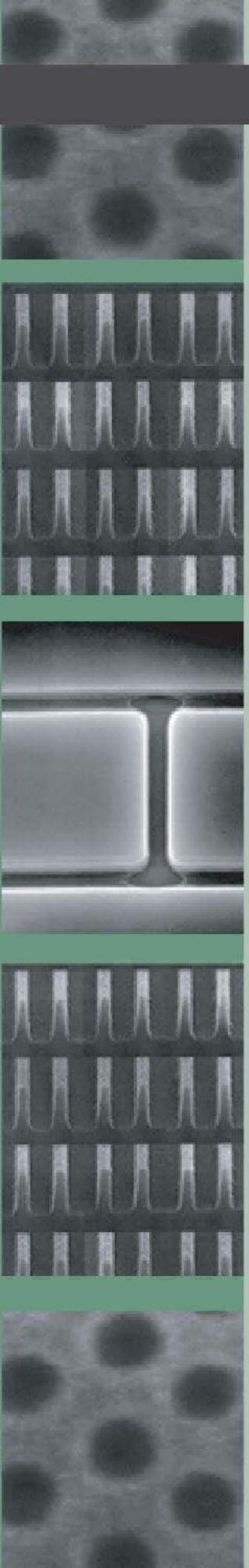


SPIE PRESS

SPIE

Principles of Lithography

Third Edition

Harry J. Levinson

Principles of Lithography

Third Edition

Principles of Lithography

Third Edition

Harry J. Levinson

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Levinson, Harry J.
Principles of lithography / Harry J. Levinson. – 3rd ed.
p. cm. – (Press monograph ; 198)
Includes bibliographical references and index.
ISBN 978-0-8194-8324-9
1. Integrated circuits–Design and construction. 2. Microlithography. I. Title.
TK7874.L397 2010
621.3815'31–dc22

2010026775

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

Printed in the United States of America.

Contents

Preface to the Third Edition

The need to update *Principles of Lithography* only five years after the release of the second edition is evidence of the quickly changing and exciting nature of lithography as applied to the production of integrated circuits and other micro- and nanoscale devices. Since the second edition was written, immersion lithography has moved from development into volume manufacturing, double patterning methods have been developed to maintain scaling using optical lithography, extreme ultraviolet (EUV) lithography has started to move from the laboratory to development pilot lines, and there has been a significant increase in skills in many other aspects of lithography. I hope that the readers of this new edition are satisfied with my efforts to explain the salient points of the newly developed technology. For students, I have added more homework problems.

As a practical matter, not every contribution to lithography described in a publication could be referenced, but I do hope that I have properly recognized the work of others, while providing the reader with a tractable and useful list of references.

I would like to thank the many people who contributed to this latest edition of *Principles of Lithography*, providing figures, papers, and explanations. Specific figures were provided by Lars Liebmann (10.18), David Kyser of Applied Materials (Fig. 7.11), Simi George of Lawrence Berkeley Laboratories (Fig. 12.18), David Brandt of Cymer (Fig. 12.17), Hiroko Takeuchi of Nuflare (Fig. 7.6), Norihiro Yamamoto of Spansion (Figs. 7.26 and 7.27), Bruno La Fontaine of Cymer (Figs. 12.1 and 12.2), Stefan Wurm of GLOBALFOUNDRIES (Fig. 12.6), Wolfgang Holzapfel of Heidenhain (Fig. 5.36), Andreas Erdmann from the Fraunhofer Institute (Fig. 5.45), Brigette Wehrmann and Leah Rickman of Suss Microtec (Fig. 5.48), Kafai Lai of IBM (Fig. 8.34), and Paul Nealey from the University of Wisconsin (Fig. 13.20).

Robert L. Brainard of the State University of New York (SUNY), Albany, provided useful information on resists for next-generation lithography. Conversations with Andrew Hazelton of Nikon regarding linewidth control and double patterning were also very useful. Christian Buergel and Fritz Gans of the AMTC supplied information on mask making and inspection. Anna Tchikoulaeva and Paul Ackmann of GLOBALFOUNDRIES provided constructive suggestions to improve the chapter on masks and reticles. Joe Daggett from Sumika provided useful information on adhesion promotion. Assistance with EUV transmission calculations from Erik Gulliksen of Lawrence Berkeley Laboratories was most appreciated. Discussions with Uzodinma Okoroanyanwu of GLOBALFOUNDRIES were helpful for understanding pinhole formation in ultrathin resists. Yuansheng Ma, also from GLOBALFOUNDRIES, helped significantly with the section on resist line-edge roughness. Vito Dai of GLOBALFOUNDRIES supplied expert advice on data handling for direct-write

lithography. Copies of useful references were provided by Banqui Wu of Applied Materials, John Burnett of NIST, and Bert Jan Kampherbeek of MAPPER. Christian Wagner of ASML helped to clarify some subtle issues regarding EUV lithography. Eli Dagan explained the status of mask-making tools at Applied Materials. Gavin Rider generously provided information on electrostatic damage to masks. Winfried Kaiser of Carl Zeiss helped to clarify issues relating to high-performance optics. Timothy Groves of SUNY, Albany, provided useful explanations and information on aspects of electron-beam lithography. Fabian Pease, Edward Sebesta, and John Burnett pointed out specific errors in prior editions, which I hope have now been corrected.

I continue to appreciate those people who provided figures and information for the earlier editions.

I would like to thank Jerry Sanders, founder of Advanced Micro Devices, whose commitment to excellence enabled many of the ideas in this book to be discovered, learned, and applied by individual engineers in service to the customer.

Finally, I would like to thank my wife, Laurie, for her continuing, saintly patience and support.

Harry J. Levinson
November 2010

Preface to the Second Edition

This book was written to address several needs, and the revisions for the second edition were made with those original objectives in mind. First, and foremost, this book is intended to serve as an introduction to the science of microlithography for people who are unfamiliar with the subject. Most papers written for journals or conference proceedings assume that the reader is familiar with pre-existing knowledge. The same can be said for compilations of chapters written by experts who are providing their colleagues and peers with useful summaries of the current state of the art and of the existing literature. Such papers and books, while quite useful to experienced lithographers, are not intended to address the needs of students, who first need to understand the foundations on which the latest advances rest. It is the intention of this book to fill that need.

For the experienced lithographer, there are many excellent books written on specialized topics, such as photoresist and resolution-enhancement techniques, and I have referenced many of those fine works. However, I have often felt that several topics have not been well addressed in the past; most notably those subjects directly related to the tools we use to manufacture integrated circuits. Consequently, this book goes into a few subjects in depth. These include such topics as overlay, the stages of exposure tools, and light sources. Finally, this text contains numerous references. These are resources for students who want to investigate particular topics in more detail, and they provide the experienced lithographer with lists of references by topic.

A wise leader once told me that one of the most challenging tasks is to transform complexity to simplicity; in other words, to make apparent the forest obscured by all of the trees. I hope that I have succeeded adequately on the subjects covered in this book. Of course, simplicity should never be confused with easiness or completeness. To assist the student in recognizing these distinctions, more problems have been added to the end of each chapter. It is expected that the reader of this book will have a foundation in basic physics and chemistry. No topics will require knowledge of mathematics beyond elementary calculus.

Lithography is a field in which advances proceed at a swift pace, and many new topics have been included in this second edition, commensurate with the learning that has taken place during the past few years, and several subjects are discussed in more detail. Optical proximity corrections and next-generation lithography are examples where the landscape looks quite different than it did just a few years ago. Other topics, such as immersion lithography, were ideas that few took seriously just a few years ago, yet today are considered quite mainstream.

It has been my good fortune to work with a number of outstanding lithographers. In addition to the people acknowledged in the preface to the first edition, I would like to thank several people who contributed to this update. These include Tim Brunner of IBM, Wolfgang Henke of Infineon, Margaret Conkling of Nikon

Precision, Nigel Farrar, Vladmir Fleurov, Palash Das and Charles Hindes of Cymer, Andreas Erdmann of Fraunhofer-Institut für Integrierte Schaltungen, Doug Resnick and John Algair of Motorola, Wilhelm Maurer of Mentor Graphics, Christian Wagner and Robert Socha of ASML, Paul Graeupner of Carl Zeiss, Johannes Nieder of Leica, John Ricardi and Harry Rieger of JMAR, Ray Morgan of Canon, USA, Walter Gibson of XOS, and Sandy Burgan of DNS. Merry Schnell and Sharon Streams of the publications staff of SPIE have been very helpful and supportive. I apologize if I have failed to mention anyone who has helped me with this update.

It has also been a privilege and joy to work on a more frequent basis with some exceptionally outstanding lithographers in my own department, as well as other lithography departments, at AMD. In particular, this includes manufacturing organizations, where the principles discussed in this book have been skillfully applied and expertly enhanced to produce high-performance nonvolatile memory and the world's most powerful Windows-compatible microprocessors. From AMD, I would like to thank Bruno La Fontaine, Jongwook Kye, Ivan Lalovic, Adam Pawloski, Uzodinma Okoroanyanwu, Rolf Seltmann, Wolfram Grundke, and Rick Edwards for useful and informative discussions on lithography. I would like to thank my wife, Dr. Laurie Lauchlan, and my daughters, Sam and Sarah, who continued to exhibit amazing patience while I worked on the second edition of this book. On September 11, 2001, the world witnessed the destructive power of the irrational mind. I hope that this book will be a small reminder of the tremendous capacity of the rational human mind to improve the world around us.

Harry J. Levinson
January 2005

Preface

This book has been written to address several needs. First, it is intended to serve as an introduction to the science of microlithography for someone who is unfamiliar with the subject. It is expected that the reader has a foundation in basic physics and chemistry. No topic requires knowledge of mathematics beyond elementary calculus. The book covers a number of advanced subjects as well, and it can be used by experienced lithographers who wish to gain a better understanding of topics that are not in their own areas of expertise. Numerous references are made to literature in optical lithography, providing a guide for both novice and experienced lithographers who want to learn about particular subjects in detail.

A number of discussions—such as thin-resist modeling, metrics for imaging, thin-film optics, and the modeling of focus effects—first appeared in Advanced Micro Design internal reports. Eventually, some parts of these reports were published elsewhere. Their tutorial nature is not coincidental, as they were analyses that I used to develop my own understanding of lithography. It is often found that complex situations are best comprehended through simple models that describe most of the relevant physics, with the remaining effects considered as perturbations. This is the approach I used in learning lithography myself, and it is the method used in this book. Students of the class on lithography science that I periodically teach will recognize many of the figures and equations. A number of these are also used in the first chapter in the SPIE *Handbook on Microlithography, Micromachining, and Microfabrication, Volume I: Microlithography*. I coauthored that chapter with Bill Arnold of ASM Lithography.

Additional topics have been added or expanded significantly, especially those concerning light sources, photomasks, and next-generation lithography. The chapter on overlay is an expanded version of a similar chapter found in my earlier text, *Lithography Process Control* (SPIE Press, 1999). The chapter on photoresists takes a different approach from that found in most books on lithography. Here, resists are approached from the perspective of the practicing lithographer, rather than the resist chemist. Some resist chemistry is discussed because this knowledge is essential for using resists properly, but the emphasis is on operational considerations.

A number of acknowledgments are in order. First, there are several people at AMD to be thanked. Jongwook Kye provided the data for Figs. 8.8, 10.4, and 10.5. Figure 12.8 shows two micrographs, generated by Janice Gray, of an EUV mask fabricated at AMD. Dr. Uzodinma Okoroanyanwu has provided the author with guidance through the world of resist chemistry. Figure 3.26 was contributed by Marina Plat.

It has been my good fortune to have had the opportunity to interact with many of the world's experts in lithography, and their works are referenced throughout this book. Quite a number of the ideas presented in this book first appeared in

papers that I coauthored with Bill Arnold, who is now Executive Scientist at ASM Lithography. I have also had a long association with Dr. Moshe Preil of KLA-Tencor, and we have written several papers together.

Lithographers at semiconductor companies are integrators. We combine optics, precision machines, photochemicals, and photomasks into working processes. While chip makers often get the glory, the lens makers, resist chemists, and tool makers are the unsung heroes of the microelectronics revolution. It has also been my good fortune to be able to interact with many of the world's experts in optics, equipment, and lithographic materials. A number of people have provided assistance in the writing of this book. Several people at Carl Zeiss (Dr. Winfried Kaiser, Dr. Christian Wagner, Dr. Bernd Geh, and Dr. Reiner Gerreis) have explained elements of lens optics and metrology to me. Bernd Geh donated the data of Fig. 4.6, Winfried Kaiser provided Fig. 5.15, and Figs. 5.21 and 5.25 were provided by Dr. W. Ulrich. Figure 5.2 was obtained from Roy Wetzel of ASM Lithography. Dr. Don Sweeney of Lawrence Livermore Laboratory provided Fig. 5.19 and offered generally useful information on EUV technology, as did Dr. Dan Bajuk, Dr. Saša Bajt, and Dr. David Attwood. Figure 5.9 was obtained from Bob Willard of Lambda Physik. Figure 2.25 was provided by Wes Brykailo of the Shipley Company. The data in Figs. 3.9, 3.10, and 3.11 were contributed by Eddie Lee, Emir Gurer, Murthy Krishna, and John Lewellen of Silicon Valley Group. Figure 3.14 was a donation from Will Conley of Motorola, and Dr. Sergey Babin of Etec provided Fig. 7.14. Dr. Rainer Kaesmaier of Infineon and Dr. Hans Löschner of IMS Vienna were particularly generous in providing information on ion-projection lithography, including Figs. 12.13 and 12.14. Kurt Heidinger of the Cyantek Corporation supplied information on wet chromium etchants. I want to thank Phil Seidel for providing the data of Fig. 11.1. Dr. Chuck Gwyn of Intel contributed Fig 12.10.

Dr. David Joy helped to improve my understanding of SEM metrology. Bill Moffat of Yield Engineering tutored me as a young engineer and more recently on adhesion promotion. The references would be far less complete without the excellent services provided by AMD's librarians, John Owen, Wendy Grimes, and Sharon Shaw. I apologize to anyone who has helped me put together this book and has not been acknowledged.

Finally, I would like to thank my wife, Dr. Laurie Lauchlan, who provided me with a true appreciation of the importance of metrology, and for her apparently infinite patience while I wrote this book. This book is dedicated to Laurie, and our daughters, Samantha and Sarah.

Harry J. Levinson
April 2001

Chapter 1
Overview of Lithography

The patterns of integrated circuits are created on wafers by lithography. The steps of this critical manufacturing process are listed in Table 1.1. Each step will be discussed at length in later chapters of this book, but a brief description of each will be given here. Most of this book is devoted to photolithography, where optical methods are used to transfer the circuit patterns from master images—called masks or reticles—to the wafers. Photolithography is the method used for patterning nearly all integrated circuits fabricated today.

Adhesion promotion: The lithography process creates the patterns of integrated circuits in films of specialized materials called resists, which are coated onto the wafers on which the circuits are made. Resists typically do not adhere properly to untreated surfaces of silicon or silicon-containing materials, such as silicon dioxide and silicon nitride. To ensure proper adhesion, the wafer surfaces are treated prior to resist coating.

Resist coat: Resists are typically comprised of organic polymers applied from a solution. To coat the wafers with resist, a small volume of the liquid resist is first

Table 1.1 The steps in the lithography process. The steps in italic are optional.

Adhesion promotion
↓
Resist coat
↓
Softbake
↓
Alignment
↓
Exposure
↓
Post-exposure bake
↓
Development
↓
Measurement and inspection
↓
Hardbake

dispensed onto a wafer. The wafer is then spun about its axis at a high rate of spin, flinging off the excess resist and leaving behind, as the solvent evaporates, a thin (0.1–2 μm, typically) film of solid resist.

Softbake: After the resist coating has been applied, the density is often insufficient to support later processing. A bake is used to densify the resist film and drive off residual solvent.

Alignment: Integrated circuits are fabricated by a series of patterning steps. These start with a lithography operation followed by an etch or ion implantation. Between patterning steps, there may be film depositions, planarizations, and other processes. Each new pattern must be placed on top of preceding layers, and proper overlay of the new layer to the circuit pattern already on the wafer is achieved during the alignment step by using specialized equipment.

Exposure: Photoresists are materials that undergo photochemical reactions when exposed to light. There are two types of photoresists—positive and negative. Positive resists are normally insoluble in the chemicals referred to as resist developers, but are made soluble by exposure to light. Negative resists have the opposite behavior: they are soluble in developer and rendered insoluble by exposure. By exposing the resist selectively in some areas and not others, the pattern of the circuit can be created in the resist film. This selective exposure is accomplished in optical lithography by the imaging of a mask. Photomasks are sheets of glass, partially covered by an opaque material, usually chromium, that has been removed according to the pattern of the circuit. By shining light onto the mask, and then projecting the transmitted image onto the resist film, the pattern of one layer of the circuit is transferred to the resist film on the wafer.

Post-exposure bake: This is an optional baking step used to drive additional chemical reactions or the diffusion of components within the resist film. The purposes of the post-exposure bake are discussed in Chapters 3 and 4.

Development: This is the step by which a resist is removed, depending upon whether or not it has been exposed.

Measurement and inspection: This is an optional operation where it is determined by measurement if the resist features on the wafer are sized correctly and properly overlay preceding patterns, and that they are sufficiently free from defects. These measurements may be used for the purposes of process control or dispositioning.

Hardbake: This is another optional process. Because wafers with photoresist patterns nearly always go into etch or ion implantation operations immediately following the lithography step, a final bake is often used to drive out volatile organic materials and water in order to preserve the vacuum integrity of the etch and ion implantation equipment. The temperature required for this step is usually

so high that it will degrade the photochemical properties of the resist. Had the high-temperature hardbake been employed prior to the development step, the resist would not have developed properly. Consequently, the hardbake is one of the last steps in the lithography process, though it may precede measurement and inspection.

The role of the lithography process in overall integrated circuit fabrication can be appreciated by considering the sequence of deposition, lithography, and etch steps used to establish electrical contacts to the transistors that make up integrated circuits (Fig. 1.1). Electrical interconnections are made after the transistors have been fabricated and covered with insulating oxide or another suitable dielectric. The wafers are then covered with resist, and the resist is exposed in places where electrical contacts to the transistors are desired. For reasons explained later, positive resists are commonly used for making electrical contacts. The exposed resist is developed out, exposing the oxide above the points where electrical contact is desired. The wafer is then put into an environment of oxide etchant. The oxide still covered by resist is protected against the etch, while the oxide is removed where electrical contact is desired. The contact holes can then be filled with metal, thus establishing electrical contact.

The primary tool used for projecting the image of the circuit from a photomask onto a resist-coated wafer is the wafer stepper. Wafer steppers exist in two configurations—step-and-repeat and step-and-scan. In a step-and-repeat system, the wafer is positioned under a lens so that a projected image of the layer to be exposed will properly overlay the patterns already on the wafer (Fig. 1.2). The systems that allow the stepper to bring the wafer into proper position are discussed in Chapter 6. Once the wafer is properly positioned and brought into focus, a shutter in an illumination system is opened, allowing light to pass through the

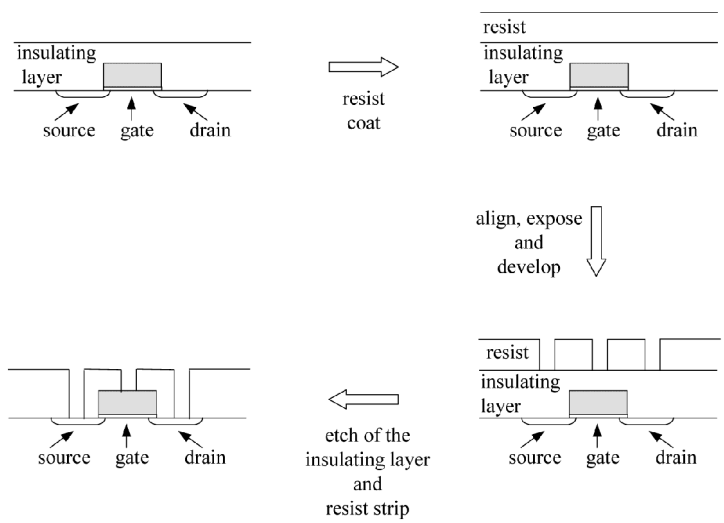


Figure 1.1 The sequence of steps used to make electrical connections in a transistor in an integrated circuit.

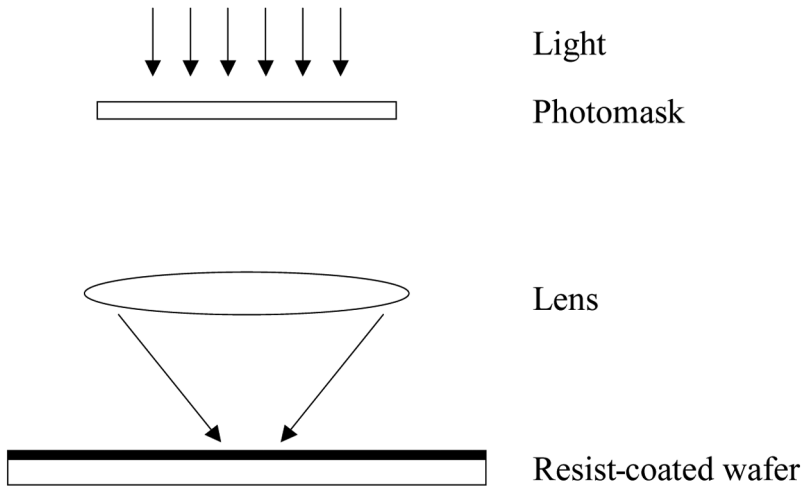


Figure 1.2 Exposure of resist-coated wafers on wafer steppers.

photomask. The pattern on the mask is then imaged by the lens onto the wafer. The image is reduced laterally by the amount N:1, where N is the lens-reduction factor, most commonly equal to 4 in leading-edge systems today. Values of N of 1, 2, and 2.5 are found on steppers that have been designed primarily for high productivity.

Large values of N are desirable to the extent that they reduce the effects of variations in linewidths and misregistration on the reticle, generally by the factor of N, as the image of the reticle is projected onto the wafer. Defects are also reduced in size to the point that they often fall below the resolution limit of the lens. The first commercially available wafer stepper, the GCA DSW4800, had $N = 10$. However, as chip sizes became larger, smaller values of N were required to avoid reticles that were bigger than what was practical. For many years, most steppers had lens-reduction factors of 5, but values of 4 began to appear in the early 1990s. They are found on all leading-edge systems today. For a given mask size, larger exposure fields on the wafer are possible with smaller values of N, and systems designed for high productivity (but not necessarily to produce the finest patterns) often have $N < 4$. With larger exposure fields, fewer fields need to be exposed in order to pattern wafers completely, which gives an opportunity for higher throughput.

Because of reduction by the projection optics, only part of the wafer (an "exposure field") is exposed at any one time on a wafer stepper. After one field is exposed, the wafer is moved on an extremely precise stage so that another part of the wafer can be exposed. This process of exposure and wafer stepping is repeated until the entire wafer is exposed. In the typical reduction stepper, the stage travels in the horizontal plane beneath a fixed, vertically mounted lens.*

In a step-and-repeat system, the entire area of the reticle to be exposed is illuminated when the shutter is opened. In a step-and-scan system, only part of

*The Micrascan systems on which wafers moved vertically were exceptions. These exposure tools were originally made by Perkin-Elmer. The Perkin-Elmer exposure tool business was purchased by SVG, Inc., which in turn was acquired by ASML.

the reticle, and therefore only part of the exposure field on the wafer, is exposed when the shutter is opened (Fig. 1.3). The area exposed on step-and-scan systems at any instant is usually a narrow rectangular region, referred to as the slit. The entire field is exposed by scanning the reticle and wafer synchronously. The reticle stage must be scanned at a speed of N times faster than the wafer, where N is the lens-reduction factor. (The motivations for scanning are discussed in Chapter 5.)

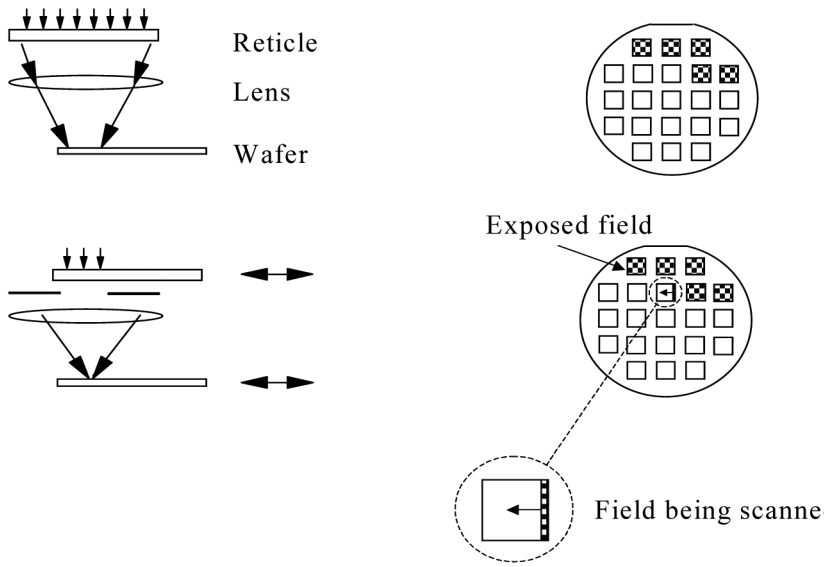


Figure 1.3 Step-and-repeat and step-and-scan configurations.

This method of fabrication is extremely efficient, since many circuit features are patterned in parallel. For example, a NAND flash memory will have two contacts per bit in the memory arrays. With 34-nm design rules, four 32-Gb NAND flash memory chips are exposed simultaneously on a step-and-scan system when the mask is illuminated. Consequently, over 256 billion contacts are imaged in a single exposure, which occurs in about 0.1 sec. If one thinks of this in terms of a transfer of design information to the wafer, where each contact is considered to be a single piece of information, this is a remarkable rate of information transfer: over 2.5 *trillion* bits of information are transferred per second. Moreover, this rate of transfer continues to increase as lithography processes become able to print smaller features. This increasing rate of transfer, driven by the ability to print features of decreasing size, is the key factor in the reduction of the cost of microelectronics.

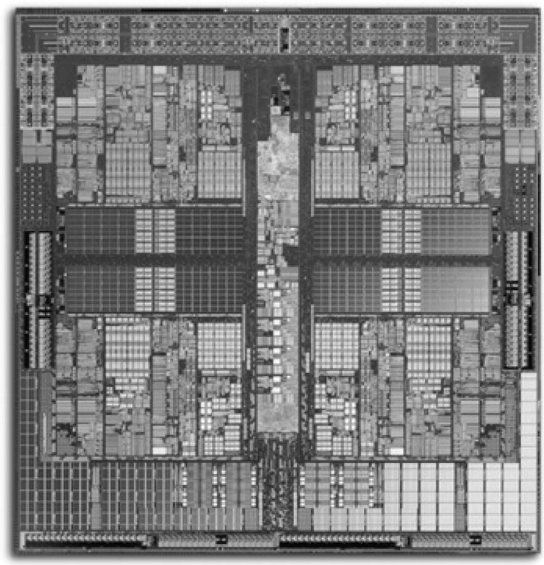
Each step of the lithographic process is discussed in this book. Pattern formation is of central importance because the great functionality of modern microelectronics has been enabled by the ability to pack large numbers of individual transistors in a unit area of silicon. The principles of optics relevant to imaging small features are covered in Chapter 2, while photoresists are discussed in Chapter 3. Methods of predicting lithographic performance are presented in Chapter 4. The primary tool

used in lithography—the wafer stepper—is described in Chapter 5, and this leads into overlay, the topic of Chapter 6. Mask technology is the subject of Chapter 7. Advanced methods of optical lithography are reviewed in Chapter 8. The problem of measuring the small features created by the lithography process is addressed in Chapter 9. The limitations imposed by the laws of physics on optical methods are discussed in Chapter 10. Lithography costs, which can be as significant as technical issues, are covered in Chapter 11. Finally, possible next-generation lithography options that might succeed optical lithography are reviewed in Chapters 12 and 13.

Problems

1.1 For a mask that is 152 × 152 mm, and assuming that a 10-mm border at the edges of the plate is required to hold the mask, show that the largest field that can be patterned on the wafer is 13.2 × 13.2 mm if the lens-reduction factor is 10×. What is the largest field if the lens-reduction factor is 5 × ? 4 × ?

1.2 The AMD Shanghai quadcore microprocessor has an area of 258 mm^2, and is approximately square in shape:



Can such a die be printed in a single exposure with an exposure tool with 10× lens reduction? 4× lens reduction? Why do you think that large lens-reduction factors are not used, even though their use would reduce the criticality of reticle quality?

1.3 What are the nine principal steps in the lithographic process? Which steps are optional?

1.4 In the lithographic process, what are the materials called in which patterns are formed on the wafers?

Chapter 2
Optical Pattern Formation

2.1 The Problem of Imaging

The basic problem of imaging is shown in Fig. 2.1. Light from an illumination source passes through a photomask, which defines the patterns. The simple photomask illustrated here consists of two types of complementary areas—one type that is opaque, while the other is transparent. In this example, the transparent (or "clear") area is a long space of uniform width, and the optical and resist profiles shown in Fig. 2.1 are cross sections for this geometry. Some of the light that passes through the mask continues through a lens, which projects an image of the mask pattern onto a wafer. The wafer is coated with a photosensitive film, a photoresist that undergoes a chemical reaction upon exposure to light. After exposure, the wafer is baked and developed, leaving regions covered by photoresist and complementary regions that are not covered. The patterning objective of microlithography is to produce well-defined resist features, sized within specifications. This is a challenge because of the shape of the light-intensity distribution produced at the wafer plane by a lens of finite resolution. This distribution lacks a clearly defined edge (Fig. 2.1). From the light-intensity distribution alone, it is not possible to know where the edges of the feature are. If the light-intensity distribution had the shape shown in Fig. 2.2, there would be no such problem, because a clear delineation would exist between areas of the resist exposed to light and areas unexposed to light.

The light distribution shown in Fig. 2.1 was not drawn arbitrarily or artistically. Because light is a form of electromagnetic radiation, it is possible to use equations that describe the propagation of electromagnetic waves to calculate optical image formation,[1] and the light-intensity distribution shown in Fig. 2.1 was generated accordingly. The physics of image formation will be discussed in more detail in this and later chapters.

Lithographic processes are easier to control the closer the actual optical images resemble the ideal ones. If the light profile at the wafer plane is represented by the distribution shown in Fig. 2.2, the edges of the feature could be clearly identified by looking at the light-intensity distribution. The photoresist on the wafer would be cleanly separated into exposed and unexposed areas. In the situations that actually occur, the photoresist receives continuously varying doses of light in the regions corresponding to features on the mask. Proper pattern definition on the wafer

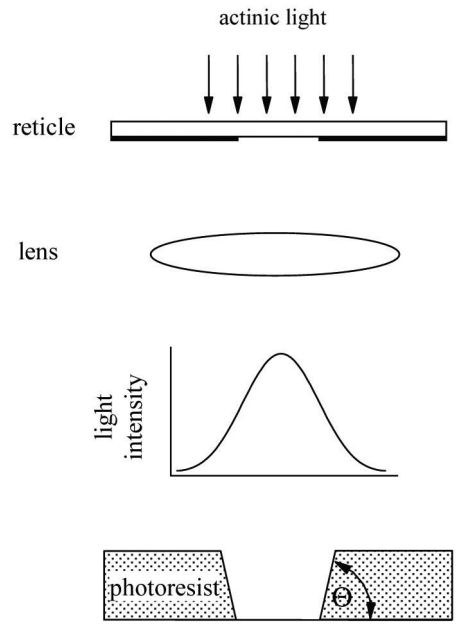


Figure 2.1 An illustration of the imaging process. Light passes through a reticle (photomask). The resulting pattern is imaged onto a photoresist-covered wafer by a lens. The finite resolution of the lens results in a light-intensity distribution that does not have clearly defined edges. The particular light-intensity distribution shown in this figure was calculated using PROLITH 1.5, for a 0.25-μm space on a 0.65-μm pitch (i.e., for a grating pattern of 0.25-μm spaces and 0.4-μm lines), where the numerical aperture (NA) of the aberration-free lens is 0.5, imaging at a wavelength of 248 nm, and with 0.5 partial coherence. The parameter "numerical aperture" will be explained later in this chapter, and partial coherence is discussed in Appendix A. *Actinic* refers to light that drives chemical reactions in the photoresist.

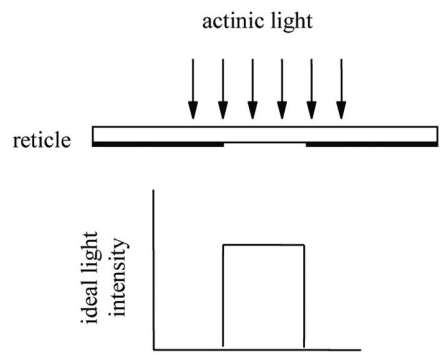


Figure 2.2 An ideal light-intensity distribution.

requires that small differences in exposure doses at the edges of pattern features be distinguished through the resist processing. Edge control will be better for light-intensity distributions that closely resemble the ideal case illustrated in Fig. 2.2.

Light-intensity profiles produced by two sets of optics, one of "higher resolution" than the other, are shown in Fig. 2.3. (The parameters that affect image sharpness and resolution will be discussed later in this chapter.) The image produced by the higher-resolution lens is closer to the ideal image than the one produced by the lens with lower resolution. For a given feature size, it is possible to have better pattern definition with higher-resolution optics. However, because of the highly competitive nature of the microelectronics industry, lithographers need to operate their processes at the limits of the best available optics. With a given set of optics, the light-intensity distributions are degraded further, relative to the ideal case, with smaller features (Fig. 2.4). As features become smaller, the edge slope of the light-intensity profiles become more sloped, and the peak intensity decreases. The challenge for lithographers is to produce the smallest possible features on the wafers, with shape and size well controlled, for a given generation of optics.

2.2 Aerial Images

The origin of the image-intensity profile of Fig. 2.1 is the physical phenomenon of diffraction, a subject that was studied extensively in the nineteenth century and is well understood.[1] Diffraction effects are inherent to the wave nature of light. The resolution of optical tools is fundamentally limited by the physical phenomenon of diffraction, and as device geometries shrink, the lithography engineer must ultimately contend with this barrier imposed by the laws of physics. This is illustrated in Fig. 2.4, where aerial images are shown for features of varying sizes.

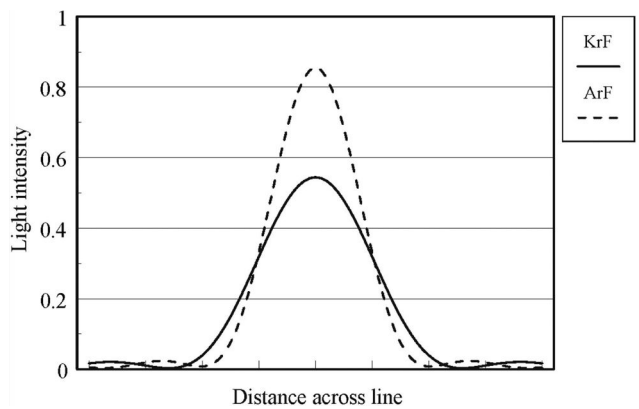


Figure 2.3 Light-intensity profiles for 150-nm isolated spaces (nominal). One image was generated for an aberration-free 0.7-NA lens, with a partial coherence of 0.6, at a wavelength of 248 nm (KrF), while the other image was generated for a 193-nm (ArF) unaberrated lens, with a numerical aperture of 0.75 and a partial coherence of 0.6. These images were calculated using the simulation program Solid-C.[2] The ArF lens has higher resolution than the KrF lens. All images are in the plane of best focus.

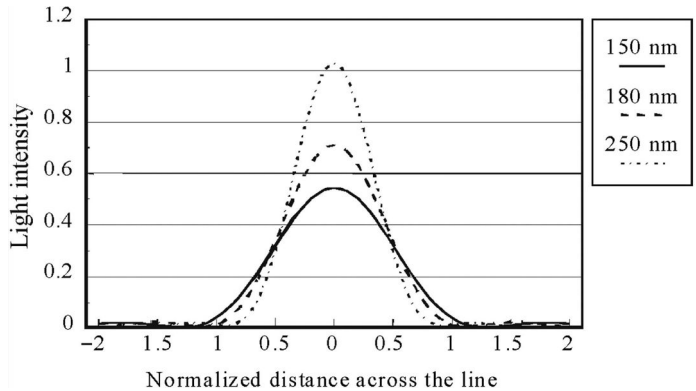


Figure 2.4 Calculated light-intensity distributions for isolated spaces of varying sizes, using parameters for an aberration-free 0.7-NA lens, with a partial coherence of 0.6, at a wavelength of 248 nm. All of the features are normalized to a width of 1.0. All images are in the plane of best focus.

As the feature size shrinks, the edge acuity of the light-intensity distribution is degraded. If even smaller features are considered, at some point one must say that features are no longer resolved, but it is clear from Fig. 2.4 that there is a gradual transition from "resolved" to "unresolved." A definition of resolution is not obvious because of this lack of a clear delineation between "resolved" and "unresolved." Simple diffraction analyses lead to the most frequently cited quantitative definition of resolution, the Rayleigh criterion, which will be introduced shortly. While sufficiently instructive to justify discussion, the Rayleigh criterion does not provide a measure directly applicable to the situation encountered in photolithography, and one should use it with care. Discussion of the Rayleigh criterion in this book emphasizes its assumptions, and therefore its applicability.

The subjects of diffraction and image formation are quite involved and, at best, can only be introduced here. The mathematical formalism is particularly complex, and the interested reader is referred to the references cited at the end of this chapter. The phenomenon of diffraction will be introduced through two examples—the diffraction grating and the circular aperture.

The origin of the resolution limits of optical systems can be appreciated by considering the system shown in Fig. 2.5. A reticle with a diffraction grating with a periodicity $2d$ (equal lines and spaces) is illuminated by a source of coherent light of wavelength λ. (For readers unfamiliar with coherence, a short introduction is provided in Appendix A.) For an angle of incidence of θ_i, the grating diffracts the light into various beams whose directions are given by[1]

$$\sin(\theta) - \sin(\theta_i) = \frac{m\lambda}{2d}, \quad m = 0, \pm 1, \pm 2, \ldots. \tag{2.1}$$

Here we are assuming that the index of refraction of air, which surrounds the mask and lens, is ≈ 1.

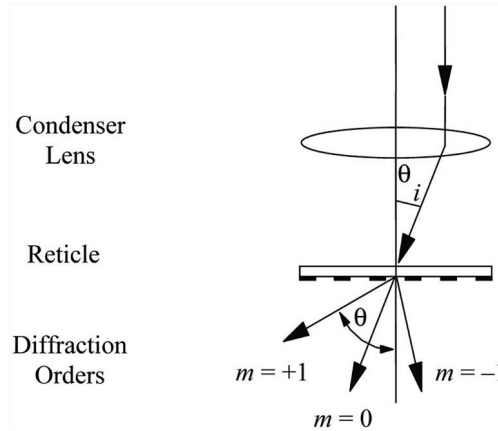


Figure 2.5 A situation with coherent light and a diffraction grating on the reticle. A ray of light is diffracted into beams that propagate in distinct directions given by Eq. (2.1).

Consider a lens with a collection angle of θ_0 (Fig. 2.6) used to image the grating. Due to Eq. (2.1), the lens collects only a finite number of diffraction beams. Coherent light illuminating a grating diffracts into beams that correspond to the object's Fourier components, and a diffraction-limited lens will recombine those beams that pass through the entrance pupil of the lens. If the lens collects all of the diffracted beams, then the grating will be fully reconstructed in the image. However, since the lens collects only a finite number of beams the image is only a partial reconstruction of the original grating pattern. As light from increasingly larger angles is collected, the image consists of more terms in the Fourier-series expansion of the light-intensity distribution of the grating (see Fig. 2.7). For on-

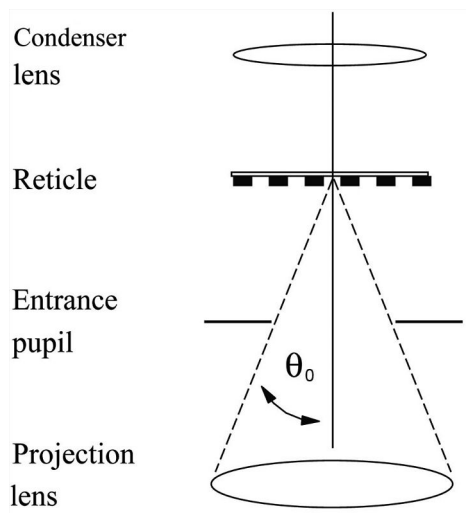


Figure 2.6 The finite size of lenses limits the angles at which diffracted beams are collected.

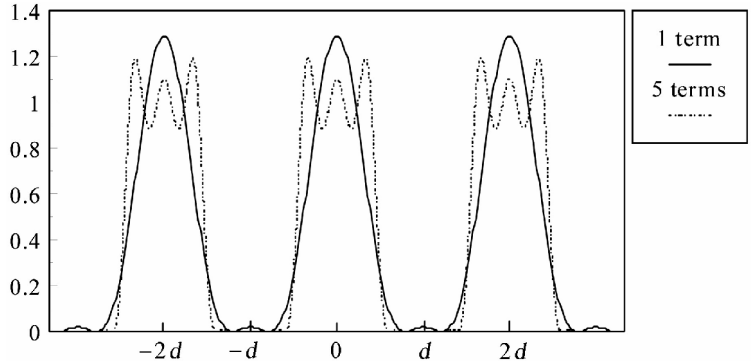


Figure 2.7 Partial sums of Eq. (2.2), with $I_0 = 1$. The image more closely approximates a rectangular grating when more terms are added.

axis illumination ($\theta_i = 0$),

$$I(x) = I_0 \left| \frac{1}{2} - \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^m}{m} \sin \frac{m\pi}{2} \cos \frac{m\pi x}{d} \right|^2 . \tag{2.2}$$

If the lens captures only a single beam, then there is no pattern in the image, since a single plane wave has no spatial variation in its intensity:

$$\text{Intensity of a single plane wave} = \left| A_0 e^{i\vec{k}\cdot\vec{x}} \right|^2 = |A_0|^2 . \tag{2.3}$$

At least two interfering plane waves are needed to generate spatial variation. To retain more than the first (constant) term in the expansion of $I(x)$, θ_0 must be large enough to capture diffracted beams for $m \geq 1$. Equation (2.1) then leads to the following expression for a minimally resolved feature:

$$d = 0.5 \frac{\lambda}{\sin \theta_0} . \tag{2.4}$$

The resolution is found to be directly proportional to the wavelength of the light and inversely proportional to $\sin \theta_0$. Resolution is limited because only a finite number of diffracted beams pass through the entrance pupil of the lens.

The phenomenon of diffraction can be appreciated further by considering the situation in which uniform illumination is normally incident on a masking screen where there is a circular hole (Fig. 2.8). Such a situation arises, for example, if the light is produced by a point source located far from the masking screen. The light that passes through the hole illuminates another screen. In the absence of diffraction, the incident light would simply pass through the aperture and produce a circular illuminated spot on the second screen. Because of diffraction, in addition to the normally incident rays of light, some light propagates at divergent angles.

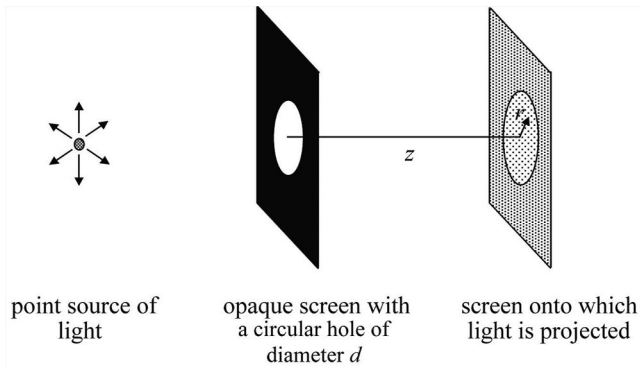


Figure 2.8 A circular aperture illuminated by a point source of light.

