They also have a long list of challenges; however, it is worth a try based on the relatively low cost to develop, the potential of becoming a workhorse, and for the absence of an increasingly difficult mask infrastructure.

Nanoimprint holds the promise of high resolution and a high height-to-width aspect ratio. Defects, template fabrication, template lifetime, throughput, and cost remain tough problems to solve.

It is not inconceivable that ArF water-immersion will be holding the last frontier for lithography, especially if double-mask exposure and high-contrast resists are realized. There may be more lithography-friendly designs, better semiconductor devices, and circuit optimization for the last frontier.

References

- 1 B. J. Lin, "Signamization," *Proc. SPIE* **2726**, pp. 71–81 (1996).
- 2 B. J. Lin, "Simulation of optical projection with polarization-dependent stray light to explore the difference between dry and immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**, pp. 9–20 (2004).
- 3 H. Kinoshita, K. Kurihara, Y. Ishii, and Y. Torii, "Soft x-ray reduction lithography using multilayer mirrors," *J. Vac. Sci. Tech. B* **7**, p. 1648 (1989).
- 4 J. E. Bjorkholm, J. Bokor, L. Eichner, R. R. Freeman, T. E. Jewell, W. M. Mansfield, A. A. Macdowell, E. L. Raab, W. T. Silfvast, L. H. Szeto, D. M. Tennant, W. K. Waskiewicz, D. L. White, D. L. Windt, O. R. Wood II, and J. H. Brunning, "Reduction imaging at 14 nm using multilayer-coated optics: Printing of features smaller than 0.1 μm ," *J. Vac. Sci. Tech. B* **8**, p. 1509 (1990).
- 5 D. Attwood, *Soft x-rays and extreme ultraviolet radiation*, Cambridge Univ. Press, p. 404 (1999).

- 6 D. A. Tichenor, G. D. Kubiak, W. C. Replogle, L. E. Klebanoff, J. B. Wronosky, L.C. Hale, H.N. Capman, J.S. Taylor, J.A. Folta, C. Montcalm, R. M. Hudyma, K. A. Goldberg, and P. Naulleau, "EUV engineering test stand," *Proc. SPIE* **3887**, pp. 48–69 (2000).
- 7 D. A. Tichenor et al., "Extreme Ultraviolet lithography machine," U.S. Patent 6,031,598 (2000).
- 8 H. Meiling, J. Benschop, U. Dinger, and P. Kurz, "Progress of the EUVL alpha tool," *Proc. SPIE* **4343**, pp. 38–50 (2001).
- 9 K. Ota, K. Murakami, H. Kondo, T. Oshino, K. Sugisaki, and H. Komatsuda, "Feasibility study of EUV scanners," *Proc. SPIE* **4343**, pp. 60–69 (2001).
- 10 H. Meiling, V. Banine, N. Harned, B. Blum, P. Kurz, and H. Meijer, "Development of the ASML EUV alpha demo tool," *Proc. SPIE* **5751**, pp. 90–101 (2005).
- 11 G. Schriever, M. Rahe, W. Neff, K. Bergmann, R. Lebert, H. Lauth, and D. Basting, "Extreme ultraviolet light generation based on laser produced plasmas (LPP) and gas discharge based pinch plasmas: a comparison of different concepts," *Proc. SPIE* **3997**, pp. 162–168 (2000).
- 12 J. Pankert, R. Apetz, K. Bergmann, G. Derra, M. Janssen, J. Jonkers, J. Klein, T. Krucken, A. List, M. Locken, C. Metzmacher, W. Neft, S. Probst, R. Prummer, O. Rosier, S. Seiwet, G. Siemons, D. Vaudrevange, D. Wagemann, A. Weber, P. Zink, P. Zink, and O. Zitzen, "Integrating Philips' extreme UV source in the alpha-tools," *Proc. SPIE* **5751**, pp. 260–271 (2005).
- 13 D. W. Myers, I. V. Fomenkov, B. A. M. Hansson, B. C. Klene, and D. C. Brandt, "EUV source system development update: advancing along the path to HVM," *Proc. SPIE* **5751**, pp. 248–259 (2005).
- 14 S. Bajt, Z. R. Dai, E. J. Nelson, M. A. Wall, J. Almeda, N. Nguyen, S. Baker, J. C. Robinson, J. S. Taylor, M. Clift, A. Aquila, E. M. Gullikson, and N. V. G. Edwards, "Oxidation resistance of Ru-capped EUV multilayers," *Proc. SPIE* **5751**, pp. 118–127 (2005).
- 15 M. Kureishi, R. Ohkubo, M. Hosoya, T. Shoki, N. Sakaya, H. Kobayashi, O. Nozawa, Y.-I. Usui, and O. Nagarekawa, "Development of low damage mask making process on EUV mask with thin CrN buffer layer," *Proc. SPIE* **5751**, pp. 158–167 (2005).
- 16 H. Seitz, F. Sobel, M. Renno, T. Leutbecher, N. Olschewski, T. Reichardt, R. Walter, H. Becker, U. Buttgereit, G. Hess, K. Knapp, C. Wies, and R. Lebert, "Recent results on EUV mask blank multilayers and absorbers," *Proc. SPIE* **5751**, pp. 190–199 (2005).
- 17 S. H. Lee, M. Chandhok, J. Roberts, and B. J. Rice, "Characterization of flare on Intel's EUV MET," *Proc. SPIE* **5751**, pp. 293–300 (2005).