For certain configurations, the light-intensity distributions resulting from diffraction can be obtained analytically, and the circular aperture, illustrated in Fig. 2.8, is one of these. The solution for the light intensity is plotted in Fig. 2.9. Rather than a circular beam propagating in a single direction, the light is distributed over angles that are functions of the wavelength of the light and the diameter of the circular aperture. For light of wavelength λ, the light-intensity distribution at radius r on the imaging screen is given by[1]

$$I(x) = I_0 \left[2 \frac{J_1(x)}{x} \right]^2, \tag{2.5}$$

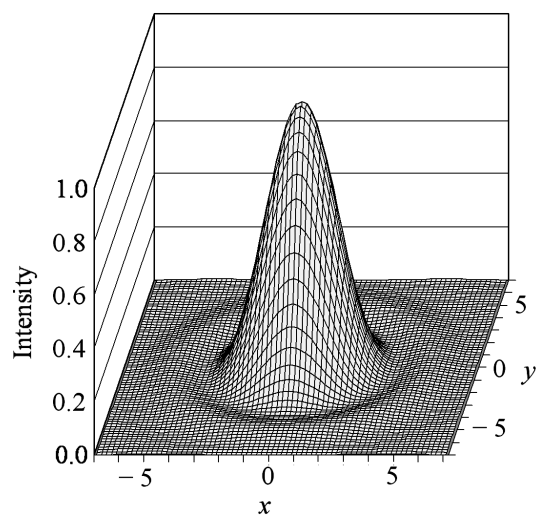


Figure 2.9 The light-intensity distribution from a point source (of unit intensity) projected through a circular aperture in an opaque screen. The dimensions x and y are given in units of $\pi d/\lambda z$, where d is the diameter of the aperture and z is the distance between the two screens.

where $x = \pi dr/\lambda z$, d is the diameter of the aperture, z is the distance between the two screens, and J_1 is the first-order Bessel function. I_0 is the intensity at the peak of the distribution. It should be noted that shorter wavelengths and larger apertures lead to less angular divergence. The light-intensity distribution given by Eq. (2.5) is called the Airy pattern, after G. B. Airy who first derived it.[3] Because of the diffraction that occurs at the entrance pupil or edges of a lens producing divergent beams, the images of point objects are focused Airy patterns.

Rayleigh used the above property of diffraction to establish a criterion for the resolving power of telescopes, using the following argument.[4] Suppose there are two point sources of light (stars) separated by a small angle. Being independent sources, the light rays from different stars do not interfere, i.e., they are mutually incoherent. Accordingly, the total light intensity is the sum of the individual intensities. The two stars are said to be minimally resolved when the maximum of the Airy pattern from one star falls on the first zero of the Airy pattern from the other star. This occurs when

$$\frac{x}{2\pi} = 0.61. \tag{2.6}$$

The resulting total light intensity due to the two stars with an angular separation given by $x = \pi dr/\lambda z = 0.61 \times 2\pi$ is plotted in Fig. 2.10, where two peaks can be clearly distinguished. It was thus learned that the resolving capability of optical instruments is limited simply because of the finite entrance pupil.

For focusing optics, the minimum resolved distance δ between the peak of the Airy distribution and its first zero can be related using criteria that must be satisfied for focusing optics, resulting in

$$\delta = 0.61\frac{\lambda}{n\sin\theta}, \tag{2.7}$$

where n is the index of refraction of the medium surrounding the bottom of the lens ($n \approx 1$ for air), λ is the wavelength of the light, and 2θ is the angle illustrated

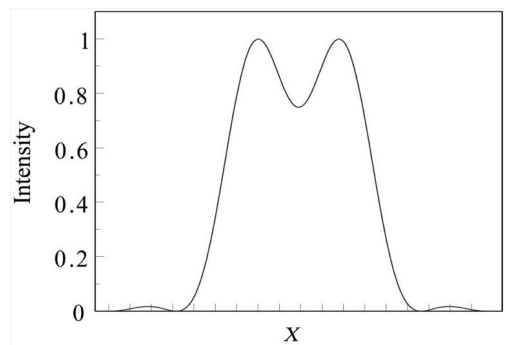


Figure 2.10 Light-intensity distribution of two light sources projected through a circular aperture.

in Fig. 2.11. The quantity $n \sin \theta$ is called the numerical aperture (NA) of the lens. Proper lens design is required for the numerical aperture, and hence the resolution, to be uniform across the exposure field. The minimum separation given by Eq. (2.7) is referred to as the Rayleigh resolution.

Examples of wavelengths and numerical apertures typically used in semiconductor processing are given in Table 2.1 along with the accompanying Rayleigh resolution. As can be seen from this table, a high resolution (small feature size) can be achieved with a large NA or a short wavelength. When focus is taken into consideration, these two paths to resolution are not equivalent, as is discussed shortly.

It is worth restating that the Rayleigh criterion for resolution is obtained by considering the imaging of point sources of light, which is not the situation encountered by photolithographers. The geometries on photomasks that must be imaged onto the front surface of wafers are extended in size, generally approximating the size and shape of a circuit element. The resolution problem is shown in Fig. 2.4. As the features become smaller, the optical profile becomes

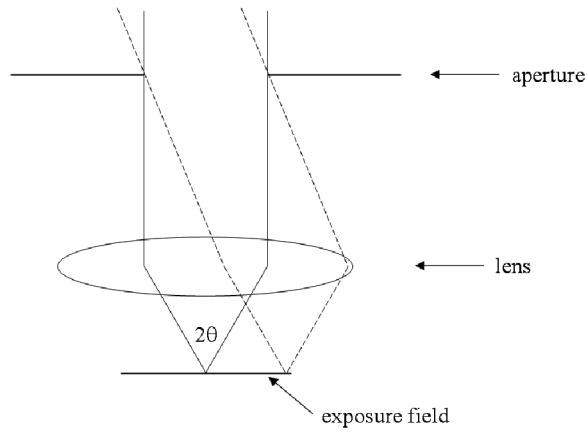


Figure 2.11 The angle defined by the aperture in a lens determines its numerical aperture.

Table 2.1 Typical parameters for stepper optics.

Wavelength (nm)	Numerical aperture	Rayleigh resolution (μm)	Year of first use	Light source
436	0.30	0.89	1982	Hg arc lamp (g line)
365	0.45	0.49	1990	Hg arc lamp (i line)
365	0.60	0.37	1994	Hg arc lamp (i line)
248	0.50	0.30	1994	Hg arc lamp or KrF excimer laser
248	0.60	0.25	1997	KrF excimer laser
248	0.70	0.22	1999	KrF excimer laser
193	0.75	0.16	2001	ArF excimer laser
193	0.92	0.13	2004	ArF excimer laser
193	1.35	0.09	2007	ArF excimer laser

flatter, and it becomes increasingly difficult to locate the "edge" of a feature. As is discussed later, linewidth control is directly related to the slope of the optical intensity distribution. Although it was derived for a situation different from those encountered by lithographers, the Rayleigh criterion for resolution is nevertheless useful as a scaling equation. For example, if one increases the numerical aperture by 20%, the optical resolution is also expected to change by ~20%, even if the exact value of the minimally printed linewidth is not given by Eq. (2.7). For the microelectronics industry, the relevant issue is not the "resolution" of point sources, but rather how small features that exist in real integrated circuits (IC) can be printed. In the context of microelectronics processing, the minimum printable feature size is also dependent upon the properties of the photoresist, while the Rayleigh criterion was derived solely on the basis of optical considerations. As resists improve, the minimum feature size that can be printed becomes smaller for fixed optics. To take advantage of the scaling property of the Rayleigh criterion, and the dependence on resist processes, Eq. (2.7) is often rewritten as

$$\text{smallest feature} = k_1 \frac{\lambda}{NA}, \tag{2.8}$$

where NA is the numerical aperture, and the prefactor of 0.61, associated with point sources, is replaced by k_1 (usually referred to as the "k factor" for a given process). For a given resist system, Eq. (2.8) provides scaling, and the prefactor changes if a different photoresist process is used.

The imaging analysis of a grating structure contains a curious feature associated with coherent illumination. For a particular set of optics, grating dimensions are resolved to a certain point and then not resolved at all for smaller features, at the point when the multiple beams no longer enter the lens. This behavior of optics, where its performance falls off rapidly as the rated resolution limits of the lenses are approached, is familiar to most photolithographers. The extremely sharp cutoff in resolving power of the grating example in this book is an artifact of the complete coherence of the light and single direction of the illumination. This cutoff led to a situation in which terms of the expansion Eq. (2.2) were retained with their complete magnitudes or not retained at all. For partially coherent and incoherent light, the coefficients of the cosines in the image-series expansion decrease more gradually as the diffraction limits are approached, and as the illumination includes light incident from multiple angles.

Consider again the mask pattern consisting of a grating of a periodicity of $2d$. Near the resolution limit, the image consists of only a constant plus a single cosine term that has the same periodicity as the mask pattern, regardless of the configuration of the optics. After passing through the optical system, the light-intensity profile appears as shown in Fig. 2.12. The minimum light intensity I_{min} is greater than zero, and the peak light intensity I_{max} is less than the illumination

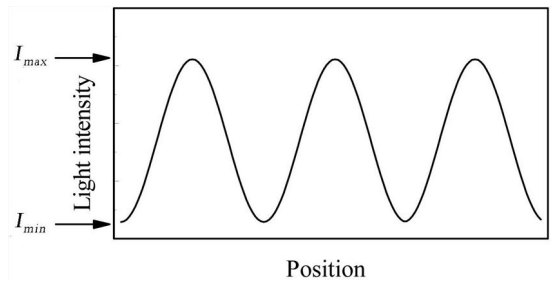


Figure 2.12 Light intensity as a function of position, for a grating pattern on the reticle.

intensity. The optical contrast or modulation is defined as [5,6]

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2.9}$$

The optical contrasts of grating structures for different types of illumination, as a function of spatial frequency, are plotted in Fig. 2.13. The spatial frequencies ν are related to feature size by the relationship

$$\nu = \frac{1}{2d}. \tag{2.10}$$

High spatial frequencies correspond to small features. In the case of diffraction-limited optics (that is, optics where the imaging performance is limited only by the phenomenon of diffraction and not by lens quality), the transfer characteristics can be plotted in terms of normalized spatial frequencies ν/ν_o, where $\nu_o = NA/\lambda$. As shown in Fig. 2.13, there is a sharp cutoff for image contrast with completely coherent illumination, with no image contrast resulting for features smaller than $0.5NA/\lambda$. For incoherent and partially coherent light, the image contrast degrades, and there is no similar distinction between features that can and cannot be imaged with good contrast. The amount by which sinusoidally varying intensities of incoherent light diminish as they pass through an optical system is called the *modulation transfer function* (MTF)[7] and has been used not only within the context of photolithography, but also to characterize the resolving capabilities of optics. For optics with circular apertures, the functional form of the modulation transfer function is known and, for aberration-free optics, is given by

$$MTF(\nu) = \frac{2}{\pi}(\phi - \cos\phi \sin\phi), \tag{2.11}$$

for $\nu \leq 2NA/\lambda$, where ν is the spatial frequency and

$$\phi = \cos^{-1}\left(\frac{\lambda\nu}{2NA}\right). \tag{2.12}$$

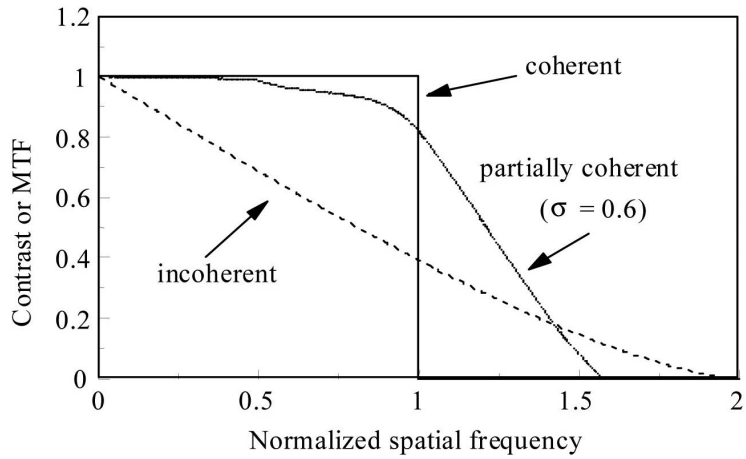


Figure 2.13 Contrast [defined in Eq. (2.9)] of the image of a grating, produced by an in-focus diffraction-limited lens. The normalized spatial frequency is given by $\lambda/2dNA$, where $2d$ is the period (or pitch) of the grating, NA is the numerical aperture of the lens, and λ is the wavelength of the light. The coherence factor σ is defined in Appendix A. For incoherent light, the contrast equals the MTF.

The circular aperture MTF = 0 for $\nu > 2NA/\lambda$. This MTF function is plotted in Fig. 2.13. For incoherent illumination, the image of a diffraction grating is given by[7]

$$I(x) = I_0\left[\frac{1}{2} - \frac{2}{\pi}\sum_{m=1}^{\infty}\frac{(-1)^m}{m}MTF\left(\frac{m}{2d}\right)\sin\frac{m\pi}{2}\cos\frac{m\pi x}{d}\right]. \tag{2.13}$$

The illumination used in microlithography is neither completely coherent nor incoherent, and the curves shown in Fig. 2.13 for coherent or incoherent light do not accurately characterize the transfer properties of mask features onto wafers. The contrast of partially coherent light, more typical of the illumination encountered in the context of optical lithography, is also shown in Fig. 2.13.

When operating in a region where the optical contrast curve is relatively flat, the imaging variations among features of different sizes are small. As the resolution limit of the lens is approached, the imaging capabilities of the lens fall off quickly. The greater the coherence of the light, the greater the optical contrast is up to the resolution limit, but there is a sharp cutoff for small features. Patterns in photoresist can be generated for optical contrasts less than 1.0, but significantly greater than zero; the exact number depends upon the resist system. The partial coherence of exposure tools is chosen to provide the optical contrast necessary for patterning minimum-sized features in single layers of photoresist. Metrics for optical-image quality other than contrast are used more extensively today, and these are discussed later in this chapter.

The preceding discussion has described several general properties of optical systems. First and foremost, the resolution of an optical system is limited by

the physical phenomenon of diffraction. This limit may not be reached in actual optical systems because of lens aberrations, which are discussed in Chapter 4. The sharpness of the resolution "edge" of a lens is dependent upon the degree of coherence of the illumination, and for exposure systems that have been used in the past, the falloff in optical contrast becomes steep as the resolution limit is approached, but it is not perfectly sharp.

Creating resist images within manufacturing tolerances, and on a repeatable basis, from poorly defined optical profiles is the challenge confronting microlithographers. Fortunately, there is more to lithographic pattern formation than optics. Photoresist is a second key ingredient for patterning. One can think of a lithographic-exposure system as essentially a camera in which the photoresist plays the role of the film. Following the development process, a pattern is formed in the photoresist. The quality of the final resist patterns is determined by the resolving power of the optics, the focusing accuracy, the contrast of the resist process, and an assortment of other variables that cause the final image to be a less-than-perfect reproduction of the features on the mask. Just as common film for personal cameras produces pictures of varying quality depending upon the film's photospeed, contrast, and graininess, as well as the focus setting of the camera, the final images produced photolithographically are affected by the resist processes and the optics used.

An understanding of the independent factors that contribute to pattern formation and their interactions is essential for the photolithographer. For example, consider the situation in which a production process has started to produce poorly defined patterns. Is the problem due to the resist process or the exposure system? An understanding of the various contributions to the final image enables the lithography engineer to resolve the problem expeditiously. Process-development engineers must also have a solid understanding of the fundamentals of lithography. For a new process, advanced optics provide high resolution at the expense of capital, while sophisticated resist processes might extend the capabilities of any optics, but with a possible increase in process complexity (and other issues discussed later). By appreciating the roles that the resist and optics play in pattern formation, the photolithography engineer can design the most cost-effective and capable manufacturing line. An overview of the lithographic process is presented in this chapter whereby the separate contributions from the optics and the resist process can be seen. Each subject is then discussed in more detail in subsequent sections.

Two characteristics of the final resist pattern are of primary interest—the size and the shape. Processes capable of producing the narrowest possible linewidths are highly desired because they enable smaller chips to be produced, allowing large numbers of devices to be produced on each wafer. Thus, there is a tight connection between lithographic capability and profits. Short gates are also needed for fast electronics. The requirements of the resist shape are usually imposed by the post-lithographic processing. Consider, for example, the situation in which the photoresist is to serve as a mask for ion implantation. If the edges of the resist are nearly vertical, normally incident implantation is clearly delineated by the

resist edge, resulting in a sharp doping profile. On the other hand, if the resist has considerable slope, high-energy ions penetrate the partial thickness of the resist at the edges of the pattern, and the resulting doping profile is graded. The relevance of the resist-edge slope and its effect on the doping profile depends upon the overall process requirements. Another common example in which the slope of the resist is important occurs when the resist is to be used as a mask for plasma or reactive ion etching, and the etching process erodes the photoresist. The slope of the photoresist can be transferred to the etched layer, which may or may not be desirable. There is no lithographic consideration that purely determines the optimum process. However, once the requirements of the final resist pattern are determined from considerations of the post-lithographic processing, the masking process can be specified. Typically, the most preferred resist profiles are ones that are nearly vertical. Since these are generally the most difficult to produce, as well as the most desired, our discussion is oriented toward such profiles.

Regardless of the specific size and shape requirements of the resist patterns, which may vary from one technology to another, all lithographic processes must be consistent and reproducible, relative to manufacturing specifications, in order to be appropriate for use in production. The parameters that affect process uniformity and consistency must be understood, as well as those that limit the ultimate performance of a lithographic tool or process. Identification of those parameters is a subject of a significant portion of this chapter.

A typical resist profile is shown in Fig. 2.14. The shape of the cross section is often idealized as a trapezoid. Three dimensions are of greatest interest: the width of the resist line at the resist-substrate interface, the slope of the sidewall, and the maximum thickness of the resist film after development.

Of course, actual resist profiles often depart significantly from the idealized trapezoid. The Semiconductor Equipment and Materials International (SEMI) standard that defines linewidth[8] accounts for such departures, and linewidth is

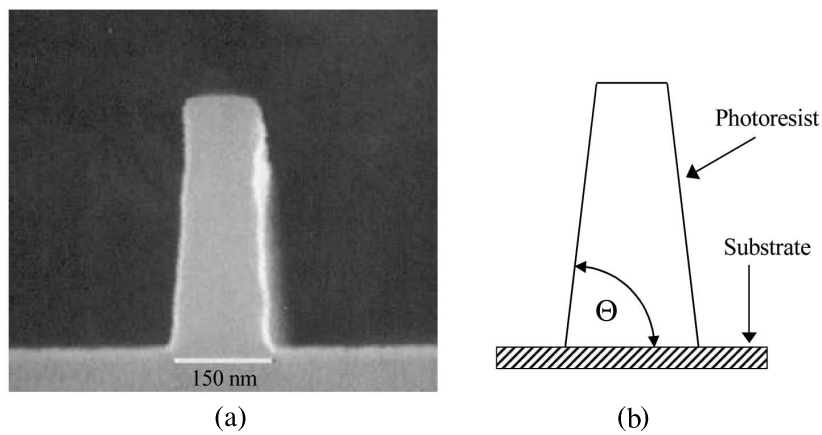


Figure 2.14 (a) Cross section of Sumitomo PAR101 resist, printed on an ASML 5500/900 193-nm step-and-scan system, which had a 0.75 NA. (b) Idealized trapezoid cross section.

defined to be a function of the height from the resist-substrate interface (Fig. 2.15). Throughout the next section, if the word "linewidth" is used with no further clarification, it is understood to be the dimension $L = y_2 - y_1$ measured at the resist-substrate interface. This definition has been chosen for three reasons. First, the width at the base of the resist line is of greatest relevance to the result after the post-lithographic processing. For example, in a highly selective etch where there is little resist erosion, all other dimensions of the resist line have negligible influence on the resultant etch. Moreover, in the presence of a sloped resist profile, the definition adopted here for linewidth is unambiguous. Finally, the value of the linewidth is decoupled from the slope of the resist line.

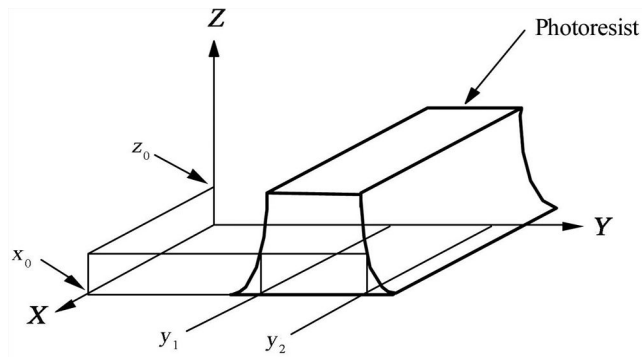


Figure 2.15 The SEMI standard definition for linewidth. Linewidths are given by the quantities $L = y_2 - y_1$ defined at the distances x_0 along the resist line and at height z_0.

At this point, the astute reader has noted that the term "resolution" has not been given a rigorous definition. Since the ability of lithography to resolve small features is a driving force in the microelectronics industry, a definition of this key quantity might appear to be in order. The use of theoretically based measures, such as the Rayleigh criterion, are useful for gaining insight into the lithography process and estimating capability, but they are not completely adequate because they fail to account for imperfect lens design and manufacture. They are also insufficient for other reasons. For example, a lithography process might be unable to reproduce a pattern coded for 50 nm on the reticle (50 nm times the lens reduction), but adjustments to exposure in the same process might allow 50-nm features to be formed on the wafer by overexposing larger features (for example 100 nm times the lens reduction) on the reticle. Depending on one's perspective, the resolution of the lens could be considered either 100 nm or 50 nm. The resulting ambiguity is a consequence of an additional variable, print bias, which is the difference between the size of the feature on the mask (divided by the lens reduction) and the size of the corresponding feature patterned in resist on the wafer. Similarly, optics have different capabilities for imaging isolated lines, isolated spaces, square contacts, or isolated islands. Should resolution be considered for one type of feature or for all features, with or without print bias? The answers to these questions

are user dependent, making it impossible to assign the parameter "resolution" to particular optics, or even a particular lithography process, in a way that meets all of the widely varying needs. Nevertheless, one can speak as to whether particular features are resolved, within context. For a particular circumstance, limits can be placed on linewidth, slope, resist thickness, etc. Once these limits are specified, a feature can be said to be resolved so long as it fits within the limits. However, the sizes and shapes of resist patterns that can be considered acceptable in the most important circumstances (integrated circuit manufacturing lines) depend upon the requirements of post-lithographic processing.

When considering the limits of optical lithography (a topic discussed in detail in Chapter 10) the pitch is typically the most relevant parameter. The principal benefit of advanced lithographic technology is the ability to pack large numbers of connected transistors onto a small area of silicon. It is this capability that enables enormous device functionality to be available at low prices. The smallest line printable with an optical system is much smaller than the half pitch. One may be able to print such small lines, but that capability does not directly address the challenge of packing a large number of connected transistors onto a small area. Such capability is useful for high-performance logic systems, so it is not without value; however, when describing the limits of lithographic systems, it is most useful to express them in terms of the half pitch.

For evaluating and comparing optical systems, these problems may be circumvented by choosing a particular resist process as a reference process, with pre-established requirements for linewidth variations, slope, etc. Other resist processes can be characterized as well by comparison with the reference process. Elegant definitions of resolution may be appropriate in a laboratory environment, while the masking engineer needs quantities that can be determined with the tools at hand. There are many ways to characterize the resolution capabilities of optical systems,[9,10] but in the end the resolving capabilities of a lithography exposure system are usually determined by inspecting photoresist patterns. By using such a method, the contributions from the resist process influence the patterning. It is appropriate at this point to discuss the role of photoresist in pattern formation.

2.3 The Contributions of Physics and Chemistry

With some basic concepts in hand, we can now start to explore the photolithographic process through some simple models.[11,12] The first step in developing a model is the assignment of numerical values to the light-intensity profiles, such as those depicted in Figs. 2.1, 2.3 and 2.4. The light intensity is modulated spatially by the mask and further modified by the optics of the lithography equipment. Because we are considering only a long space of glass in the photomask in this illustrative example, the light intensity I at the plane of the wafer is a function $I(x)$ of only one variable, where x is the distance at the wafer plane along the direction perpendicular to the space. To simplify the discussion, we also initially assume the substrate, which is coated by the resist film, to be flat and nonreflecting. Effects due to substrate topography and reflectivity are addressed

later. The basic assumption for this simple model is that the thickness $T(x)$ of photoresist remaining after development is determined by the exposure energy dose $E(x) = I(x) \cdot t$, where t is the exposure time:

$$T(x) = T_E(E) = T_E[E(x)]. \tag{2.14}$$

This is a reasonable assumption for very thin or high-contrast photoresist. For positive photoresist, greater exposure results in thinner resist films, while negative resist remains thicker with higher exposures. The current discussion is restricted to positive photoresists.

A typical plot of $T_E(E)$, normalized to the initial thickness T_0, is shown in Fig. 2.16 and is called the "characteristic curve" of the resist, following similarly named curves used for characterizing photographic films.[13] Such a curve may be obtained by exposing large areas on a wafer with different exposure energies and then measuring the residual thicknesses in those areas after development. To be precise, the characteristic curve is not only a property of the resist but also of the entire resist process, and it depends upon the developer and process parameters such as bake temperatures.

Several features of the curve should be noted. First, there is a flat region at low exposure where little resist is removed. In this part of the curve, the final thickness may not be equal to the thickness T_0 of the resist film before exposure, since the developer may remove some of the resist even though it is unexposed. Typically, resist-thickness loss in unexposed or less exposed regions is 20 to 200 Å. Second, there is an exposure dose E_0 above which the photoresist film is completely removed. This dose is referred to as the *dose to clear* or *E-zero*. Finally, the curve is linear (with a logarithmic scale for the abscissa) in the region around E_0:

$$\frac{T_E(E)}{T_0} = \gamma \ln\left(\frac{E_0}{E}\right), \tag{2.15}$$

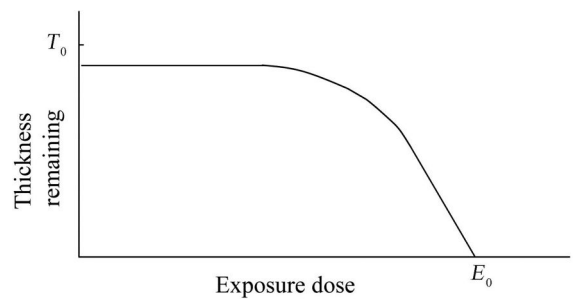


Figure 2.16 A typical characteristic curve for photoresist, showing photoresist thickness remaining after development as a function of exposure dose. Resist thickness is plotted on a linear scale, while the exposure dose is plotted on a logarithmic scale to produce a curve that is approximately linear in the vicinity of E_0.

for $E < E_0$. The slope γ is called the "contrast" of the resist process and is defined by Eq. (2.15). Typical values for γ are shown in Table 2.2.

The dependence of the shape of the final resist pattern on the resist process can be seen in the following analysis. The slope of the resist sidewall is simply

$$\frac{dT}{dx} = \tan\theta, \tag{2.16}$$

where the derivative is evaluated at the point x_0 where the resist profile and the substrate intersect (Fig. 2.17). Note also that

$$E_0 = E(x_0). \tag{2.17}$$

Equation (2.16) can be expressed as[14]

$$\frac{dT}{dx} = \left(\frac{dT_E}{dE}\right)\left(\frac{dE}{dx}\right). \tag{2.18}$$

Equation (2.18) neatly divides the factors that determine the resist profile. The first factor, dT_E/dE, is characteristic of the photoresist and development process, independent of the exposure tool, while the second factor, dE/dx, is completely determined by the optics of the exposure system. One obtains dT_E/dE by differentiating a curve such as the one shown in Fig. 2.16, but on a linear scale. The function $E(x)$ is discussed shortly.

In the vicinity of x_0, $T_E(E)$ is described by Eq. (2.15). This results in the following expression for the slope of the resist profile:

$$\tan\theta = \frac{dT}{dx} = -T_0\gamma\left[\frac{1}{E(x)}\right]\left[\frac{dE(x)}{dx}\right]. \tag{2.19}$$

Steeper resist profiles are produced with high-contrast resists (large values of γ) and steep optical profiles, which have large values for the normalized derivative, $(1/E)(dE/dx)$. Our simple model is based upon the assumption that the resist development behavior that has been measured in large exposed areas can be

Table 2.2 Resist contrast for selected photoresists.[15–18] As noted, contrast is a property of the resist process, not just the resist material. The contrasts listed in this table are for typical optimized processes.

Resist	Exposure wavelength (nm)	Contrast (γ)	Reference
AZ 1470	436	5.8	15
Shipley 511	365	6.9	16
Apex E	248	4.7	17
TOK IP3000	365	6.8	18

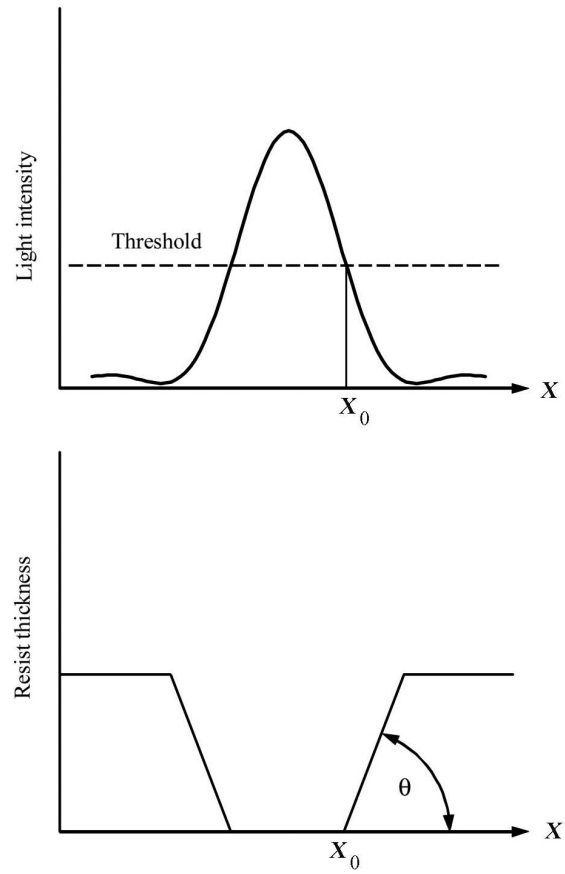


Figure 2.17 A space in (positive) resist is created by exposure and development. The resist thickness $T(x)$ is a function of x.

applied directly to situations where the light intensity is modulated over small dimensions. This is equivalent to assuming that $T(x)$ is solely a function of $E(x)$. Within the limits to which our assumption is valid, the dependence of the profile slope is cleanly separated into the contributions from the optics [the factor $(1/E)(dE/dx)$] and from the resist process (represented by γ), and each can be studied independently of the other.

Theoretical models of image formation in projection optics have been developed, usually starting from the Hopkins formulation for partially coherent imaging.[19] The Hopkins theory is based on scalar-wave diffraction theory from physical optics. Imaging models provide the engineer with the ability to calculate $E(x)$ for various configurations of optics. Let us consider the optics term $E(x)$ and its normalized derivative, $(1/E)(dE/dx)$. $E(x)$ was calculated for a "perfect" lens using a commercially available simulation program, Solid-C,[2] and the results are shown in Figs. 2.4 and 2.18. All profiles are shown in normalized dimensions so that for each feature size the mask edges occur at +/−0.5. The normalized derivative is shown only around the region of interest, close to the edge of the mask

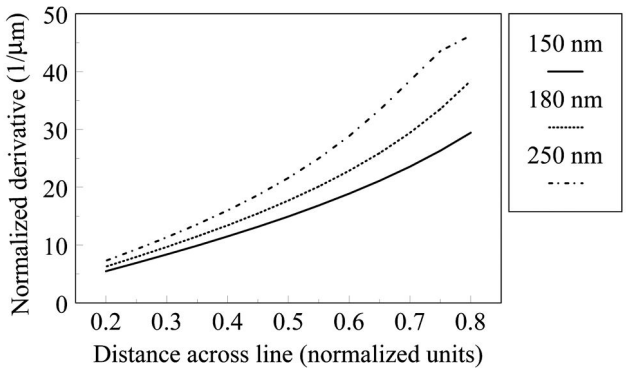


Figure 2.18 Absolute values of calculated normalized derivatives (image log slope) $(1/E)(dE/dx)$ of the intensity distributions shown in Fig. 2.4, near the position in the image that corresponds to the mask edge at $x = 0.5$. The center of the feature occurs at $x = 0$. Overexposure corresponds to values of $x > 0.5$.

feature. As one can see, for fixed optics the "sharpness" of the optical-intensity profile degrades and the value of $(1/E)(dE/dx)$ at the edge of the mask feature is clearly reduced as the feature size is decreased. Lens aberrations or defocus would also reduce the magnitude of $(1/E)(dE/dx)$, while higher-resolution optics would increase its value. Since Eq. (2.19) is nonlinear, substantial degradation of the optical contribution to the profile slope can be tolerated if the contrast γ of the resist process is large enough.

It should be noted that

$$\frac{1}{E}\left(\frac{dE}{dx}\right) = \frac{1}{I}\left(\frac{dI}{dx}\right), \tag{2.20}$$

where E is the exposure dose and I is the corresponding light intensity, since

$$E = I \times t, \tag{2.21}$$

where t is the exposure time. All of the expressions involving the normalized derivative of the image can be written equivalently in terms of exposure dose or light intensity.

One can express the normalized derivative of the intensity or dose as

$$\frac{1}{E}\left(\frac{dE}{dx}\right) = \frac{d}{dx}\left[\ln\left(E\right)\right], \tag{2.22}$$

if one is willing to disregard technical rigor. (A logarithm of a parameter with dimension, such as length, mass, etc., is not a well-defined quantity.) Because of Eq. (2.22), the normalized derivative of the dose (or intensity) is often referred to as the *image log slope* (ILS).

Another advantage of a high-contrast resist process becomes apparent when one is trying to print over a surface that is not flat. Since resist, when spun on the wafer, tends to planarize,[20–22] masking over an underlying feature is equivalent to patterning photoresist of varying thickness. This can be seen in Fig. 2.19. If the thickness of the photoresist is changed from T_0 to $T_0 + \Delta T$, with all other aspects of the lithographic process remaining equal, then the edge of the photoresist line moves from x_0 to $x_0 + \Delta x$, and the minimum exposure energy required to clear the resist to the substrate changes from E_0 to $E_0 + \Delta E$. After exposing the thicker resist with energy E_0, the amount of resist remaining is ΔT.

By drawing a new characteristic curve for thicker photoresist, as seen in Fig. 2.20, the slope in the linear region is given as a result of:

$$\gamma = \frac{\dfrac{\Delta T}{T_0 + \Delta T}}{\ln\left(1 + \dfrac{\Delta E}{E_0}\right)}, \tag{2.23}$$

$$\approx \frac{\Delta T}{T_0 \dfrac{\Delta E}{E_0}}, \tag{2.24}$$

for small variations. (It is assumed here that the resist is on a nonreflecting substrate. The more complex situation in which the substrate is reflecting is discussed in Chapter 4.)

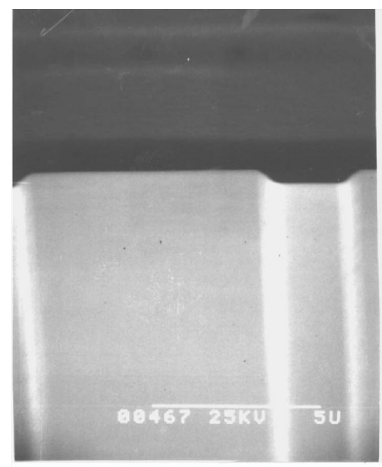


Figure 2.19 Resist (dark material at the top of the figure) spun over topography has a relatively planar top surface.

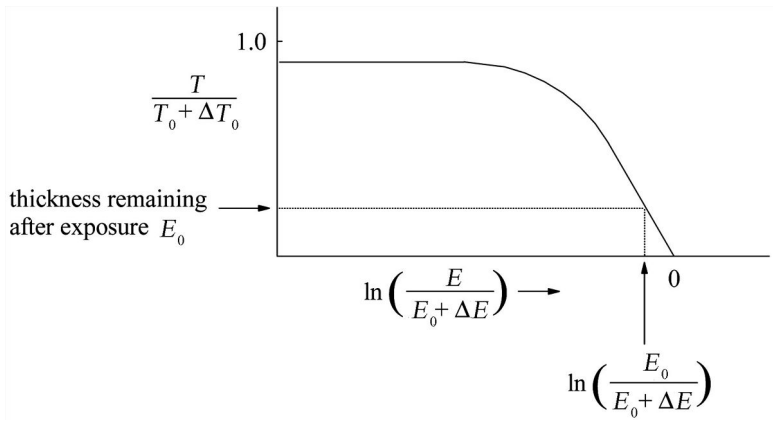


Figure 2.20 Characteristic curve for thicker photoresist. The ordinate is the thickness of the resist remaining after development, normalized to the resist thickness $T_0 + \Delta T_0$, while the abscissa is the natural logarithm of the dose normalized to $E_0 + \Delta E$.

The shift in the edge of the photoresist line that occurs due to the change in the thickness ΔT of the photoresist is given by

$$\Delta x = \frac{\Delta E}{dE/dx}, \tag{2.25}$$

where the derivative is evaluated at the point x_0. From Eq. (2.24), this becomes

$$\Delta x = \left(\frac{\Delta T}{\gamma T}\right)\left[\left(\frac{1}{E_0}\right)\left(\frac{dE}{dx}\right)\right]^{-1} \tag{2.26}$$

$$= \frac{\Delta T}{\gamma T}\,(ILS)^{-1}\,. \tag{2.27}$$

The change to linewidth as a function of resist thickness is kept small by having resists of high contrast, and optics with large image log slopes. Again, there is a separation between the optical term, which reappears in the form of the normalized derivative (or image log slope) $(1/E)(dE/dx)$, and the photoresist, whose contribution is expressed through the quantity γ.

It is also interesting to explore exposure latitude, i.e., the sensitivity of the linewidth to the exposure dose, in terms of our simple model. If the exposure dose is changed fractionally from E to $(1 + f)E$, then the edge of the resist line is moved from x_0 to $x_0 + \Delta x$ (Fig. 2.21).

The edge of the resist line is still determined by Eq. (2.17):

$$(1 + f)E(x_0 + \Delta x) = E_0 = E(x_0). \tag{2.28}$$

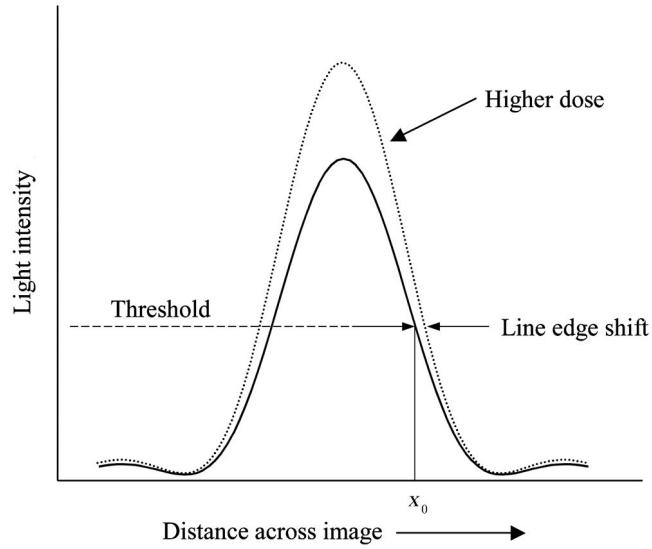


Figure 2.21 With a change in dose, there is a shift in position where the light intensity equals a threshold value.

Accordingly,

$$\Delta x = -fE(x_0)\left(\frac{dE}{dx}\right)^{-1}, \tag{2.29}$$

letting

$$f = \frac{\Delta E}{E}. \tag{2.30}$$

We then obtain an expression for the change in linewidth ($\Delta L = 2\Delta x$):

$$\Delta L = 2\frac{\Delta E(x)}{E(x)}\left[\frac{1}{E(x)}\left(\frac{d}{dx}\right)E(x)\right]^{-1}, \tag{2.31}$$

$$\Delta L = 2\frac{\Delta E(x)}{E(x)}(ILS)^{-1}, \tag{2.32}$$

where all expressions are evaluated for x at the line edge. Our expression for exposure latitude [Eq. (2.31)] contains only factors involving the exposure optics and is independent of the photoresist process and its characteristics. Since it is observed empirically that exposure latitude can be modulated to some degree through the resist process, it is clear that the model presented here is not exactly valid, but is good to the first order and best within the limits of thin- or high-contrast photoresist.

In several expressions—those for resist slope [Eq. (2.19)], variation of linewidth with changing resist thickness [Eq. (2.27)], and exposure latitude [Eq. (2.32)]—it can be seen that the lithography process is more controllable with a large image log slope (see Table 2.3). For this reason, the image log slope was introduced as a metric of image quality[11] and continues to be widely used.[23] Additional critical parameters related to the image log slope are discussed in later chapters.

Higher resolution optics can affect the printing of large objects, as well as enable the printing of smaller features. Consider the light-intensity profiles for a 150-nm feature, shown in Fig. 2.3. The higher-resolution optics provide sharper images, which results in larger image log slopes (Fig. 2.22), at least for the best focus. The normalized derivative at the edge of the space increases from 15 μm^{-1} for the 0.7-NA lens with 248-nm wavelength illumination to 27 μm^{-1} for the 0.75-NA lens at 193-nm wavelength illumination, an increase of nearly 2×. According to the results of the previous section, this translates into direct increases in exposure latitude and linewidth control. Such behavior has led to the proposal that resolution be defined by exposure latitude.

From Fig. 2.22, it can be seen that the image log slope is not determined uniquely by the parameters of the optics, such as exposure wavelength, numerical

Table 2.3 Several key lithography parameters in relation to the image log slope.

Parameter	Relation to the image log slope (ILS)
Resist profile	$\tan\theta = -T_0\gamma\,(ILS)$
Exposure latitude	$\Delta L = [2\Delta E(x)/E(x)]\,(ILS)^{-1}$
Linewidth change with resist thickness	$\Delta x = (\Delta T/\gamma T)\,(ILS)^{-1}$

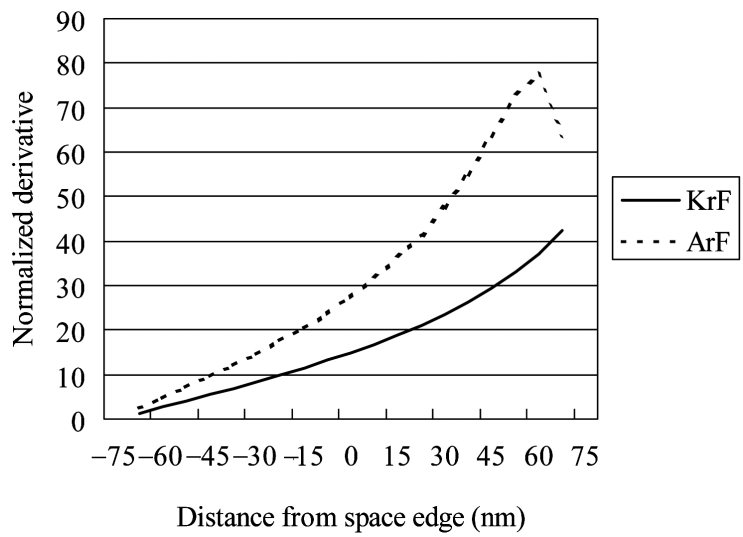


Figure 2.22 Absolute values of normalized derivatives of the light-intensity distributions shown in Fig. 2.3. The normalized derivative is given in units of μm^{-1}.

aperture, degree of partial coherence, aberrations, etc. The value of the normalized derivative $(1/E)(dE/dx)$, which affects the exposure latitude, resist wall profile, and linewidth control over steps, depends not only on the resolution of the optics but also on the print bias as well, that is, the point at which the derivative is evaluated. The value of $(1/E)(dE/dx)$ at the mask edge is different from that at other positions. The normalized derivative is evaluated at the position x_0, which is the edge of the resist line, and may not correspond to the edge of the feature on the mask. The resist patterns may have dimensions different from the ones on the mask. If the spaces in (positive) resist are printed larger than the ones on the mask, then they are termed "overexposed," and if they are printed smaller, they are called "underexposed." Resist patterns having the same dimensions as on the mask are termed "nominal." Resulting from overexposure or underexposure, the quantities that depend upon the normalized derivative, such as exposure latitude, are functions of the print bias. This is seen quite clearly in Fig. 2.22, where the best exposure latitude occurs for overexposure. As is explained in the next section, this result is true only so long as focus is well maintained, and the situation in which the optics are significantly defocused is more complex.

Higher-resolution optics improve the image log slope, and hence many parameters, such as resist profiles and exposure latitude. Unfortunately, higher-resolution optics are never acquired simply to make the photolithographer's job easier. When higher resolution optics become available, the challenge is simply shifted to smaller features, where the normalized derivative remains small. From the preceding, an expression for the exposure latitude is given by

$$\frac{\Delta E}{E} = \frac{\Delta L}{L}\left\{\frac{L}{2}\left[\frac{1}{E(x)}\left(\frac{d}{dx}\right)E(x)\right]\right\}. \tag{2.33}$$

In terms of the image log slope, one can rewrite this equation as

$$\frac{\Delta E}{E} = \frac{\Delta L}{2L}\left(L \times ILS\right). \tag{2.34}$$

The quantity in parenthesis is referred to as the normalized log slope. This is important to recognize, since image log slopes need to increase as feature sizes decrease in order to maintain constant exposure latitude. When discussing exposure latitude particularly among different feature sizes, the normalized log slope is the appropriate parameter. As a consequence, a related metric, with an additional normalization related to the nominal linewidth L is often used, namely, the normalized image log slope (NILS):

$$NILS = L\left(\frac{1}{E}\right)\left(\frac{dE}{dx}\right). \tag{2.35}$$

The above results show how the optics and resist processes play separate and somewhat independent roles in image formation. Of course, the performance

of the lithographic process is also determined by practical issues not captured in theoretical models, such as resist adhesion, and many of these problems are discussed in later chapters. In the analysis presented in this section, the role that photoresist plays is related to its development rate as a function of exposure. The study of optics and its role in photolithography reduces to the analysis of $E(x)$ and the parameters that determine its values. The focusing of the image has a significant effect on $E(x)$, as is seen in the next section.

2.4 Focus

In earlier discussions, the light-intensity distributions were considered only in the planes of best focus. In optical systems, defocus occurs when the separation between the imaging plane and the lens is different from the optimum. From common experience, one knows that defocus reduces the clarity of optical images. The image produced by a common overhead projector may be slightly out of focus, but the presentation may remain readable. If the defocus becomes too great, the image is unintelligible. In lithography the depth-of-focus concept is the range of lens-wafer distances over which linewidths are maintained within specifications, and resist profiles are adequate. The problem of defocus is particularly acute in optical lithography, where the depth-of-focus is becoming so small that it is a concern as to whether optical wafer steppers are capable of maintaining the image in focus.

The problem of defocus can be appreciated by considering the imaging of a point source of light. In purely geometrical optics, where it is assumed that light travels in straight lines, a point object will be imaged by ideal optics to a point located in the focal plane (Fig. 2.23). However, in other planes, the point source of light is broadened into a circle. When imaging, there is generally an amount of broadening that is considered acceptable. Conceptually, the range of values of z over which this broadening remains acceptable is the *depth-of-focus*. At a given distance z from the focal plane, it can be seen that the broadening will be greater for larger values of θ.

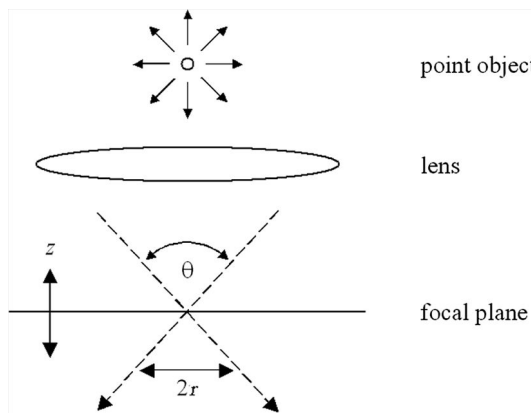


Figure 2.23 In geometrical optics, a point source imaged by an ideal lens is focused to a point in the focal plane. Away from the plane of best focus, the image is a circle of radius r.

It might therefore be expected that lenses will have decreasing depths-of-focus with increasing numerical aperture. (Compare Figs. 2.11 and 2.23.) In the discussion of the imaging of a grating structure, it was noted that gratings with smaller lines and spaces created diffracted beams at larger angles [Fig. 2.5 and Eq. (2.1)]. From this consideration, one can also expect that objects with smaller feature sizes will have smaller depths-of-focus. This shows that there is a tradeoff between resolution and depth-of-focus. Lenses with larger numerical apertures may be used to resolve small features, but the resulting images will have smaller depths-of-focus than the corresponding images of larger features imaged by lower-NA lenses.

The diffraction of a point source of light by a circular aperture discussed earlier (Fig. 2.9) has been extended to the situation in which a lens is used to image the point source of light. The resulting calculated light-intensity profiles in different focal planes are shown in Fig. 2.24. With defocus, the peak intensity diminishes and more light is diffracted away from the center spot. Using the criteria that the peak intensity should not decrease by more than 20%, the following expression results for a depth-of-focus (DOF):[1]

$$DOF = \pm 0.5 \frac{\lambda}{(NA)^2}, \tag{2.36}$$

which is valid for small to moderate values of NA and imaging in air. Over this range of focus, the peak intensity of a point object focused by a lens remains within 20% of the peak value for best focus, where NA is the numerical aperture of the lens and λ is the wavelength of the light. This expression is usually referred to as the Rayleigh depth-of-focus; one Rayleigh unit of defocus is $0.5\lambda/NA^2$. The situation for large values of NA will be discussed in Chapter 10.

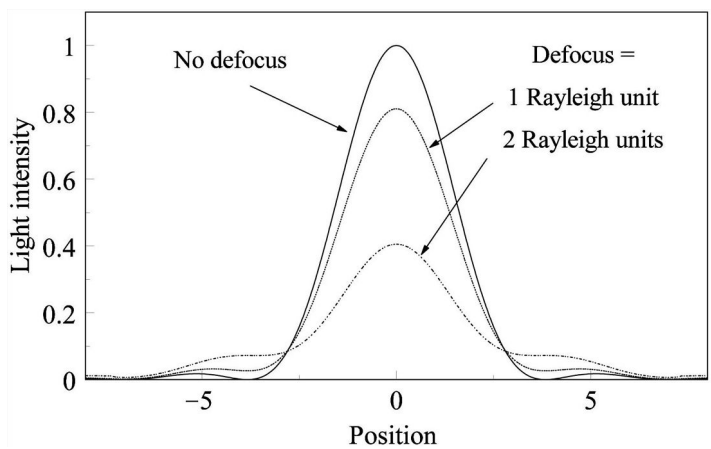


Figure 2.24 Light-intensity profile of a point source of light imaged by a circular diffraction-limited lens at different planes of focus. The horizontal axis is given in units of $\pi dNA/\lambda$, where d is the diameter of the lens aperture. One Rayleigh unit of defocus is $0.5\lambda/NA^2$.

It is useful to use the Rayleigh criterion for estimating the depth-of-focus encountered in photolithography. Typical parameters for 180-nm lithography are $\lambda = 248$ nm and $NA = 0.6$, which results in a Rayleigh depth-of-focus of ± 0.34 μm. That is, the distance between the lens and the wafer needs to be controlled to this very small amount. The small depth-of-focus in optical lithography, and the associated process-control problem, is the reason that focus is given a great deal of attention by lithographers (including the author of this book).

Just as the Rayleigh criterion for resolution was generalized, Eq. (2.36) is often written in the form

$$DOF = k_2 \frac{\lambda}{(NA)^2}. \tag{2.37}$$

Equation (2.37) is expressed as a total range, instead of distances from a midpoint, because the symmetry of the aerial image is broken by the photoresist or optical aberrations in actual application, and the depth-of-focus is often asymmetric about the plane of best focus in high-resolution optical lithography. This asymmetry is discussed later in this section. The Rayleigh expressions for resolution and focus both depend upon only two lens parameters—the wavelength and the numerical aperture—and a pair of constants, k_1 and k_2. One can then relate resolution R and depth-of-focus:

$$DOF = (\text{constant}) \frac{R^2}{\lambda}. \tag{2.38}$$

Within the framework of this analysis, one can see that the depth-of-focus diminishes as one attempts to print smaller features (smaller R). From the Rayleigh criterion for resolution, there are two options for improving resolution: decreasing the wavelength or increasing the numerical aperture. Doing either of these also decreases the depth-of-focus, at least according to Rayleigh. However, improving resolution by decreasing the wavelength has less effect on the depth-of-focus than by increasing the numerical aperture. This is the primary reason that leading-edge optical lithography has involved the use of decreasing wavelengths in the pursuit of smaller features.

The Rayleigh criteria might lead one to a false conclusion that the depth-of-focus diminishes whenever one attempts to improve the resolution of the optics. From the following argument, one can appreciate that this is not true. Suppose one has a lens that has a resolution of 1.0 μm, i.e., it does not image features smaller than this. One could therefore say that such a lens has zero depth-of-focus for submicron features. A higher-resolution lens capable of submicron resolution therefore has greater depth-of-focus for smaller features. This line of reasoning leads to the observation that there is an optimum numerical aperture that is a function of feature size.[12,24–28]

At one time, improper interpretations of the Rayleigh criteria led to the conclusion that optical lithography would not be able to break the 1.0-μm-

resolution barrier. However, photolithography has produced, in volume production, fully functional, large-scale devices with 140-nm pitches, i.e., minimum features ~70 nm. Furthermore, it continues to be the dominant method of patterning feature design rules to 45 nm and even smaller (using techniques that will be described in Chapter 8). Just as the Rayleigh criterion was insufficient for defining the resolution of optics, in the context of photolithography there are additional considerations to be examined for understanding focus and depth-of-focus.

The depth-of-focus of the lens depends on the feature size and type being imaged. As the NA is increased, the limiting resolution becomes finer, but the depth-of-focus for larger features becomes smaller. The optimum NA and partial coherence values depend on the size and type of geometry being imaged, as well as the type of illumination and degree of coherence. The degree of partial coherence σ is defined in Appendix A. For example, the optimum NA to image a contact hole is higher than what is needed for a grating of equal lines and spaces of the same size on the reticle, and greater depth-of-focus (for lines and spaces) is achieved with higher values of σ. The dependence of imaging on feature size, shape, and surroundings has led to wafer steppers with variable parameters, such as numerical aperture and partial coherence. Because resolution is not useful without adequate depth-of-focus, the concept of practical resolution[29,30]—the resolution achieved with some specified depth-of-focus—is helpful for defining the optimum NA, σ, and other process parameters. For the situation encountered in lithography, the effects of defocus can be readily observed. Shown in Fig. 2.25 are cross sections of lines and spaces of resist as focus is varied. Clearly, the sidewall slope increases for large defocus. As shown in Fig. 2.25, linewidths also vary with focus, and the sensitivity of linewidths to defocus is a function of the exposure dose and print bias.[31]

As defined earlier, the depth-of-focus is the distance by which the separation between the lens and the wafer can be changed while maintaining linewidths within specifications, with adequate resist profiles, and sufficient resist thickness.

The numerical value for the depth-of-focus can be compared to the capability of the exposure tools for controlling the distance between the lens and the wafer. As long as the exposure tools and the methodology used to control focus errors can maintain this distance tighter than the depth-of-focus, the process is capable.

There are a number of detractors to the depth-of-focus relative to the diffraction limit. Even with an aberration-free lens, the depth-of-focus found in practice may be different than expected from consideration of the aerial image alone because the finite thickness of the resist influences the depth-of-focus. This is a consequence of refraction at the air-resist interface and the need to maintain a good aerial image throughout the entire thickness of the resist film, not just in a single plane. Typical resist thicknesses (0.1–0.2 μm) are on the order of the Rayleigh depths-of-focus of contemporary high-resolution lenses. Consequently, one might expect that variations in the light-intensity distributions in the direction parallel to the optical axis should be relevant to how resist profiles are generated. In this case, the thin-resist model discussed earlier does not provide accurate predictions. This can be appreciated from simple geometrical optics.

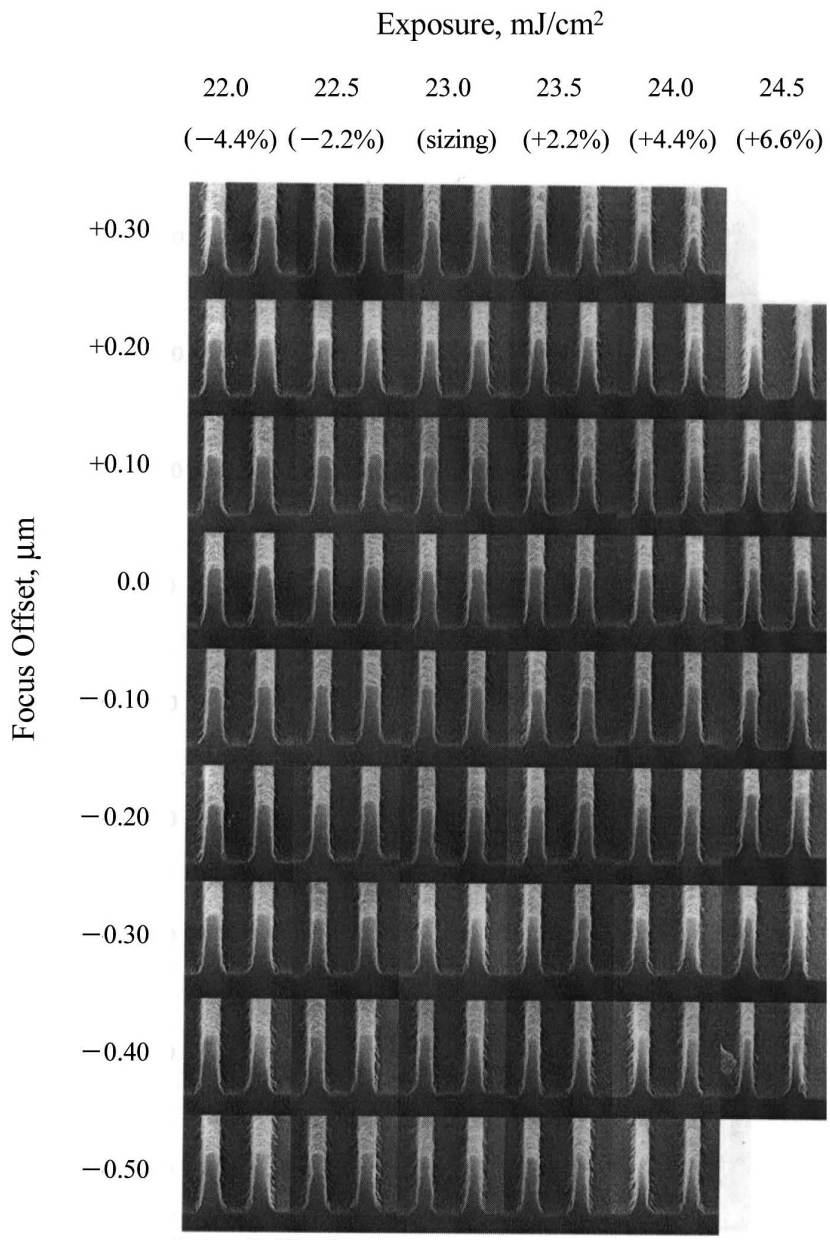


Figure 2.25 Resist profiles for imaging 105-nm lines (320-nm pitch) at best focus and various amounts of defocus. The KrF resist was UV110, patterned on a SiON antireflection coating, using annular illumination ($\sigma = 0.75$ outer, 0.50 inner) on an ASML 5500/300 ($NA = 0.60$). Annular illumination is described in Chapter 8.

Consider a point source of light focused by a lens, as shown in Fig. 2.26. In the limit of geometrical optics, where diffraction effects are ignored, the light is focused by the lens to a point in the plane of best focus. In any plane (perpendicular to the optical axis) other than the plane of best focus, the image is a circle. The distance along the optical axis in which such circles have an acceptably small diameter is the depth-of-focus. If diameters of the focused image vary appreciably throughout the thickness of the resist film, then a more sophisticated model for the optical image than has thus far been discussed needs to be considered.

To understand the effects of thick resist on imaging, let us refer to Fig. 2.26. Suppose that a maximum image radius r is acceptable. For an optical image in air, the depth-of-focus would then be

$$DOF = \frac{2r}{\tan \theta_0}, \tag{2.39}$$

where θ_0 is the largest angle for incident rays of light and is related to the numerical aperture by

$$\sin \theta_0 = NA. \tag{2.40}$$

When a thick layer of resist is placed in the vicinity of best focus, refraction becomes a factor. In Fig. 2.26, two focused light rays converge to the point F_1 instead of the focal point F_0 because of refraction at the air-resist interface.

The angle θ_1 between the normal and the refracted ray is related to θ_0 by Snell's law:

$$\sin(\theta_1) = \frac{\sin(\theta_0)}{n}, \tag{2.41}$$

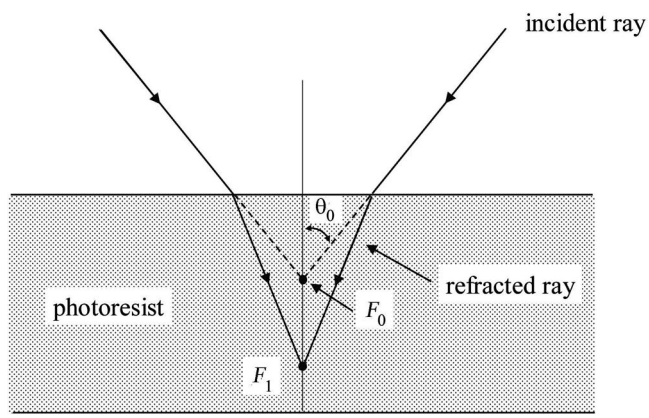


Figure 2.26 Refraction of light at the air-resist interface, its effect on the position of focus, and the spreading of the image due to defocus.

where n is the index of refraction of the photoresist. The depth into the photoresist through which these light rays can travel before they diverge a lateral distance greater than $2r$ is increased from Eq. (2.39) to

$$DOF = \frac{2r}{\tan \theta_1}. \tag{2.42}$$

This leads to an effective increase in the depth-of-focus, over the aerial image, by a factor of

$$\frac{\tan (\theta_0)}{\tan (\theta_1)}. \tag{2.43}$$

For small angles, this is approximately equal to the photoresist's index of refraction n. That is, the reduction in the depth-of-focus due to resist thickness T is T/n. For photoresist, typical values for n are in the neighborhood of 1.7. The refraction at the resist-air interface also breaks the symmetry of an unaberrated image about the plane of best focus, and is an effect seen in detailed lithography simulations.[32]

For well-defined and controlled patterns, good imaging needs to be maintained throughout the thickness of the resist film that may vary over topography. A rigorous capturing of the effect of device topography on the depth-of-focus is tricky because there are linewidth variations due to thin-film optical effects (to be discussed in Chapter 4) that reduce the overall process window. Even on nonreflective substrates, there are changes in linewidths because of variations in resist thickness, and these would occur even for aerial images with infinitely large depths-of-focus. Consequently, the reduction in the depth-of-focus due to topography can be considered to be T/n, where T is the largest-value resist thickness across the topography.

It is important that linewidths are good simultaneously throughout the entire exposure field. The concept of *usable depth-of-focus* follows from the understanding that, in the practical world of integrated circuit manufacturing, imaging needs to be good for all points within a field to ensure complete circuit functionality. This philosophy, where all points within a field need to be good simultaneously, was first discussed in detail in the context of overlay,[33] but the basic concepts are applicable to all lithographic parameters, including focus. The depth-of-focus of a given feature in a particular orientation can be determined for a given set of optics and resist process. However, this does not represent the depth-of-focus that can be used for making integrated circuits because integrated devices typically have geometries of many different orientations distributed over the area of the chip. It has become accepted practice to distinguish between the *usable depth-of-focus* (UDOF) and the *individual depth-of-focus* (IDOF).[34] The IDOF is the depth-of-focus at any one point in the image field, for one particular orientation. The UDOF is the amount by which the separation between the wafer and the lens can be changed and still keep linewidths and resist profiles adequate throughout the entire exposure field. UDOF is the common range of all IDOFs over the printable

area and all orientations. This is generally much smaller than the depth-of-focus at any individual point in the field (IDOF).

There are many factors that cause the UDOF to be less than the IDOF. Astigmatism is one such factor, and it refers to situations in which lines of different orientation have different planes of best focus. An example of the consequence of astigmatism is shown in Fig. 2.27. In this figure, the vertical spaces are well resolved, but the horizontal spaces are so defocused that the resist pattern is bridged. This is an extreme example, patterned on an early 5× stepper. For modern exposure systems, the distances between focal planes for objects of different orientation are less than 50 nm and typically lead only to differences in linewidth, rather than extreme consequences, such as bridging. For a cylindrically symmetric lens, such as those typically found in lithography exposure tools, one would expect astigmatism to occur between geometries parallel to a radius vector (sagittal) and perpendicular to the radius vector (tangential). This is certainly true for astigmatism that results from the design. However, astigmatism can occur between geometries of any orientation in any part of the field because of lens manufacturing imperfections. For example, the many lens elements that comprise a stepper lens may not all be centered perfectly on a common optical axis, thereby breaking the symmetry in the center of the lens field and potentially causing astigmatism where none would be possible for a truly cylindrically symmetric system.

Another focus-related aberration is field curvature. As a result of this aberration, the surface of best focus is a section of the surface of a sphere rather than a plane, and focus is not constant throughout an exposure field. As seen in Fig. 2.28, good focus can be maintained at some points within the exposure field, but most of the field suffers from some defocus because of field curvature. For both astigmatism and field curvature, the rays of light from a given point on the reticle are all focused to a single point in the image, but the collection of such points are not always in a single plane. Astigmatism and field curvature reduce the usable depth-of-focus,

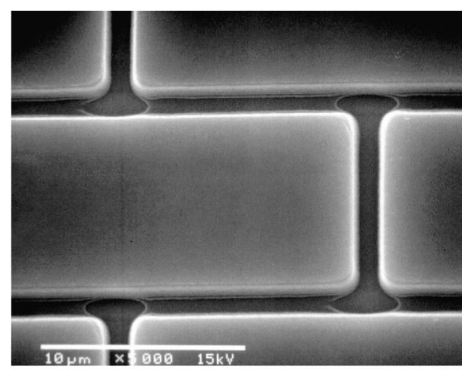


Figure 2.27 An example of astigmatism. This is a top-down micrograph of rectangles of resist separated by trenches. The vertical spaces in the resist are well resolved, but the horizontal lines are so defocused that they are bridged.

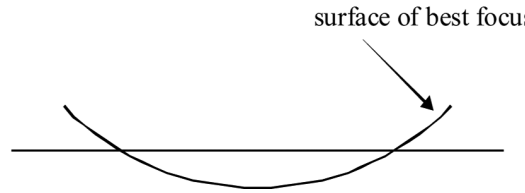


Figure 2.28 Field curvature.

which is the amount that the distance between the lens and wafer can be varied and still maintain good imaging, as illustrated in Fig. 2.29.

Typical specified depths-of-focus for various wafer steppers are given in Table 2.4. The amount of defocus that results in significant changes in either slope or linewidth should be noted: it ranges between 1.5 and 0.4 μm, and has been decreasing over time. This should be compared to the amount of focus variation that can be induced within a process (Table 2.5). All parameters in Table 2.5, except for stepper focus control and the metrology figure,[35] refer to variations across an exposure field. In the preceding discussion, it was shown that the relevant value for circuit topography is the actual height of circuit features divided by the index of refraction of the photoresist.

As lenses improved, with significant reductions in field curvature, astigmatism, and other aberrations, the UDOF began to approximate the Rayleigh DOF. More

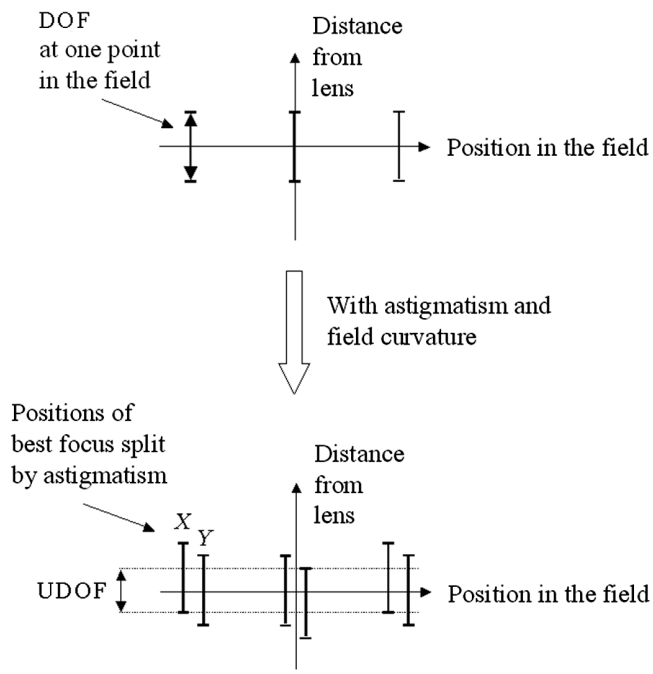


Figure 2.29 Usable depth-of-focus is smaller than the individual depth-of-focus because of astigmatism and field curvature.

Table 2.4 Typical depths-of-focus for various wafer steppers. Resolution and depth-of-focus are given for approximately comparable conditions across technologies and exposure platforms. As will be shown in Chapter 8, techniques have been developed over time to improve resolution and depth-of-focus significantly.

Stepper	Resolution (μm)	Usable depth-of-focus (μm)	NA	Wavelength	Year
GCA 4800	1.25	1.50	0.28	g-line	1980
Nikon	1.00	1.50	0.35	g-line	1984
Canon	0.80	1.20	0.43	g-line	1987
ASML 5000/50	0.50	1.00	0.48	i-line	1990
Micrascan 2+	0.30	0.80	0.50	KrF	1995
ASML 5500/300	0.25	0.70	0.57	KrF	1996
Canon ES2	0.15	0.40	0.68	KrF	1999
ASML 5500/950	0.15	0.40	0.63	ArF	2000

Table 2.5 Parameters in a typical focus budget.

Parameter	Typical values (range, in nm)
Chuck nonflatness	30
Lens field curvature, astigmatism	20
Wafer nonflatness	40
Stepper focus control	20
Circuit topography	30
Metrology for determining focus	20
Total (sum)	160
Total (rss)	68

recently, new imaging methods, referred to as *resolution-enhancement techniques* (RET), have been adopted. These new techniques have made it possible to achieve larger depths-of-focus than the Rayleigh DOF. After the year 2000, the UDOF specified by exposure-tool suppliers usually takes into account that some of these techniques are used. Resolution-enhancement techniques will be discussed in more detail in Chapter 8.

The actual focal range over which imaging must be good is typically somewhere between the sum total of the parameters in Table 2.5 and the root-sum-of-squares (rss) total, since some of the components tend to act randomly (wafer nonflatness, stepper focus control), while others are systematic components (circuit topography). As one can see, the values for these parameters are comparable to the depths-of-focus for high-resolution optics.

One common misconception is that optical lithography is limited by diffraction when the depth-of-focus becomes less than the thickness of photoresist. Such a misunderstanding results from not rigorously distinguishing between the DOF of the aerial image and the operationally measured DOF, which is performed using resist of finite thickness. The operationally measured DOF automatically captures the effects of resist thickness. Consequently, it is quite possible to speak of having a process with a 0.5-μm-thick resist layer and a 0.4-μm DOF. The 0.4 μm refers to

the range over which the 0.5-μm layer of resist can be moved while maintaining resist linewidths and profiles within specifications.

Changes in linewidth as a function of focus and dose are shown in Fig. 2.30. At a given dose, the linewidths change with defocus. It should be noted that the amount of linewidth change as a consequence of defocus depends upon the exposure dose. A dose at which there are minimal linewidth changes when the image is defocused is referred to as the *isofocal dose*.

The thin-resist models used in the preceding section to illustrate the concepts of resolution can be extended to considerations of focus. The light-intensity distribution can be calculated for planes other than the plane of best focus; examples are shown in Fig. 2.31. As can be seen, the optical-intensity profiles degrade with defocus. However, there are positions where the light intensity changes little, or not at all, with defocus. Processes that are set up to have the edges of developed resist features correspond to these positions generally have small changes in linewidth. These positions are known as the conjugate or isofocal points,[37] and they are usually close to the nominal line edges, i.e., the resist linewidths have approximately the same dimensions as those on the reticle (adjusted for the reduction ratio of the projection optics). They correspond to the isofocal dose. Earlier, it was seen that the normalized derivative of the optical-intensity profile at best focus was increased by having a substantial exposure bias. As derived, parameters such as exposure latitude, linewidth control over topography, and resist sidewall slope are improved with a large normalized derivative. In the absence of defocus, one would conclude that a large print bias would lead to an optimized process.

In the presence of defocus, the normalized derivative behaves differently, as shown in Fig. 2.32. At best focus, the image log slope increases for a significant

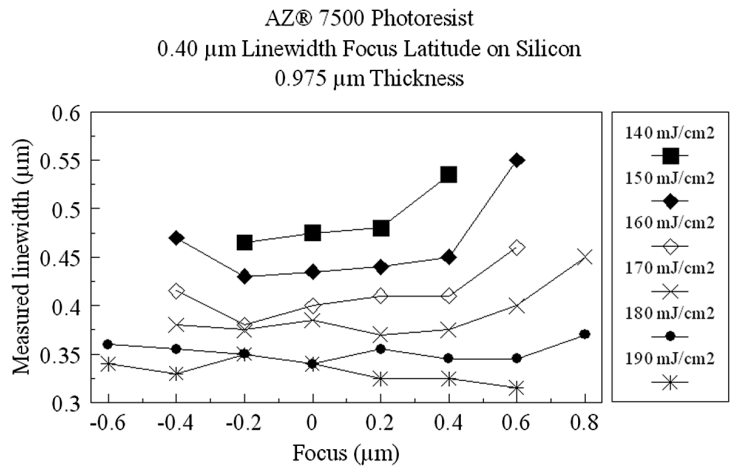


Figure 2.30 Linewidth as a function of focus for dense lines.[36] The process conditions were: softbake and post-exposure bake at 110 °C for 60 sec, and puddle development for 52 sec in AZ 300 MIF developer at 22 °C exposed on a Nikon 0.54-NA i-line wafer stepper. Curves such as these are referred to as Bossung curves.[31]

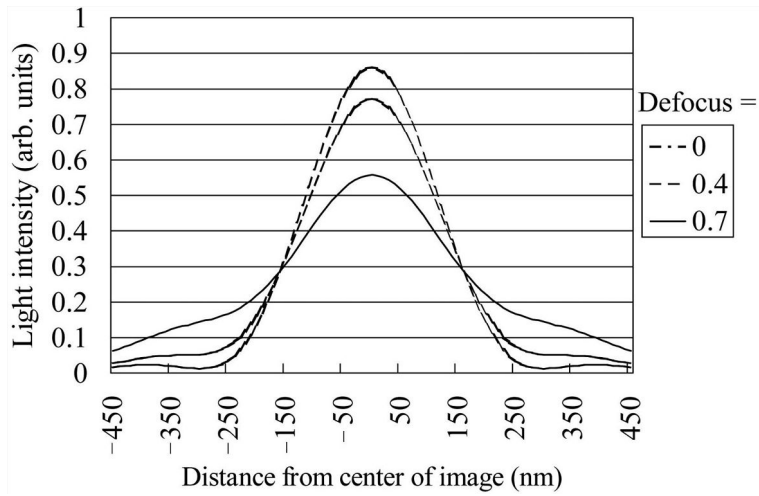


Figure 2.31 Calculated light-intensity distributions at different focus settings for a 300-nm space exposed at $\lambda = 248$ nm, $NA = 0.5$, and $\sigma = 0.6$. Defocus is measured in units of microns (μm).

amount of print bias. The magnitude of the image log slope diminishes at all print biases for defocus, but the degradation is greater away from the conjugate point. For sufficiently large defocus, the image log slope actually decreases with print bias. Thus, one is led to a significantly different conclusion concerning the advantage of large print bias when defocus is a significant concern. It was shown earlier in this chapter that the process window is increased with a large image log slope. However, it is often advantageous to compromise the image log slope if there is a limited depth-of-focus.

The thin-resist model of the preceding section can be extended to include defocus, in which case the light-intensity profile becomes a function of two variables:

$$E(x) \rightarrow E(x, \xi), \tag{2.44}$$

where x refers to the distance parallel to the wafer plane and perpendicular to the (long) resist lines and spaces, and ξ is the amount of defocus. The condition that determines the position of the edge of the photoresist is

$$E_0 = E(x_0, \xi) = E(x_0 + \Delta x, \xi + \Delta\zeta). \tag{2.45}$$

Expanding the right side of Eq. (2.45) in a Taylor series about the plane of best focus, the linewidth change due to defocus $\Delta\zeta$ is obtained:

$$\Delta x = \frac{-1}{2}(\Delta\zeta)^2 \left(\frac{1}{E_0} \times \frac{\partial^2 E}{\partial \zeta^2}\right)\left(\frac{1}{E_0} \times \frac{\partial E}{\partial x}\right)^{-1}. \tag{2.46}$$

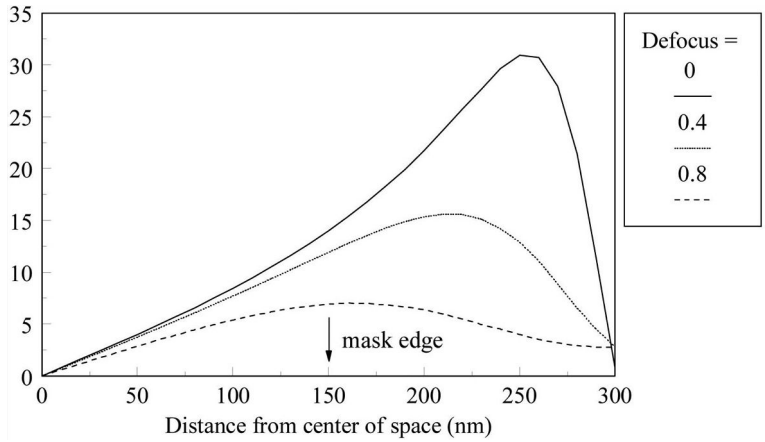


Figure 2.32 Absolute values of the normalized derivative (image log slope) of the light-intensity profiles shown in Fig. 2.31. The normalized derivative is in units of μm^{-1} and defocus is measured in μm.

Derivatives in x are evaluated at the point x_0, and derivatives in ζ are evaluated at best focus, $\zeta = 0$. In the Taylor expansion, the linear terms in ζ are zero because

$$\frac{\partial E}{\partial \zeta} = 0 \tag{2.47}$$

in the plane of best focus, and higher-order terms in Δx were dropped from Eq. (2.46). The symmetry of the aerial image about the plane of best focus is expressed by Eq. (2.47). In the presence of photoresist, or certain optical aberrations, this symmetry is broken. This is not particularly problematic where the thin-resist model is a valid approximation, but it can be relevant for situations involving high numerical apertures and thick (relative to the Rayleigh DOF) photoresist.

In prior discussions, process latitude was improved when the image log slope (the normalized derivative),

$$\frac{1}{E_0} \times \frac{\partial E}{\partial x}, \tag{2.48}$$

is increased. For example, as seen in Fig. 2.18, the image log slope at best focus may be maximized for overexposed spaces, an operating point often distant from the isofocal point. Consequently, there may be large variations in linewidths because of defocus, offsetting any benefits of a large image log slope at best focus.

For situations involving large NAs, there is not a single definition for "best focus." For low-NA optics, where the Rayleigh DOF greatly exceeds the resist thickness, most effects of defocus, such as linewidth and resist-slope variation, behave symmetrically about some plane. Moreover, the smallest features are typically imaged at that plane. In such a circumstance, it is reasonable to define

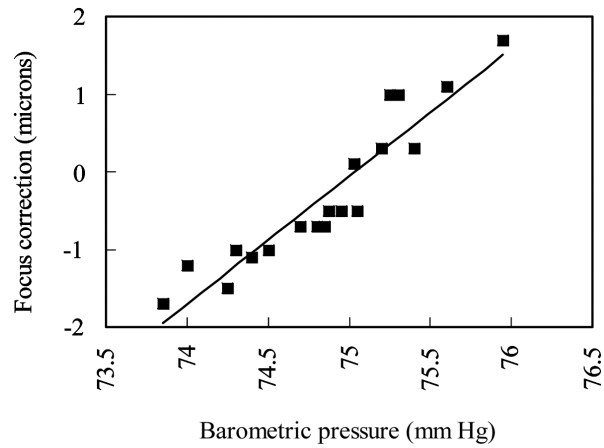


Figure 2.33 Focus correction versus pressure for a stepper equipped with a Zeiss 10-78-46 lens (0.38 NA, g line).

that plane as the plane of best focus. However, for high-NA optics, asymmetry is introduced, and the choice of best focus becomes less obvious. In particular, the middle of the region over which linewidths are within specifications is no longer the place at which the smallest features are resolved, nor the resist slope maximized.

Because of the significant effects on lithographic performance, focus needs to be well controlled. Unfortunately, a number of factors can cause focus to vary. For example, a change in barometric pressure can induce a shift in focus. Lenses with glass elements create images by means of the physical phenomenon of refraction where the direction of a beam of light changes at the interface between materials of different indices of refraction. The angle of refraction, and hence the focal distance, depends upon the indices of refraction of the glass and air. While the index of refraction of the air is ≈ 1, it differs slightly, and changes as barometric pressure changes. As the air pressure inside of wafer fabricators rises and falls with the outside barometric pressure, the focal length of lenses containing refractive elements also changes.[38] The focus adjustment required to keep an exposure tool at best focus while barometric pressure changes for an exposure system that does not have automatic focus for pressure compensation is shown in Fig. 2.33. Modern exposure tools control focus automatically for changes in air pressure. Some systems use pressurized lenses, where the air pressure inside the lens is kept at a fixed value regardless of the pressure of the air in the fab. An alternative method of control involves the measurement of barometric pressure and compensation of the focus through software. In addition to adjusting the focus when barometric pressure changes, exposure tools also need to move wafers up and down, to adjust their tilt, and to correct for changes in wafer thickness and flatness if the resist films are to be maintained in focus. Automatic focusing systems are discussed in detail in Chapter 5. Barometric pressure affects the reduction of the projection optics[39]as well as focus, and requires similar compensation.

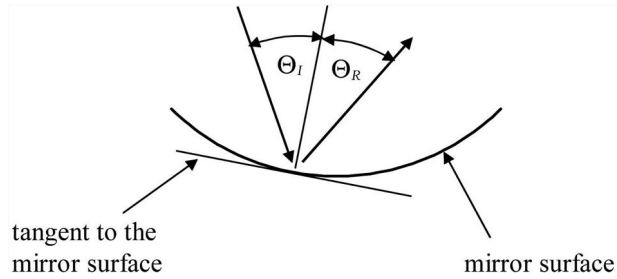


Figure 2.34 Reflections from mirror lens elements are purely geometrical and independent of the index of refraction of the air.

Mirrors with curved surfaces can also be used for imaging. The imaging properties from reflective surfaces are purely geometrical (Fig. 2.34), and are independent of the refractive index of the air or wavelength of the light:

$$\Theta_I = \Theta_R, \tag{2.49}$$

where Θ_I and Θ_R are the angles of incidence and reflection, relative to the normal, respectively. Lenses that use mirrors have less sensitivity of focus to barometric pressure than lenses that use only glass elements. Stepper lenses will be discussed in more detail in Chapter 5.

FLEX (Focus-Latitude Enhancement eXposure) has been proposed as a method for increasing the DOF; this method involves partial exposures at different focus positions.[40,41] With FLEX sharply focused images are superimposed with defocused ones throughout the depth of the resist film. The contact holes are extended by doing partial exposures at multiple locations along the optical axis. FLEX is also known as *focus drilling*. The appropriate choice of exposure and focus steps can lead to a 3× or 4× increase in contact-hole DOF, but the DOF for line features is typically not improved. FLEX has been demonstrated to improve the DOF, but with a loss of ultimate resolution by the multiple defocused exposures.[42] As we shall see in a later section, this is a general property of most attempts to improve the DOF that do not address the fundamental issue of diffraction at the reticle.

Problems

2.1 Assuming diffraction-limited optics, show that the optical resolution of a 0.54-NA ArF lens is approximately the same as a 0.7-NA KrF lens. Which has the greater depth-of-focus?

2.2 Derive Eq. (2.29).

2.3 The following table gives wavelengths and numerical apertures used in volume manufacturing over time:

Year	Wavelength	NA
2001	248 nm	0.80
2003	193 nm	0.75
2005	193 nm	0.85
2007	193 nm	0.92

Calculate the Rayleigh DOF for each year. What is the trend for depth-of-focus over time?

2.4 In Fig. 2.18, for a diffraction-limited 0.7-NA lens exposing at a wavelength of 248 nm, the normalized derivatives of the aerial images (image log slopes) at the position of the nominal line edge are as follows:

Feature size	Normalized derivative (1/μm)
250 nm	21.6
180 nm	17.7
150 nm	15.0

Using Eq. (2.32), show that the changes in linewidth for a 10% change in exposure dose are 9, 11, and 13 nm for 250-, 180-, and 150-nm features, respectively. Show that the dose control needs to be 27%, 16% and 11% for 250-, 180-, and 150-nm features, respectively, for linewidths to change by no more than 10% of the feature size. How does the challenge of process control change as smaller features are produced?

2.5 Assuming normally incident coherent light on a grating of equal lines and spaces, show that the angles of incidence onto the wafer of the first-order diffracted beams in air are 29.7, 43.5, and 55.8 deg for 250-, 180-, and 150-nm lines, respectively. What are the minimum numerical apertures required to image these features?

2.6 Show that the depth-of-focus is reduced by ~59 nm by increasing the resist thickness from 200–300 nm for a resist with an index of refraction of 1.7.

References

1. M. Born, and E. Wolf, *Principles of Optics*, 7th ed., Cambridge Univ. Press, Cambridge (1999). Note that the sign convention for the angles employed by Born and Wolf (and in this text) is not used universally. As used here, the angles indicated in Fig. 2.5 are positive.
2. Solid-C was a simulation program produced by SIGMA-C GmbH of Munich, Germany. SIGMA-C was acquired by Synopsys, Inc., in 2006. Synopsys produces lithography simulation software called Sentaurus Lithography™.
3. G. B. Airy, "On the diffraction of an object-glass with circular aperture," *Trans. Camb. Phil. Soc.* **5**(3), 283–291 (1835).
4. Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *Phil. Mag.* **8**, 261–274 (1879).