- 18 M. Sugawara, M. Ito, T. Ogawa, E. Hoshino, A. Chiba, and S. Okazaki, “Pattern printability of off-axis incident light in EUV lithography,” *Proc. SPIE* **4688**, pp. 277–288 (2002).
- 19 P. Harris and M. MacCallum, “Compensation for imaging errors in EUV lithography,” *Proc. SPIE* **5374**, pp. 833–838 (2004).
- 20 A. Chiba, M. Sugawara, and I. Nishiyama, “Estimation of EUV mask flatness for allowable pattern shift,” *Proc. SPIE* **5037**, pp. 841–849 (2003).
- 21 M. C. Lam and A. R. Neureuther, “A 3-D substrate and buried defect simulator for EUV mask blanks,” *Proc. SPIE* **5751**, pp. 455–465 (2005).
- 22 T. Liang, E. Frendberg, D. Bald, M. Penn, and A. Stiveres, “E-beam repair: fundamental capability and applications,” *Proc. SPIE* **5567**, pp. 456–66 (2004).
- 23 M. Lowisch, U. Dingeer, U. Mickan, and T. Heil, “EUV imaging – an aerial image study,” *Proc. SPIE* **5374**, p. 53 (2004).
- 24 M. Weiss et al., “Progress in EUVL, optics lifetime expectations,” *2nd Int. EUV Lithog. Sym.*, Antwerp, Belgium (2003).
- 25 *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association ITRS (2005).
- 26 B. J. Lin, “NGL comparable to 193-nm lithography in cost, footprint, and power consumption,” *Microelectron. Eng.* **86**, pp. 442–447 (2008).
- 27 G. N. Taylor, R. S. Hutton, S. M. Stein, C. H. Boyce, O. R. Wood II, and D. R. Wheeler, “Extreme UV resist technology: the limits of silylated resist resolution,” *Proc. SPIE* **2437**, pp. 308–330 (1995).
- 28 T. E. Jewell, M. M. Becker, J. E. Bjorkholm, J. Bokor, L. Eishner, R. R. Freeman, W. M. Mansfield, A. A. Macdowell, M. L. O’Malley, E. L. Raab, W. T. Silfvast, L. H. Szeto, D. M. Tennant, W. K. Waskiewicz, D. L. White, D. L. Windt, O. R. Wood II, and J. H. Bruning, “20:1 projection soft x-ray lithography using tri-level resist,” *Proc. SPIE* **1263**, pp. 90–98 (1990).
- 29 B. J. Lin, “Quarter- and sub-quarter micrometer photolithography,” *Proc. Symp. Pattern. Sci. Tech. II*, ECS, **92**(6), p. 3 (1991).
- 30 H. Meiling, V. Banine, P. Kurz, and N. Harned, “Progress in the ASML EUV program,” *Proc. SPIE* **5374**, pp. 31–42 (2004).
- 31 T. Kohler, R. L. Brainard, P. P. Naulleau, D. van Steenwinckel, J. H. Lammers, K. A. Goldberg, J. F. Mackevich, and P. Trefonas, “Performance of EUV photoresists on the ALS micro exposure tool,” *Proc. SPIE* **5753**, pp. 754–764 (2005).
- 32 G. F. Lorusso, I. S. Kim, “Shadowing effect characterization and compensation,” IMEC Program Technical week H1, L111 (2007).
- 33 W. R. Hunter, L. Ephrath, W. D. Grobman, C. M. Osburn, B. L. Corwder, A. Cramer, and H. E. Luhn, “1 μm MOSFET VLSI Technology: Part V – A