5. W. J. Smith, *Modern Optical Engineering*, 2nd ed., McGraw Hill, Boston (1990).
6. W. Oldham, P. Jain, A. Neureuther, C. Ting, and H. Binder, "Contrast in high-performance projection optics," *Proc. Kodak Microelectron. Sem.*, 75–80 (1981).
7. J. W. Goodman, *Introduction to Fourier Optics*, McGraw Hill, New York (1968).
8. SEMI Standard P19-92, "Specification for metrology pattern cells for integrated circuit manufacture," *Semiconductor Equipment and Materials International* (2007).
9. A. J. den Dekker and A. van den Bos, "Resolution: a survey," *J. Opt. Soc. Am.* **14**(3), 547–557 (1997).
10. F. Schellenberg, "Resolution enhancement technology: the past, the present, and extensions for the future," *Proc. SPIE* **5377**, 1–20 (2004).
11. H. J. Levinson and W. H. Arnold, "Focus: the critical parameter for submicron lithography," *J. Vac. Sci. Technol. B* **5**, 293–298 (1987).
12. W. H. Arnold and H. J. Levinson, "Focus: the critical parameter for submicron lithography, part 2," *Proc. SPIE* **772**, 21–34 (1987).
13. L. Larmore, *Introduction to Photographic Principles*, 2nd ed., Dover Publications, New York (1965).
14. P. D. Blais, "Edge acuity and resolution in positive type photoresist systems," *Solid State Technol.* **20**, 76–85 (1977).
15. W. H. Arnold, and H. J. Levinson, "High resolution optical lithography using an optimized single layer photoresist process," *Proc. Kodak Microelectron. Sem.*, 80–92 (1983).
16. Shipley Co., Newton, Massachusetts, Private Communication (1996).
17. H. J. Levinson, unpublished.
18. Nick Eib, Private Communication (2000).
19. H. H. Hopkins, *Wave Theory of Aberrations*, Clarendon Press (1950).
20. L. E. Stillwagon, R. G. Larson, and G. N. Taylor, "Planarization of substrate topography by spin coating," *J. Electrochem. Soc.* **134**, 2030–2037 (1991).
21. D. B. LaVergne and D. C. Hofer, "Modeling planarization with polymers," *Proc. SPIE* **539**, 115–122 (1985).
22. L. K. White, "Approximating spun-on, thin film planarization properties on complex topography," *J. Electrochem. Soc.* **132**, 168–172 (1985).
23. H. J. Levinson, Y. Ma, M. Koenig, B. La Fontaine, and R. Seltmann, "Proposal for determining exposure latitude requirements," *Proc. SPIE* **6924**, 69241J (2008).

24. C. Nölscher, L. Mader, S. Guttenberger, and W. Arden, "Search for the optimum numerical aperture," *Microelectron. Eng.* **11**, 161–166 (1990).

25. K. Yamanaka, H. Iwasaki, H. Nozue, and K. Kasama, "NA and σ optimization for high-NA I-line lithography," *Proc. SPIE* **1927**, 310–319 (1993).

26. W. N. Partlo, S. G. Olson, C. Sparkes, and J. E. Connors, "Optimizing NA and sigma for sub-half-micrometer lithography," *Proc. SPIE* **1927**, 320–331 (1993).

27. B. Lin, "The optimum numerical aperture for optical projection microlithography," *Proc. SPIE* **1463**, 42–53 (1991).

28. A. Suzuki, S. Yabu, and M. Ookubo, "Intelligent optical system of a new stepper," *Proc. SPIE* **772**, 58–65 (1987).

29. H. Ohtsuka, K. Abe, Y. Itok, and T. Taguchi, "Quantitative evaluation method of conjugate point for practical evaluation of wafer stepper," *Proc. SPIE* **1088**, 124–133 (1989).

30. H. Fukuda, A. Imai, T. Terasawa, and S. Okazaki, "New approach to resolution limit and advanced image formation techniques in optical lithography," *IEEE Trans. El. Dev.* **38**(1), 67–75 (1991).

31. J. W. Bossung, "Projection printing characterization," *Proc. SPIE* **100**, 80–84 (1977).

32. M. S. Yeung, "Modeling high numerical aperture optical lithography," *Proc. SPIE* **922**, 149–167 (1988).

33. H. J. Levinson and R. Rice, "Overlay tolerances for VLSI using wafer steppers," *Proc. SPIE* **922**, 82–93 (1988).

34. B. Katz, J. Greeneich, M. Bigelow, A. Katz, F. van Hout, and J. Coolsen, "High numerical aperture I-line stepper," *Proc. SPIE* **1264**, 94–126 (1990).

35. 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**, 541–549 (1994).

36. AZ Electronic Materials, Branchburg, New Jersey (2000).

37. S. A. Lis, "Processing issues and solutions to conjugate lithography," *Proc. Kodak Microelectron. Sem.*, 117–136 (1986).

38. K. Hale, and P. Luehrmann, "Consistent image quality in a high performance stepper environment," *Proc. Kodak Microelectron. Sem.*, 29–46 (1986).

39. F. Sporon-Fiedler and J. Williams, "Atmospheric pressure induced reduction errors in reduction stepper lenses," *Proc. SPIE* **538**, 86–90 (1985).

40. T. Hayashida, H. Fukuda, and N. Hasegawa, "A novel method for improving the defocus tolerance in step-and-repeat photolithography," *Proc. SPIE* **772**, 66–71 (1987).

41. H. Fukuda, N. Hasegawa, T. Tanaka, and T. Hayashida, "A new method for enhancing focus latitude in optical lithography: FLEX," *IEEE Electron. Dev. Lett.* **8**(4), 179–180 (1987).

42. C. A. Spence, D. C. Cole, and B. B. Peck, "Using multiple focal planes to enhance depth-of-focus," *Proc. SPIE* **1674**, 285–295 (1992).

Chapter 3
Photoresists

As discussed in Chapter 1, the mask pattern is transferred to the wafer by means of optical projection onto photosensitive materials known as photoresists. These materials clearly play a critical role in lithography. The chemistries of the most common classes of photoresists are outlined in this chapter. However, the emphasis is on operational performance, i.e., the ways in which resists behave, as observable by the practicing lithographer (in contrast to the resist chemist).

3.1 Positive and Negative Resists

Resists are broadly classified as positive or negative. Unexposed positive resists normally have very low solubility in developer and become soluble by exposure to light. Negative resists behave in the opposite manner; unexposed negative resists are soluble in developer and lose their solubility upon exposure to light. Negative resists were used predominately prior to the advent of wafer steppers. Resists most widely used in the early years of the semiconductor industry, such as Kodak's thin-film resist (KTFR), were based upon the photoinduced cross-linking of low-molecular-weight cyclized polyisoprene[1,2] (Fig. 3.1). Azide additives were used to improve their sensitivity to light and to facilitate cross-linking (Fig. 3.2). Cross-linking reduced the material's solubility in organic solvents such as xylene. After exposure, an organic solvent was used to develop the resist by removing the low-molecular-weight unexposed resist and leaving behind cross-linked, high-molecular-weight material.

These early negative resists had a number of deficiencies. During development, the cross-linked material would absorb some of the organic solvent used as the developer, causing swelling. Because of the swelling, closely spaced geometries would come into contact during develop, and would sometimes remain stuck together, resulting in a patterning defect. This swelling limited the thickness and resolution potential of these resists. The use of organic solvents for developing also created safety and disposal problems that required engineering solutions. The resists were also desensitized when exposed to air because of competing reactions with oxygen that could reduce the amount of cross-linking. The wafers needed to be kept in nitrogen ambient during exposure, complicating exposure-tool design.

With the advent of wafer steppers, the industry shifted to positive photoresists for most critical applications. The earliest commercially available wafer steppers

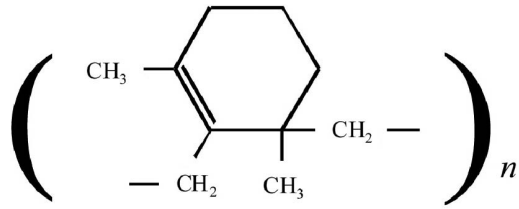


Figure 3.1 Cyclized polyisoprene.[1]

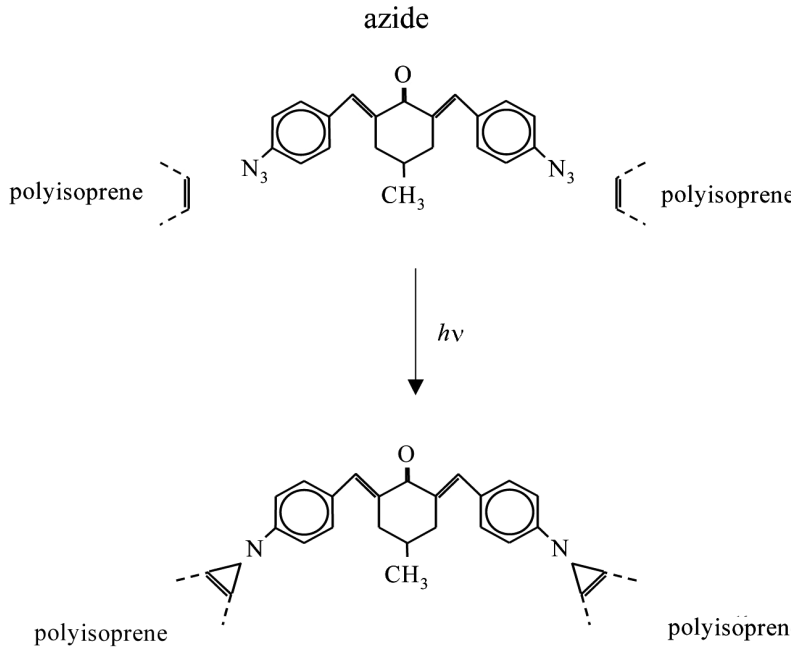


Figure 3.2 The cross-linking of cyclic polyisoprene, by means of a commonly used azide, 2, 6-bis(4'-azidobenzal)-4-methylcyclohexanone.[3,4]

were introduced in the late 1970s with optics that functioned at the mercury g line (λ = 436 nm). The negative resists used at that time were not very sensitive at such a long wavelength. There were good reasons to use g-line illumination, such as it facilitated the design and fabrication of the stepper's projection optics. The mercury g line is visible (blue) light. Many lens designers were familiar with designing at visible wavelengths, and a great deal of lens metrology involved visual assessment of optical interference patterns. A transition to shorter wavelengths would have required significant development by the lens makers (something that occurred later). Fortunately, there were g-line-sensitive positive resists, based on novolak resins, commercially available at the time steppers were first introduced. This, along with other advantages of the positive resists, resulted in an industry-wide transition from the use of negative resists primarily to the predominant use of positive resists. With the exception of contact-hole patterning (discussed

shortly), there were no fundamental reasons that positive resists should have been chosen over negative ones. However, the available positive resists had other necessary attributes, such as the lack of swelling. The transition to positive resists occurred not simply because of intrinsic advantages of positive resists (with the exception of contact-hole patterning), but rather because of specific advantages of the existing (novolak) positive resists relative to the existing (azide/isoprene) negative resists[5] (see Table 3.1). Recently, very high-resolution negative resists that can be developed in the same aqueous developers as positive resists have been introduced.[6–8]

In principle, there are no compelling reasons to use positive resists in preference to negative resists, with some exceptions. Consider the aerial image of square features—contact holes—shown in Fig. 3.3. For producing contacts with negative resists, the feature on the mask is a square of chrome surrounded by glass, while the mask feature for a positive-resist contact process consists of a square of glass surrounded by chrome. As can be seen in Fig. 3.3, the negative-resist process suffers from having undesirable exposure at the center of the contact region, which can lead to closed contacts. The creation of contact holes on wafers is thus facilitated by the use of positive resists, because the aerial image is superior.[9] For other patterns, the advantages of one tone of resist over the other are less clear. For example, on a grating with lines and spaces of nearly equal size, the clear and dark patterns are optically a rough equivalent, so one tone of resist is not going to be fundamentally preferable to the other tone. There are recently developed negative resists with very high performance.[8]

Resists are usually operational over specific wavelength ranges, and they are usually optimized for application at very specific and narrow ranges of wavelength. To work well at a given wavelength, there are several requirements:

(1) The resist photochemistry must take place efficiently. This is important for cost-effective manufacturing, since lengthy times to expose the resists reduce exposure-tool productivity. It should be noted that there is such a thing as an exposure dose that is too low, leading to dose-control problems and poor pattern fidelity because of shot noise.[10,11] (Shot noise is discussed in Section 3.8 and Chapter 12.)
(2) The resist must not be too optically absorbent. For high-resolution imaging, resist films need to be exposed with a high degree of uniformity from the

Table 3.1 Properties of azide-negative resists, compared to those of novolak-positive resists.

Azide/isoprene negative resists	Novolak resists
Swells during develop	No swelling during develop
Marginal step coverage	Good step coverage
Organic solvent developer	Aqueous developer
Toxic strippers	Environmentally benign resist strippers
Sensitive to ambient oxygen	Operates well in air

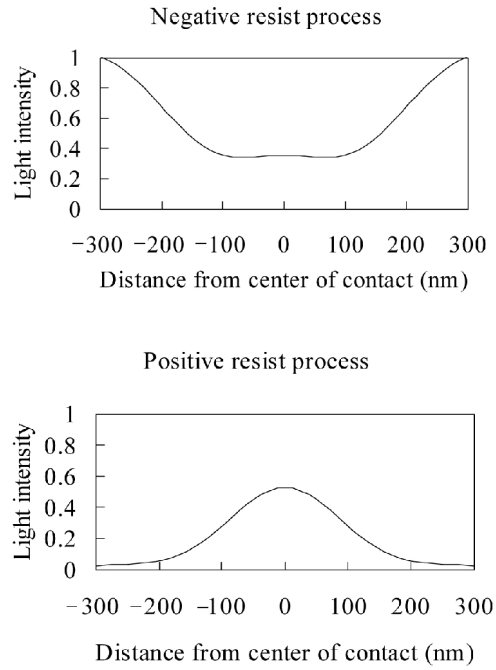


Figure 3.3 Simulated cross sections of the aerial images of 250-nm contact holes. The simulations were performed using Prolith2 and using the high-NA full-scalar model. The key parameters were wavelength = 248 nm, $\sigma = 0.5$, and $NA = 0.7$. A chrome/glass binary mask was assumed.

top of the resist film to the bottom, and light will not penetrate the resist film adequately if it is too optically absorbent.

(3) There must not be competing positive and negative reactions. These can occur when the light drives desired photochemical reactions in a positive resist but may also induce polymer cross-linking, a negative-resist characteristic.

These are some of the broad issues related to resists. The remainder of this chapter discusses specific technical issues in greater detail.

3.2 Adhesion Promotion

Adhesion promotion is required for photoresist to stick to surfaces of silicon, silicon dioxide, silicon nitride, etc., during resist development and subsequent etches. Silicon-dioxide surfaces are usually covered by a layer of physisorbed water that remains after resist coating if not first removed[12] (Fig. 3.4). This adsorbed water layer enables water or solvent to penetrate the interface between the resist and the substrate during resist development, causing a loss of adhesion. An adhesion-promotion process—often referred to as priming—is required to avoid this.

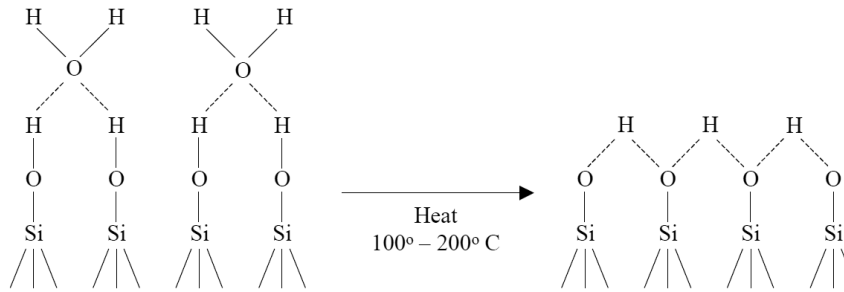


Figure 3.4 Thermal dehydration of a hydrated native oxide on a silicon substrate.

The first step in the adhesion-promotion process is dehydration. This is accomplished most frequently by heating the wafers in a vacuum to temperatures typically between 100 °C and 200 °C. The wafers are then introduced to a vapor of an adhesion-promoting agent, the most common of which is hexamethyldisilazane, $[(CH_3)_3Si]_2NH$, usually referred to by its acronym HMDS. This method of priming, where the priming agent is applied by exposing the wafer to a vapor, is referred to as vapor priming. The HMDS forms a film on the surface (Fig. 3.5) to which water does not adhere and prevents developer from undercutting the resist during the development step. HMDS has a vapor pressure of ~20 Torr at room temperature, which facilitates its application to the wafer surface. The ambient around the wafers can be reduced to $\leq$ 1 Torr, and then vapor can be introduced from a container of HMDS kept under vacuum. Keeping the HMDS under vacuum has the added advantage of extending its lifetime by avoiding reactions with atmospheric wafer and oxygen, which can lead to HMDS decomposition.[13] Alternatively, HMDS can be applied by bubbling nitrogen through the HMDS. Maintaining the HMDS under a nitrogen atmosphere also prevents degradation.

A likely decomposition reaction for HMDS is $[(CH_3)_3Si]_2NH + 2H_2O = [(CH_3)_3Si]_2O + NH_3$, and is very similar to the surface reaction. An ammonia smell is an indicator that the HMDS has degraded. Other materials besides HMDS are used for adhesion promotion, such as trimethylsilyldiethylamine (TMSDEA), N, N-diethylaminotrimethylsilane (DEATS)—$[(CH_3)_3SiN(C_2H_5)_2]$, or mixtures of DEATS and HMDS. Adhesion to untreated metals is usually not a

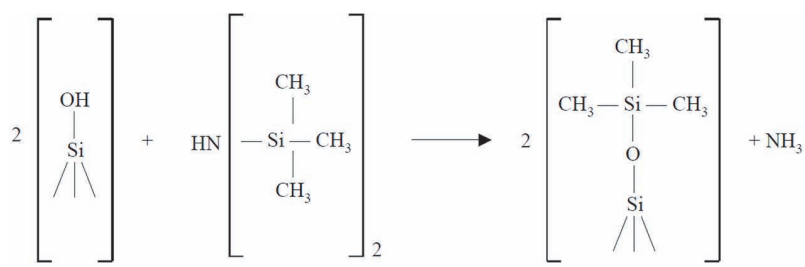


Figure 3.5 The adhesion-promoting agent, hexamethyldisilazane (HMDS), forms a water-resistant film on the surface of a wafer that prevents developer from undercutting the resist during the development process.

problem for semiconductor processing, but adhesion promoters are available for use on metals. For metal surfaces, chelating-silane adhesion promoters, such as trimethylsilylacetamide, have been proposed.[14,15]

The adhesion-promotion process can be monitored by measuring surface wettability. This is usually accomplished by placing a drop of water on a primed wafer and measuring the contact angle θ (Fig. 3.6). Forces are balanced in equilibrium such that

$$\gamma_{LV} \cos \theta = \gamma_{SV} - \gamma_{SL}, \tag{3.1}$$

where γ_{LV} is the surface tension of the liquid in equilibrium with its saturated vapor, γ_{SV} is the surface tension of the substrate in equilibrium with a saturated atmosphere of water, and γ_{SL} is the surface tension between the water and the primed substrate.[16] The angle θ is shown in Fig. 3.6, and a properly primed surface typically has values of 50–70 deg for the contact angle of water. Higher angles indicate greater degrees of priming, i.e., more hydrophobic surfaces and smaller values of γ_{SL}. From thermodynamic arguments, the work of adhesion of the liquid to the solid has been shown to be[17]

$$W_A = \gamma_{LV}(1 + \cos \theta). \tag{3.2}$$

There is such a thing as overpriming.[18] While priming has the positive effect of ensuring resist adhesion during resist development, it often has the effect of reducing the adhesion between the resist and the substrate. That is, the primary purpose of the priming is to prevent liquid from penetrating the resist/substrate interface during development[19] (Fig. 3.7). Large amounts of priming reduce this penetration, but may do so at the expense of true adhesion of the resist to the substrate. With overpriming, the resist may not wet the surface of the wafer adequately during the resist-coating operation, resulting in parts of the wafer that are not covered by resist. The resist may delaminate—"pop"—during exposure, because the resist is not adhering adequately to the wafer surface.

Vapor priming can be performed in stand-alone priming ovens or in single-wafer modules. With stand-alone ovens, it is possible to bake the wafers for a long time, thereby achieving maximum dehydration. On the other hand, full integration with the rest of the lithography equipment is possible when priming takes place in single-wafer modules. As might be expected, adhesion is a property that is

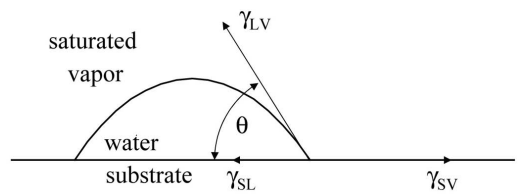


Figure 3.6 A drop of water on a surface.

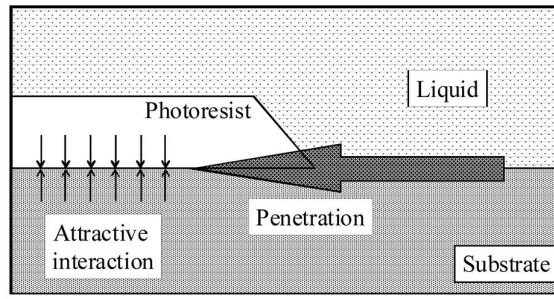


Figure 3.7 The penetration of liquid at the interface between the resist and the substrate.[19]

modulated by the resist composition. Some polymers will adhere better to surfaces than others.

3.3 Resist Spin Coating, Softbake, and Hardbake

Resist is applied to the wafer by a spin coating. In this method, photoresist is dispensed in liquid form onto the wafer (Figs. 3.8 and 3.9). The wafer is then spun about its axis at several thousand revolutions per minute (rpm). The resist is subjected to centrifugal forces, spreading the resist, and leaving a thin uniform layer adhering to the surface. The faster the spin speed, the thinner the resulting layer of resist (Fig. 3.10). The attractive molecular forces that make the resist viscous reduce the rate of thinning caused by centrifugal forces. Consequently, resists with higher viscosity produce thicker films than resists based on similar chemical platforms but of lower viscosity. As the large centrifugal forces near the edge of the wafer pull on the resist, the attractive molecular forces within the resist pull on resist material near the center of the wafer, where centrifugal forces are smaller.

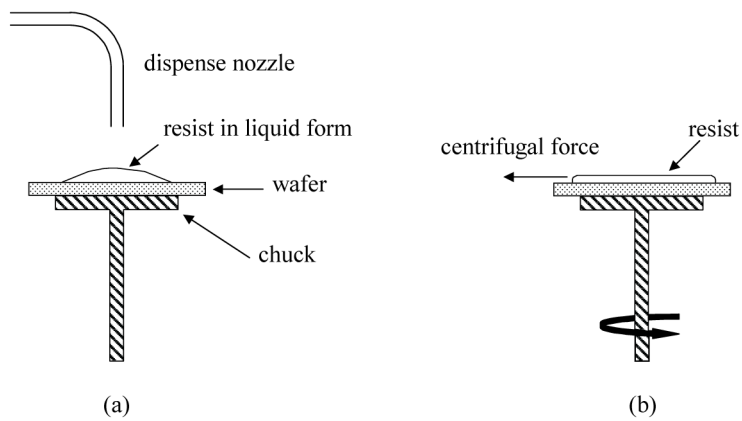


Figure 3.8 Schematic of the process of spin coating. (a) Resist is dispensed onto the wafer in liquid form. (b) The spreading and thinning of resist during spinning.

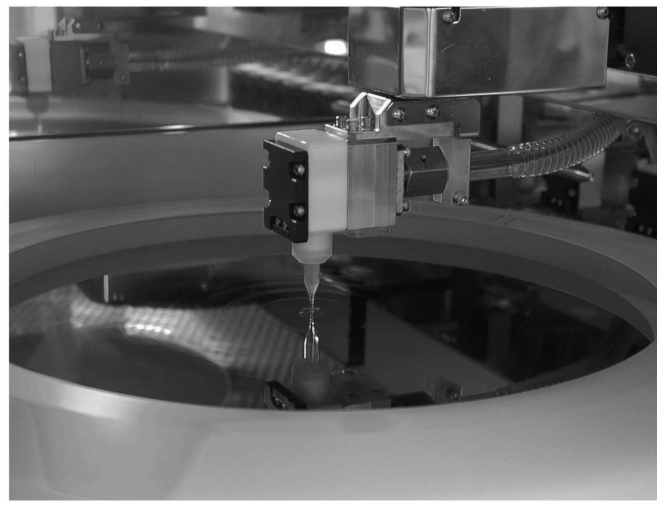


Figure 3.9 A bare silicon wafer being coated with resist in a DNS RF3 coat module. (Photo courtesy of Dainippon Screen Mfg. Co. Ltd.)[20]

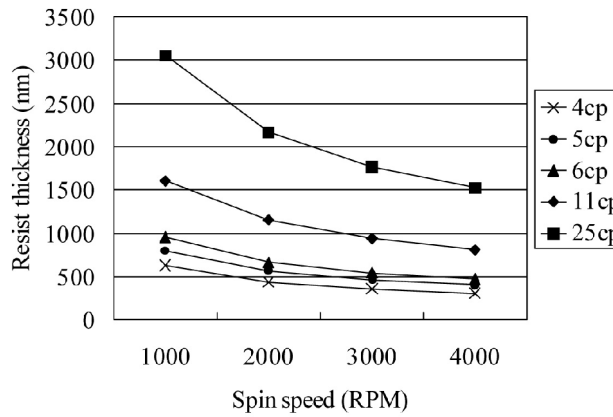


Figure 3.10 Resist thickness versus spin speed for resists of different viscosities from the TDUR-DP069 family of resists, spun on 200-mm wafers, and using a ProCell from Silicon Valley Group.[21]

Users select the rate of spin by first determining the final thickness desired. Average thickness can be varied by adjusting the rate of spin and/or resist viscosity. While a coater may be capable of spinning over a wide range of speeds, there is usually a smaller range over which the coatings are uniform. Typically, spinning either too fast or too slow produces nonuniform coatings, as illustrated in Fig. 3.11. Approximate ranges over which uniform coatings are produced are listed in Table 3.2. This range of spinning speed is also somewhat dependent on resist material and solvent. The viscosity of the resist is chosen to provide the desired thickness within the range of spin speed where uniform coatings are obtained. Fine-tuning average resist thickness is accomplished by the final rate of spin.

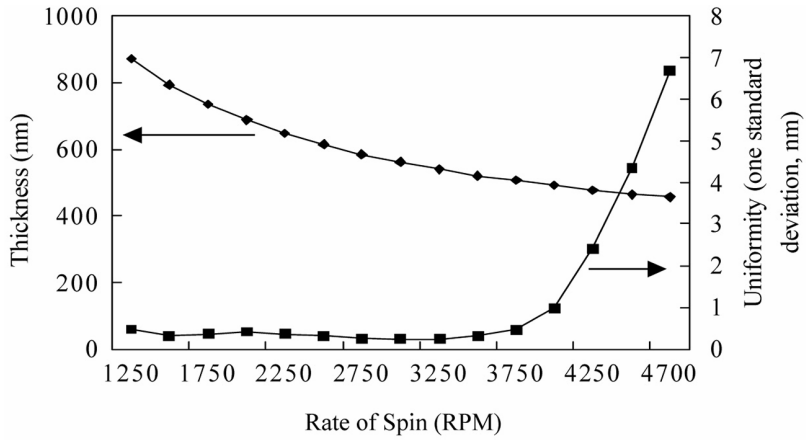


Figure 3.11 Average resist thickness and uniformity of Shin-Etsu 430S resist, coated on 200-mm wafers on an SVG ProCell.[21]

Table 3.2 Optimum spin speeds for coating photoresist.

Wafer diameter (mm)	Range of optimum spin speeds (rpm)
100	3000–5000
200	1500–3000
300	1000–2000

The process of fluid spreading on a spinning disk has been the subject of considerable study since spin coating has emerged as the primary method for applying resists. It is possible to gain insight into the spin-coating process by first considering a simple situation in which a flat wafer (at rest) initially has a uniform layer of fluid at a thickness of h_0. It is assumed that the fluid has constant viscosity during the spin-coating process. If the wafer is then spun at a rotation rate of f, starting at time $t = 0$, then the balance of forces is[22]

$$-\eta \frac{\partial^2 v}{\partial z^2} = 4\pi^2 \rho f^2 r, \tag{3.3}$$

where η is the fluid's viscosity, v is the velocity of material at radial position r and height z, and ρ is the mass density. There is also a conservation-of-mass equation:

$$\frac{\partial h}{\partial t} = \frac{-4\pi^2 \rho}{3\eta} \left[\frac{\partial \left(r^2 h^3 \right)}{\partial r} \right], \tag{3.4}$$

where h is the thickness of the resist at radius r and time t. If any other forces or physical processes are ignored, these equations result in[23]

$$h = \frac{h_0}{\sqrt{1 + \dfrac{16\pi^2 f^2}{3\eta} h_0^2 t}}, \tag{3.5}$$

where the thickness is independent of radius. This equation shows a film that thins to zero thickness as $t \rightarrow \infty$, while resists tend asymptotically to finite thicknesses. What is missing from the preceding analysis is solvent evaporation, which causes the viscosity to increase over time to the point where resist solidifies. It is this effect that leads to resist thickness that is fairly independent of spin time, following the initial transient thinning.[23] Solvent evaporation causes the resist to cool, thus affecting its viscosity and subsequent rates of evaporation. This evaporation takes place nonuniformly over the wafer surface and is a function of spin speed, exhaust, and solvent type. This variation in solvent evaporation rates can lead to resist-thickness nonuniformity. It is possible to compensate for nonuniform rates of solvent evaporation by choosing suitable initial temperatures for the resist, wafer, and ambient atmosphere, which are independently adjustable on modern resist coaters (see Fig. 3.12 in Color Plates). Since the resist thickness will depend upon the solvent evaporation, the coating process depends upon the resist's particular solvent.[24]

Resist can be dispensed while the wafer is motionless or spinning. These two methods are referred to as static and dynamic dispense, respectively. The dispense nozzle can be held fixed above the center of the wafer during dispense, or, with some hardware configurations, it can be moved. Movable dispense arms are useful for obtaining good resist coatings while using a small amount of photoresist on large-area wafers.

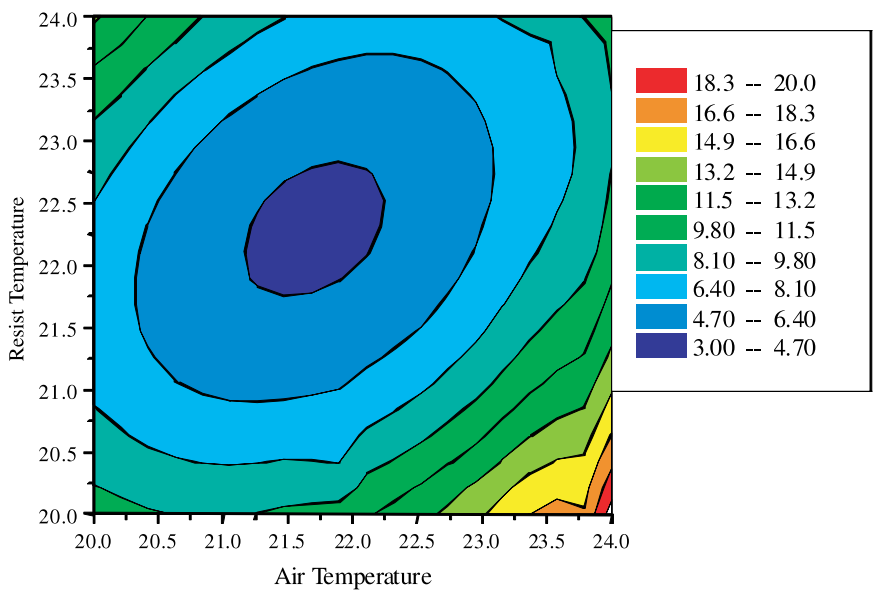


Figure 3.12 Resist uniformity contours as a function of air and resist temperatures for Shin-Etsu 430S resist coated on an SVG ProCell. The data are single-standard deviations, in units of Angstroms.[21] The initial wafer temperature was 22 °C. The most uniform coatings are produced with resist and air temperatures slightly different from the initial wafer temperature (see Color Plates).

Minimizing resist consumption is important for controlling the cost of lithography because photoresist is expensive. Resists used for noncritical layers can cost as little as $100 per liter, but deep ultraviolet (DUV) resists (particularly for 193-nm exposures) can cost five times that, or more, while resists for emerging technologies can cost as much as $2500 per liter. Since only a small percentage of the dispensed resist actually remains on the wafer after spinning (see Problem 3.1), it is useful to minimize the amount of resist that is actually dispensed. Because of this inefficiency, alternative methods of resist coating, such as the use of microinjectors[25] (such as found in ink-jet printers[26]) or spray coating,[27] have been pursued, but the quality of resist coatings produced by these techniques thus far has not justified their adoption.

The pumps that dispense photoresists are critical components of every resist coater. The pump ensures that defined and reproducible amounts of resist are dispensed onto each wafer. Because it is impossible to produce and package resist that is absolutely defect free, the resist is typically filtered at the point of use, from within the resist pump. When packaged, resists are filtered to levels that are one-half to one-third of the minimum feature size, which currently represents filtration down to sizes of 20 nm and smaller. Resists are filtered to similar levels at the point of use.[28,29] As features become smaller, so do the sizes at which resists are filtered. Because polymer solutions do not pass readily through filters with pore sizes significantly smaller than 50 nm, additional engineering is required to implement point-of-use filtration at very small pore sizes. Regardless, resists have been successfully filtered at the point of use down to 10 nm, and antireflection coatings (see Chapter 4) have been filtered to a size of 5 nm. An additional mechanism for removing defects can be introduced by using filter materials that have an affinity for defect-causing constituents in the resist.[30,31] In addition to particulates, defects in resist films can originate as bubbles in the resist. State-of-the-art pumps must also remove bubbles, which are particularly problematic when new bottles of resist are installed. When high volumes of resist are dispensed rapidly, the pump pulls a significant quantity of resist from the bottle. If the pull on the resist is too great, the resist cavitates, with the resist solvent forming bubbles. This can be avoided by using a suitably large diameter tube between the pump and resist bottle. Bubbles can also be pulled into the resist as the pump draws out the last of the resist from the bottle. This can be avoided by installing a new bottle of resist well before the prior one is empty, but this wastes expensive resist. This problem has been addressed by improved packaging where the resist containers consist of pouches, similar to those used to dispense blood. These pouches are often contained within a bottle[32] to ease storage and transport, and to secure the pouches against puncture.

Another task that the pump needs to perform in order to avoid defects is suck-back. After resist is dispensed, a meniscus of resist will form at the end of the resist nozzle (Fig. 3.13). If the pump does not actively provide some draw on the resist back into the dispense nozzle, a small quantity of resist may drop on the wafer. The suck-back mechanism of the pump prevents this from happening. If the suck-back is too large, there is the potential for bubble formation on subsequent dispenses, or

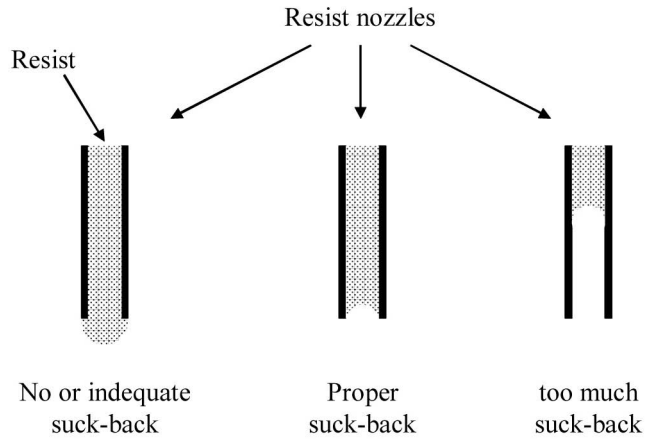


Figure 3.13 Various levels of suck-back in resist dispense nozzles.

the amount of resist dispense may not be well controlled. An optimum amount of suck-back is necessary, and is dependent upon the viscosity of the resist.

Resists are spun to different thicknesses to satisfy different requirements. The best resolution is often obtained with thin resists; however, resist films need to be thick enough to serve as etch masks or to block ion implantations. High-resolution lithography today typically employs resists with thicknesses in the range of 0.1–0.2 μm. Very thin coatings have not been used for several reasons, in spite of the resolution potential. First, the resist film needs to be intact throughout the entire etch process, during which the resist film erodes. In typical plasma-etch processes, the etch rate of the resist is less than one-third that of the underlying film, but in some instances can have essentially the same etch rate as the substrate film. The resist film must be thick enough to have resist remaining at the end of the etch process over those areas of the substrate that are not supposed to be etched. This includes areas involving substrate topography, where the minimum resist thicknesses often occur.[33] A profile of resist over topography is shown schematically in Fig. 3.14. Prior to the widespread use of chemical-mechanical polishing, requirements imposed by wafer topography often determined resist thicknesses. In particular, the resist thickness, as measured in large flat areas, must generally be much thicker than the underlying topography.

Besides the problems resulting from substrate topography and etch selectivity, ultrathin resist coatings were long avoided because of concerns with defects, particularly pinholes. Early studies showed large increases in pinhole density for resist coatings thinner than 150 nm.[34,35] However, subsequent studies[36,37] showed that low-defect ultrathin resist coatings could be obtained by applying the same careful attention to the coating parameters for ultrathin resists that was applied to coating production resists where yields are critical. However, ultrathin resist coatings are fundamentally susceptible to pinhole formation. Insight into the problem was provided by Okoroanyanwu,[38] who (referencing earlier work) showed that the thickness at which instabilities occur in spin-coated films is material-

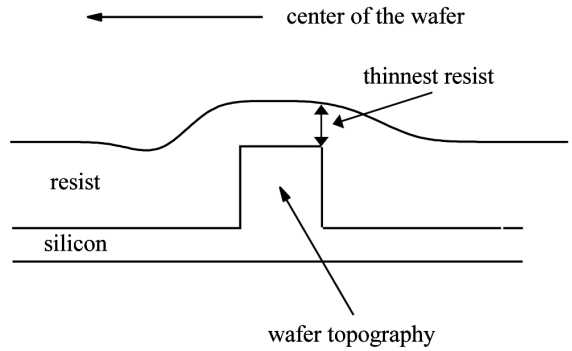


Figure 3.14 Substrate topography results in regions with thinner resist than found on flat wafers coated under identical conditions.

dependent,[39] and also depends on the interaction between the liquid resist and the substrate.[40]

This can be understood as follows. Consider a liquid film of nominal thickness h_0 on a flat solid surface. Perturbations in the surface, such as those that arise during spin coating, will induce a pressure gradient.[38] At position $\tilde{x}$ and time t these perturbations induce modulations in the thickness $h(\tilde{x}, t)$, which will be described by sums of terms of the form[40]

$$h(\tilde{x}, t) = h_0 + ue^{i\tilde{q}\cdot\tilde{x}}e^{-t/\tau}. \tag{3.6}$$

The time constant τ of the perturbation is given by[40]

$$\frac{1}{\tau} = \frac{q^2}{\eta}\left[\gamma q^2 h_0^3 - \left(\frac{A}{2\pi h_0}\right)\right], \tag{3.7}$$

where η is the fluid's viscosity, γ is the surface tension, and A is a material-dependent parameter. For a given wave number q and sufficiently thick film (large values of h_0), the time constant is positive, and the waves are damped. However, for films thinner than

$$h_c = \left(\frac{A}{2\pi\gamma q^2}\right)^{\frac{1}{4}}, \tag{3.8}$$

the instabilities will grow with time, until the depth of the instability is the same as the nominal resist thickness. At this point, a pinhole forms (Fig. 3.15). Note that the thicknesses at which the instabilities occur are material dependent. Thus, some early observations of pinhole formation were a consequence of intrinsic material properties, not simply the failure to ensure clean coating processes. For future technologies that may require ultrathin resist layers, resists need to be formulated so that they are robust with respect to pinhole formation.

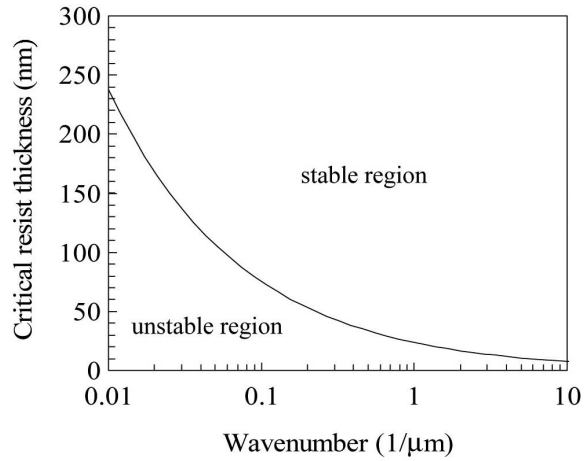


Figure 3.15 Critical resist thickness versus wavenumber for polystyrene.[39] $A = 8 \times 10^{-20}$ J and $\gamma = 40 \times 10^{-3}$ J/m^2.

Somewhat thick coatings (0.5–3 μm) tend to be required for implant masks, but the resolution required for these layers is usually less than for the most critical layers. Extremely thick coatings, where the thickness can be as thick as 100 μm,[41] are used for processing thin-film heads for disk drives and for micromachining applications.[42,43] The coating of very thick resists involves special considerations. One approach that reduces material consumption and improves uniformity is to use a closed coater bowl which has a solvent-saturated atmosphere.[44] This approach is also advantageous for coating square substrates, such as those used for fabricating photomasks, where obtaining good resist uniformity at the corners has always been a challenge.[45]

After the resist is dispensed and the wafer has completed its spinning, there is a thick ring of resist at the edge of the wafer referred to as an *edge bead*.[46] The edge bead exists not only on the upper surface of the wafer, but also on the wafer edge. It may even extend to the bottom surface of the wafer, near the wafer edge (Fig. 3.16), and because it can chip and flake, leading to defects, it is desirable to remove the edge bead prior to post-lithographic processing. In the steps that follow lithographic patterning, the wafer may be subjected to elevated temperatures, which will harden the resist and cause the remaining edge bead to be particularly difficult to remove. Consequently, it is best to remove the edge bead during the lithography process. Oftentimes, immediately following resist coat or softbake, the edge bead can be removed by using solvent streams or by exposure. Solvent edge-bead removal is usually applied in the resist-coating module, and the edge bead is typically removed by this method following resist coating. Alternatively, a separate unit may be used to expose the resist at the edge of the wafer, and the edge bead will be removed during the develop step. Optical edge-bead removal avoids solvent splashing that can sometimes occur with solvent edge-bead removal. However, there are some films applied by spin coating, such as organic antireflection coatings (to be discussed later), that are not photosensitive, and their edge bead cannot be

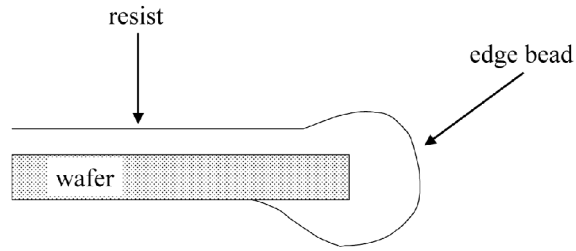


Figure 3.16 At the edge of wafers there is extra resist, known as the edge bead.

removed by optical exposure and subsequent develop.[47] One of the advantages of coating substrates in a solvent-saturated atmosphere is the reduction of the edge bead.

Recently there has been increased attention on edge-bead removal. One reason is the highly competitive nature of the semiconductor electronics business, which compels minimizing the width of the edge-bead removal in order to maximize the number of potentially good die on the wafers. As the width of the edge bead becomes smaller than 2 mm, control requirements become increasingly stringent. The introduction of immersion lithography, which is discussed in more detail in Chapter 10, injects additional motivation for a high-quality edge-bead-removal process.

Following coating, the resist film on the wafer usually does not have the desired dissolution properties in developer. For a positive resist, the develop rate of unexposed resist should be very low, but the resist, as coated, is often porous, retains considerable solvent, and has a high develop rate as a consequence. The most common method of densifying the resist is baking. The bake that immediately follows resist coating is referred to as the softbake, prebake, or post-apply bake.