- single level polysilicon technology using electron-beam lithography," *IEEE Trans. Electron. Dev.* **ED-26**, pp. 353–359 (1979).
- 34 H. C. Pfeiffer, "Advanced e-beam systems for manufacturing," *Proc. SPIE* **1671**, pp. 100–110 (1992).
 - 35 J. P. Ballantyne, "Electron-beam fabrication of chromium master masks," *J. Vac. Sci. Tech.* **12**, pp. 1257–1260 (1975).
 - 36 J. P. Beasley and D. G. Squire, "Abstract: electromask generator," *J. Vac. Sci. Tech.* **12**, p. 1251 (1975).
 - 37 H. C. Pfeiffer, "The mask making challenge," *Opt. Eng.* **26**, p. 325 (1987).
 - 38 P. A. Allen and P. A. Warkentin, "A comparison of optical Gaussian beam lithography with conventional e-beam and optical lithography," *Proc. SPIE* **1088**, pp. 12–20 (1989).
 - 39 I. Brodie and J. J. Muray, "The physics of microfabrication," Plenum Press, New York, p. 80 (1982).
 - 40 T. H. P. Chang, "Proximity effect in electron-beam lithography," *J. Vac. Sci. Tech.* **12**, pp. 1271–1275 (1975).
 - 41 B. J. Kampherbeek, "High throughput maskless lithography," CPL Workshop, Vienna (Sep. 2005).
 - 42 B. J. Kampherbeek, M. J. Wieland, G. de Boer, G. F. ten Berge, A. M. C. Houkes, R. Jager, T. van de Peut, J. J. M. Peijster, S. W. H. K. Steenbrink, T. F. Teepen, and A. H. V. van Veen, "High throughput maskless lithography," Sematech 2008 Litho Forum (2008).
 - 43 A. S. Tremsin, D. R. Beaulieu, H. F. Lockwood, "Advances in microchannel amplifier for maskless lithography," *Microelectron. Eng.* **83**(4–9), pp. 990–993 (2006).
 - 44 C. Barndstatter, H. Loeschner, G. Stengl, G. Lammer, H. Buschbeck, E. Platzgummer, H.J. Doring, T. Elster, and O. Fortagne, "Projection maskless lithography," *Proc. SPIE* **5374**, pp. 601–609 (2004).
 - 45 H. J. Doering, T. Elster, J. Heinitz, O. Fortagne, C. Brandstaetter, E. Haugeneder, S. Eder-Kapl, G. Lammer, H. Loeschner, K. Reimer, J. Eichholz, and J. Saniter, "Proof-of-concept tool development for projection mask-less lithography (PML2)," *Proc. SPIE* **5751**, pp. 355–365 (2005).
 - 46 C. Klein, E. Platzgummer, H. Loeschner, and G. Gross, "PML2: The mask-less multi-beam solution for the 32nm node and beyond," Sematech 2008 Litho Forum (2008).
 - 47 N. W. Parker, "Maskless lithography using multiple high-current-density, independently controlled shaped e-beams," Sematech 2008 Litho Forum (2008).
 - 48 P. Petric, "Reflective electron beam lithography(REBL), a novel approach to high speed maskless e-beam direct write lithography," *52nd International*

- Conference on Electron, Ion, and Photon Beam Technology and Nanofabrication* (2008).
- 49 S. D. Berger and J. M. Gibson, “New approach to projection-electronlithography with demonstrated 0.1mm linewidth,” *Appl. Phys. Lett.* **57**(2), p. 153 (1990).
 - 50 H. C. Pfeiffer, “PREVAIL: IBM’s e-beam technology for next generation lithography,” *Proc. SPIE* **3997**, p. 206 (2000).
 - 51 K. Suzuki et al., “Nikon EB stepper: its system concept and countermeasures for critical issues,” *Proc. SPIE* **3997**, pp. 214–224 (2000).

Index

- γ , 75
- 157-nm, 166
- 157-nm lithography, 422
- 193-nm, 166
- 1D features, 119
- 2D features, 119
- 2nd-order component, 54
- A, B, C coefficients, 68
- aberrated wavefront, 36
- absorption
- coefficient, 189
 - distribution, 64
 - effect, 187
- active compensation, 317
- adhesion, 239, 251
- advancing contact angle, 378
- aerial image, 29
- measurement system (AIMS), 149
- after-etch image, 308
- air
- bubbles, 371, 376
 - gauge, 224
 - knife, 392
- alignment
- accuracy, 204
 - mark, 204
- all-reflective, 11
- system, 161
- alternating phase-shifting mask
- (AltPSM), 143, 144, 151, 253, 277, 285, 287
- amine, 252
- anisotropic etch, 88, 184
- anisotropical, 167
- antireflection coating (ARC), 2, 190, 215
- bottom (BARC), 3, 190, 351
- top (TARC), 3, 190, 347
- aperture
- limited, 30
 - ratio, 46
- ArF, 136
- astigmatism, 40, 165
- attenuated phase-shifting mask
- (AttPSM), 44, 144, 154, 159, 262
- available DOF, 329, 381, 418
- AZ1350, 179
- azide, 173
- backscattering, 443
- bandwidth
- narrowing, 140
 - requirement, 137
- base polymer, 171
- baseline distance, 205
- basis functions, 313, 318
- beamsplitter, 18, 163, 204
- Bessel function, 46
- best focus, 220
- BigMaC, 56, 245
- mask, 280
- binary
- complex mask (BCM), 144
 - intensity mask (BIM), 143, 151, 158
 - mask, 143
 - phase mask (BPM), 144
- blanket exposure, 179
- bleaching, 181
- Bossung curve, 225

- bottom
 antireflection coating (BARC), 3, 190, 351
 layer, 177
- boundary matrix, 192
- Brewster angle, 347
- bright field, 207
- broadband, 13
- bubbles, 182, 372, 375, 378, 391
- bulk imaging, 175
- bull's-eye illumination, 270
-
- CaF₂, 166, 388
- capacitive gauge, 224
- capillary pull, 251
- capping layer, 429
- CAR, 317
- catadioptric
 lens, 163
 system, 161
- critical dimension
 centered (CDC), 113
 optimizer, 317
 range (CDR), 113
 uniformity (CDU), 309
 improvement, 316
- central scattering bars (CSB), 289
- chain scission, 170
- chemical mechanical polishing (CMP), 296
- chemically amplified resist (CAR), 174, 252
- child molds, 461
- chip, 11
- chromatic aberration, 137, 162
- chrome
 masks, 9
 on glass (COG), 143
- chromium, 145
- coherent
 imaging, 41
 length, 151
- coma, 40, 165
- common
 E-D region, 100
 window, 221
- complex index of refraction, 199
- concentric mirror, 13
- conformal deposition, 89
- conic mating wedges, 142
- constant
 edge-position contours, 99
 line-width contour, 101
- contact angle, 371, 378
- contamination, 437
- contrast enhancement resists, 64
- conversion efficiency, 428
- covered phase-shifting (CPS) edge, 155, 158
- C-Quad illumination, 270
- critical-dimension uniformity (CDU), 309
 improvement, 316
- crosslinking, 171
- crystal axis, 202
- cyclized rubber, 173
-
- D_{eff} , 76
- dark-field alignment, 206, 207
- data rate, 456
- deep submicrometer, 16
- deep UV, 9, 14, 133
 lithography, 144
- defect, 9, 11
 defect source, 240
- depth of focus (DOF), 4
 diffraction, 329, 334
 required, 329, 334
- developed
 image, 67, 188
 resist image, 249
- development effect, 187
- diazoquinone
 novolak, 184
- sensitized novolak (DQN), 171, 175