Several methods of baking resist-coated wafers have been used. Convection ovens were used commonly in the early days of semiconductor processing, but they provided poor consistency and were not amenable to integrated, single-wafer processing. Several other methods were used, including infrared ovens, microwave ovens, and hot plates. Of all the baking methods tried, hot plates have provided the best temperature control, and they are used for all leading-edge resist processing today. On state-of-the-art tools, across-wafer temperature uniformity is $< 0.1\ ^\circ C$,[48] while the mean temperature can be controlled to within $\pm 0.05\ ^\circ C$ of the set point. Convection ovens continue to be used to bake very thick (5–100 μm) resists, where long bakes are required to evolve solvent from the depth of the resist film. Batch processing is most efficient in such cases. Thick resists are used in the making of thin-film heads for disk and tape drives, and for micromachining applications.[49]

Once wafers are removed from hot plates, they begin to cool, and this cooling needs to occur in a controlled way. Wafers simply removed from hot plates cool in an irregular fashion, depending upon the airflow, proximity to other still-warm wafers, and handling. The most effective way to ensure consistent cooling is to transfer baked wafers to chill plates after baking where they can be returned to room temperature consistently. However, consistency with chill plates still requires

attention. Consider the data shown in Fig. 3.17, which shows the temperature of wafers as they are placed on a hot plate, removed, and finally set on a chill plate.[50] When the wafers are placed at time equal to 0 on the hot plate, their temperatures rise quickly to the hot-plate temperature of 105 °C. After 60 sec on the hot plate, the wafers are lifted above the plate on pins, so that the wafers can then be picked up by a wafer-transfer arm and moved to a chill plate. The temperature of the wafers declines slowly while the wafers are sitting on the pins. Because the wafer-transfer arm may or may not be occupied with moving other wafers, the wafers may sit on the pins for variable times. The change in critical dimensions was measured for a 10-sec variation in time on the pins and was found to be 7.4 nm when the relative humidity was 42%, and 10.8 nm when the relative humidity was 49% for a conventional g-line resist.[50] Chill plates were introduced to semiconductor resist processing in order to provide reproducible bake processes,[51] and the example shown in Fig. 3.17 illustrates that even subtle aspects of the bakes need to be considered in order to squeeze out the last few nanometers of process control.

Hot plates need to be well designed in order to ensure good process control. A hot plate receives a thermal shock when a wafer is first placed on it, and the heating elements respond accordingly. If the hot plate controller is not well designed, there may be a temperature overshoot before the wafer temperature is returned to the desired value. This overshoot can cause statistically measurable shifts in linewidth.[52] In order to have a well-controlled process, the wafer temperature needs to be controlled across each wafer, wafer-to-wafer, and during the entire heating and cooling cycle for each wafer. State-of-the-art hot-plate designs control temperature throughout the heating cycle and meet uniformity specifications even while the wafer is heating, not just during the steady state.

Solvent evolves from the resist film during the softbake, and the rate of solvent evolution depends upon the transfer rate at the resist-gas interface.[53,54] Solvent content is an important characteristic because resist-development rates and diffusion during post-exposure bakes (discussed later in this chapter) depend upon

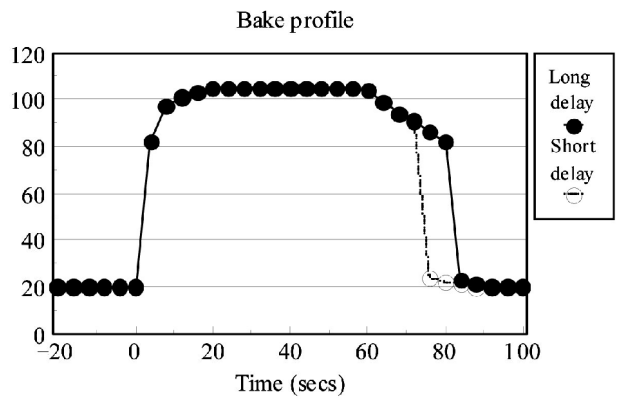


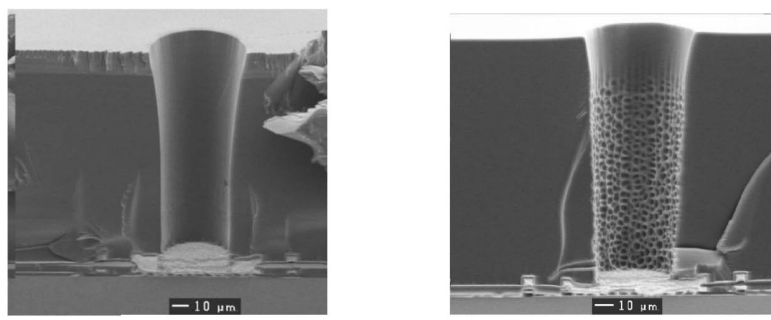
Figure 3.17 Changes in effective baking because of variable times between bake and chill.[50]

the solvent content of the resist. Thus, exhaust and airflow that play a role in resist baking need to be properly controlled.

For very thick resists, the need for long bakes is illustrated in Fig. 3.18, which shows scanning electron micrographs of cross-sectioned contacts formed in 100-μm-thick Shin-Etsu SIPR 7120 i-line photoresist. The contact on the left [Fig. 3.18(a)] received a hot-plate bake at 100 °C for 8 minutes, followed by a lengthy two-hour oven bake at 110 °C, while the contact on the right received much less baking. Pitting occurred in the developed resist because too much solvent remained in the resist as a consequence of inadequate softbake. Because it takes a long time for solvents to percolate through the polymer matrix comprising the resist, long bakes are required for very thick resists in order to remove solvent adequately from the portions of the resist film closer to the substrate.

After the lithography patterning process is completed, the wafers usually go into another processing step, such as reactive ion etch or ion implantation, that takes place in a vacuum. To preserve the vacuum integrity of the post-lithographic processing equipment, it is usually desirable to bake the wafers one last time in order to drive off low-molecular-weight materials that might otherwise outgas in the equipment used for post-lithographic processing. To achieve the desired goal, the required temperature for this bake is often above the decomposition temperature of key constituents of the photoresist, such as photoactive compounds or photoacid generators. For this reason, the high-temperature bake—usually referred to as a hardbake—is deferred until after the critical photochemistry and pattern formation has been completed.

The tools used for resist processing are often referred to as "tracks" because, in their earliest configurations, the wafers were moved along a track from the input cassette, to resist coat, then to bake, and finally to the output cassette. Modern resist-processing equipment often contains several, or all, of the modules needed to address various aspects of resist processing: vapor priming, resist coating, edge-bead removal, bakes, and develop. The straight-line track configuration has been replaced, and the sequence of processing can now be chosen more flexibly by the



(a) (b)

Figure 3.18 Cross sections of contacts in 100-μm-thick resist. (a) Softbake consisting of a hot-plate bake at 100 °C for 8 min, followed by a two-hour oven bake at 110 °C; (b) inadequate bake.[55]

user. Wafers are moved by robots from station to station, and operations that occur sequentially in time no longer need to be processed in adjacent processing modules. Nevertheless, lithographers still often refer to the resist-processing equipment as "tracks."

A number of companies, listed in Table 3.3, produce equipment for resist processing. Most of these companies produce tools that have capability for baking, developing, and removing edge bead, in addition to resist coating. Others make equipment that is more specialized or designed for laboratory applications, in which resist materials may be hand-dispensed.

3.4 Photochemistry of Novolak/DNQ g- and i-line Resists

Nearly all g- and i-line resists are based on a particular photochemistry.[56] These resists contain novolak resin (Fig. 3.19) and a diazonaphthoquinone (DNQ) photoactive compound. There are numerous papers and books that discuss the photochemistry of DNQ resists in detail,[57] and the interested reader can refer to such publications for additional details. The essentials are presented in this section.

Novolak resin is soluble in aqueous basic solutions, such as NaOH, KOH, or tetramethylammonium hydroxide [$(CH_3)_4NOH$] in water, that comprise the developers for DNQ resists. Tetramethylammonium hydroxide is usually referred to by its acronym TMAH. Diazonaphthoquinone is an insoluble material in such solutions, and when a sufficient amount of DNQ is mixed with novolak resin, the resulting mixture dissolves at extremely slow rates in basic solutions. However, DNQ reacts photochemically under exposure to light of a proper wavelength (300–450 nm). After absorbing a photon, the DNQ molecule evolves molecular nitrogen and forms a ketocarbene, a short-lived intermediate that rearranges into a more stable ketene structure (Fig. 3.20). This ketene is reactive with water, which is usually available in adequate amounts within the resist film for such reactions to occur, and the ketene ultimately converts to indenecarboxylic acid. This acid, unlike unexposed DNQ, is soluble in basic solutions and does not inhibit the dissolution of the novolak. This transformation of the DNQ under exposure to light from a compound that is insoluble in developer to one that is soluble is the foundation for the photochemistry of g- and i-line resists. Because the DNQ is a substance that reacts to light, it is referred to as *photoactive compound*. It is important that there be adequate water for the ketene to ultimately form indenecarboxylic acid, for the ketene may otherwise react with the novolak resin,

Table 3.3 Resist-coating equipment suppliers.

Company	Location
Tokyo Electron (TEL)	Tokyo, Japan
Sokudo	Kyoto, Japan
Headway	Garland, Texas
Suss MicroTec	Vaihingen an der Enz, Germany

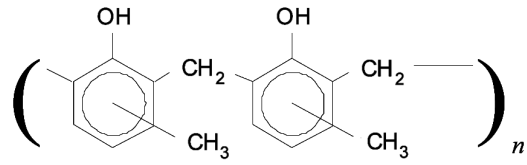


Figure 3.19 Novolak resin (ortho-ortho coupling).

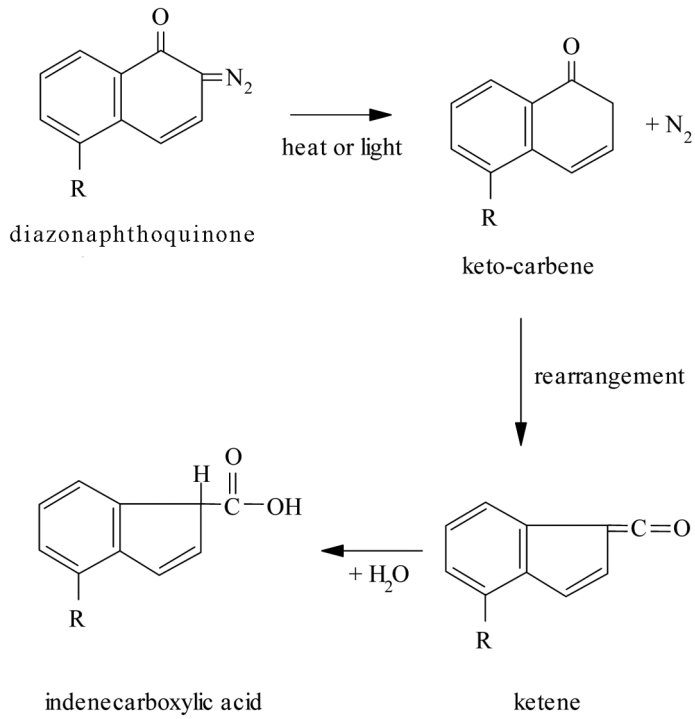


Figure 3.20 The basic photochemistry of diazonaphthoquinone photoactive compounds.

causing cross-linking. Such cross-linking is a negative resist behavior, rather than the desired positive one.

The original DNQ resists consisted of three constituents: novolak resin, photoactive compound, and solvent. It was found that coatings produced by such resists are susceptible to thickness nonuniformities, called striations, induced by the spin-coating process. Later families of resists contained *leveling agents*, which are materials that improve coating uniformity.[58]

Manufacturers of DNQ resists distinguish their products through different novolak structures and molecular weight distributions, and by using varying forms of the photoactive compound. The substituent R shown in Fig. 3.20, is usually a sulfonate (SO_3R')[59] and may be placed at different, or multiple, positions[60] on the DNQ molecule to improve their absorption characteristics at specific wavelengths. Changing the specific form of the sulfonate also modifies resist-dissolution characteristics. The novolak resin can be varied by using different

isomers of the basic cresol constituents, and by using different couplings between the cresol elements. Examples of different couplings are shown in Fig. 3.21. Polymer molecular weight distributions also affect resist dissolution properties.

The evolution of nitrogen can have significance. Very sensitive resists may evolve nitrogen rapidly—perhaps too fast for the nitrogen to diffuse to the surface of the resist—and bubbles may form in the resist. This may also occur when very intense light sources are used. Following spin coating and bake, the resist may contain residual stress, which may become relieved when the nitrogen is evolved, leading to delamination of the resist from the wafer surface after exposure.

3.5 Acid-Catalyzed DUV Resists

Until the advent of high-repetition-rate KrF excimer lasers in the late 1990s, DUV light sources were typically much weaker than g-line and i-line light sources. Early DUV lithography employed mercury-xenon arc lamps, which have spectral output between 240 and 260 nm and have low intensity relative to the g and i lines. Consequently, DUV photoresists needed to be significantly more sensitive to light than DNQ resists in order to achieve sufficiently short exposure times to maintain efficient exposure-tool throughput. This problem was solved by the invention of *chemical amplification*,[61,62] typically involving acid catalysis. These resists work as follows. The dissolution in developer of the backbone polymer of unexposed resist is prevented by protecting groups (Fig. 3.22). In addition to the protected polymer, the chemically amplified resist contains photoacid generators (PAGs) that represent the photoactive components of the resist. Examples of photoacid generators are shown in Fig. 3.23.[63] Upon exposure, the PAG produces an acid. The acid can catalyze the cleaving of protecting groups from the backbone polymer, though often requiring that the wafer be heated. This cleaving reaction is catalytic, in the sense that the acid still remains after the reaction, and is therefore available to promote the cleaving of additional protecting groups. With prolonged heating, a single photoacid is able to diffuse and cleave multiple protecting groups. By these means, a single exposure event is capable of producing multiple-deprotection

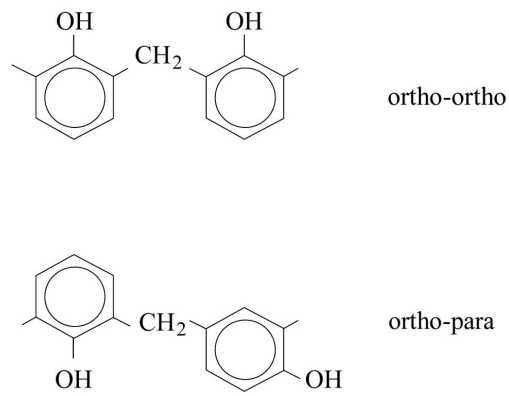


Figure 3.21 Different couplings in novolak that can lead to different resist characteristics.

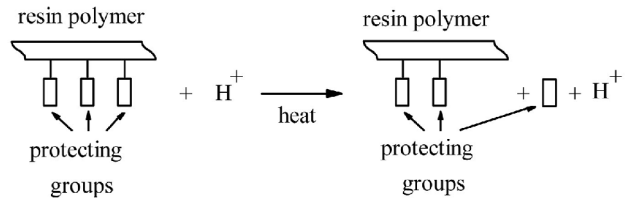


Figure 3.22 Basic mechanism for chemically amplified resists.

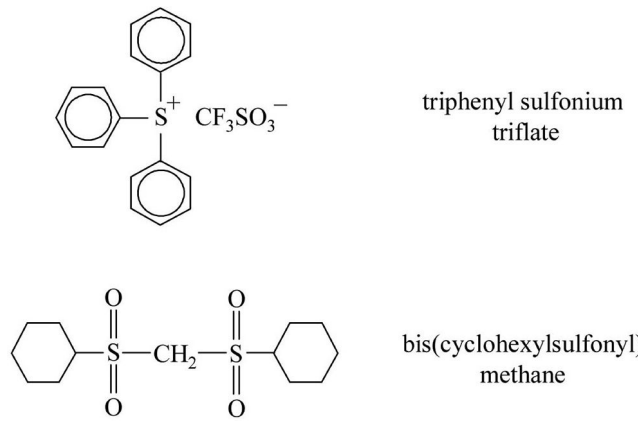


Figure 3.23 Examples of photoacid generators.

events (gain), hence the term *chemical amplification*. Resist behavior depends on the strength of the photoacid and its size, which affect its catalytic efficiency and diffusion characteristics, respectively.[64]

One of the other problems with developing useful resists for DUV lithography was the high level of optical absorption by novolak resins at DUV wavelengths. The problem of highly absorbing resist is shown in Fig. 3.24. Light absorption, as indicated by photoacid concentration, is modeled for a resist with parameters appropriate for a commercially available DUV resist, and hypothetical resists with significantly more absorption. For highly absorbing resists, the light fails to reach the bottom of the resist film, so the space could not be developed out. For less-absorbing resists, such complete failure of the resist process does not occur, but the resist profiles would be very sloped.

The problem of novolak's high DUV absorption was overcome by using polymers transparent at wavelengths $\leq$ 260 nm. Poly(hydroxystyrene) (Fig. 3.25) was one polymer known to dissolve in aqueous bases[65] and to be transparent at DUV (~250 nm) wavelengths. It is the polymer backbone of a number of positive resists used today for KrF lithography. As lithography is developed for shorter wavelengths, the problem of optical absorption reoccurs, and new chemistries need to be developed. No single chemical platform has been adopted for ArF lithography, and a number of different materials have been developed successfully.[66]

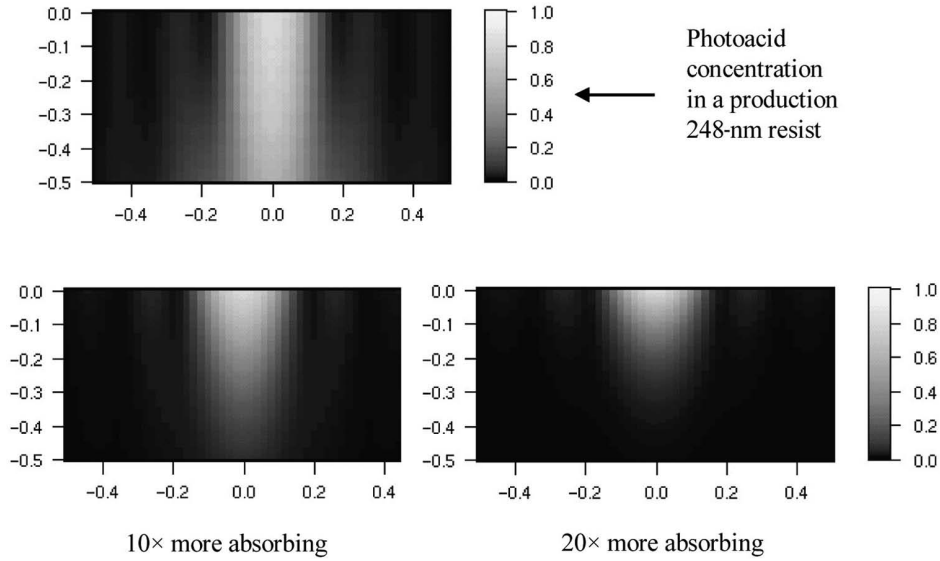


Figure 3.24 Photoacid concentration in a 200-nm space, modeled using Solid-C. The resist Dill parameters for the "production"-type resist are $A = 0.048\ \mu m^{-1}$ and $B = 0.054\ \mu m^{-1}$. (The Dill parameters are defined in Chapter 4.) The index of refraction for the resist is $n = 1.76$, and the resist thickness is 500 nm. The optics are 0.75 NA, $\lambda = 0.75$, and $\sigma = 0.5$. Light-shaded regions indicate areas where light has penetrated.

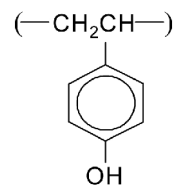


Figure 3.25 Poly(hydroxystyrene).

In one of the first DUV resists widely used at IBM, where chemical amplification was invented and advanced significantly, the first protecting group was t-BOC, shown in Fig. 3.26. Later, IBM developed APEX-E, which became the first chemically amplified positive DUV resist that was commercially available (from Shipley[67]).

Three-component DUV resists, most notably the acetal-based materials, that have functionality analogous to novolak/DNQ resists[68–71] have also been generated. The three-component resists consist of a matrix polymer, dissolution inhibitor, and a photoacid generator. These materials differ from the t-BOC and related resists in that the chemical species responsible for dissolution inhibition are not necessarily chemically bonded to the resin.

At short wavelengths, it is often difficult to find polymers for the resist backbones that have all of the desired attributes, such as transparency, etch resistance, and good adhesion to substrates. Consequently, resists can be

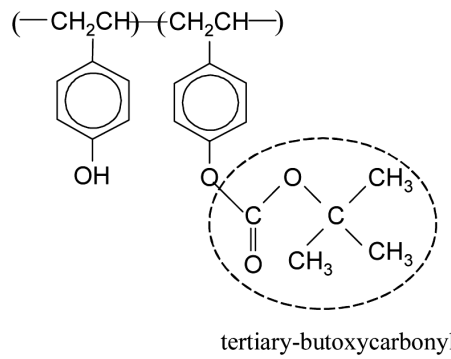


Figure 3.26 One of the earliest chemically amplified resists, with a t-BOC protecting group providing functionality and a poly(hydroxystyrene) polymer backbone.

constructed from functional subgroups,[72] as illustrated in Fig. 3.27. This breaks the problem down into separate functions that can be addressed separately by the resist chemists.

Chemical amplification enabled DUV lithography to have enough productivity to be a production-worthy technology. Unfortunately, there is no such thing as a free lunch. Following exposure of a chemically amplified resist, only a small quantity of photogenerated acid is produced. Prior to post-exposure bake, while the photoacids perform the task of deprotection, these acids are susceptible to neutralization by ambient bases.[74] Part-per-billion levels of amines, such as ammonia, or an amide such as N-methylpyrilidone (NMP) in the air have proven sufficient to affect the photochemistry significantly.[75] Such small quantities of basic chemicals can have a significant effect on a chemically amplified resist because only small quantities of photoacids are generated. Airborne amines affect the top of the resist film first. Short exposure to a low level of amines first leads to a T-topping of the resist profile [Fig. 3.28(a)]. This occurred for the resist shown in

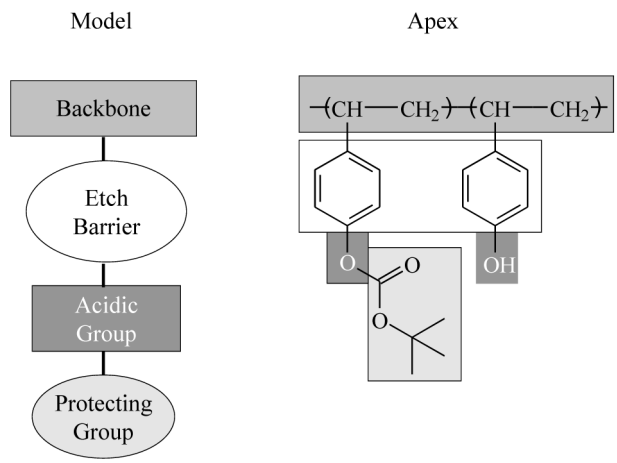
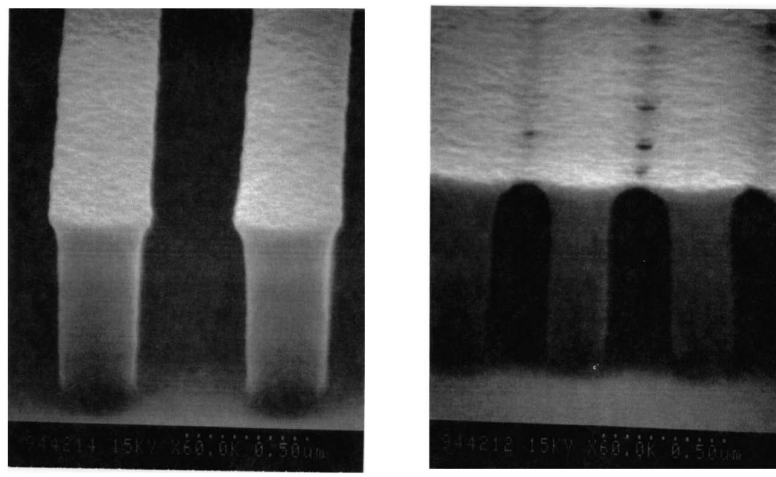


Figure 3.27 Model for constructing a chemically amplified resist.[73]



(a) (b)

Figure 3.28 A chemically amplified resist that has been exposed to amine-containing air during the time between exposure and post-exposure bake. With short exposure to amines, (a) a small amount of T-topping is induced, while (b) severe poisoning of the top of the resist occurs with longer exposures.

Fig. 3.28 because photoacid was neutralized in the top layer of the resist and was therefore unable to fully deprotect the top of the resist. With extended exposure to amines, the top layer almost failed entirely to develop [Fig. 3.28(b)].

Several methods have been employed to address this problem. The first—the brute force method—involves air filtration in the enclosed environment of the resist-processing equipment and exposure tools. Activated charcoal has been used to remove complex organic amines, such as NMP,[74] while weak acids, such as citric acid, have been used to neutralize ammonia.[76] Newer polymeric filters are capable of removing amines of low and high molecular weights[77] and have an improved lifetime before saturation,[78] an important consideration for maintaining a productive lithography operation. Cluster operations are used where resist-processing equipment is linked to the exposure tool. This makes it possible to process a wafer completely through the lithography operation (with the exception of most post-patterning metrology), without having the wafers leave a filtered environment, and it minimizes the critical time between exposure and post-exposure bake.

In another approach, a diffusion barrier is coated over the resist. This is typically another organic film spin coated onto the wafer immediately following the resist coating or softbake.[79–81] These overcoating materials, called topcoats, are typically designed to be water soluble and can be rinsed off the wafer prior to development. Topcoats can provide considerable immunity to ambient amines, but do so at additional expense, increased process complexity, and the potential for added defects. Topcoats also address the problem of T-topping that results from evaporation of photoacid from chemically amplified resist.[82] Additionally,

supplementary chemicals such as bases are also sometimes added to chemically amplified resist in order to improve their immunity to poisoning by amines.[83,84] In one particular approach, the bases are photoreactive and are converted to neutral materials upon exposure.[85]

A fourth approach involves the design of resists intrinsically less susceptible to photoacid neutralization. An example of this is the environmentally stable chemically amplified positive (ESCAP) chemistry.[86] Inhibition of base diffusion is the concept underlying the ESCAP approach to photoresist. It was observed that reduced free volume that inhibited diffusion caused polymers with lower glass transition temperatures to absorb less NMP.[87] Through the addition of metaisomers, it proved possible to reduce the glass transition temperature of t-BOC-related resists, providing reduced sensitivity to ambient amines.[88,89] The first widely used resist based on this chemistry was UVIIHS, sold by Shipley.

None of the approaches to reducing the susceptibility of chemically amplified resist through molecular design has proven capable of eliminating the poisoning of the resists by bases. Even the chemically amplified resist most resistant to poisoning by bases requires additional action. The approaches taken most commonly are air filtration coupled with cluster operation, where exposure tools are interfaced directly with the resist-processing equipment, as described previously. The reductions in the intrinsic susceptibility of the resists to poisoning by bases have proven useful, since linewidth-control requirements could not be met without the combination of built-in resilience coupled with a favorable environment.

Since none of the methods devised to reduce the susceptibility of chemically amplified resists to poisoning are completely effective, it is important to minimize the level of bases in the air. This is challenging since there are many potential sources of bases, and filtering is never completely effective. Amine levels in wafer fab areas can rise to >10 ppb levels when small quantities of developer or adhesion promoters are spilled. Many paints, plastics, and epoxies also outgas amines. Materials that initially evolve large quantities of volatile bases, but quickly reach low-outgassing states, can be used in the fabrication of lithography tools; however, their use can be problematic in an operating wafer fab. Some materials continue to evolve amines over long periods of time and should not be used in wafer fabs or in the construction of lithography equipment. Companies that manufacture lithographic processing equipment have developed lists of materials suitable for use in tools used for DUV lithography, and wafer-fab organizations have compiled similar lists of construction materials and paints.

The air is not the only source of bases that can have an impact on the resist profiles and linewidths. On many substrate materials, notably nitrides, chemically amplified resist has been found to form a “foot,” similar to the T-top shown in Fig. 3.28, except inverted.[90] Titanium nitride is particularly notorious for inducing footing of chemically amplified resists, but many silicon nitrides and oxides can also lead to footing depending upon the films’ deposition conditions. Similarly, acidic films can lead to a pinching of the resist width at the resist-substrate

interface.[91] There are several solutions to the problems of footing and bottom pinching:

(1) If possible, modify the substrate material.
(2) Insert a coating between the resist and the substrate layer. This is often an antireflection coating and is discussed in the next chapter. However, with resists that are extremely sensitive to bases, substantial amounts of bases can penetrate many coatings, and footing problems remain.
(3) Use a resist intrinsically robust against poisoning by bases.

Since the first two approaches are often unacceptable, resists with built-in resistance to poisoning are very desirable.

An interfaced operation of tracks and exposure tools has stimulated improvements in equipment reliability since track down time also keeps the stepper out of operation, and vice versa.[92] In addition to reducing the equipment failures that cause unscheduled down time, photocluster productivity can be improved through good management of preventive maintenance schedules that consider requirements of both the resist-processing equipment and the exposure tool. Stepper-track integration also requires a high degree of training for operators and technicians who need to be skilled in operating two pieces of equipment.

With chemically amplified resist, there is the potential for the evolution of organic material immediately following exposure. The photoacid generator decomposes into the photoacid and another material that may be sufficiently volatile to outgas from the resist. With resists where the deprotection can occur at room temperature, there is yet another chemical mechanism that can result in the outgassing of organic vapors during exposure.[93] The organic material evolving from the resist during exposure has the potential to coat the bottom of the projection optics. Bottom-optic contamination of multimillion-dollar projection lenses is clearly undesirable, and resists that have low levels of outgassing are preferable. Well-designed airflows between wafers and lenses can also mitigate the effects of resist outgassing.

3.6 Development and Post-Exposure Bakes

Basic, aqueous solutions are used to develop nearly all photoresists in use today for microelectronics fabrication. Most frequently, these are tetramethylammonium hydroxide (TMAH) solutions because they are free of metal ions that can degrade device performance or reliability. Sodium and potassium are particularly undesirable, since these alkali metals are known to reduce gate oxide integrity,[94] making NaOH and KOH solutions unsuitable as resist developers for IC manufacturing. The penetration of the developer cation into the resist depends upon the molecular structure of the resist and the developer type because the TMAH cation is more than twice the size of K^+.[95] An interesting effect of the large TMAH cation size is the decrease in development rate observed with increasing developer temperature, opposite the trend seen in NaOH and KOH developers.[96] Surfactants are often added to the developer to facilitate wetting of the resist.

Since surfactants may also modify development rates,[97] processes can often differ significantly depending upon whether the developer contains surfactants or not.

The concentration of developers is usually measured in terms of normality (N).[98] While resist processes can work in principle over wide ranges of developer normality, 0.26 N TMAH solutions appear to have become somewhat of an industry standard. This standardization has been instituted primarily to reduce developer costs, rather than because 0.26 N developer normality has been found to provide the greatest lithography capability.[99,100] Some resist-processing equipment has point-of-use developer dilution capability, which enables the use of more dilute developer while maintaining the cost advantages of a standard developer normality.

In the 1970s and early 1980s, wafers were typically developed by immersing batches of wafers in tanks of developer. This method has been largely superseded by single-wafer processes because of superior uniformity and control, relative to batch processes. Moreover, single-wafer processes are amenable to cluster operation, where the resist-processing tracks are interfaced to the exposure tools. The "puddle process" is the most common method for resist developing, where 100–200 cc of developer are dispensed on each wafer, forming a puddle that covers the entire wafer surface.[101] The developer is left on the wafer for the desired length of time, typically 30–75 sec. The developer is then rinsed off with deionized water. Additional puddles are sometimes applied to wafers in order to reduce defects and to reduce loading effects, where critical dimensions have a dependence on local pattern density, because the developer becomes depleted in regions where a lot of material is to be removed during the development process.[102]

Puddle development on a track system is the most commonly used method today. The tubes carrying the developer to the wafer are jacketed, and temperature is controlled actively to within ±0.2 °C. Since the process of development involves chemical reactions, resist-development rates are temperature dependent,[103] and temperature stability is required for a controlled process. Once the developer is on the wafer, evaporation causes the wafer to cool. Because the rate of evaporation is greatest on the outside of the wafer, thermal gradients are generated leading to different rates of development from the center to the edge of the wafer. (TMAH developers become more aggressive at lower temperatures.) Evaporation also causes developer concentration to increase. These effects can be minimized by using developer that is dispensed at a low temperature to begin with and by a low airflow around the wafer while the resist is developing.

For puddle development, the nozzle through which the developer is ultimately dispensed can be a simple tube or a more complex device. The developer nozzle must provide uniform application of developer and not be a significant source of defects. These two requirements have not always been easy to achieve simultaneously.

Developer has also been sprayed onto wafers, either continuously during the entire development process, or as a means to form a puddle of developer. Continuous spraying has the potential to consume large amounts of developer, and a clever ultrasonic nozzle has been devised that atomizes the developer in a jet of nitrogen;[104] however, this approach is not without susceptibility to defects.

Aqueous developers introduce effects that limit aspect ratios—the height of a resist feature divided by its width. As developer and rinse water are removed from developed wafers, surface-tension forces pull closely adjacent lines together (Fig. 3.29). This phenomenon fundamentally limits the aspect ratio of the resist features. The ultimate limitation involves a number of parameters, including the pitch and the structural properties of the resist,[105] but resist collapse is common when aspect ratios exceed 4/1 (height/width) and can frequently occur for even smaller values.

During the drying step, after the resist has been developed and rinsed, the level of the rinse liquid reaches a condition similar to that shown in Fig. 3.30, where the space between adjacent resist lines is partially filled with fluid. The fluid meniscus exhibits curvatures due to the differences in pressure across the fluid interface that result from surface tension in the confined geometry. These differences in pressure exert forces on the resist features that act perpendicularly and inwardly from the resist sidewall. These resulting capillary forces principally caused by pressure differences across the meniscus are given by:[106]

$$\Delta P = \frac{\gamma}{R}, \tag{3.9}$$

where γ is the surface tension and R is the radius of curvature (Fig. 3.30). R is related to the space width between the two resist lines:

$$R = \frac{d}{2\cos\theta}. \tag{3.10}$$

Thus, the force on the resist line is given by:[107]

$$F = hD\frac{2\gamma\cos\theta}{d}, \tag{3.11}$$

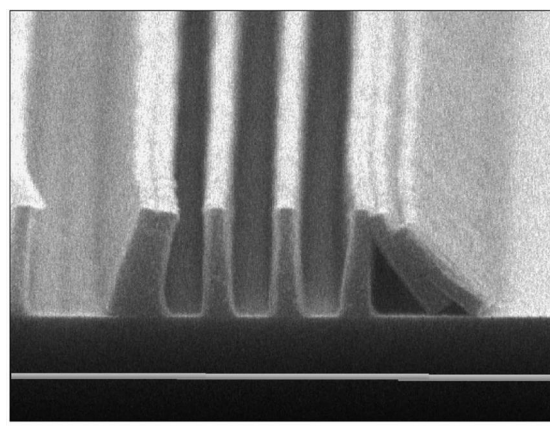


Figure 3.29 Closely spaced, high-aspect-ratio lines of resist that have collapsed after develop.

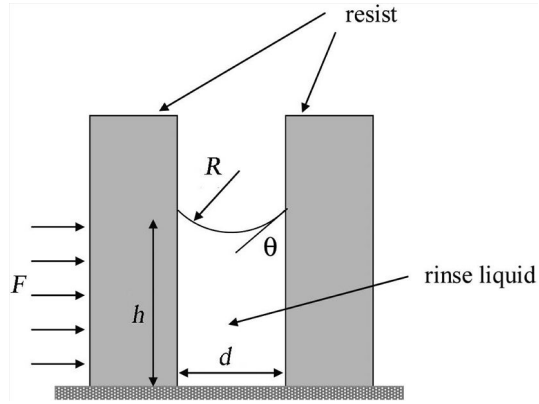


Figure 3.30 Forces on resist due to developer rinse liquid.

where D is the length of the resist line and h is the height of the rinse liquid. From Eq. (3.11), one can see that pattern collapse is lessened when the space d between resist lines is large.

Pattern collapse also depends upon the height of the resist line h. Since wider resist lines will be stiffer and more mechanically resistant to deformation (for a given applied force) it is to be expected that pattern collapse depends upon the height relative to the width, that is, the resist aspect ratio.[108]

Resist collapse can be reduced by adding a surfactant to the rinse liquid.[109–112] Surfactants are beneficial because they reduce the surface tension. Heating the resist during development has also been proposed as a way to reduce pattern collapse,[113] by hardening the resist and therefore increasing its structural stability.

Frequently, wafers are baked after exposure but before development. One reason this might be required is chemical amplification. Other motivations for incorporating a post-exposure bake into the resist process are discussed in Chapter 4. The post-exposure bake is usually a very critical process, and linewidths are typically more sensitive to this bake than to the softbake.

Linewidth sensitivities to post-exposure bake temperature for commercially available KrF resists are shown in Table 3.4. Resist suppliers have responded to industry requests for materials providing better linewidth control, as evidenced by reductions in post-exposure bake linewidth sensitivities in the progression of Apex-E → UVIIHS → UV6. These sensitivities have also driven hot-plate specifications since resists with essentially no post-exposure linewidth variation are not universally applicable to all feature types or have provided the best resolution.

As one might expect, the dependence of linewidth on hot-plate temperature control will be influenced by the quality of the optical image. This can be understood as follows.[116] With a post-exposure bake (PEB) at temperature τ for a controlled period of time, there will be an effective blurring of the exposure dose $E(x)$. If we assume that this blurring is described by a simple Fickian diffusion mechanism, then this "diffused image" $E_\tau(x)$ is related to the exposure dose $E(x)$

Table 3.4 Linewidth sensitivities of commercially produced KrF resists.[114,115] Not all of these resists are still being produced.

Resist	Supplier	Linewidth sensitivity (nm/°C)
Apex E	Shipley	16.0
UVIIHS	Shipley	7.5
UV6	Shipley	2.6
TM-461	JSR	2.6
DP-024	TOK	1.8
ARCH 2	Arch	~0

as follows:

$$E_\tau(x) = \frac{1}{\lambda_D \sqrt{2\pi}} \int E(z) e^{\frac{-(x-z)^2}{\lambda_D^2}} dz, \tag{3.12}$$

where λ_D is the diffusion length associated with the PEB process. If the temperature changes $\tau \rightarrow \tau*$, then

$$E_{\tau*}(x) = E_\tau(x) + \frac{\partial E_\tau(x)}{\partial \tau} \Delta\tau + \cdots. \tag{3.13}$$

At the different temperature τ^* the feature edge moves $x \rightarrow x + \Delta x$. Since

$$E_{\tau*}(x + \Delta x) = E_\tau(x) \tag{3.14}$$

$$= E_\tau(x + \Delta x) + \frac{\partial E_\tau(x)}{\partial \tau} \Delta\tau, \tag{3.15}$$

this rearranges to

$$\frac{E_\tau(x + \Delta x) - E_\tau(x)}{\Delta x} = \frac{-1}{\Delta x} \left[\frac{\partial E_\tau(x)}{\partial \tau} \right] \Delta\tau. \tag{3.16}$$

Since $\Delta CD = 2\,|\Delta x|$,

$$\Delta CD = 2\Delta\tau \left[\frac{1}{E_\tau(x)} \right] \left(\frac{\partial E_\tau(x)}{\partial \tau} \right) \left(\frac{1}{ILS} \right). \tag{3.17}$$

Hence, changes in linewidth induced by variations in hot-plate temperatures are also inversely proportional to the image log slope, indicating dependence on the initial optical images.

In order to satisfy the extraordinarily tight linewidth-control requirements of today's and future processes, hot plates must have excellent across-plate uniformity—less than 0.1 °C, or better—and must be matched to even tighter levels. This is a challenge in terms of achieving capability, and it also poses a formidable metrology challenge—it must be possible to measure hot-plate

temperatures and uniformities much less than control requirements. Resistance temperature detectors (RTD) are the best available thermometers today. In application, these are usually bonded to a silicon wafer, which are then placed onto the hot plates.

3.7 Operational Characterization

In Chapter 2, resist properties were incorporated into a simple model through a single parameter, the resist contrast γ. It is useful to revisit contrast and explore some of its properties. Consider once again the situation where a resist film is exposed by uniform illumination and then developed. The development rate R is directly related to the time rate of change of the resist thickness T:

$$R = \frac{-dT}{dt}. \tag{3.18}$$

With T as the resist thickness remaining after development,

$$T = T_0 - \int_0^t R(M)dt', \tag{3.19}$$

where T_0 is the initial resist thickness, and M is the fraction of photoactive compound or photoacid generator remaining after exposure. The development rate is a function of M. By definition, γ is given by

$$\gamma = \frac{-1}{T_0}\frac{\partial T}{\partial (\ln \varepsilon)} \text{ at } \varepsilon = 1, \tag{3.20}$$

where $\varepsilon = E/E_0$, E is the exposure dose, and E_0 is the minimum dose to clear the resist. This is the same parameter discussed in Chapter 2. From Eqs. (3.19) and (3.20),

$$\gamma = \frac{-\partial}{T_0 \partial (\ln \varepsilon)} \int_0^t R(M)dt' \text{ at } \varepsilon = 1, \tag{3.21}$$

$$\gamma = \frac{-1}{T_0} \int_0^t \left(\frac{\partial R}{\partial M}\right)\left(\frac{\partial M}{\partial \varepsilon}\right) dt' \text{ at } \varepsilon = 1. \tag{3.22}$$

This is an interesting expression. The integrand is divided into two factors. The first is governed by the development properties of the resist, and the second factor is determined by the exposure. For example, a dyed resist would be more absorbing, reducing the change in M near the resist-substrate interface during exposure, thereby reducing γ. This also explains the curious observation that measured contrast shows variations with resist thickness,[117] which is not characteristic of an intrinsic property of the resist chemistry.

Alternative definitions for contrast have been proposed. An expression similar to Eq. (3.22) has been derived:[118–120]

$$\gamma = \frac{R(T_0)}{T_0} \int_0^{T_0} \left(\frac{\partial \ln R}{\partial \ln E} \right) \frac{dz}{R(z)}, \tag{3.23}$$

where z is the depth in the resist film. (Again, technical rigor has been less than complete with the taking of logarithms of nondimensionless quantities.) The expression

$$\frac{\partial \ln R}{\partial \ln E} = \gamma_{th} \tag{3.24}$$

for the "theoretical contrast" has gained acceptance as a measure of resist performance.[121,122] Although there are difficulties associated with using this metric, such as its dependence on exposure dose, it has been found to be a good indicator of how well resists will perform.[123] Equation (3.24) has several advantages over the earlier expression for contrast as introduced in Chapter 2. It is dependent solely on the dissolution properties of the photoresist and does not have the dependencies on resist film thickness found in the earlier definitions. However, in the limit of very transparent or thin resists typical of KrF resists, the two expressions for contrast are roughly equivalent.[124]

3.8 Line-Edge Roughness

An issue for resists that is particularly significant for feature sizes smaller than 100 nm is *line-edge roughness* (LER).[125] For very small lines, structure at the molecular level contributes to roughness that can be a significant fraction of the linewidth (Fig. 3.31). A measure of edge roughness is the standard deviation of the actual line edge relative to an average line edge. This measure can be understood in more detail from Fig. 3.32. The actual resist line edge deviates from an ideal straight edge at position x by an amount $y(x) - \bar{y}$ from the average line edge $\bar{y}$. Typically, measurements are taken at regularly spaced intervals at discrete locations denoted by x_m. The most common measure of LER derives from the standard deviation of $y(x)$. An estimated value of LER from a line of resist measured at N points is given by

$$s_{LER} = \sqrt{\frac{1}{N-1} \sum_{m=1}^{N} [y(x_m) - \bar{y}]^2}. \tag{3.25}$$

Values for LER are typically given as 3× the standard deviation.

A parameter related to line-edge roughness is linewidth roughness (LWR). Similar in concept to line-edge roughness, LWR is a measure of the variation in linewidth along the length of resist lines. The linewidth at position x_m is $L(x_m)$

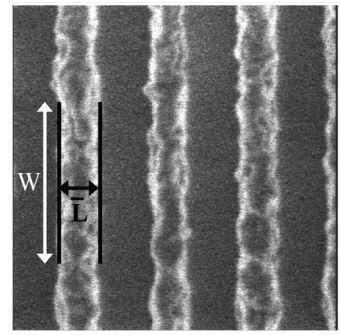


Figure 3.31 Lines of photoresist exhibiting LER.