- diazoquinone (*continued*)
 sensitized novolak resist, 247
 sensitizer, 171
- die-to-database inspection, 11
- die-to-die inspection, 11
- diffraction DOF, 329, 334, 381, 418
- diffraction-limited, 30
 image, 36
 system, 29
- diffusion, 175
- dioptric system, 161
- dipole illumination, 271
- directly sum, 315
- discharge-produced plasma (DPP), 426
- disk illumination (DKI), 217, 270, 276, 277
- dissections, 303
- dissolution rate, 167, 171
 distribution, 65
- distortion, 165
- depth of focus (DOF), 10, 149
 budget, 222
 consumed, 222
 provided, 222
 DOF_{avail}, 329, 381, 418
 DOF_{diffrac}, 329, 334, 381, 418
 DOF_{required}, 329, 334, 382
 sinθ curves, 380
- double
 exposure, 157, 256
 patterning, 459
 telecentric lens system, 147
- double-dipole illumination (DPL), 271
- dyeing, 86, 181
- E95, 137, 138
- e-beam
 directwrite, 23
 proximity effects, 443
- E-D, 5
 branches, 99
 elliptical E-D windows, 110
- E-D (*continued*)
 forest, 106
 region, 99
 tools, 95
 tree, 99, 101, 213
 window, 144, 146, 213, 220
 rectangular, 110
- edge
 bead, 240
 contrast, 152
 correction, 298
 placement, 99
 scattering bars (ESB), 289
- eigenfunctions, 301
- EL-versus-DOF curves, 111
- electric
 field, 31
 power, 438
- electron-beam lens array (EBLA), 448
- electroplating, 91
- elliptical E-D windows, 110
- emulsion mask, 9
 absorber, 145
- environmental chamber, 166
- etalons, 139
- etch, 1, 7, 22
- evanescent waves, 354
- excimer laser, 133, 136, 139
 lifetime, 141
 maintenance, 141
 safety interlocks, 140
- exposure
 center, 221
 field size, 11
 scanning time, 231
 system, 7
 time, 231
 tools, 213
- exposure defocus (E-D), 5
 branches, 99
 elliptical E-D windows, 110
 forest, 106
 region, 99
 tools, 95

- exposure defocus (*continued*)
tree, 99, 101, 213
window, 144, 146, 213, 220
- exposure gap (E-G), 8
- exposure latitude (EL), 4
versus-DOF curves, 111
- extreme ultraviolet (EUV), 163, 418
resists, 438
- F₂, 136
- fabrication tolerance, 257
- field size, 7, 160, 372, 377
- field-by-field alignment, 206
- field-emitter array (FEA), 447
- film nonuniformity, 223
- flare, 167
- flat, 202
- flatness variation, 151
- FLEX system, 64
- focal center, 221
- focal plane–optical proximity correction (FP-OPC), 306
- focus sensing, 224
- focusing error, 224
- forward scattering, 443
- Fourier
integral, 47
transform, 42, 46
- Fresnel
approximation, 8
formulae, 346
number, 8
- full depth–optical proximity correction (FD-OPC), 306
- full-size scattering bar (FSB), 292
- full-wafer field exposure system, 7, 11
- full-width half-maximum (FWHM), 137
- fused silica, 136, 145
- g-line system, 135
- gate length, 158
- ghost lines, 169
- global alignment, 206
- glues, 167
- gratings, 139
- gray levels, 144
- heat, 424, 428, 437, 439, 441, 453
- hexamethyldisilazane (HMDS), 239
- hierarchy, 303
- higher-index
coupling fluid, 401
materials, 401
- higher-order spatial frequencies, 54
- h-line system, 135
- hollow subresolution scattering bar (HSSB), 294
- homogeneous wave equation, 199
- Hopkins, 45
- hot plate, 314, 317
- hydrogen silsesquioxane (HSQ), 444
- hydrophilic, 393
- i-line system, 135
- illuminator, 141
- image plane deviation, 222
- imaging lens, 159
- immersion
developing, 249
fluids, 325
hood (IH), 392
lithography, 323, 325
microscopy, 325
- implant, 7
- incoherent imaging, 44
- indene, 175
- indenecarboxylic acid, 171, 175
- index of refraction, 181
complex, 199
- interferometer, 204
- interferometric lithography, 325
- interfield signature, 311
- inverse Fourier transform, 42, 47

- ion
 implantation, 1, 184
 stopping mask, 167
- isotropic etch, 87, 184
- isotropical, 167
- k_1 , 95, 97
 low, 213
 reduction, 252
- k_2 , 35, 96
- k_3 , 95, 97
- kernel, 44, 300
- KMER, 173
- Köhler illumination system, 141, 142
- KPR, 173
- KrF, 136
- KTFR, 173
- laminar flow, 376
- Langmuir-Blodgett, 239
- laser speckles, 139
- laser-produced plasma (LPP), 426
- latent image, 64, 177, 184, 187
- Legendre polynomials, 312
- lens
 aberration, 165
 distortion, 4
- lens-based configuration (LBC), 367, 398
- Li, 427
- lift off, 1, 7, 89, 184
 mask, 167
- light
 frequency, 324
 rays, 29
 source, 133
 velocity, 324
 waves, 29
- line-edge roughness (LER), 175
- line-end shortening, 295
- linewidth
 derived (LWD), 79
 LWD- β , 84
 roughness (LWR), 440
- local dosage-optical proximity correction (LD-OPC), 304
- log slope, 57
- long micrometer, 16
- low k_1 , 213
- low-contact-area vacuum chucks, 203
- low-pass filter, 47, 53
- lumped parameter model (LPM), 71, 266
- LWD- β , 84
- magnetic field, 31
- MAPPER, 444
- mask, 1
 absorber, 145
 bias, 146
 blank, 143
 error enhancement factor (MEEF), 23, 252, 265
 flatness, 149, 222, 435
 reflectivity, 147
 sagging, 18
 substrate, 144
 transmission function, 46
- masking level, 2
- mask-to-wafer
 distance, 9
 gap, 8, 9
- mating wedges, 142
- matrix formulation, 192
- mechanical shock, 166
- MEEF, 23, 252, 265
- mercury
 arc lamp, 133
 line, 9
- metal-oxide semiconductor (MOS), 158
- microchannel amplifier (MCA), 447
- microloading, 296, 308

- microscopy, 97
- mid-UV, 9, 133
- middle layer, 177
- midmicrometer, 16
- minimum printable feature size, 11
- mirror, 12
 - image, 10, 13
 - scheme, 13
- misnomer, 143
- model-based OPC, 299
- molecular weight, 170
- MoSi, 262
- MoSi₂, 145
- multibeam systems, 450
- multilayer
 - resist systems, 215
 - stack, 422, 428
- multilevel
 - intensity mask (MIM), 144
 - phase mask (MPM), 144
- multiple
 - exposure, 391
 - illumination, 217
 - reflection, 63, 182, 185, 192
- nanoimprint, 461
- near UV, 133, 163
- negative resist, 168, 174, 187
- nitrene, 173
- nonpolar solvent, 174
- nonuniform exposure, 187
- no-PDS (NPDS), 358
- normalized feature size, 99
- notch, 202
- numerical aperture (NA), 4, 31, 48, 159
 - at the mask side, 31
 - NA/ σ combinations, 117
- numerical half aperture (NHA), 35, 328
- O₂ RIE, 177
- off-axis
 - alignment, 205
 - illumination (OAI), 142, 268, 275, 285
- one-component
 - negative resist, 171
 - positive resist, 170
- onium salt sensitizer, 174
- optical
 - CD (OCD) metrology, 230
 - lithography galaxy, 417
 - proximity correction (OPC), 294
 - proximity effect (OPE), 295
- optimum
 - σ , 218
 - NA, 213
- outgasses, 167, 182, 378
- overcut, 189
 - profile, 66, 80, 187
- overlay
 - accuracy, 204
 - budget, 23
- pack-and-seal technique, 385
- pack-unpack, 406
- paraxial, 11
- partial coherent light, 269
- partially coherent, 45
- particulate, 167
- pattern
 - shadowing, 440
 - transfer, 7
- PBOCST, 174
- pellicle, 10, 145, 436
- permeability, 199
- permittivity, 199
- phase
 - error, 130
 - grating focus monitor, 228
 - shift focus monitor, 227
- phase-shifted edge (PS edge), 144, 155