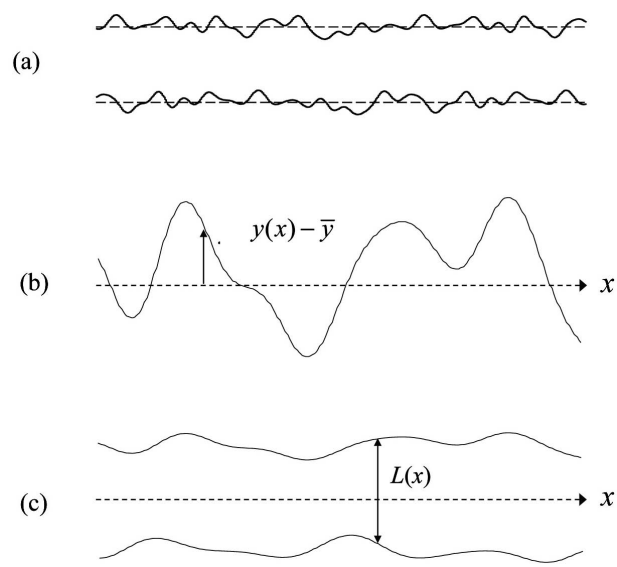


Figure 3.32 Illustrations of line-edge and linewidth roughness. (a) The ideal resist line edges are dashed lines, while the actual line edges follow the solid lines. (b) At position x the deviation of the actual line edge from the average line edge $\bar{y}$ is $y(x)-\bar{y}$. (c) The linewidth at position x is $L(x)$.

[see Fig. 3.32(c).] The measure of LWR is three times the linewidth's standard deviation:

$$LWR = 3s_{LWR} = 3\sqrt{\frac{1}{N-1}\sum_{m=1}^{N}\left[L(x_m)-\bar{L}\right]^2}, \tag{3.26}$$

where $\bar{L}$ is the average linewidth along the length over which LWR is measured. Because linewidth roughness has direct impact on transistor performance,[126–131] and gate linewidth control is a critical parameter for microprocessor performance, resist requirements in the International Technology Roadmap for Semiconductors

(ITRS) are stated in terms of LWR. When the roughness along the two opposite edges of a line are uncorrelated,

$$LWR = LER\sqrt{2}. \tag{3.27}$$

Edge roughness with $3\sigma \sim 3$–5 nm has been measured typically for leading-edge ArF resists. It has proven useful to look at the characteristics of line-edge roughness in more detail, beyond the total variation $3\sigma_{LER}$. In particular, it is useful to consider the roughness as a function of spatial frequency. This is done conveniently through the power spectral density (PSD) of the roughness:[132,133]

$$S(k) = \lim_{W\to\infty} \frac{1}{W} \left| \int_{-W/2}^{W/2} [y(x) - \bar{y}]\ e^{-2\pi ikx} dx \right|^2, \tag{3.28}$$

where $S(k)$ is the power spectral density at spatial frequency k. Large values of $S(k)$ indicate that there is a large amount of LER at the spatial frequency k. High values of k represent variations that occur over short distances. There may be a reticle contribution to the low-spatial-frequency component of the LER,[125] while high-spatial-frequency roughness is related more to the resist process.[134,135] LER has been measured as a function of spatial frequency, and representative results from an insightful paper by Yamaguchi and coworkers are shown in Fig. 3.33.[136] The data from several resists showed similar behavior for LER as a function of spatial frequency. In all cases, when plotted on a log-log scale the largest contributions to LER came from lower spatial frequencies (although not necessarily at the very lowest spatial frequency).

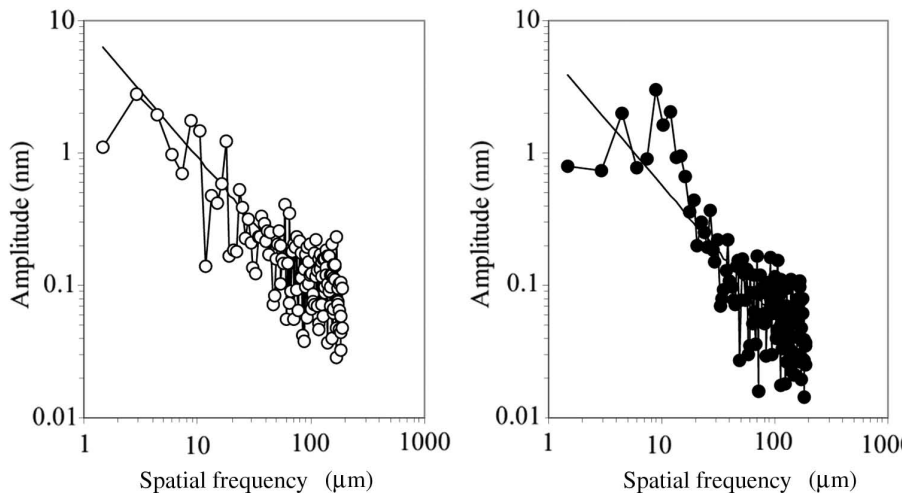


Figure 3.33 LER power spectral density as a function of spatial frequency. The results on the left were from e-beam exposures, while the data on the right were from ArF exposures. The electron beam was intentionally blurred to induce greater LER.

Because SEMs measure over finite distances, either as a practical matter or because the resist line has finite width W over which there is interest (such as for a transistor gate), it is useful to use a discrete and approximate version of Eq. (3.28):

$$S(k) = \frac{d}{N}\left|\sum_{m=-N/2}^{(N/2)-1}\left[y(x_m) - \bar{y}\right] e^{-2\pi i k x_m}\right|^2, \tag{3.29}$$

where $W = Nd$ and $x_m = md$. This discrete form also reflects the common practice of collecting images of resist patterns on grids of uniform spacing using scanning electron microscopes. Reconstruction of a line edge requires knowledge of the magnitude and phase of roughness at every spatial frequency, but $S(k)$ provides only the magnitude. However, simulated line edges can be generated assuming random phases:

$$y(x_m) - \bar{y} = \frac{1}{\sqrt{Nd}}\sum_{j=-N/2}^{(N/2)-1}\sqrt{S\left(k_j\right)}e^{\frac{2\pi i m j}{N}} e^{i\theta_j}, \tag{3.30}$$

where $kj = j/Nd$ and θ_j is a uniformly distributed random phase. Conventionally, the phase for $j = -N/2$ is set to zero.

A source of variation such as LWR will lead to across-chip and across-wafer linewidth variation. Consider, for example, gate lengths. Suppose we consider transistors that extend over an active region of width W (see Fig. 3.31). Each gate will have a linewidth $\bar{L}$ averaged over the dimension W of the gate. Because of LWR there will be gate-to-gate variations in $\bar{L}$. The magnitude of such gate-to-gate variability may be calculated from Eq. (3.30). A measured power spectral density can be obtained from SEMs by using Eq. (3.29) and the phases θ_j can be generated by using a random number generator to obtain values for the random phases.[137]

It is noted that the power spectral density typically follows the functional form:[138]

$$S(k) = \frac{2\sigma_{LER}^2 L_c}{\left(1 + k^2 L_c^2\right)^{0.5+\alpha}}, \tag{3.31}$$

where σ_{LER} is the standard deviation of the roughness. The parameter L_c is referred to as the correlation length. Hence, the LER for a large number of resists can be characterized by just three numbers, σ_{LER}, L_c, and α.

There is another function, the autocorrelation function, that is useful. If $y(x)$ is sampled at evenly spaced points $x_m = md$, then the autocorrelation function R at lag $x_{i+j} - x_i = x_j$ is calculated as:

$$R(x_j = x_{i+j} - x_i = jd) = \sum_{i=1}^{N-j}\left[y\left(x_{i+j}\right) - \bar{y}\right]\left[y(x_i) - \bar{y}\right], \tag{3.32}$$

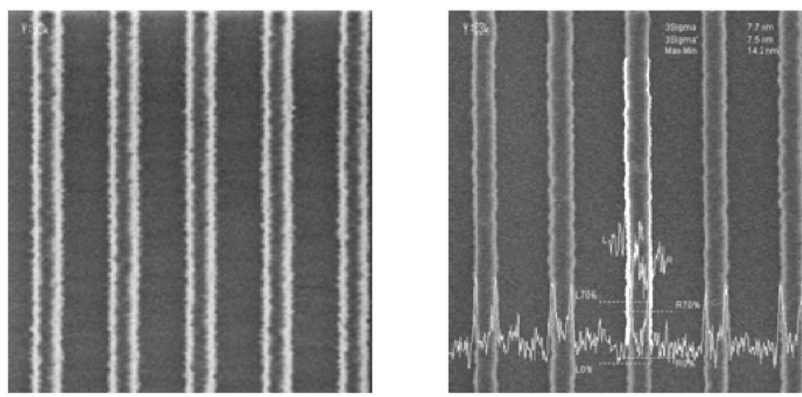
which may be computed from the measurements $y(x)$ of the deviation of the actual resist edge from a straight line [see Fig. 3.32(b)]. When $\alpha = 0.5$,

$$R(x) = \sigma^2_{LER}\, e^{\frac{-x}{L_c}}. \tag{3.33}$$

Hence, when $\alpha = 0.5$, the value of x at which $R(x) = \sigma^2_{LER}/e$ is the correlation length L_c. The values for the line edge $y(x)$ are typically obtained from scanning electron micrographs, using analysis software to ascertain the edge positions. Such analysis software is often available from the makers of scanning electron microscopes; alternatively, stand-alone software packages can be used.[139] From these values the autocorrelation function R can be calculated, which can then be used to determine the correlation length, as illustrated in the following example.

Figure 3.34 illustrates SEMs showing LER. Using data from these SEMs the autocorrelation functions are calculated, and these are shown in Fig. 3.35, normalized to σ^2_{LER}. According to Eq. (3.33), the values for x at which the normalized autocorrelation functions $= 1/e = 0.368$ provides the correlation lengths. For these examples, $L_c \approx 13.0$ for the resist patterns, while $L_c \approx 22.3$ for the patterns after etch.

The characteristics of line-edge roughness and its impact on critical-dimension variation has been studied as a function of the three parameters σ, L_c, and α. The amplitude of the line-edge roughness is directly proportional to σ_{LER},[137] and this behavior is illustrated in Fig. 3.36, where $y(x) - \bar{y}$ is plotted for two values of $3\sigma_{LER}$. The influence of L_c is shown in Fig. 3.37. For larger values of L_c the variations occur over longer distances and lower spatial frequencies. The parameter α also has an impact on the distribution of spatial frequencies, as can be seen in Fig. 3.38. As can be seen, more high-spatial-frequency roughness is associated with smaller values of α.



(a) (b)

Figure 3.34 SEM images for nominal 65-nm lines on a 190-nm pitch:[137] (a) resist (b) after etch into a layer of Si_3N_4.

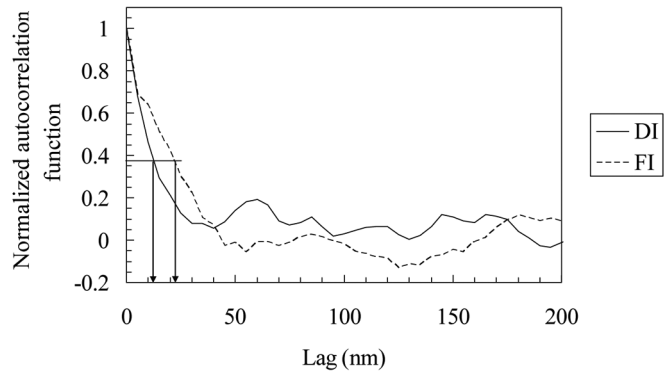


Figure 3.35 Plots of R/σ^2_{LER} for resist (DI) patterns and patterns after etch (FI).

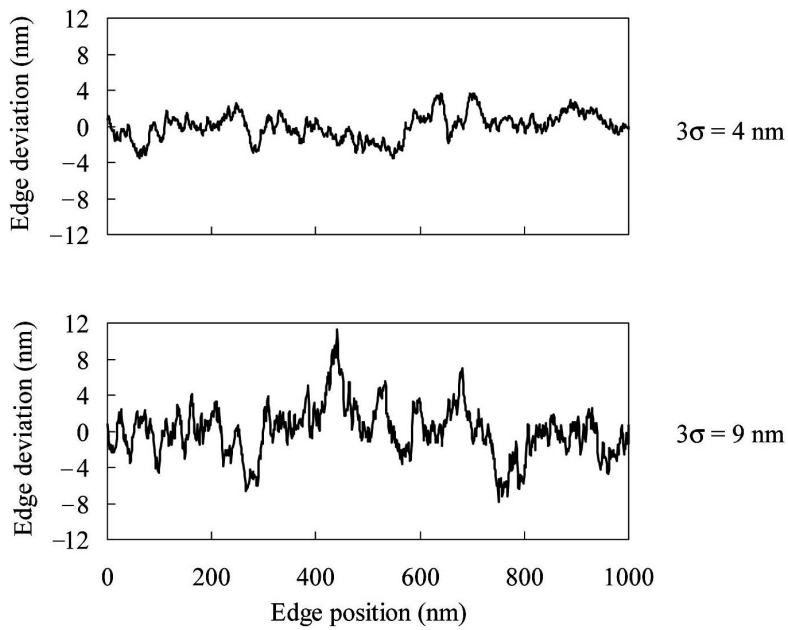


Figure 3.36 Simulated line-edge deviations, assuming L_c = 25 nm and α = 0.5.[141]

One of the impacts of LER is critical dimension (CD) variation. What is meant by a critical dimension is the distance $\bar{L}$ between two lines that are best fits over a distance W to the edges of resist (Fig. 3.31). Due to LER, the values of $\bar{L}$ will vary from location to location, resulting in critical-dimension variation σ_{CD} even in the absence of any other sources of critical-dimension variation. The amount of variation σ_{CD} is directly proportional to the standard deviation of the line-edge roughness σ_{LER}. Simulations have shown that the critical-dimension variation is a weak function of the parameter α, so it is useful to assume a value of $\alpha = 0.5$, and then the analytical form for the autocorrelation function [Eq. (3.33)] can be used. The dependence of σ_{CD} on L_c is shown in Fig. 3.39. The amount of critical

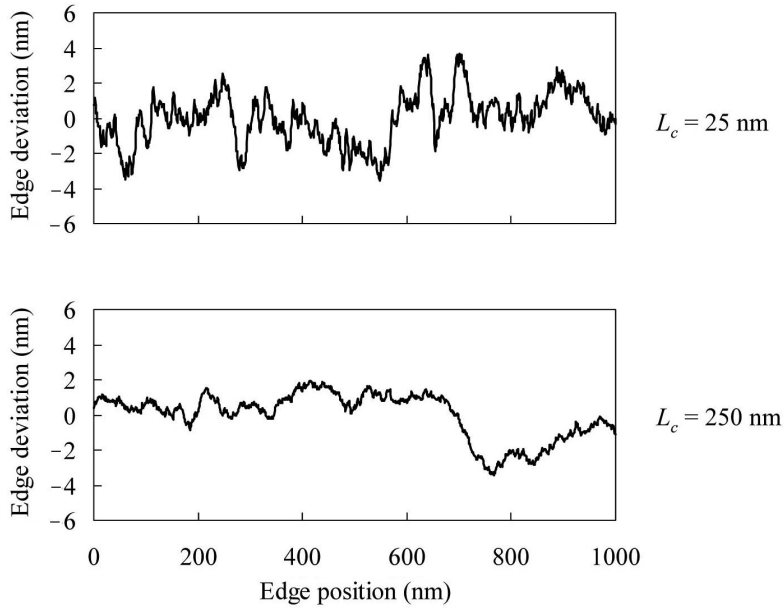


Figure 3.37 Simulated line-edge deviation, assuming $3\sigma = 4$ nm and $\alpha = 0.5$.[141]

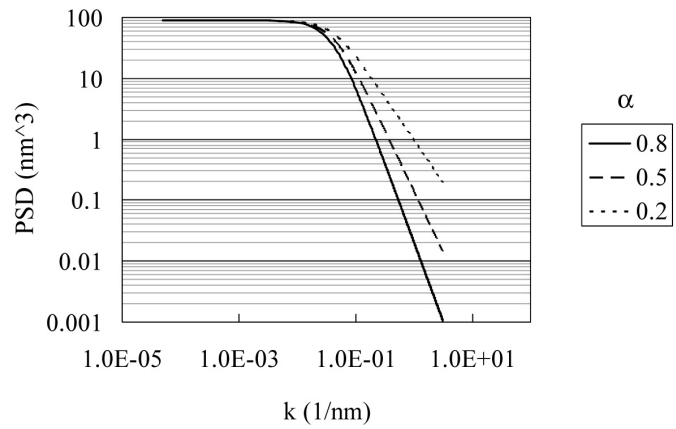


Figure 3.38 Graph of Eq. (3.25) showing the impact of α on the power spectral density of LER.[141]

dimension variation is larger for bigger values of the correlation length. Also, there is more variation for smaller values of W, for given values of σ_{LER}, L_c, and α. This is simply a manifestation of lower variation when averaging over a greater quantity of data. However, this does have a significant impact on scaling. It might be expected that the amount of critical dimension variation should decrease in proportion to shrinks in dimensions, such as gate lengths L. As can be inferred from Fig. 3.39, it is also necessary to compensate for decreasing values of W as well.

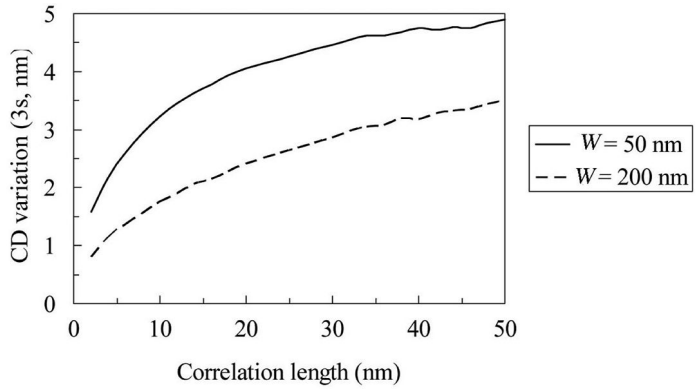


Figure 3.39 The dependence of CD variation on the correlation length. The graphs were generated by simulating CD variation for lines of resist of two values of W, 50 nm and 200 nm. For both cases it was assumed that $\alpha = 0.5$ and $3\sigma_{LER} = 4$ nm.

Fitting to simulated data, the dependence of σ_{CD} on σ_{LER} and L_c was found to be

$$\sigma_{CD} = \left(\frac{\log L_c + 0.71}{11.2}\right)\sigma_{LER}, \tag{3.34}$$

where the logarithm is base 10, and L_c is in units of nanometers. This equation shows the direct relationship between σ_{CD} and σ_{LER}, as well as the weaker, logarithmic dependence of σ_{CD} on L_c.

There are a number of factors that contribute to line-edge roughness,[142] not all of which are directly related to the resist materials and chemistry. An optical image that has a diffuse line edge might be expected to lead to resist patterns with edges that are not sharply defined, and this is indeed the case. It has been observed that LER increases as the image log slope (a good measure of optical edge acuity, introduced in Chapter 2) becomes small.[143,144] However, from Fig. 3.40, it does appear that LER reaches a nonzero minimum value, even for images with large edge slope. This indicates more fundamental sources of line-edge roughness. Some LER certainly originates at the molecular level, but substantial low-spatial-frequency LER can be seen in Fig. 3.33, occurring over long distances relative to molecular-length scales. Some of the low-frequency LER on the wafer may get transferred from LER on the mask.[145,146] Naturally, the high-frequency LER from the mask is filtered by the lens.

When exposures become low, statistical variations in the number of photons involved in exposing the resist can contribute to LER.[148] This results from basic photon statistics (shot noise), where the root-mean-square (rms) variation in the number of photons Δn is related to the average number of photons $\hat{n}$ by the following expression:[149]

$$\frac{\Delta n}{\hat{n}} = \sqrt{\frac{1}{\hat{n}} + 1}. \tag{3.35}$$

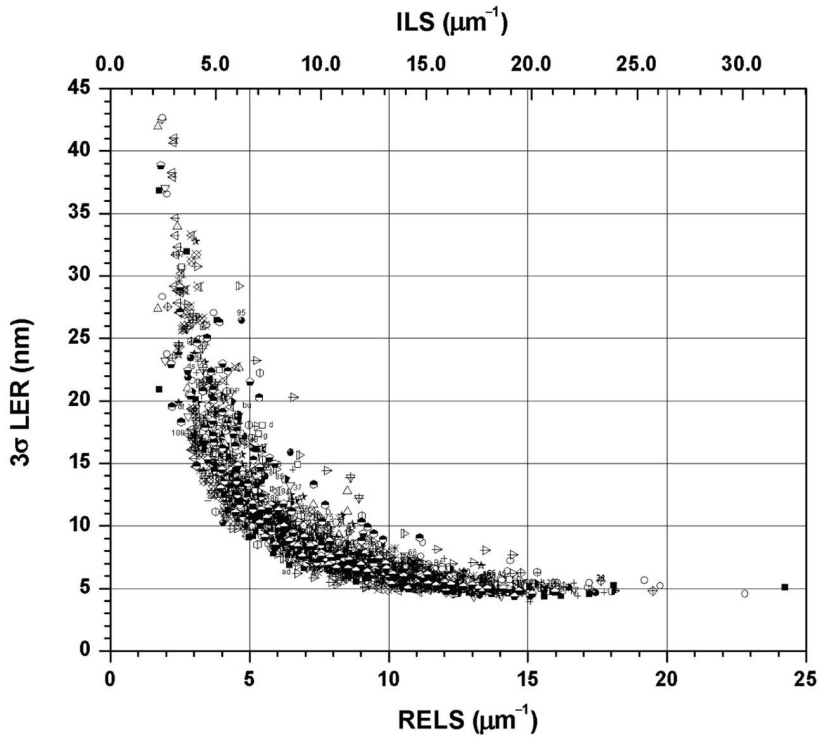


Figure 3.40 LER versus image log slope and resist-edge log slope (RELS) for Sumitomo PAR735 resist exposed on a 0.75-NA ArF scanner at the isofading condition. Fading is explained in Chapter 5, and details on the isofading condition can be found in Ref. 147. The RELS parameter can be considered an empirically derived image log slope. The data involved a wide range of linewidths, pitches, doses, and focus conditions.[147]

When the number of photons becomes small, the statistical fluctuation can be appreciable. Consider, for example, ArF light passing through the top of a resist film in an area that is 10×10 nm. For an intensity of 1 mJ/cm^2, approximately 1000 photons on average enters this area. From Eq. (3.35), this light intensity will fluctuate approximately ±3%. At the edge of features, the light intensity is typically $\sim 1/3$ that found in the middle of large features, so the effect of photon statistics is greater at line edges. An increase in LER at low doses has been observed and is shown in Fig. 3.41. As dimensions shrink, and tolerable LER becomes smaller, the area over which fluctuations become significant also decreases.[150] See Problem 3.3 to pursue this further.

As one might expect, diffusion (which occurs during post-exposure bake) serves to reduce LER. This is related to the original motivation for chemical amplification—addressing the productivity of exposure tools with low light intensity. In ArF chemically amplified resists, the photoacid diffuses with a diffusion length that is typically 20–30 nm. The amount of line blurring in KrF ESCAP resists was measured to be ~50 nm (FWHM),[151,152] although about half that of other KrF resists.[153] For simple Fickian diffusion, the diffusion length =

$2\sqrt{Dt}$, where D is the diffusivity and t is the time during which diffusion takes place. Diffusivity typically increases with higher temperatures, so diffusion will be less at low temperatures and with short bake times. This behavior is shown in Fig. 3.42. Diffusion over such length scales may serve to smooth out the roughness, but it is problematic when the amount of blurring becomes comparable to the half pitch.[154] At this point, the optical image originally projected into the resist will not be seen in a developed resist pattern, since the diffusion during post-exposure will blur the pattern excessively.

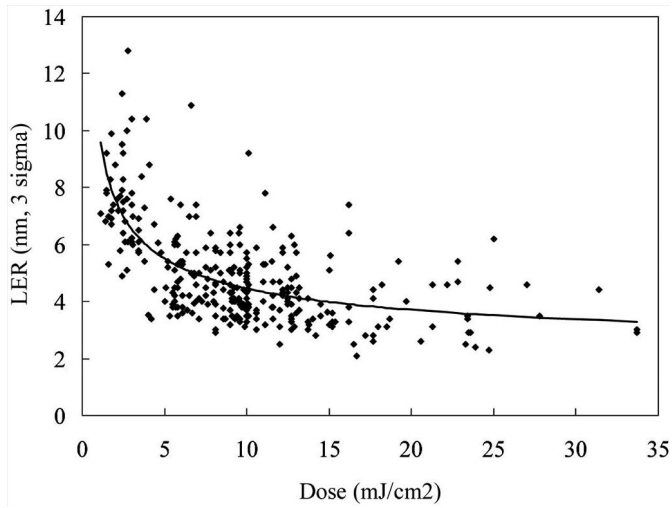


Figure 3.41 LER versus exposure dose for resists exposed on an extreme-ultraviolet (EUV) 0.3-NA microexposure tool with a synchrotron light source[161,162] and corrected for mask-absorber contributions to LER.[163] The solid line is a fit of $LER = \left(a/\sqrt{dose}\right) + b$ to the data.

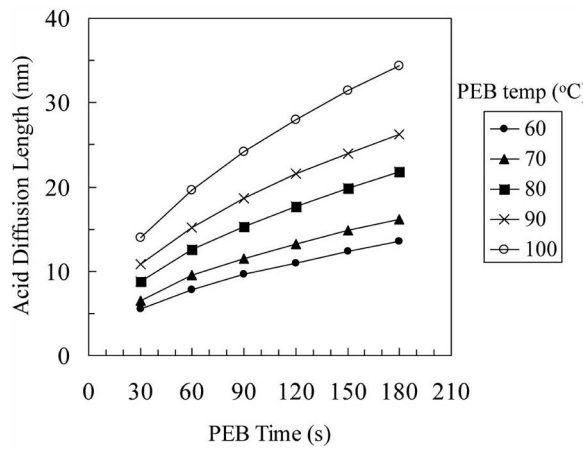


Figure 3.42 Measured acid diffusion length in a KrF resist.[156]

Although diffusion is not a useful way to reduce LER for sub-50-nm lithography, there are some other methods that appear to have some beneficial effects. Directly addressing line-edge roughness can be difficult, particularly when there is a conflict between a need for low LER and low exposure doses for exposure-tool throughput. There has been work to apply post-develop processing to reduce LER, including bakes,[155] etches, vapor treatments, ozonation, and chemical rinses,[156] all of which provide some reduction in LER.[157,158] Of these, vapor smoothing and surface-conditioning rinses appear to have the most benefit and can easily be implemented. However, most of this improvement occurs at high spatial frequencies and not at the low-spatial-frequency LER that has the greatest impact on CD variability.[159,160]

3.9 Multilayer Resist Processes

To address problems associated with significant substrate topography, highly reflective substrates, or highly absorbing imaging layers, multilayer resist processes have been developed. The most basic of these, the bilayer resist process, is outlined in Fig. 3.43. In this technique, the substrate is first coated with an organic planarizing underlayer. In a second coating operation, an imaging photoresist is spun on top of the underlayer. The optical properties of the underlayer are usually set to minimize reflections from the substrate back into the imaging

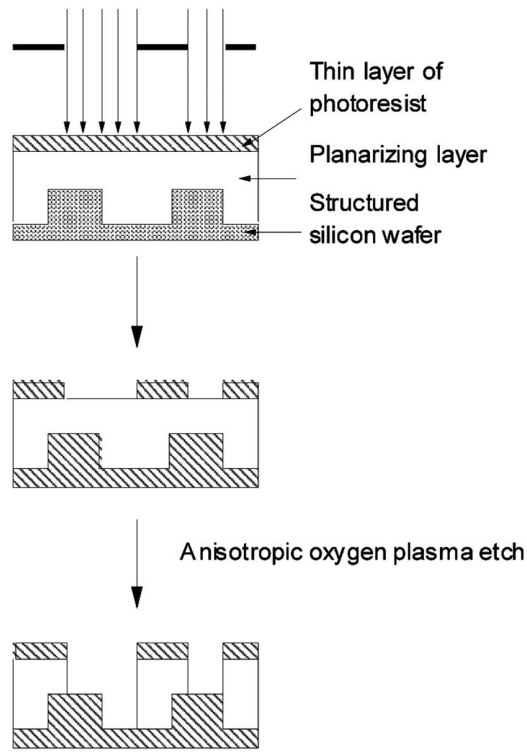


Figure 3.43 Bilayer resist process.

layer. The imaging photoresist generally contains a significant amount of silicon.[164] After the resist is exposed and developed, the wafers are placed into an anisotropic oxygen-reactive ion etch that transfers the pattern into the underlayer.[165] The silicon in the imaging layer provides good etch resistance to the oxygen.

The bilayer process described above differs somewhat from the first bilayer process, as it was originally proposed.[166] As first conceived, the imaging layer was a conventional novolak resist coated on a DUV resist material, such as polymethylmethacrylate. The resist was exposed at the dominant wavelength of the time, the mercury g line. The pattern transfer was accomplished by flood exposing the wafer to uniform DUV light, which was strongly absorbed by the novolak resist. As features became smaller, and shorter wavelengths were used for imaging the top layer, the imaging resists became transparent to DUV light. Alternative methods, such as reactive ion etch, are needed for the pattern transfer into the underlayer.

Because of their greater process complexity relative to single-layer resist processes, bilayer processes have been used primarily in technology development. One early application of bilayer-resist processes produced 0.5-μm features and smaller on a stepper that had a rated resolution of only 1.25 μm.[167] However, bilayer processes have been used in production as well.[168]

In trilayer resist processes, a third film, usually a sputtered inorganic material, is deposited on top of the organic planarizing film prior to coating the imaging layer. There are a number of reasons for using such a film:

(1) During the coating of the imaging layer, there is the potential for the solvent to dissolve the planarizing layer, unless the two films have been engineered in advance for compatibility.[169] The insertion of an inorganic layer obviates this additional materials engineering.

(2) To act as an etch mask for the organic planarizing layer, the imaging layer must contain silicon. The use of an inorganic middle layer enables the use of ordinary photoresists for the imaging layer, with the sputtered films acting as a hard mask.

(3) Inorganic hardmasks often provide superior etch selectivity to oxygen etches, thereby providing capability for thicker underlayers.

Trilayer processes are obviously more complex than bilayer processes (which are already complex enough!), but prove useful in several situations, such as the manufacturing of recording heads for magnetic data storage.[170] To reduce the complexity of multilayer resist processing, alternative techniques have been proposed that have many elements in common with the bilayer resist process described above. One class of alternative methods is referred to as top-surface imaging, often designated by its acronym, TSI.[171,172] The basic top-surface imaging process is outlined in Fig. 3.44. A thick resist layer is coated on the substrate. The effect of exposure for this resist material is to modify the diffusivity of silicon-containing molecules in the resist. In the negative-tone embodiment of

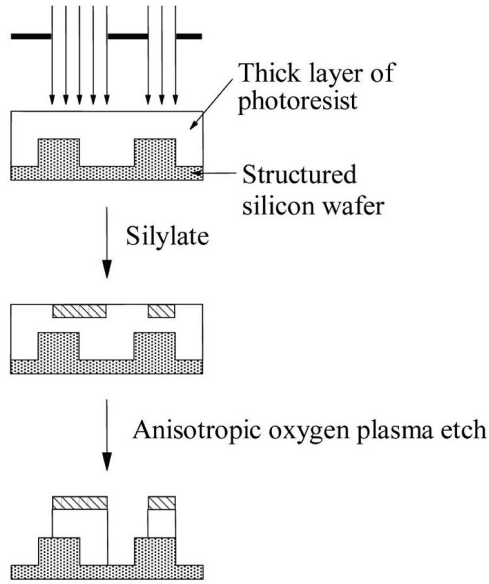


Figure 3.44 Top-surface imaging resist process.

top-surface imaging the diffusivity is increased in the exposed areas, while it is decreased in the positive-tone version of the process. The negative-tone process is described here and illustrated in Fig. 3.44. After exposure, the resist is placed in an ambient vapor of silicon-containing molecules, and the wafer is usually heated. The silicon-containing molecules are absorbed selectively in the exposed regions during this step, which is referred to as silylation. After the exposed regions have absorbed a suitable amount of silicon, the wafers can go to the oxygen-reactive ion-etch operation. The silicon blocks the oxygen etch while the unsilylated material will be etched. One of the earliest implementations of top-surfacing imaging was called DESIRE, standing for diffusion-enhanced silylating resist.[171] The resist, having suitable properties used in this early work, was called PLASMASK.[172] The TSI process contains the basic steps of the bilayer process, but requires only a single resist coating thereby reducing process complexity—at least in principle. Unfortunately, TSI processes have proven difficult to control, and have generally not found their way to use in manufacturing.

A simple but effective resist process involving multiple coatings is contrast enhancement. In this process, a highly bleachable film is coated over the resist-imaging layer.[173] The effect of these films is to enhance the contrast of the resist process.[174] By using contrast-enhancement materials that are water soluble, the process is simple to integrate. After exposure, and often before post-exposure bake, the contrast-enhancement layer is rinsed off with water.[175] Unlike other multilayer resist processes there are no additional etches required. While contrast-enhancement layers have been used in semiconductor processing, their level of use has diminished with the advent of high-contrast resists. In some applications, such

as processing thick resist layers for magnetic recording or micromachining, there is still a challenge to produce nearly vertical resist sidewalls. Contrast-enhancement films are used in these applications.

In this chapter, the basic chemistry of resists has been outlined. To be useful to process engineers, resist behavior needs to be captured in models that enable prediction. Such models do not need to be traceable to chemical first principles, but must be as compact as possible, consistent with providing accurate predictions. Modeling of lithographic patterning is the subject of the next chapter.

Problems

3.1 Suppose that 2 cc of resist are dispensed onto a 200-mm-diameter wafer, resulting in a resist coating that is 1.0 μm thick. Show that over 92% of the resist is spun off the wafer, assuming that the resist is 20% solids.

3.2 Diazonaphthoquinone photoactive compounds will undergo photochemical reactions to light with wavelengths as long as ~ 450 nm,[176] which is visible blue light. Explain the relationship between this fact and the practice of using yellow lights in the lithography areas of cleanrooms in which chips are manufactured.

3.3 Suppose that one exposes an ArF resist with a dose 10 mJ/cm^2 (in large open areas). At the edge of the line, the dose is 0.3× this. Show that the shot noise along the line edge due to photon statistics is ±1.8% (1σ) in 10 × 10 nm.

3.4 For ArF lithography, show that a dose of > 25 mJ/cm^2 is required to maintain ±3% dose control (shot-noise limited) through an area 2 × 2 nm.

3.5 What is the primary purpose for the resist softbake?

3.6 Why don't diazonaphthoquinone resists work well when processed in very dry ambient conditions?

3.7 What problems occur when chemically amplified resists are processed under conditions where bases are present? Why do bases cause these problems?

3.8 What is the most commonly used chemical for adhesion promotion?

References

1. H. A. Levine, L. G. Lesoine, and J. A. Offenbach, "Control of photoresist materials," *Kodak Photores. Sem. Proc.*, 15–17 (1968).
2. R. K. Agnihotri, D. L. Falcon, F. P. Hood, L. G. Lesoine, C. D. Needham, and J. A. Offenbach, "Structure and behavior of cyclized rubber photoresist," *Photo. Sci. Engr.* **16**(6), 443–448 (1972).
3. T. Iwayanagi, T. Ueno, S. Nonogaki, H. Ito, and C. G. Willson, "Materials and processes for deep UV lithography," in *Electronic and Photonic Applications in Polymers*, M. J. Bowden and S. R. Turner, Eds., American Chemical Society, Washington, DC (1988).
4. J. J. Sagura, and J. A. Van Allen, "Azide sensitized resin photographic resist," U.S. Patent No. 2,940,853 (1960).
5. P. S. Gwozdz, "Positive versus negative: a photoresist analysis," *Proc. SPIE* **275**, 156–163 (1981).
6. W. Thackeray, G. W. Orsula, E. K. Pavelchek, D. Canistro, L. E. Bogan, A. K. Berry, and K. A. Graziano, "Deep UV ANR photoresists for 248-nm eximer laser photolithography," *Proc. SPIE* **1086**, 34–47 (1989).
7. G. Gruetzner, S. Fehlberg, A. Voigt, B. Loechel, and M. Rother, "New negative-tone photoresists avoid swelling and distortion," *Sol. State Technol.*, 79–84 (January, 1997).
8. J. D. Shaw, J. D. Delorme, N. C. LaBianca, W. E. Conley, and S. J. Holmes, "Negative photoresists for optical lithography," *IBM J. Res. Develop.* **41**(1–2), 81–94 (1997).
9. C. A. Mack and J. E. Connors, "Fundamental differences between positive and negative tone imaging," *Proc. SPIE* **1574**, 328–338 (1992).
10. W. Henke and M. Torkler, "Modeling of edge roughness in ion projection lithography," *J. Vac. Sci. Technol.* **17**(6), 3112–3118 (1999).
11. G. P. Patsis, N. Glezos, I. Raptis, and E. S. Valamontes, "Simulation of roughness in chemically amplified resists using percolation theory," *J. Vac. Sci. Technol.* **17**(6), 3367–3370 (1999).
12. C. A. Deckert, and D. A. Peters, "Adhesion, wettability and surface chemistry," in *Adhesion Aspects of Polymer Coatings*, 469–499, Plenum Press, New York.
13. Bill Moffat, Yield Engineering Systems, private communication (2000).
14. J. N. Helbert, "Process for improving adhesion of resist to gold," U.S. Patent No. 4,497,890 (1983).
15. J. N. Helbert, "Photoresist adhesion promoters for gold metallization processing," *J. Electrochem. Soc.* **131**(2), 451–452 (1984).
16. S. Wu, *Polymer Interface and Adhesion*, Marcel Dekker, New York (1982).

17. W. A. Zisman, *Relation of equilibrium contact angle to liquid and solid constitution*, in *Contact Angle Wettability and Adhesion, Advances In Chemistry Series*, vol. 43, American Chemical Society, Washington, DC (1964).

18. M. C. B. A. Michielsen, V. B. Marriott, J. J. Ponjée, H. van der Wel, F. J. Towlager, and J. A. H. M. Moonen, "Priming of silicon substrates with trimethylsilyl containing compounds," *Microelectron. Eng.* **11**, 475–480 (1990).

19. H. Yanazawa, T. Matsuzawa, and N. Hashimoto, "Chemical characterization of photoresist to SiO_2 adhesion in integrated circuit technologies," *Kodak Microelectron. Sem.,* 153–160 (1977).

20. The resist-processing equipment business of Dainippon Screen Mfg. Co., Ltd. was spun off in 2006 to form Sokudo Co., Ltd.

21. ProCell coating data were provided by Eddie Lee, Emir Gurer, Murthy Krishna, and John Lewellen, then of Silicon Valley Group. Silicon Valley Group was acquired by ASML, and their resist-coater operations were eventually shut down.

22. A. G. Emslie, F. T. Bonner, and L. G. Peck, "Flow of a viscous liquid on a roating disk," *J. Appl. Phys.* **29**(5), 858–862 (1958).

23. D. Meyerhofer, "Characteristics of resist films produced by spinning," *J. Appl. Phys.* **49**(7), 3993–3997 (1978).

24. B. T. Chen, "Investigation of the solvent-evaporation effect on spin coating of thin films," *Polym. Engr. Sci.* **23**(7), 399–403 (1983).

25. U. Demirci, G. G. Yaralioglu, A. S. Ergun, B. T. Khuri-Yakub, and G. Perçin, "Acoustically-actuated flextensional 2D micromachined ejector array for environmentally-benign deposition of photoresist and low-κ dielectrics," *IEEE Trans. Semicond. Manuf.* **17**(4), 176–184 (January, 2004).

26. G. Perçin, L. Levin, and B. T. Khuri-Yakub, "Piezoelectrically actuated droplet ejector," *Rev. Sci. Instrum.* **68**(12), 4561–4563 (1997).

27. C. Brubaker, M. Wimplinger, P. Lindner, and S. Pargfrieder, "Investigating the use of spray-coating technology in MEMS applications," *Micro Magazine*, 45–55 (March, 2004).

28. N. Vitorino, E. Wolfer, Y. Cao, D. Lee, and A. Wu, "POU Filters reduce ArF lithography defects," *Semicond. Int.*, 20–23 (October, 2009).

29. B. Gotlinsky, J. Beach, and M. Mesawich, "Measuring the effects of sub-0.1μm filtration on 248-nm photoresist performance," *Proc. SPIE* **3999**, 827–834 (1999).

30. T. Umeda, M. Momota, N. Iguchi, and S. Tsuzuki, "Research of appropriate filter membrane for reducing defects in ArF lithography," *Proc. FUJIFILM Interface*, 1–10 (2005).

31. B. Gotlinsky, M. Mesawich, and D. Hall, "Filtration reduces defects in advanced lithography," *Semicond. Int.*, 63–66 (April 2004).

32. M. Cole, R. Van Ausdal, and J. Waldman, "Improved container and dispense system leads to reduced defects," *Microcontamination*, 37–61 (1989).

33. L. M. Manske, D. B. Graves, and W. G. Oldham, "Dynamic measurements of film thicknesses over local topography in spin coating," *Appl. Phys. Lett.* **56**(23), 2348–2350 (1990).

34. K. P. Muller and H. S. Sachdev, "Defect studies on single and bilayer resist systems," *J. Vac. Sci. Technol. B* **10**, 2560–2564 (1992).

35. K. Early, D. M. Tennant, D. J. Jeon, P. P. Mulgrew, A. A. MacDowell, O. R. Wood II, G. D. Kubiak, and D. A. Tichenor, "Characterization of AZ PN114 resist for soft x-ray projection lithography," *Appl. Optic.* **32**(34), 7044–7049 (1993).

36. K. B. Nguyen, C. Lyons, J. Schefske, C. Pike, K. Phan, P. King, H. Levinson, S. Bell, and U. Okoroanyanwu, "Characterization of the manufacturability of ultrathin resist," *J. Vac. Sci. Technol. B* **17**(6), 3039–3042 (1999).

37. C. Pike, S. Bell, C. Lyons, M. Plat, H. Levinson, and U. Okoroanyanwu, "Lithography using ultrathin resist films," *J. Vac. Sci. Technol. B* **18**(6), 3360–3363 (2000).

38. F. B. Wyart and J. Daillant, "Drying of solids by thin liquid films," *Can. J. Phys.* **68**, 1084–1088 (1990).

39. U. Okoroanyanwu, "Thin film instabilities and implications for ultrathin resist processes," *J. Vac. Sci. Technol. B* **18**(6), 3381–3387 (2000).

40. G. Reiter, "Unstable thin polymer films: rupture and dewetting processes," *Langmuir* **9**, 1344–1351 (1993).

41. W. W. Flack, S. White, and B. Todd, "Process characterization of one-hundred-micron thick photoresist films," *Proc. SPIE* **3678**, 474–490 (1999).

42. S. Furukawa, H. Miyajima, M. Mehregany, and C. C. Liu, "Electroless plating of metals for micromechanical structures," 7^{th} *Int. Conf. Sol. State Sensor. Actuator.*, 66–69 (1993).

43. M. Allen, "Polyimide-based processes for the fabrication of thick electroplated microstructures," 7^{th} *Int. Conf. Sol. State Sensor. Actuator.*, 60–65 (1993).

44. A. Curtis, P. Bisson, C. Hamel, and E. Hughlett, "Minimum consumption wafer coating," *Proc.*8^{th} *Int. Sym. Adv. Packing Mater.*, 302–310 (2002).

45. G. A. Luurtsema, "Spin coating for rectangular substrates," MS Thesis, University of California, Berkeley (1997).

46. M. Chan, "Another look at edge bead," *Proc. Kodak Microelectron. Sem.*, 16–32 (1979).

47. I. Jekuac, M. Watt, T. Hornsmith, and J. Tiffany, "Necessity of chemical edge bead removal in modern-day lithographic processing," *Proc. SPIE* **5376**, 1255–1263 (2004).

48. Q. Zhang, P. Friedberg, K. Poolla, and C. Spanos, "Enhanced spatial PEB uniformity through a novel bake plate design," *Proc. AEC-APC Symp. XVII*, 9/26–28 (2005).