- phase-shifting
conflicts, 257
mask (PSM), 143, 150, 253
subresolution-assisted (SA PSM), 154, 294
PHOST, 174
photoacid, 378
generator, 378
photomask(mask), 142
photon energy, 324, 442
photoresist, 145
phototransfer resist system, 179
pitch splitting, 419
placement
error, 129, 146
tolerance, 23
plating, 7
polar solvent, 174
polarization
dependent stray light (PDS), 340, 350
effects, 325
polarized, 139
polarizing beamsplitter, 164
polychromatic illumination, 206
polymethylmethacrylate (PMMA), 179, 247, 445
portable conformable mask, 179
positive resist, 168, 174, 187
postapplication bake, 2, 244
postdevelopment bake, 243
postexposure bake (PEB), 2, 174, 184, 198, 243, 245
Poynting
theorem, 197
vector, 196
precoating bake, 243
pre-exposure bake, 243
prisms, 139
programmed defect mask (PDM), 149
propagation
angle, 48
matrix, 192
protons, 174
proximity
dispense-in-vapor, 242
printing system, 7
PSM, 143, 150, 253
subresolution-assisted (SA PSM), 154, 294
puddle developing, 249
pupil function, 44
quadrupole, 270, 279
illumination (QRI), 277, 283, 287, 292
quartz, 144
radiation
induced damage, 140
spectrum, 133
random phase shifting, 430
rate image, 188
reactive ion etching (RIE), 177
 O_2 , 177
transfer resist system, 177
receding contact angle, 378
rectangular E-D window, 110
reduction
of spatial frequencies, 152
ratio, 24, 161
step-and-repeat, 7
step-and-scan systems, 7
reflected e-beam lithography (REBL), 451
reflection, 4, 253
variation, 187
refracted image, 67, 188
refractive index, 323, 369, 379
distribution, 64
repeating defects, 145
required DOF, 329, 334, 382
resist
bleaching, 86
coating, 2, 238
collapse, 251
development, 2

- resist (*continued*)
hardening, 3
thickness, 186, 222, 239
- resolution, 10
enhancement technique (RET), 253
scaling equation, 95, 213
- restricted pitch, 289
- reticle, 16
- RIE-transfer resist system, 177
- rim PSM, 154, 262
- ring field, 12
- ring illumination (RGI), 270, 276, 287
- ripples, 283
- root-sum-square, 315
- rule-based OPC, 298
- scattering bars (SB), 288
central (CSB), 289
edge (ESB), 289
full-size (FSB), 292
hollow subresolution (HSSB), 294
subresolution (SSB), 288, 292
- scatterometric CD (SCD)
metrology, 230
- seal ring, 398
- Seidel
aberration coefficients, 36, 165
- sensitizer, 171
- short submicrometer, 16
- shrinkage, 170
- Signamizer, 356, 417
- silica, 136
- silylation, 177
- SiO₂, 166
- slit, 18, 163, 424
- slot, 18, 424
- Snell's law, 326
- solid
angle, 31
immersion mask, 402
- space-charge effect, 443
- spatial
coherence, 139
frequency, 47, 50, 253, 268, 271, 285, 294
- special routing, 394
- speckles, 136
- spherical, 40
aberration, 165, 330
wavefront, 29
- spin coating, 238
- spontaneous emission, 136
- spray developing, 250
- spray-and-spin-in-vapor, 241
- stage acceleration, 231
- standing waves, 183
- step-and-repeat, 14
and-scan, 18, 325
- stepping time, 231
- stimulated emission, 136
- stray
light, 167, 185, 419, 436, 440
reflection, 147
- striation, 240
- strong phase-shifting scheme, 156
- subresolution-assisted phase-shifting mask (SA PSM), 152, 154, 294, 294
- subresolution scattering bar (SSB), 288, 292
- substrate reflection, 185
- system stray light (SSL), 340
- target point, 302
- t-BOC, 174
- telecentric, 161
lens system, 147
- temporal coherence, 139
- thermal expansion coefficient, 144, 147
- third-order beam shaping, 450
- three-beam
imaging, 343
interference, 353

- three-layer system, 177
threshold leveling, 304
throughput, 231
through-the-lens alignment, 205
tin, 427
top antireflection coating (TARC),
 3, 190, 347
topcoat, 379
topography, 177, 187
top-surface imaging (TSI), 177
transfer function, 300
transmission
 error, 130
 function, 44
transverse electric (TE), 42, 341,
 347, 348, 383, 403
transverse magnetic (TM), 42, 340,
 347, 348, 403
twin stage, 236
two-beam interference, 353
two-component
 negative resists, 173
 positive resist, 171
two-layer system, 177

undercut profile, 89
unpolarized, 43
UttPSM, 157

vacuum wavelength, 51, 324
vibration, 253
viscosity, 370

wafer
 backside, 398
 chuck, 202
 edge, 398
 flatness, 4, 202, 223
 stage, 202
wafer-based configuration (WBC),
 368
wall power, 438, 441, 456
wave number, 30
wavefront interferometer, 166
wavelength
 of electrons, 442
 reduction, 4
weak phase-shifting scheme, 156
working distance, 9, 11, 161

x ray, 2
 proximity printing, 9, 422

Zernike
 aberration coefficients, 38
 polynomials, 38, 165, 312



Burn J. Lin has been a senior director at TSMC, Inc., since 2000. He has been the president of Linnovation, Inc., since 1992. Prior to that, he held various technical and managerial positions in the field of microlithography at IBM. He has been extending the limits of optical lithography for close to four decades.

Dr. Lin is the Editor in Chief of the *Journal of Micro/Nanolithography, MEMS, and MOEMS*, a member of the U.S. National Academy of Engineering, an IEEE Life Fellow, an SPIE Fellow, and a visiting professor at the National Taiwan University. He is the recipient of the 2009 IEEE Cleo Brunetti Award,

2009 Benjamin G. Lamme Meritorious Achievement Medal, 2007 Industrial Technology Advancement Award, 2006 Distinguished Optical Engineering Award, 2005 Most Valuable Player in VLSI Research Inc.'s All-Stars of the Chip Making Industry, 2005 Two Best R&D Managers in Taiwan, 2004 Outstanding Research Award from PWY Foundation, 2004 1st recipient of SPIE Fritz Zernike Award, 2003 Outstanding Scientific and Technological Worker Award, and 2002 Ten Best Engineers in Taiwan. Throughout his career, he received 2 TSMC Innovation Awards, 10 IBM Invention Awards, and an IBM Outstanding Technical Contribution Award.

He pioneered deep-UV lithography starting from 1975, multilayer resist systems from 1979, simulation of 2D partially coherent imaging from 1980, exposure-defocus methodology from 1980, scaling equations of resolution and depth of focus from 1986, k_1 reduction from 1987, proofs of 1X mask limitations from 1987, vibration in optical imaging from 1989, electrical measurement of contact holes from 1989, E-G tree for x-ray proximity printing from 1990, experimental demonstration of the impacts of mask reflectivity on imaging from 1990, optimum lens NA from 1990, attenuated phase-shifting mask from 1991, signamization from 1996, LWD- η and LWD- β from 1999, nonparaxial scaling equations of resolution and depth of focus from 2000, 193-nm immersion lithography from 2002, and polarization-induced stray light from 2004. He is currently developing cost-effective lithography technology for the 22-nm node.

He has written 2 book chapters, published over 100 mostly first-authored articles, and has more than 50 U.S. patents.