49. S. Ray and M. Mehregany, "Design, fabrication, and characterization of electrostatic microrelays," *Proc. SPIE* **2642**, 64–73 (1995).

50. J. M. Kulp, "CD shift resulting from handling time variation in the track coat process," *Proc. SPIE* **1466**, 630–640 (1991).

51. G. MacBeth, "Prebaking positive photoresists," *Proc. Kodak Microelectron. Sem.*, 87–92 (1982).

52. O. D. Crisalle, C. L. Bickerstaff, D. E. Seborg, and D. A. Mellichamp, "Improvements in photolithography performance by controlled baking," *Proc. SPIE* **921**, 317–325 (1988).

53. S. H. Lin, B. T. Liu, W. C. Chen, and J. P. Hsu, "Mathematical modeling of soft baking in photoresist processing," *J. Electrochem. Soc.* **145**(12), 4256–4259 (1998).

54. J. P. Hsu, S. W. Huang, and S. Tseng, "Mathematical simulation of soft baking in photoresist processing," *J. Electrochem. Soc.* **147**(5), 1920–1924 (2000).

55. B. K. Avrit, E. W. Maxwell, L. M. Huynh, and E. Capsuto, "Characterization of an ultra-thick positive photoresist for electroplating applications," *Proc. SPIE* **5376**, 929–938 (2004).

56. J. C. Strieter, "The chemical behavior of positive working systems," *Proc. Kodak Microelectron. Sem.*, 116–122 (1976).

57. R. Dammel, *Diazonathoquinone-based Resists*, SPIE Press, Bellingham, Washington (1993).

58. P. Frasch and K. H. Saremski, "Feature size control in IC manufacturing," *IBM J. Res. Dev.* **26**(5), 561–567 (1982).

59. L. Blum, M. E. Perkins, and A. W. McCullough, "A study of the dissolution kinetics of a positive photoresist using organic acids to simulate exposed photoactive compounds," *Proc. SPIE* **771**, 148–159 (1987).

60. P. Trefonas and B. K. Daniels, "New principle for image enhancement in single layer positive photoresists," *Proc. SPIE* **771**, 194–210 (1987).

61. H. Ito, C. G. Willson, and J. M. J. Fréchet, "New UV resists with negative or positive tone," *Dig. Tech. Papers Symp. VLSI Technol.*, 86–87 (1982).

62. H. Ito, "Chemical amplification resists: history and development within IBM," *IBM J. Res. Develop.* **41**(1–2), 69–80 (1997).

63. K. Shimomura, Y. Okuda, H. Okazaki, Y. Kinoshita, and G. Pawlowski, "Effect of photoacid generators on the formation of residues in an organic BARC process," *Proc. SPIE* **3678**, 380–387 (1999).

64. J. F. Cameron, S. L. Ablaza, G. Xu, and W. Yueh, "Impact of photoacid generator structure on DUV resist performance," *Proc. SPIE* **3678**, 785–799 (1999).

65. M. J. Hanrahan and K. S. Hollis, "A comparison of the dissolution behavior of poly(p-hyroxystyrene) with novolac," *Proc. SPIE* **771**, 128–135 (1987).

66. U. Okoroanyanwu, H. Levinson, A.-M. Goethals, and F. Van Roey, "Progress in 193-nm photoresist and related process technologies," *Olin Microlithog. Sym.*, 1–34 (1998).

67. This material was produced originally by Shipley. That company was acquired by Rohm and Haas Electronic Materials, which was, in turn, acquired by Dow Chemical.

68. G. Pawlowski, K. J. Przybilla, W. Speiss, H. Wengenroth, and H. Roeschert, "Chemical amplification and dissolution inhibition: a novel high performance positive tone deep UV resist," *J. Photopolym. Sci. Technol.* **5**(1), 55–66 (1992).

69. H. Roeschert, K. J. Przybilla, W. Speiss, H. Wengenroth, and G. Powlowski, "Critical process parameters of an acetal based deep UV photoresist," *Proc. SPIE* **1672**, 33–45 (1992).

70. M. Padmanaban, Y. Kinoshita, T. Kudo, T. Lynch, S. Masuda, Y. Nozaki, H. Okazaki, G. Pawlowski, K. J. Przybilla, H. Roeschert, W. Speiss, N. Suehiro, and H. Wengenroth, "Structure-property relationship in acetal-based chemically amplified three component DUV resist," *Proc. SPIE* **2195**, 61–73 (1994).

71. T. Fujimori, S. Tan, T. Aoai, F. Nishiyama, T. Yamanaka, M. Momota, S. Kanna, Y. Kawabe, M. Yagihara, T. Kokubo, S. Malik, and L. Ferreira, "Structural design of a new class of acetal polymer for DUV resists," *Proc. SPIE* **3999**, 579–590 (2000).

72. K. Patterson, M. Somervell, and C. G. Willson, "The challenges in materials design for 157-nm photoresists," *Sol. State Technol.*, 41–48 (March, 2000).

73. This figure was provided by Will Conley.

74. S. A. MacDonald, W. D. Hinsberg, H. R. Wendt, N. J. Clecak, C. G. Willson, and C. D. Snyder, "Airborne contamination of a chemically amplified resist. 1. Identification of problem," *Chem. Mater.* **5**, 348–356 (1993).

75. K. R. Dean, D. A. Miller, R. A. Carpio, and J. S. Petersen, "Airborne contamination of DUV photoresists: determining the new limits of processing capability," *Olin Microlithog. Sem.*, 109–125 (1996).

76. J. C. Vigil, M. W. Barrick, and T. H. Grafe, "Contamination control for processing DUV chemically amplified resists," *Proc. SPIE* **2438**, 626–643 (1995).

77. D. Kinkead, "The value of airborne base contamination measurement in DUV lithography," *Microlithography World*, 22–25 (Autumn, 1999).

78. *The Handbook of Airborne Molecular Contamination Control*, Extraction Systems, Franklin, Massachusetts (1996).

79. O. Nalamasu, M. Cheng, A. G. Timko, V. Pol, E. Reichmanis, and L. F. Thompson, "An overview of resist processing for deep-UV lithography," *J. Photopolym. Sci. Technol.* **4**, 299–318 (1991).

80. T. Kumada, Y. Tanaka, A. Ueyama, S. Kubota, H. Kuezuka, T. Hanawa, and H. Morimoto, "Study on the over-top coating supressing surface insoluble layer generation for chemical amplification resist," *Proc. SPIE* **1925**, 31–42 (1993).

81. A. Oikawa, H. Santoh, S. Miyata, Y. Hatakenaka, H. Tanaka, and N. Nakagawa, "Effect of using a resin coating on KrF chemically amplified positive resists," *Proc. SPIE* **1925**, 92–100 (1993).

82. J.-B. Kim, Y.-G Kwon, J.-H. Choi, and M.-H. Jung, "Acid diffusion and evaporation in chemically amplified resists," *Proc. SPIE* **3678**, 536–543 (1999).

83. Y. Kawai, A. Otaka, A. Tanaka, and T. Matsuda, "The effect of an organic base in chemically amplified resist on patterning characteristics using KrF lithography," *Jpn. J. Appl. Phys.* **33**, 7023–7027 (1994).

84. K. Asakawa, T. Ushirogouchi, and M. Nakase, "Effect of basic additives on sensitivity and diffusion of acid in chemical amplification resists," *Proc. SPIE* **2438**, 563–570 (1995).

85. K. J. Przybilla, Y. Kinoshita, T. Kudo, S. Masuda, H. Okazaki, M. Padmanaban, G. Pawlowski, H. Roeschert, W. Spiess, and N. Suehiro, "Delay time stable chemically amplified deep UV resist," *Proc. SPIE* **1925**, 76–91 (1993).

86. W. Conley, G. Breyta, B. Brunsvold, R. DePietro, D. Hofer, S. Holmes, H. Ito, R. Nunes, and G. Fichtl, "The lithographic performance of an environmentally stable chemically amplified resist," *Proc. SPIE* **2724**, 34–60 (1996).

87. W. Hinsberg, S. MacDonald, N. Clecak, C. Snyder, and H. Ito, "Influence of polymer properties on airborne chemical contamination of chemically amplified resists," *Proc. SPIE* **1925**, 43–52 (1993).

88. H. Ito, W. P. England, N. J. Clecak, G. Breyta, H. Lee, D. Y. Yoon, R. Sooriyakumaran, and W. D. Hinsberg, "Molecular design for stabilization of chemical amplification resist toward airborne contamination," *Proc. SPIE* **1925**, 65–75 (1993).

89. H. Ito, W. P. England, R. Sooriyakumaran, N. J. Clecak, G. Breyta, W. D. Hinsberg, H. Lee, and D. Y. Yoon, "Approach toward environmental stabilization of chemical amplification resist," *J. Photopolym. Sci. Technol.* **6**, 547–562 (1993).

90. K. R. Dean, R. A. Carpio, and G. K. Rich, "Investigation of deep ultraviolet photoresists on TiN substrates," *Proc. SPIE* **2438**, 514–528 (1995).

91. C. P. Soo, S. Valiyaveettil, A. Huan, A. Wee, T. C. Ang, M. H. Fan, A. J. Bourdillon, and L. H. Chan, "Enhancement or reduction of catalytic dissolution reaction by chemically amplified resists by substrate contaminants," *IEEE Trans. Semicond. Manufact.* **12**(4), 462–468 (1999).

92. P. Burggraaf, "Stepper-track integration: solving the problems," *Semicond. Int.*, 30–33 (1992).

93. R. R. Kunz and D. K. Downs, "Outgassing of organic vapors from 193-nm photoresists: impact on atmospheric purity near the lens optics," *J. Vac. Sci. Technol.* **17**(6), 3330–3334 (1999).

94. R. Sze, *Physics of Semiconductor Devices*, 2nd ed., Chapter 7, John Wiley and Sons, New York (1981).

95. M. K. Templeton, C. R. Szmanda, and A. Zampini, "On the dissolution kinetics of positive photoresists: the secondary structure model," *Proc. SPIE* **771**, 136–147 (1987).

96. C. M. Garza, C. R. Szmanda, and R. L. Fischer, "Resist dissolution kinetics and submicron process control," *Proc. SPIE* **920**, 321–338 (1988).

97. S. Kawada, Y. Tamai, S. Omae, and T. Ohmi, "Effect of surfactant-added developer on development of chemically amplified photoresist," *Proc.* **3999**, 717–725 (2000).

98. A normal solution ($N = 1$) contains a one-gram equivalent of the dissolved substance per liter of solution.

99. T.-Y. Chu and K.-P. Cheng, "Developer concentration influence on DUV process," *Proc. SPIE* **3678**, 448–454 (1999).

100. M. Touky, B. Maxwell, and S. Chanthalyma, "The effect of developer normality on the resist dissolution rate and performance," *Proc. SPIE* **3678**, 721–726 (1999).

101. R. F. Leonard and J. M. McFarland, "Puddle development of positive photoresists," *Proc. SPIE* **275**, 187–191 (1981).

102. T. Huang, W. Wang, C. Huang, N. Tseng, T. Guo, C. Shih, and W. Wu, "Developing loading effect on lithography i-line process," *Proc. SPIE* **7140**, 71403J (2008).

103. J. M. Shaw and M. Hatzakis, "Developer temperature effects on e-beam and optically exposed positive photoresist," *J. Electrochem. Soc.* **126**(11), 2026–2031 (1979).

104. D. DeBruin, M. R. Hannifan, M. Ha, and K. Sautter, "Linewidth uniformity: a critical evaluation of spray geometry and process interactions in ultrasonic nozzles," *Proc. SPIE* **771**, 306–313 (1987).

105. W. D. Domke, V. L. Graffenberg, S. Patel, G. K. Rich, H. B. Cao, and P. F. Nealey, "Pattern collapse in high aspect ratio DUV- and 193-nm resists," *Proc. SPIE* **3999**, 313–321 (2000).

106. H. B. Cao, P. F. Nealey, and W. Domke, "Comparison of resist collapse properties for deep ultraviolet and 193 nm resist platforms," *J. Vac. Sci. Technol.* **18**(6), 3303–3307 (2000).

107. T. Tanaka, M. Morgami, and N. Atoda, "Mechanism of resist pattern collapse during development process," *Jpn. J. Appl. Phys.* **32**, 6059–6064 (1993).

108. H. Lee, J. Park, Y. Kwon, and H. Oh, "Simulation of complex resist pattern collapse with mechanical modeling," *Proc. SPIE* **5039**, 976–983 (2003).

109. S. Hien, G. Rich, G. Molina, H. B. Cao, and P. F. Nealey, "Collapse behavior of single layer 193 nm and 157 nm resists: use of surfactants to realize the sub 130-nm nodes," *Proc. SPIE* **4690**, 254–261 (2002).

110. K. Tanaka, R. Naito, T. Kitada, Y. Kiba, Y. Yamada, M. Kobayashi, and H. Ichikawa, "Improvement of pattern collapse issue by additive-added D.I. water rinse process," *Proc. SPIE* **5039**, 1366–1381 (2003).

111. M. Watanabe, Y. Tomo, M. Yamabe, Y. Kiba, K. Tanaka, and R. Naito, "Improvement of resist pattern collapse," *Proc. SPIE* **5037**, 925–933 (2003).

112. P. Zhang, M. Jaramillo, D. King, B. Ross, D. Witko, T. Paxton, and T. Davis, "The impact of surfactant in developer and rinse solution on 193 nm lithography performance," *Proc. SPIE* **5039**, 1409–1415 (2003).

113. T. Tanaka, M. Morigami, H. Oizumi, T. Soga, T. Ogawa, and F. Murai, "Prevention of resist pattern collapse by resist heating during rinsing," *J. Electrochem. Soc.* **141**(12), L169–171 (1994).

114. K. G. Kemp, D. J. Williams, J. W. Cayton, P. Steege, S. Slonaker, and R. Elliott, "Effects of DUV resist sensitivities on lithographic process window," *Proc. SPIE* **3049**, 955–962 (1997).

115. J. M. Parker, K. R. Dean, and D. A. Miller, "Direct temperature metrology helps minimize CA-resist CD variation," *Sol. State Technol.*, 139–146 (September, 2000).

116. H. J. Levinson, Y. Ma, M. Koenig, B. La Fontaine, and R. Seltmann, "Proposal for determining exposure latitude requirements," *Proc. SPIE* **6924**, 69241J (2008).

117. P. Luehrmann, and G. Goodwin, "Photoresist process optimization and control using image contrast," *KTI Microelectron. Sem.*, 279–292 (1989).

118. C. A. Mack, "Photolithographic optimization using photoresist contrast," *KTI Microelectron. Sem.*, 1–12 (1990).

119. S. V. Babu and S. Srinivasan, "Optical density and contrast of positive resists," *Proc. SPIE* **539**, 36–43 (1985).

120. V. Srinivasan and S. V. Babu, "Effect of process parameters on contrast of positive photoresists: calculation," *J. Electrochem. Soc.* **133**, 1686–1690 (1986).

121. M. P. C. Watts, "Optical positive resist processing. II. Experimental and analytical model evaluation of process control," *Proc. SPIE* **539**, 21–28 (1985).

122. P. Spragg, R. Hurditch, M. Toukhy, J. Helbert, and S. Malhotra, "The reliability of contrast and dissolution rate-derived parameters as predictors of photoresist performance," *Proc. SPIE* **1466**, 283–296 (1991).

123. R. Hurditch, and J. Ferri, "Investigation of positive resist performance based on the model of a dissolution switch," *OCG Microelectronics Seminar*, 71–90 (1993).

124. T. A. Brunner, "Relationship between the slope of the HD curve and the fundamental resist process control," *J. Vac. Sci. Technol. B* **17**(6), 3362–3366 (1999).

125. S. C. Palmateer, S. G. Cann, J. E. Curtin, S. P. Doran, L. M. Eriksen, A. R. Forte, R. R. Kunz, T. M. Lyszczarz, M. B. Stern, and C. Nelson, "Line edge roughness in sub-0.18-μm resist patterns," *Proc. SPIE* **3333**, 634–642 (1998).

126. S. Xiong, J. Bokor, Q. Xiang, P. Fisher, I. Dudley, and P. Rao, "Study of gate line edge roughness effects in 50 nm bulk MOSFET devices," *Proc. SPIE* **4689**, 733–741 (2002).

127. J. A. Croon, G. Storms, S. Winkelmeier, I. Polentier, M. Ercken, S. Decoutere, W. Sansen, and H. E. Maes, "Line edge roughness: characterization, modeling and impact on device behavior," *Dig. Int. Electron. Dev. Meeting*, 307–310 (2002).

128. T. Linton, M. Chandhok, B. J. Rice, and G. Schrom, "Determination of the line edge roughness specification for 34 nm devices," *Dig. Int. Electron. Dev. Meeting*, 303–306 (2002).

129. A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Trans. Electr. Dev.* **50**(5), 1254–1260 (2003).

130. A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electr. Dev* **50**(9), 1837–1852 (2003).

131. K. Choi, R. Dittrich, M. Goldbach, C. Hohle, K. Keil, T. Marschner, M. Tesauro, F. Thrum, R. Zimmerman, and J. Kretz, "Gate edge roughness in electron beam write and its influence to device characteristics," *Proc. SPIE* **6921**, 69210J (2008).

132. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, in *Signal Processing Series*, Prentice-Hall, Upper Saddle River, New Jersey (1989).

133. G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*, Holden-Day, San Francisco (1968).

134. T. Yamaguchi, H. Namatsu, M. Nagase, K. Kurihara, and Y. Kawai, "Line-edge roughness characterized by polymer aggregates in photoresists," *Proc. SPIE* **3678**, 617–624 (1999).

135. Q. Lin, R. Sooriyakmaran, and W.-S. Huang, "Toward controlled resist line edge roughness: material origin of line edge roughness in chemically amplified positive-tone resists," *Proc. SPIE* **3999**, 230–239 (2000).

136. A. Yamaguchi, R. Tsuchiya, H. Fukuda, O. Komuro, H. Kawada, and T. Iizumi, "Characterization of line-edge roughness in resist patterns and estimation of its effect on device performance," *Proc. SPIE* **5038**, 689–698 (2003).

137. Y. Ma, H. J. Levinson, and T. Wallow, "Line edge roughness impact on critical dimension variation," *Proc. SPIE* **6518**, 651824 (2007).

138. G. Palasantzas, "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," *Phys. Rev. B* **48**(19), 473–478 (1993).

139. An example of a stand-alone software package for analyzing LER is the SEM Metrology Interactive Toolbox (SuMMIT) from EUV Technology (Martinez, California).

140. In Ref. 137 the authors actually plot the height-height correlation function (HHCF) rather than the autocorrelation function R. However, these two quantities are related as: $\mathrm{HHCF}^2(x) = 2\sigma^2 - R(x)$.

141. The simulated results were provided by Dr. Yuansheng Ma.

142. G. M. Gallatin, "Continuum model of shot noise and line edge roughness," *Proc. SPIE* **4404**, 123–132 (2003).

143. W. Hinsberg, F. A. Houle, J. Hoffnagle, M. Sanchez, G. Wallraff, M. Morrison, and S. Frank, "Deep-ultraviolet interferometric lithography as a tool for assessment of chemically amplified photoresist performance," *J. Vac. Sci. Technol. B* **16**(6), 3689–3694 (1998).

144. M. I. Sanchez, W. D. Hinsberg, F. A. Houle, J. A. Hoffnagle, H. Ito, and C. Nguyen, "Aerial image contrast using interferometric lithography: effect on line edge roughness," *Proc. SPIE* **3678**, 160–171 (1999).

145. P. P. Naulleau and G. G. Gallatin, "Line-edge roughness transfer function and its application to determining mask effects in EUV resist characterization," *Appl. Optic.* **42**(17), 3390–3397 (2003).

146. H. Tanabe, G. Yoshizawa, Y. Liu, V. L. Tolani, K. Kojima, and N. Hayashi, "LER transfer from a mask to wafers," *Proc. SPIE* **6607**, 66071H (2007).

147. A. R. Pawloski, A. Acheta, I. Lalovic, B. La Fontaine, and H. J. Levinson, "Characterization of line edge roughness in photoresist using an image fading technique," *Proc. SPIE* **5376**, 414–425 (2004).

148. H. I. Smith, "A statistical analysis of ultraviolet, x-ray and charged particle lithographies," *J. Vac. Sci. Technol. B* **4**(1), 148–153 (1986).

149. F. Reif, *Fundmentals of Statistical and Thermal Physics*, McGraw Hill, New York (1965).

150. E. W. Scheckler, T. Ogawa, H. Yamanashi, and T. Soga, "Resist pattern fluctuation limits in extreme-ultraviolet lithography," *J. Vac. Sci. Technol. B* **12**(40), 2361–2371 (1994).

151. The resist data were provided by Dr. Patrick Naulleau and Dr. Tom Wallow.

152. P. P. Naulleau, C. N. Anderson, L.-M. Baclea-an, P. Denham, S. George, K. A. Goldberg, M. Goldstein, B. Hoef, R. Hudyma, G. Jones, C. Koh, B. La Fontaine, B. McClinton, R. H. Miyakawa, W. Montgomery, J. Roller, T. Wallow, and S. Wurm, "The SEMATECH Berkeley microfield exposure tool: learning at the 22-nm node and beyond," *Proc. SPIE* **7271**, 72710W (2009).

153. Dr. Patrick Naulleau has estimated that 1.43 nm of LER results from the mask absorber LER, and this has been factored out of the data by assuming that the measured LER is the rss of the mask LER and the LER due to all other factors. It is the LER due to nonmask sources that is plotted.

154. W. Hinsberg, F. Houle, M. Sanchez, J. Hoffnagle, G. Wallraff, D. Medeiros, G. Gallatin, and J. Cobb, "Extendibility of chemically amplified resists: another brick wall?" *Proc. SPIE* **5039**, 1–14 (2003).

155. G. M. Schmid, M. D. Stewart, C.-Y. Wang, B. D Vogt, V. M. Prabhu, E. K. Lin, and C. G. Willson, "Resolution limitations in chemically amplified photoresist systems," *Proc. SPIE* **5376**, 333–342 (2004).

156. T. Itani, H. Yoshino, S. Hashimoto, M. Yamana, N. Samato, and K. Kasama, "Polymer structure effect on dissolution characteristics and acid diffusion in chemically amplified deep ultraviolet resists," *J. Vac. Sci. Technol. B* **15**(6), 2541–2544 (1997).

157. U. Okoroanyanwu, "Resist road to the 22 nm node," *Future Fab Int.* **17**, (2004).

158. M. Padmanaban, D. Rentkiewicz, C. Hong, D. Lee, D. Rahman, R. Sakamuri, and R. R. Dammel, "Possible origins and some methods to minimize LER," *J. Photopolymer Sci. Technol.* **18**(4), 451–456 (2005).

159. D. L. Goldfarb, S. D. Burns, R. L. Burns, C. J. Brodsky, and M. C. Lawson, "Rinse additives for line-edge roughness control in 193 nm lithography," *Proc. SPIE* **5376**, 343–351 (2004).

160. M. Chandhok, K. Frasure, E. S. Putna, T. R. Youkin, W. Rachmady, U. Shah, and W. Yueh, "Improvement in linewidth roughness by postprocessing," *J. Vac. Sci. Technol. Vol. B* **26**(6), 2265–2270 (2008).

161. P. Zhang, M. Jaramillo Jr., S. Cassell, T. Wallow, A. Acheta, A. R. Pawloski, S. Bell, and R. H. Kim, "Post-etch LER performance of novel surface conditioner solutions," *Proc. SPIE* **6153**, 61533Y (2006).

162. A. Yamaguchi, H. Fukuda, H. Kawada, and T. Iizumi, "Impact of long-period line-edge roughness (LER) on accuracy in CD measurement," *Proc. SPIE* **5752**, 1362–1370 (2005).

163. T. I. Wallow, R. Kim, B. La Fontaine, P. P. Naulleau, C. N. Anderson, and R. L. Sandberg, "Progress in EUV photoresist technology," *Proc. SPIE* **6533**, 653317 (2007).

164. P. Foster, T. Steinhäusler, J. J. Biafore, G. Spazione, S. G. Slater, and A. J. Blakeney, "Second-generation 193-nm bilayer resist," *Proc. SPIE* **3678**, 1034–1039 (1999).

165. M. Neisser, G. Gozev, M. Maenhoudt, M. Lepage, and H. Stuyf, "Applying a thin imaging resist system to substrates with topography," *Sol. State Technol.*, 127–132 (August, 2000).

166. B. J. Lin, "Portable conformable mask—hybrid near-ultraviolet and deep-ultraviolet patterning technique," *Proc. SPIE* **174**, 114–121 (1979).

167. H. Santini, and N. S. Viswanathan, "Submicron optical fabrication with projection and PCM techniques," *Proc. Kodak Microelectron. Sem.*, 47–59 (1982).

168. K. Bartlett, G. Hillis, M. Chen, R. Trutna, and M. Watts, "A two layer photoresist process in a production environment," *Proc. SPIE* **394**, 49–56 (1983).

169. J. Wijdenes and M. J. H. J. Geomini, "Effect of photoresist composition on the origin of the interfacial layer in the bilayer system," *Proc. SPIE* **539**, 97–102 (1985).

170. F. Werbaneth, "Tri-level resist processing for GMR heads," *Data Storage*, 35–38 (April, 2000).

171. F. Coopmans and B. Roland, "DESIRE: a new route to submicron optical lithography," *Sol. State Technol.*, 93–98 (June,1987).

172. B. L. Yang, J. M. Yang, and K. N. Chiong, "Top-imaged single layer resist process," *J. Vac. Sci. Technol. B* **7**, 1729 (1989).

173. P. R. West and B. F. Griffing, "Contrast enhancement—route to submicron optical lithography," *Proc. SPIE* **394**, 16–17 (1983).

174. W. G. Oldham, "The use of contrast enhancement layers to improve the effective contrast of positive photoresist," *IEEE Trans. Electr. Dev.* **ED-34**(2), 247–251 (1987).

175. M. Endo, M. Sasago, Y. Hirai, K. Ogawa, and T. Ishihara, "New water-soluble contrast enhancing material for i-line lithography," *J. Electrochem. Soc.* **136**, 508–511 (1989).

176. C. G. Willson, "Organic resist materials—theory and chemistry," in *Introduction to Microlithography*, L. F. Thompson, C. G. Willson, and M. J. Bowden, Eds., American Chemical Society, Washington, DC (1983).

Chapter 4
Modeling and Thin-Film Effects

Imagine if the only way to determine whether bridges were designed with sufficient strength to support the weight of fully loaded trucks was to build each bridge in its desired location according to an engineer's best guess, drive trucks across to see if the bridge collapsed, and redesign the bridge if it failed. This approach to bridge design is clearly impractical. What has made it possible to build large bridges in a practical way are the mathematical theories of beams and plates that enable civil engineers to predict whether their designs will work or not, prior to construction. Similarly, advancements in lithographic technology have been facilitated by the availability of predictive theoretical models and tools. Today, imaging performance can be simulated on personal computers and engineering workstations in order to determine what the design parameters should be for lenses that cost millions of dollars to fabricate. Process options can be explored extensively without building large numbers of extremely expensive photomasks. In short, modeling of the lithographic process has contributed substantially to the progress in microelectronic technology.

Lithography simulations involve several key steps:

(1) The calculation of optical images. These are intensities $I(x, y)$ in the plane of the wafer that are applicable for low-NA optics, or they can be fully three-dimensional intensities $I(x, y, z)$ that may be needed for accurate simulation of exposures using high-NA lenses.
(2) Prediction of the photochemical reactions resulting from the exposure of photoresist to the previously calculated light distributions. This provides a calculated state-of-exposure at every point (x, y, z) of interest in the resist film.
(3) Computation of changes in chemical distributions within the resist as a consequence of chemical reactions and diffusion that occur during post-exposure bakes.
(4) Calculation of resist profiles following resist development.

Models for all of these key steps are discussed in this chapter.

4.1 Models of Optical Imaging

In Chapter 2 some insight into the lithographic process was provided by considering some special examples, such as the illumination of a grating by

monochromatic, coherent, and normally incident light. Such situations can be studied analytically. The images produced by more general situations can be calculated, but require more sophisticated theory. The general imaging problem occurring in lithography is shown in Fig. 4.1. Light is produced by an extended source and is incident on the reticle from multiple angles. The light is then diffracted by the pattern on the reticle. Light propagating through the entrance pupil of the lens is transmitted through the lens to form an image at the wafer plane.

Because light travels as a wave, the transmission of the light through apertures is described by the wave equation. The presence of an aperture modifies the propagation of such waves relative to transmission through free space or air. Mathematically, these modifications are the boundary conditions for the wave equation. The result of wave propagation subject to boundary conditions is summarized by Huygen's principle: "A light wave falling on an aperture σ propagates as if every element $d\sigma$ emitted a spherical wave, the amplitude and phase of which are given by that of the incident wave."[1] That is, light does not propagate directly through an aperture, but behaves as though each point on the aperture is a source of spherical waves. The net result of all such spherical waves emitted from an aperture of interest is obtained by integration over the aperture. The effects of the wave nature of light are most pronounced when features are small and comparable to the wavelength of light—precisely the situation that arises in optical lithography. Light does not pass straight through a photomask, but diffracts through many angles, and only part of the diffracted light is collected by the lens. What results are light-intensity distributions such as the one shown in Fig. 2.1 and not the ideal distribution of Fig. 2.2.

To be useful for calculating images, Huygen's principle must be expressed mathematically. Consider monochromatic light, for which solutions to the wave equation are of the form

$$U(x, y, z)e^{-i\omega t}, \tag{4.1}$$

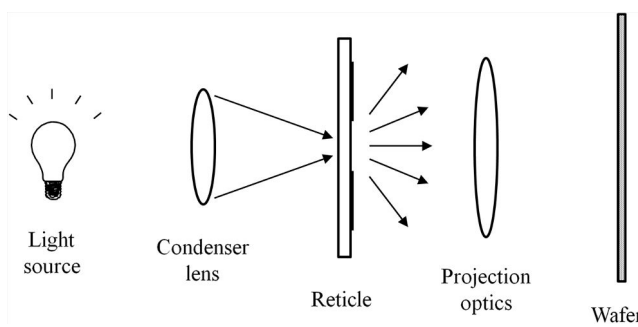


Figure 4.1 The general imaging problem. Note that not all light rays diffracted by the pattern on the reticle will pass through the projection optics.

and U will satisfy the time-independent wave equation,

$$(\nabla^2 + k^2)U = 0, \tag{4.2}$$

where $k = \omega/c = 2\pi/\lambda$, with c being the speed of light and λ being the wavelength of the light. U could be a component of the electric or magnetic field of the light wave as long as it satisfies the wave equation. Consider light originating at a point S and diffracted by an aperture σ (Fig. 4.2). The light from a point source must satisfy the wave equation and will be a spherical wave:

$$U_{source} = \frac{Ae^{ikr}}{r}, \tag{4.3}$$

where r is the radial distance from the source. At a point P on the other side of the aperture, after some analysis, it can be shown that U is obtained by an integration over the aperture σ,[1]

$$U(P) = \left(\frac{-Ai}{\lambda}\right)\left(\frac{\cos\delta}{rs}\right)\iint_{\sigma} e^{ik(r+s)}d\sigma, \tag{4.4}$$

where δ is the angle between the line S-σ and the normal vector $\tilde{n}$ angle to the aperture. It has further been assumed that r and s are large compared to the dimensions of the aperture. (This is known as the Fraunhofer approximation.) The diffraction pattern generated by an extended source can be obtained by further integrations of Eq. (4.4) over the source. Diffraction has a significant effect on the image produced by a complete optical system. While light propagates at all angles on the imaging side of the aperture, not all of the diffracted light is collected by the lens used to project the image onto the wafer. This partial collection of light leads to finite resolution and was illustrated in Chapter 2 with the example of a diffraction grating. The lens itself further modifies the light, ultimately restoring the light to a spherical wave, one that converges on the focal plane of the lens.

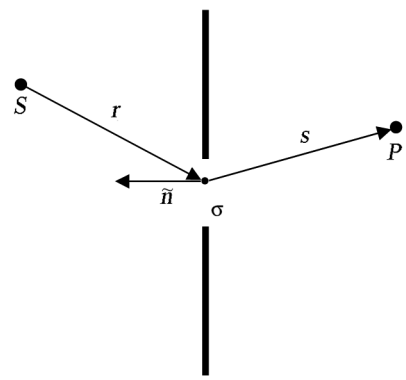


Figure 4.2 Basic configuration for diffraction by light at an aperture σ.

The integrations describing diffraction can be applied to general situations involving lines and spaces of arbitrary sizes, with illumination of multiple angles of incidence and levels of coherence, but a computer is required to perform the integrations numerically. A major advance for the semiconductor industry occurred when computer programs, capable of performing these integrations and with user-friendly interfaces, became widely available for simulating the types of imaging used for lithography. The first generally available program for calculating optical images and the resulting resist profiles was SAMPLE,[2] developed by various researchers at the University of California at Berkeley.

As with all physical modeling, simplifying assumptions are useful for making computation time practical. Models of varying complexity have been derived over the years to account for a range of physical effects. In many situations, certain physical parameters can be ignored and very accurate results obtained, while in other situations, these parameters must be included in the model in order to obtain results at the desired level of accuracy. The simplest model of imaging is the scalar diffraction theory, where the polarization of the light is ignored. This is a legitimate approximation for NAs less than ~0.7. Vector models are needed for accurate image modeling with higher NAs.

In the original SAMPLE program, there were further approximations. Light focused into a resist film has a range of incident angles (Fig. 2.23). In SAMPLE, and many subsequent modeling programs, the light-intensity distribution of the mask pattern was calculated in a chosen plane of defocus. It was then assumed that this light distribution propagated in a direction normal to the plane of the wafer. As it turned out, this was not too bad an approximation, even for moderate NAs (0.3–0.5), because light refracts at the air-resist interface in a more normal direction.[3]

As geometries smaller than the wavelength of light are printed with optical lithography, the details of the apertures on the mask become important. In the description of diffraction sketched above, the material defining the aperture was assumed to be infinitely thin and opaque. Upon closer examination of light propagation through apertures, it was revealed that the finite thicknesses and shapes of the features comprising the apertures have an effect on the resulting image. Another modeling program, TEMPEST, was developed at the University of California at Berkeley for modeling polarization and three-dimensional effects at the mask.[4] Programs such as TEMPEST have proven useful for analyzing the effects of mask topography associated with phase-shifting masks (discussed in Chapter 8). Building on the earlier SAMPLE work, another modeling program originally in the public domain, PROLITH,[5] was developed at the United States Department of Defense. After the military decided that lithography modeling was not quite within its charter, later generations of PROLITH became the products of a commercial enterprise, FINLE Technologies, later acquired by KLA-Tencor. An updated version of PROLITH is available from that company today. There are currently several companies that sell lithography modeling programs, and simulations can even be performed over the internet using programs at a website at the University of California at Berkeley.[6]

In the simplest lithography models, the optical image can be calculated with reasonable accuracy without consideration of how the resist modifies the light distribution parallel to the wafer. An aerial image is calculated and then assumed to propagate parallel to the optical axis into the resist. At high NAs, this is not an accurate approximation. Consider the situation depicted in Fig. 4.3. Two light rays are incident on a wafer. As discussed in Chapter 2, image formation requires the interference of at least two waves of light. For S-polarized light, the polarization vectors for the two waves are both perpendicular to the plane of the paper and interference between the two waves occurs. However, for P-polarization, the polarization vectors have orthogonal components that do not interfere. For light rays with 45-deg angles of incidence, the two light rays have completely orthogonal polarizations, and there is no interference at all between the two rays. This corresponds to an NA of 0.71, suggesting that there would be negligible benefits for increasing NAs to such a value since the optical contrast is not enhanced by the addition of oblique rays. Fortunately, refraction of the light at the resist-air interface returns the polarization vectors from the two rays to more parallel directions (see Problem 4.1). The refraction enables high-NA lithography. It is essential to include the effects of refraction in order to model accurately the imaging that occurs in lithography.[7]

At high NAs, it also becomes important to consider polarization-dependent effects at both the wafer and mask. First consider reflection at the wafer. Shown in Fig. 4.4 is reflectance as a function of the angle of incidence for S-polarized and P-polarized light. At low angles of incidence, the reflectance is not strongly dependent upon polarization, so it is not important to consider polarization when modeling low-NA lithography. However, at larger angles of incidence the difference becomes significant. As noted above, there is stronger interference for S-polarized light than for P-polarized light at oblique angles, so this difference in reflectivity will have a significant impact on imaging.[8]

For lithography using very high NAs, features on the mask can become comparable in size to the wavelength of light. When this happens, the treatment of the opaque areas on the mask as infinitely thin absorbers is no longer a suitable approximation. There are polarization effects at the mask that become relevant at very high NAs. For very fine features, the mask will act as a wire-grid polarizer.[9] To account for these effects properly, sophisticated physical models are needed, and desktop computers are often inadequate for performing the computations.

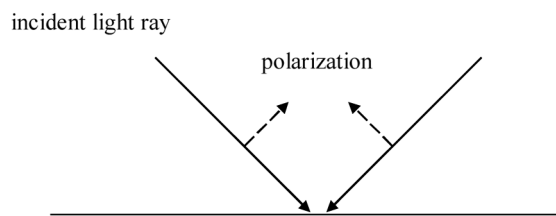


Figure 4.3 Two P-polarized rays of light, incident on the wafer surface.

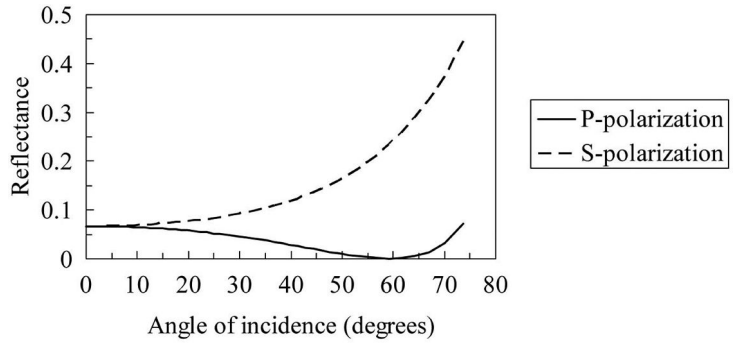


Figure 4.4 Reflectance from a resist surface into air for S-polarized and P-polarized light as a function of the angle of incidence. The index of refraction of the resist is assumed to be 1.7, and substrate effects are ignored.

4.2 Aberrations

The Rayleigh criterion found in Eq. (2.7) assumes the degradation of resolution results entirely from diffraction, i.e., the lenses are free from aberrations and imperfections. Such optics are called "diffraction limited" because the design and manufacture of the lenses are assumed to be so good that the dominant limit to resolution is diffraction. However, real optical systems are never perfect, and accurate calculations of lithographic performance need to account for these deviations from perfection. Departures of the optics from diffraction-limited imaging are caused by "aberrations," which can be understood in the context of geometrical optics as situations in which all light rays originating from a single object point do not converge to a single image point—or converge to the wrong point (Fig. 4.5). Aberrations can arise from a number of sources:

(1) Imperfect design
(2) Lens and mirror surfaces that depart from design
(3) Lens material inhomogeneity or imperfection
(4) Imperfect lens assembly.

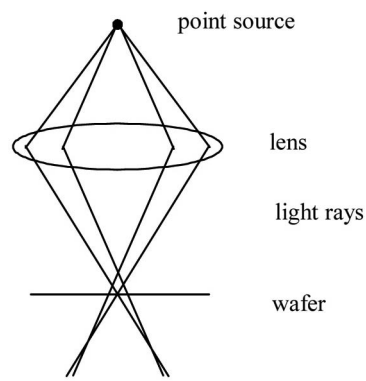


Figure 4.5 For a lens with aberrations, light rays do not all converge to the same point.

Because light rays pass through different parts of glass elements and across various points on lens-element surfaces, varying glass optical constants or errors in the curvature of lens-element surfaces cause light rays to miss the focal point of the lens. Aberrations vary from point to point across the image field, contributing to across-field linewidth variations.[10] A full accounting of imaging must include the effects of aberrations. The needs of lithography impose stringent requirements on lens materials and the surfaces of lens elements. To meet the requirements of modern lithography, the index of optical-materials refraction must be uniform to less than one part per million, and lens surfaces may deviate from design values by no more than a few nanometers (rms).

To appreciate the impact of aberrations on optical lithography, consider the illumination of a grating pattern of pitch d by coherent illumination that is normally incident on the mask. The light will be diffracted by the grating. Consider the situation in which only the 0^{th}- and $\pm 1^{\text{st}}$-order beams are collected by the lens and imaged onto the wafer. However, suppose that because of imperfections in the lens that the $+1^{\text{st}}$-order beam acquires a phase error $\Delta\phi$ relative to the 0^{th}- and -1^{st}-order beams. Consequently, the light amplitude at the image plane is given by[11]

$$U(x) = c_0 + c_1 e^{i\Delta\phi} e^{i2\pi x/d} + c_{-1} e^{-i2\pi x/d} \tag{4.5}$$

$$= c_0 + 2c_1 e^{i\Delta\phi/2} \cos\left(\frac{2\pi x}{d} + \frac{\Delta\phi}{2}\right), \tag{4.6}$$

where use has been made of the fact that $c_1 = c_{-1}$. The resulting light intensity at the wafer plane is given by:

$$I(x) = c_0^2 + 4c_1^2 \cos^2\left(\frac{2\pi x}{d} + \frac{\Delta\phi}{2}\right) + 4c_0 c_1 \cos\left(\frac{\Delta\phi}{2}\right) \cos\left(\frac{2\pi x}{d} + \frac{\Delta\phi}{2}\right). \tag{4.7}$$

The net impact of $\Delta\phi$ on the image can be seen from Eq. (4.7). First, the peak intensity is shifted by $d\Delta\phi/4\pi$. This illustrates that phase errors will result in misregistration. Additionally, the peak intensity is reduced by a factor of $\cos(\Delta\phi/2)$. Another consequence of the phase error is a reduction in image contrast.

The effects of aberrations on lithographic performance are assessed more generally by including aberrations in the imaging models. Consider light emerging from a point on the mask, passing through the optical system and ultimately focused onto the wafer. The situation at the wafer plane is considered as the reverse of what happened at the mask—the light can be thought of as a wave converging onto a point of the image. Ideally, this would be a spherical wave, but because of lens aberrations, the actual wavefront of the image may deviate by small amounts from that of a perfect spherical wave (Fig. 4.6). A wavefront is a surface of constant phase. Aberrations are incorporated into imaging models by replacing the expressions representing spherical waves, such as the exponential in Eq. (4.4), by the aberrated wavefront.

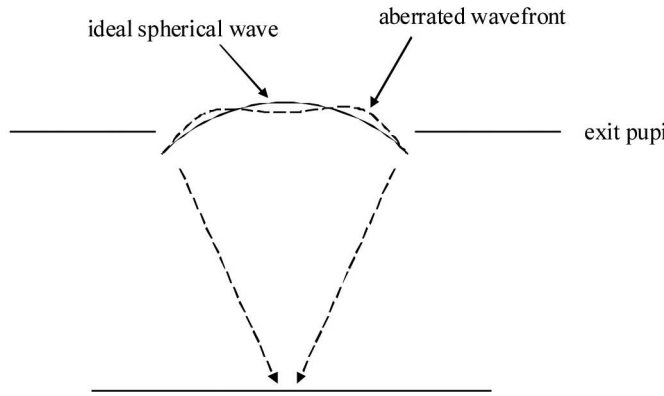


Figure 4.6 Light waves converging onto a wafer. An unaberrated image is formed by a spherical wave, while imperfect images result from nonspherical wavefronts.

The deviations of an aberrated wavefront from the ideal spherical wavefront are expressed as the phase errors at points in the exit pupil. The resulting aberration function is decomposed into sums of mathematical functions, enabling the classification of aberrations. For this decomposition, it has proven convenient to use particular sets of orthogonal polynomials, the most common being the Zernike polynomials. With this choice of expansion functions, each polynomial represents a particular type of aberration having its own characteristic effect on imaging and optical lithography.

The Zernike polynomials are functions of the polar coordinates (ρ, θ) of positions in the exit pupil, where $\rho = r/a$ and a is the pupil radius. Because the same point in the exit pupil is specified by θ and $(\theta + 360 \text{ deg})$, it is convenient for the polynomials to be functions of $\sin(n\theta)$ and $\cos(n\theta)$, where n is an integer. The first 16 Zernike polynomials are listed in Table 4.1. It is common to measure aberration levels by the coefficient of the Zernike polynomial in the expansion of the wavefront error, typically given in terms of the fraction of the wavelength. For the lenses used in optical lithography, each Zernike term contributes less than 0.05λ wavefront error.[12] Examples of two aberrations, coma and spherical aberration, are shown in Fig. 4.7 (see Color Plates).[13] The aberrations most important to lithographers are now discussed.

Coma: Coma is an aberration where light rays from a point on the reticle do not converge to a single point in the image. The consequence of coma in lithography can be understood by studying light-intensity contours of the aerial image when the lens has coma. An example of such contours is shown in Fig. 4.7. The object for these images was an isolated clear space on the mask, surrounded by chrome. For an unaberrated lens, the light-intensity contours are symmetric about the plane of best focus at the center of the image and have left-right symmetry. In the presence of coma, the contours become curved. The magnitude of this characteristic curvature is referred to as *bananicity*.[12]

One of the interesting consequences of coma is illustrated in Fig. 4.8. Consider two lines of equal width on the mask that are close to each other but far removed

Table 4.1 The first 16 Zernike polynomials.[12]

	Polynomial form	Aberration name	Primary lithographic effects
1	1	Piston	none
2	$\rho \cos\theta$	x tilt	transverse pattern shift
3	$\rho \sin\theta$	y tilt	transverse pattern shift
4	$2\rho^2 - 1$	focus	average focus shift
5	$\rho^2 \cos 2\theta$	3rd-order x astigmatism	line-orientation-dependent focus shift, elliptical contacts
6	$\rho^2 \sin 2\theta$	3rd-order 45-deg astigmatism	line-orientation-dependent focus shift, elliptical contacts
7	$\left(3\rho^3 - 2\rho\right)\cos\theta$	3rd-order x coma	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
8	$\left(3\rho^3 - 2\rho\right)\sin\theta$	3rd-order y coma	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
9	$6\rho^4 - 6\rho^2 + 1$	3rd-order spherical aberration	pattern-size-dependent average best-focus shift, depth-of-focus decrease
10	$\rho^3 \cos 3\theta$	3-foil x	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
11	$\rho^3 \sin 3\theta$	3-foil y	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
12	$\left(4\rho^4 - 3\rho^2\right)\cos 2\theta$	5th-order x astigmatism	line-orientation-dependent focus shift, elliptical contacts
13	$\left(4\rho^4 - 3\rho^2\right)\sin 2\theta$	5th-order 45-deg astigmatism	line-orientation-dependent focus shift, elliptical contacts
14	$\left(10\rho^5 - 12\rho^3 + 3\rho\right)\cos\theta$	5th-order x coma	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
15	$\left(10\rho^5 - 12\rho^3 + 3\rho\right)\sin\theta$	5th-order y coma	transverse pattern shift, depth-of-focus decrease, pattern asymmetry
16	$20\rho^6 - 30\rho^4 + 12\rho^2 - 1$	5th-order spherical aberration	pattern-size-dependent average best focus shift, depth-of-focus decrease

from other geometries. For a lens with a significant amount of coma aberration, the printed lines will have different widths.[14] Good linewidth control clearly requires low levels of lens aberrations, such as coma.

Spherical aberration: While coma affects left-right symmetry of the image, spherical aberration affects the symmetry of the image above and below the focal plane. This is illustrated in Fig. 4.9. As might be expected, the lithographic implications of spherical aberration involve focus. A particularly interesting manifestation of spherical aberration is a pitch dependence for the plane of best focus.[10] This follows immediately from the nature of diffraction and the situation depicted in Fig. 4.9. From Eq. (2.1), the angle at which a light ray is diffracted by a grating is a function of the grating pitch. When the optical system suffers from spherical aberration, light rays with different angles of diffraction will focus onto different planes (Fig. 4.9). For circuit patterns with multiple pitches, spherical aberration reduces the overall depth of focus.

Resist will also induce spherical aberration.[15] Suppose that one has an optical system that produces aberration-free images in air. As seen in Fig. 4.10, when

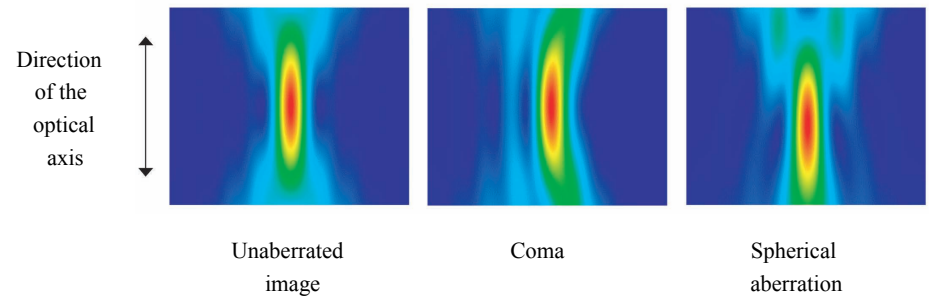


Figure 4.7 Aerial image-intensity contours for a 0.4-μm space on the mask.[10] For an unaberrated lens, the intensity contours have left-right symmetry and are also symmetric across the plane of best focus. Images produced by lenses with coma (Z_7) lose the left-right symmetry, while spherical aberration (Z_9) breaks the symmetry across the plane of best focus. The pictures in this figure were simulated with Solid-C for an i-line tool with $NA = 0.6$ and $\sigma = 0.5$. For the aberrated images, 50 nm were assumed for each aberration (see Color Plates).

Figure 4.8 Lines of equal width ($R_L = R_R$) on the reticle print with different widths ($W_L \neq W_R$) on the wafer in the presence of coma.

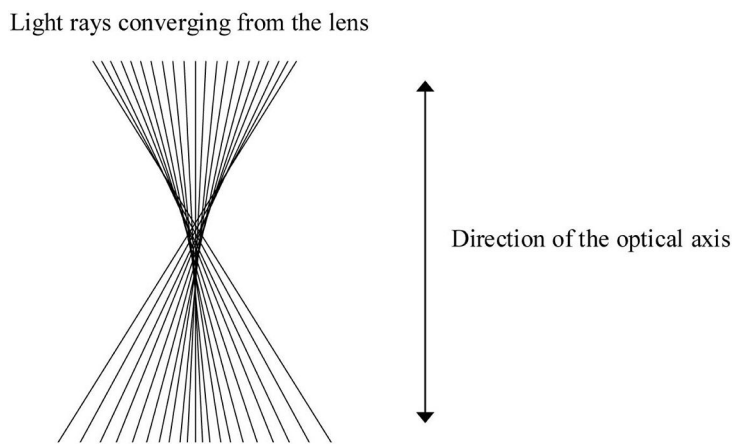


Figure 4.9 Lines with different angles of incidence have different focal points when the optical system has spherical aberration.

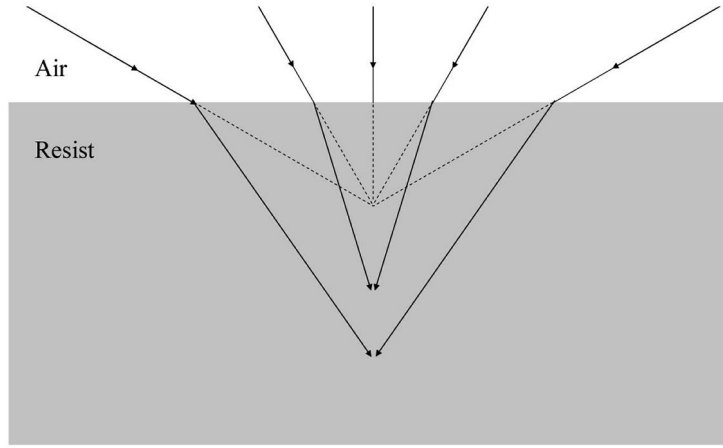


Figure 4.10 Lines with different angles of incidence refract by different amounts, leading to induced spherical aberration.

the light encounters the air-resist interface, the light rays refract. The amount of refraction depends upon the angle of incidence. It can also be seen that the amount of induced spherical aberration is greater for high-NA lenses than for low-NA ones.

Astigmatism and field curvature: (These aberrations were discussed in Chapter 2.) For these aberrations, light rays originating from a point on the reticle converge to a single point, but this point is not necessarily in the same focal plane as other features on the reticle. In the case of astigmatism, the plane of best focus depends upon the orientation of the features, while field curvature causes best focus to vary across the lens field.

Distortion: Distortion is another type of aberration in which the light rays from a point on the reticle converge to a single point on the wafer. In the case of astigmatism and field curvature, these points fall above or below a single plane. With pure distortion, the points are all in a plane that is perpendicular to the optical axis, but the distance from the optical axis is incorrect. Distortion refers to deformation of circuit patterns over long distances (millimeters), in contrast to poor definition of individual features. With distortion, the individual features are well defined, but misplaced. This misplacement leads to distortion of the pattern at the macro level. An example of distortion can be seen in the mirrors often found in amusement parks. These mirrors make people appear to be very tall or fat, but individual features are clearly discernible when observed closely. Pattern distortion that occurs in microlithography is a similar phenomenon, but on a different length scale (nm). Since distortion can lead to overlay errors, the topic of distortion is discussed in more detail in Chapter 6.

Aberrations in general vary across image fields causing variations in linewidths as well as geometry placement. With respect to linewidths, the result is usually referred to as across-field linewidth variation. The implications of across-field variations in distortion are also discussed in Chapter 6.

The extent to which stepper lenses approach the diffraction limit can be determined by comparing calculated and measured optical profiles. There are a number of commercially available software packages for calculating image profiles, and several are listed in Table 4.2. It is also possible to measure optical images directly. Several methods for measuring optical images have been developed.[16–18] There is also the aerial image measurement system (AIMS),[19] which uses a very small-field optical system to mimic actual projection optics. This type of system is employed by a number of users to evaluate mask quality. Modern lenses are typically found to be very near the diffraction limit, but deviations from the limit are observable at levels relevant to lithography. Consequently, there is considerable interest in measuring the magnitudes and effects of aberrations.

By the very nature of certain aberrations, the degradation of resolution or misregistration that they cause is greatest for points far from the optical axis. Field curvature and distortion are the most commonly encountered of such aberrations. As a result of this characteristic behavior, there has been a tendency for lenses to have higher resolution in the center of their fields—a characteristic that must be addressed by lens makers. It is challenging to design a high-resolution lens for which the resolution is well maintained over a large field. This problem has been well addressed in modern lenses, which do not show significantly better resolution in the center of the imaging field relative to the edges of the exposure field.

It is an extremely difficult task to design a lens free from the above aberrations, even at a single wavelength. Another problem arises when the light source has a range of wavelengths. The refractive indices of glass materials vary with wavelength, causing imaging to change at different wavelengths. The inability of a lens to focus light over a range of wavelengths, which results from this variation in the refractive indices of the glasses used to make the lens, is called chromatic aberration. An advantage of reflective optics is the freedom from chromatic aberration since the focusing properties of mirrors are independent of the wavelength. For systems that use refractive optics, the light sources and illuminators must be designed consistently with the bandwidth requirements of the projection optics. This must be done while keeping in mind that most sources of light, such as mercury arc lamps, produce a broad spectra of light. Early

Table 4.2 Commercially or publicly available modeling programs.

Program	Company or organization	Web site
PROLITH™	KLA-Tencor, San Jose, CA	www.kla-tencor.com
Sentaurus Lithography™	Synopsys	www.synopsys.com
EM-Suite™	Panoramic Technology	www.panoramictech.com
Calibre	Mentor Graphics, Wilsonville, OR	www.mentorgraphics.com
SAMPLE, SPLAT, TEMPEST	University of California, Berkeley, CA	http://cuervo.eecs.berkeley.edu/Volcano/
Optolith	Silvaco Data Systems, Santa Clara, CA	www.silvaco.com

lithography systems, such as the Perkin-Elmer Micralign (the Micralign will be discussed in the next chapter), used substantially reflective optics that could image over a broad range of wavelengths, including multiple peaks of the mercury arc lamp emission spectrum. However, it proved to be very difficult to build reflective optics with high numerical apertures, and a transition was made to exposure tools with refractive optics.

The problem of producing all-reflection optics with high numerical apertures is illustrated in Fig. 4.11, where a low-NA lens designed for use at a very short wavelength is shown. Increasing the numerical aperture would require an increase in the diameters and acceptance angles for the lens elements. It is clear that increasing the acceptance angle of M1, for example, would result in that mirror blocking some of the light from the mask. To avoid this, the angles between lenses have to be enlarged, increasing the difficulty of fabrication. There are other consequences of large angles, such as increasing nontelecentricity, a subject discussed in Chapter 6. It has proven difficult to design and build all-reflective lenses with numerical apertures larger than 0.3.

Lenses using refractive optics that are corrected at two wavelengths corresponding to emission peaks of mercury arc lamps have been designed. Unfortunately, the first of such lenses to be characterized was so poorly corrected for the continuum between the peaks that no advantage was realized.[20] Subsequent lenses were made that were superior in the continuum region,[21] but two-wavelength imaging with refractive optics had acquired such a bad reputation by that time the approach was soon abandoned.

For g-line and i-line lithography, the problem of chromatic aberration is not a problem that generally needs to concern the practicing lithographer. The emission from mercury arc light sources is stable with respect to wavelength over the rated lifetime of the lamps. The lenses are designed to operate over predetermined wavelengths, and the problem of chromatic aberrations is one that must be addressed primarily by the lens designers. Filters are necessary in the illuminators to remove unwanted wavelengths and sometimes slightly narrow the line spectrum,

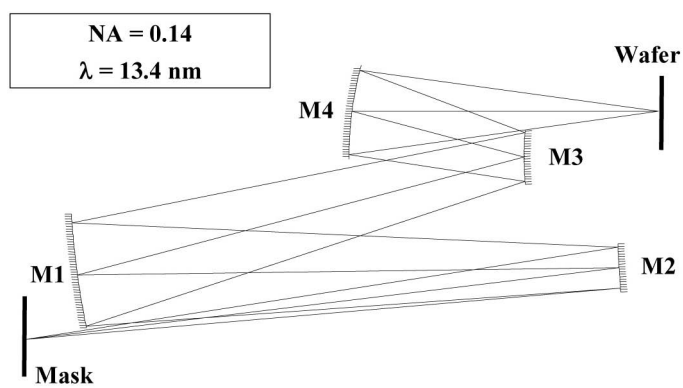


Figure 4.11 An all-reflective lens designed for use in extreme ultraviolet (EUV) lithography. EUV lithography is discussed in detail in Chapter 12.

and these filters may sometimes degrade. This is a rare occurrence, and chromatic aberration is usually not a problem that is encountered directly in g-line or i-line lithography.

In the deep UV, the small number of materials that can be used for making lenses has made it difficult to correct for chromatic aberrations, and extremely narrow bandwidths are required for refractive DUV lenses.[22] These lenses typically must operate only over a very small portion of the range of wavelengths at which KrF and ArF lasers may potentially lase. This imposes requirements for good wavelength control in DUV lithography. This is discussed in further detail in Section 5.2 on excimer laser light sources. The problem of chromatic aberration also reappears in Chapter 6, where it presents a problem for through-the-lens imaging for the purposes of alignment. Another difficulty with exposure systems that use monochromatic illumination is their susceptibility to substrate thin-film optical effects, a topic that is also discussed later in this chapter. Highly accurate simulations need to take into account the polychromatic nature of the light source and the behavior of the optical system throughout the full bandwidth of the light.

After aerial images are calculated, the light-intensity distributions within the resist films need to be calculated. The calculations of aerial images $I(x, y, z)$ provide the modulation of the light intensity along directions parallel to the xy plane of the resist film, and there is a dependence in the z direction due to the effects of defocus. Within the resist film, there are further variations in the light intensity in the z direction as a consequence of absorption in the resist and reflection from the substrate. It is this variation in the z direction (for causes other than defocus) that is the subject of the next two sections.

4.3 Modeling Photochemical Reactions

After determining the image produced by the lens in air—the aerial image—the next step in modeling the lithography process is to describe the propagation of light through the resist film. This propagation is complicated by the reaction of the resist to light. Because of the resist's photochemistry, the absorption by the resist evolves during exposure. Typically, resists become less absorbing and more transparent following exposure to light. A practical theory describing the effects of resist exposure was developed by F. H. Dill et al.,[23] who helped to establish lithography firmly as an engineering science.

Photoresists are multicomponent materials. The active ingredients in resists—the photoactive compounds and photoacid generators—undergo chemical reactions upon exposure to light. Other components may absorb light but do not undergo photochemical reactions. The intensity of light I passing in the z direction through a homogeneous material varies according to Lambert's law:

$$\frac{dI}{dz} = -\alpha I(z), \tag{4.8}$$

where α is the absorption coefficient of the material. Integrating this equation leads to

$$I(z) = I_0 e^{-\alpha z}, \tag{4.9}$$

where I_0 is the intensity of the light at $z = 0$. If absorption is caused by the presence of an absorbing molecule, α depends linearly on the concentration c of that molecular constituent over a wide range of circumstances:

$$\alpha = ac, \tag{4.10}$$

where a is the molar absorption coefficient. When a layer of photoresist is exposed to light, several things happen. First, because of the photochemical reaction in the resist, the absorption coefficient changes:

$$\alpha = a_{PAC}c + a_P(c_0 - c) + \sum a_R c_R, \tag{4.11}$$

where a_{PAC} is the molar absorption coefficient of the photoactive compound or photoacid generator, a_P is the molar absorption coefficient for the material that results from the photochemical reaction of the photoactive material, and the $\{a_R\}$ are the molar absorption coefficients for all of the other materials in the photoresist, which are assumed to be unchanged upon exposure to light. The initial concentration of the photoactive compound or photoacid generator is represented by c_0, while $\{c_R\}$ represents the concentrations of all other constituents. The concentration of photoactive compound after exposure is c. In Dill et al., Eq. (4.11) was written in the form of

$$\alpha = AM + B, \tag{4.12}$$

where

$$A = (a_{PAC} - a_P)\, c_0, \tag{4.13}$$

$$B = \sum a_R c_R + a_P c_0, \tag{4.14}$$

$$M = \frac{c}{c_0}. \tag{4.15}$$

For unexposed resist, $M = 1$, and M decreases following exposure. $M = 0$ for completely bleached resist. The "Dill" parameters, A and B, are easily measurable for a given photoresist. A represents the optical absorption that changes with exposure, while B represents the absorption by the components of the resist that do not change their absorption upon exposure to light.

The resist exposure model is complete with the addition of one final equation. Not all photons absorbed by the photoactive compound or photoacid generator

result in a chemical reaction. This quantum efficiency is taken into account by the coefficient C:

$$\frac{\partial M}{\partial t} = -CIM, \tag{4.16}$$

where t represents time and I is the intensity of the light. The parameters A, B, and C have become known as the Dill parameters for the resist. Typical values are given in Table 4.3. The contribution of Dill was the reduction of complex photochemistry to a set of three easily measured parameters, A, B, and C. The extent of photochemical reaction is captured by $M(x, y, z, t)$, where (x, y, z) is a point within the resist film, and t is the exposure time.

Once a film of resist is exposed to light, the absorptivity, represented by $AM + B$, is no longer uniform throughout the film because resist at the top of the film is more bleached (on average) than the resist near the resist-substrate interface, having received more light than the bottom. That is, following exposure, M is no longer uniform in the resist film. Bleaching is represented by values of M that are less than 1.0. Because of this bleaching, the determination of resist exposure throughout the depth of the resist film is not amenable to a closed-form solution, but the above equations can be used to calculate $M(x, y, z, t)$ iteratively for any specified values of A, B, and C using a computer. The variation in M in the z direction is determined by how light propagates through the resist thin film and is reflected from the substrate. Reflections from the substrate are discussed shortly. The variation of M in the xy plane results from the pattern on the mask and subsequent diffraction and imaging. The modeling of imaging was outlined in Section 4.1. Putting together the optical image and the subsequent exposure of resist, $M(x, y, z, t)$ can be calculated.

An interesting observation was made by integrating Eq. (4.16) with respect to time, giving[27]

$$M = e^{-CIt}, \tag{4.17}$$

and then differentiating this with respect to x. This results in the following:

$$\frac{\partial M}{\partial x} = M \ln(M) \left(\frac{1}{I}\right) \left(\frac{\partial I}{\partial x}\right). \tag{4.18}$$

Table 4.3 Dill parameters for some commercially produced photoresists.[24–26] (Not all of these resists have remained commercially available.)

Resist	A (μm^{-1})	B (μm^{-1})	C (cm^2/mJ)	Exposure wavelength	Reference
Shipley 511-A	0.85	0.04	0.016	i line	24
AZ 1470	0.56	0.03	0.010	g line	25
AZ 1470	0.88	0.22	0.010	i line	25
TOK IP-3400	0.75	0.11	0.016	i line	26
Apex-E	– 0.01	0.36	0.012	248 nm	24

This expression represents the variation of the photoactive compound following exposure. The modulation in x results from the pattern on the photomask. What is interesting about this equation is that the factor $M \ln(M)$ has a maximum, when $M = 0.37$. Resists and processes that are designed to have line edges coincident with this maximum have an enhanced opportunity for improved performance. This explains, in part, how the photoresist and resist process can affect exposure latitude—something that was not predicted by the thin-resist model in Chapter 2. The changes in linewidth depend upon the gradient in the photoactive compound, which are modulated by more than the light-intensity distribution in Eq. (4.18).

4.4 Thin-Film Optical Effects

If the substrate has optical properties identical to photoresist, the light intensity decreases gradually from the top to the bottom of the resist due to optical absorption. This decrease follows a simple exponential, at least at the beginning of exposure when the resist has homogeneous optical properties:

$$I(z) = I_0 e^{-\alpha z}, \tag{4.19}$$

where I_0 is the intensity of the incident light, α is the optical absorption coefficient, and z is the depth within the resist film, with $z = 0$ being the top surface. Typical values for α are determined from the A and B parameters given in Table 4.3 along with the relationship of Eq. (4.12). From the values for A and B in Table 4.3, one can see that the spatial scale for variation of the light intensity, due to conventional absorption of light propagating through an absorbing medium, is on the order of tenths of a micron.

The phenomenon of resist bleaching during exposure does not change this behavior significantly. Given the limit where $B \to 0$, Eqs. (4.8), (4.12), and (4.16) are solved exactly to give the light intensity and amount of remaining photoactive compound at depth z in the resist, after t seconds of exposure:[28]

$$I(z,t) = \frac{I_0}{[\,1 - e^{-CI_0t}\,(\,1 - e^{Az})\,]}, \tag{4.20}$$

$$M(z,t) = \frac{1}{[\,1 - e^{-Az}(\,1 - e^{CI_0t})\,]}, \tag{4.21}$$

where I_0 is the intensity of the incident light. At $t = 0$, Eq. (4.20) reduces to Eq. (4.9). For the case of $B \neq 0$, $\alpha(z)$ must be determined iteratively on a computer.[29]

The presence of the substrate underneath the photoresist has a significant effect on the light-intensity distribution within the photoresist film compared to the picture just presented. This can be appreciated by considering the situation depicted in Fig. 4.12. Light illuminates a substrate covered by photoresist. Typical substrates are silicon or silicon covered with various films, such as silicon dioxide, silicon nitride, aluminum, assorted silicides, titanium nitride, and other materials used to fabricate integrated circuits. In this example the films are uniform and have a large

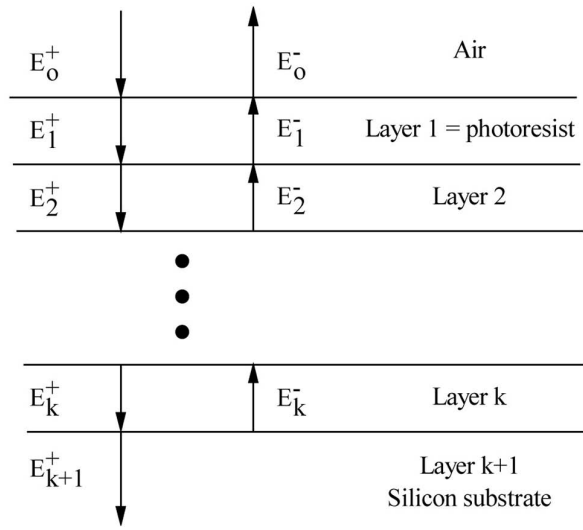


Figure 4.12 Configuration of the thin-film optical problem.[21]

spatial extent. (Issues associated with topography will be discussed shortly.) At the interface between each pair of films some of the incident light is reflected, while the remainder is transmitted.

It is highly relevant to lithography that, within the photoresist film, the light consists of an incident component as well as a reflected one. There are two significant consequences of this geometry:

(1) The light intensity varies rapidly, in the vertical direction, within the photoresist film.
(2) The amount of light energy coupled into the photoresist film has a strong dependence on the thickness of the various films in the stack.

The first property, the rapid variation of the light intensity within the photoresist film, results from the interference between the incident and reflected light within the resist. The variation in light intensity is sinusoidal, where the spacing between adjacent maxima and minima are separated by very near to $\lambda/4n$, where n is the real part of the refractive index of the photoresist. For typical values of λ and n, this quarter-wave separation is on the order of hundredths of a micron, a full order of magnitude smaller than one predicts from simple optical absorption. The effect of this is shown in Fig. 4.13, where the light intensity is plotted as a function of depth in the photoresist film for two situations—one for a silicon substrate and the other where the substrate is matched optically to the photoresist. The rapidly varying light distribution within the depth of the photoresist is referred to as a "standing wave." The consequence of standing waves of light intensity throughout the depth of the photoresist film is alternating levels of resist with high and low exposure. For positive resist, the high-exposure regions develop quickly, while the low-exposure regions develop more gradually. Manifestations of standing waves

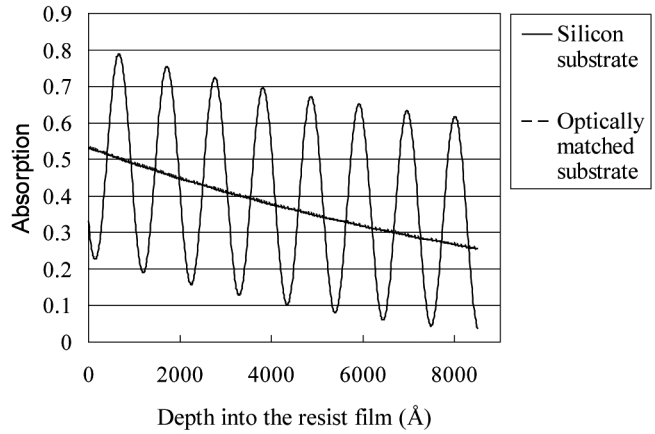


Figure 4.13 Light intensity (λ = 365 nm) throughout the depth of an 8500-Å-thick film of photoresist. The calculation of absorption was performed for the resist at the initiation of exposure; i.e., the resist was unbleached and had uniform optical properties. Depth = 0 represents the air-resist interface.

are visible in micrographs of resist features, where the resist sidewalls have ridges because alternating layers of resist have developed at different rates (Fig. 4.14).

Standing waves constrain the process because there must be sufficient exposure for the resist in the least-exposed standing wave to develop out. Otherwise, features are bridged. For highly reflective films, such as aluminum, this dose may be several times greater than the resist might require in the absence of standing-wave effects. This exaggerates the problems of substrates with topography because the resist has varying thickness over steps. Regions requiring low doses are overexposed in the attempt to clear out the last standing waves in other areas on the wafers.

In addition to the standing wave throughout the depth of the photoresist film, there is another consequence of thin-film interference effects: the amount of light absorbed in the resist has a strong functional dependence on the thickness of the substrate and resist films. Consider the total light energy E_{total} in the photoresist

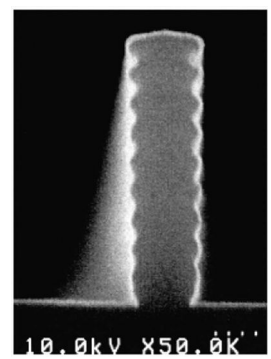


Figure 4.14 A cross section of patterned photoresist exposed on a reflective substrate. Serrations on the resist sidewalls are evident and are a consequence of standing waves.

integrated through the depth of the photoresist film:

$$E_{total} = \int_0^{T_0} E(z)dz, \tag{4.22}$$

where $E(z)$ is the energy absorbed at the height z in the resist film. T_0 is the total thickness of the resist film. The integrated light absorption E_{total} is plotted in Fig. 4.15 as a function of resist thickness for resist on an oxide layer on silicon. For 1.0-μm-thick resist, a change in the oxide thickness from 2500 Å to 2800 Å resulted in a 14% decrease in the integrated light absorption. This situation has been simulated, and the corresponding linewidths shown. Again there are oscillations, with the $\lambda/4n$ minimum-to-maximum spacing characteristic of standing-wave phenomena with linewidth minima corresponding to light-absorption maxima.

The curves shown in Fig. 4.15 are known as "swing curves." The phenomena exemplified in Fig. 4.15 have significant consequences for microlithography. Because linewidths vary with resist thickness variations on the order of a quarter wave, resist thickness must be controlled to levels much less than $\lambda/4n$, which is less than 100 nm for typical exposure wavelengths and photoresist. This is

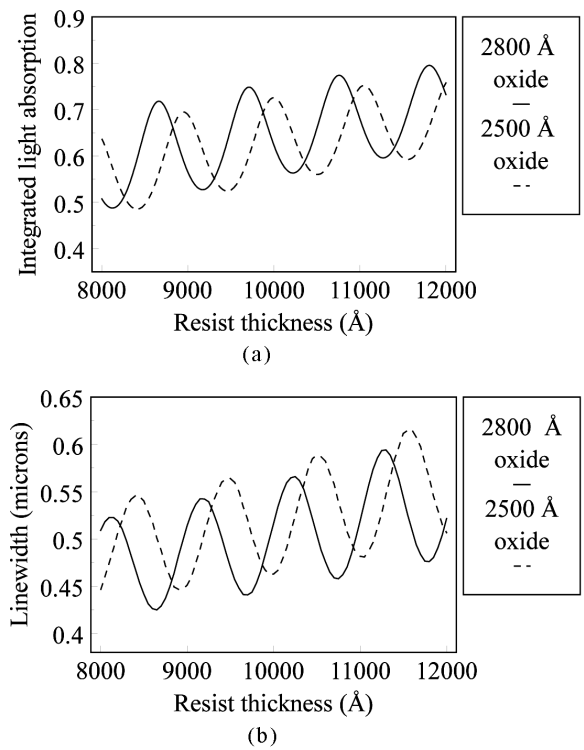


Figure 4.15 Standing-wave effects as a function of resist thickness for $\lambda = 365$ nm (i line) at the initiation of exposure. (a) Integrated light absorption [Eq. (4.22)], and (b) linewidths for 0.5-μm nominal features exposed at a fixed exposure dose, calculated using PROLITH.

an impossible task for process architectures that have virtually any topography. The standing-wave phenomenon places severe requirements on control of thin-film thickness, as seen from Fig. 4.15, where a small change in the thickness of the oxide layer changes the amount of integrated light energy absorbed and also the linewidth as a consequence. This constraint on substrate films is particularly severe for materials that have a large index of refraction, such as polysilicon. For the Hg lamp i line, the index of refraction for polysilicon is 4.9, so the quarter-wave maximum-to-minimum spacing is only 15 nm for a polysilicon film. It may not be practical to control films to tolerances that are a fraction of that. While post-exposure bakes can reduce the variations of $M(z)$ within the resist film (to be discussed in the next section), they have no effect on the integrated light energy [Eq. (4.22)].

There are a number of solutions used to address the problem of varying integrated light intensity caused by the standing-wave effect:

(1) Centering the resist thickness at a standing-wave extremum
(2) Adding dye to resists
(3) Applying bottom antireflection coatings
(4) Applying top antireflection coatings
(5) Using multiple-wavelength light sources
(6) Utilizing multilayer resist processes
(7) Utilizing surface-imaging resists.

One clearly has the minimal sensitivity to resist-thickness variations when the process is centered at an extremum of the curves shown in Fig. 4.15. On flat surfaces, choosing the resist thickness to correspond to a swing-curve maximum or minimum results in the smallest linewidth variations due to variations in resist thickness. When the topography step heights are a significant fraction of the quarter wavelength or greater, operating at a swing-curve extremum is not an option for immunizing the process against standing-wave effects.

Operating at a minimum of a swing curve has certain advantages. Swing curves such as those shown in Fig. 4.15 are usually dominated by light rays at near-normal incidence. By choosing a minimum, oblique rays are more strongly coupled into the resist since they travel further through the resist film and represent higher points on the swing curve. These oblique rays contain the higher spatial frequencies of the image. Other differences between swing-curve maxima and minima can be appreciated by examining Fig. 4.16 where the light intensity is plotted as a function of depth in resist films corresponding to a swing-curve absorption maximum (resist thickness = 8900 Å for the parameters used to calculate the results shown in Figs. 4.14 and 4.15) and a swing-curve minimum (resist thickness = 8365 Å). At a swing-curve maximum, the light intensity coupled into the resist is greatest, thereby minimizing exposure time. For a swing-curve minimum on a silicon substrate, the average dose is similar to that obtained on an optically matched substrate. The amplitudes of the intensity variations through the depth of the resist film relative to the average intensity are fairly similar. In both cases, there are absorption minima at the resist-substrate interface.

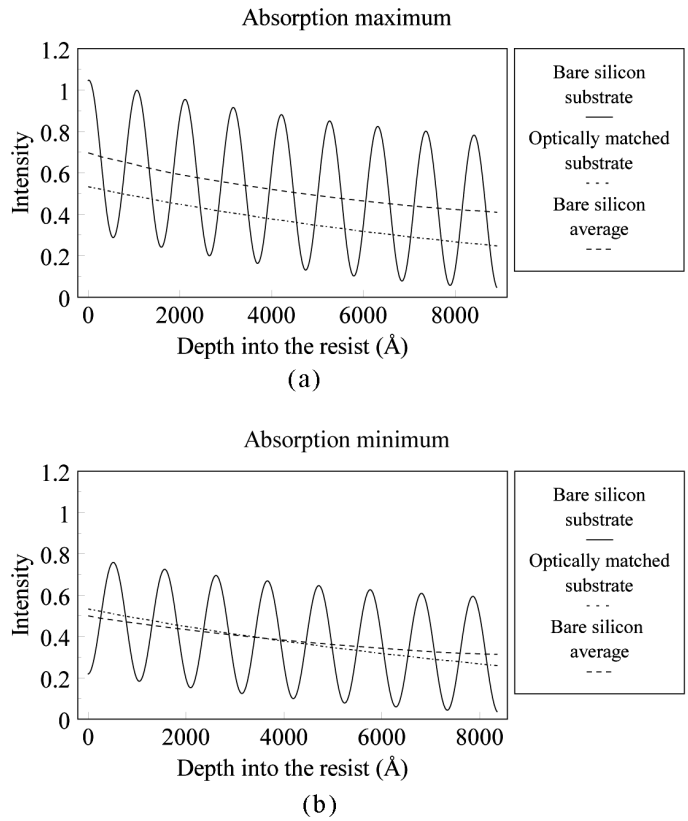


Figure 4.16 Light intensity through the depth of resist films representing swing-curve absorption (a) maxima and (b) minima.

4.5 Post-Exposure Bakes

The most common method for reducing the effects of standing waves is the post-exposure bake.[21,30,31] Following exposure, the photoactive compound or photoacid generator varies in concentration in direct proportion to the light energy. In photoresists where the photoactive compound is not bonded to the resin, diffusion of the photoactive compound may be induced by baking the resist. Diffusion takes place between regions of high and low density of photoactive compound in the direction of the largest concentration gradient.[32,33] This is typically between standing-wave maxima and minima because of the short distances ($\lambda/4n$) over which the standing waves occur. Photoacids in chemically amplified resists are typically free moieties, and they also diffuse during the post-exposure bake. Because standing waves severely degrade lithographic performance, most resists today are designed to be processed with a post-exposure bake.

One might think that this diffusion also degrades the resist profile because the photoactive compound also diffuses laterally between exposed and unexposed regions. However, the lateral gradient is usually comparable to the one across standing waves for highly reflective, or even moderately reflective, substrates. One

can estimate the point at which image degradation occurs due to lateral diffusion as follows. For image degradation, the horizontal gradients of the photoactive compound must be greater than the gradients across standing waves. As a first approximation, these gradients can be estimated from the gradients in the light intensity. At the edge of a feature, such as a line or space, the gradient in the photoactive compound is

$$\frac{\partial M}{\partial x} \approx 3 \text{ to } 7\ \mu\text{m}^{-1} \tag{4.23}$$

based on Eq. (4.18) and the data from Figs. 2.18 and 2.22. Throughout much of the image this derivative is smaller. For example, it is identically zero in the middle of an isolated line or space. On the other hand, if the standing-wave light intensity differs by 20% from maximum to minimum, then the vertical gradient is approximately

$$\frac{0.2}{\lambda/4n} = 7\ \mu\text{m}^{-1}, \tag{4.24}$$

for $\lambda = 193$ nm and $n = 1.7$. This shows that a very small standing wave can give a larger vertical gradient than results from the lateral optical image profile. As features decrease in size, the gradient at the feature edge, Eq. (4.18), increases and eventually exceeds that of the standing waves. However, shorter wavelengths help to maintain the diffusion gradient in the vertical direction. The primary reason that diffusion during post-exposure bake can lead to image degradation is the tendency toward substrates with very low reflectivity. For such situations, photoresists need to be designed so that images are not degraded significantly during post-exposure bake. This is an issue for sub-50-nm lithography.

The efficiency of the post-exposure bake depends upon the diffusion coefficient of the photoactive compound or photoacid, which depends upon many variables—the size of the photoactive compound or photoacid, the resin, the amount of residual solvent, etc.[34] This is particularly apparent for post-exposure bake temperatures in the neighborhood of the resist's glass transition temperature.[35] Because the density of the resin and the amount of residual solvent are affected by the softbake of the resist, there is often a significant interaction between the softbake and the post-exposure bake processes.[36]

Modeling the post-exposure bakes of novolak resists is reasonably straightforward. Diffusion is described by the differential equation

$$\frac{\partial n}{\partial t} = \vec{\nabla} D \times \vec{\nabla} n, \tag{4.25}$$

where n is the concentration of diffusing material and D is the coefficient of diffusion. This is reasonably straightforward to solve when D is a constant. For diazonapthaquinone/novolvak resists, this is a good approximation. Diffusion is

determined largely by the novolak matrix, which changes little during the post-exposure bake.

For chemically amplified resists, the situation is considerably more complicated. Many chemically amplified resists contain bases that can diffuse and will neutralize the photoacids. Thus, the diffusion of two species and their neutralization reactions must be accounted for in post-exposure bakes of chemically amplified resists, greatly complicating the modeling. In the modeling of chemically amplified resists it is necessary to determine the degree of deprotection at all points within the resist. A typical equation describing this is[37]

$$\frac{\partial [M]}{\partial t} = -k_1 [M]^p [A]^q , \tag{4.26}$$

where M is the extent of deprotection, A is photoacid concentration, k_1 is a rate constant, t is time, and p and q are reaction orders. The evolution of the acid concentration and the base concentration B is described by

$$\frac{\partial [A]}{\partial t} = -k_3 [A]^r + \vec{\nabla}\left(D_A \vec{\nabla} [A]\right) - k_4 [A] [B] \tag{4.27}$$

$$\frac{\partial [B]}{\partial t} = -k_5 [B]^s + \vec{\nabla}\left(D_B \vec{\nabla} [B]\right) - k_4 [A] [B] , \tag{4.28}$$

where D_A and D_B are diffusion coefficients for the photoacid and base, respectively. Instead of a single diffusion parameter, as in Eq. (4.25), there are numerous parameters required to describe the post-exposure baking of chemically amplified resists. As a result of chemical reactions that modify the porosity of the resist during the bake, the diffusion parameters are functions of photoacid concentration and the extent of deprotection. The deprotection reactions during the post-exposure bake may generate volatile materials. As these reaction products evaporate, the free volume increases, thereby increasing the diffusivity. At the same time, another effect of baking is densification of the resist, a mechanism that reduces diffusivity, the opposite effect.

To complicate matters further, there appears to be some indication that photoacid becomes trapped during the diffusion process.[38] Modeling the post-exposure bake in chemically amplified resist is today an area of active research, and it is quite important. Reasonable models for the post-exposure exposure bake of chemically amplified resists can be obtained by fitting data for lines and spaces,[39] but these models do not describe the behavior of more complex geometries very well. Acid diffusion appears to have a significant influence on patterns in two dimensions, such as contact holes[40] and line ends. Good models require consistency with the actual physics and chemistry of post-exposure bakes, and such models are necessarily complex.

4.6 Methods for Addressing the Problems of Reflective Substrates

The calculations of light intensity shown in Figs. 4.13, 4.15, and 4.16 were for a resist that was just beginning to be exposed. The standing-wave variations are less for the unexposed resist, as compared to the completely bleached situation. This effect motivated people to intentionally add nonbleaching dyes into the resist as a means of reducing standing-wave effects.[41] There are other reasons to use dyed resists, in addition to reducing standing-wave effects, as will be discussed shortly. However, adding a dye to the resist reduces its performance in terms of its ability to produce vertical sidewalls and form small features, which compromises the utility of this approach. Numerous references to the early work on dyed resist can be found in Ref. 42.

The best way to avoid standing-wave effects is to have a nonreflecting substrate. If the process architecture requires the use of reflective materials, such as metals for electrical interconnections, the substrate can be coated with an antireflection coating (ARC) to suppress reflections. Commonly used antireflection coatings are amorphous silicon on aluminum[43,44] for g-line lithography, titanium nitride for i line,[45,46] silicon nitride and silicon oxynitride for i line and DUV,[47] and various spin-coated organic materials.[48–50]

Consider again the situation where a resist film is coated over a silicon-dioxide film on a silicon surface and is exposed to uniform illumination (Fig. 4.15). As one can see in Fig. 4.15, the amplitude of the intensity variations changes little with oxide thickness; only the phase is shifted. In this case, silicon dioxide is unsuitable as an antireflection coating. This is a general property when substrate films have large indices of refraction, particularly with large imaginary components: antireflection coatings must absorb light to be effective. The requirement that antireflection coatings must absorb light results from two factors:

(1) The antireflection coating must function over a range of resist thickness, since variations in resist thickness result from circuit topography as well as variations in the process.
(2) The refractive indices of metallic substrates have large imaginary components, which introduce another condition in the equations that must be satisfied by antireflection coatings.

Consider the situation in Fig. 4.12, where layer 2 is an antireflection coating and layer 3 is a metal. Then the reflected light is given by

$$\frac{\left|E_1^-\right|}{\left|E_1^+\right|} = \frac{\left|\rho_{1,2} + \rho_{1,2} \exp\left[-i4\pi n_2 d_2/\lambda\right]\right|}{\left|1 + \rho_{1,2}\rho_{2,3} \exp\left[-i4\pi n_2 d_2/\lambda\right]\right|}, \tag{4.29}$$

where

$$\rho_{i,j} = \frac{n_i - n_j}{n_i + n_j}. \tag{4.30}$$

The condition for an antireflection coating is therefore

$$\rho_{1,2} + \rho_{2,3} \exp\left[-i4\pi n_2 d_2/\lambda\right] = 0. \tag{4.31}$$

When all refractive indices are real, the above equation leads to the well-known conditions for an antireflection coating:

$$n_2 = \sqrt{n_1 n_3}, \tag{4.32}$$

$$d_2 = \frac{(2m+1)\lambda}{4n_2}, \tag{4.33}$$

where m is any nonnegative integer. These equations are used for antireflection coatings on refractive optical elements. When the refractive index of layer 3 has a large imaginary component, Eq. (4.31) becomes two equations—one for the real part and one for the imaginary part. This necessitates another factor that can be varied, such as the imaginary component of the refractive index of layer 2, the antireflection coating.

If a material is too optically absorbing, it becomes a reflective surface. Consequently, there is an optimal level of absorption for antireflection coatings. This optimum is estimated from the following considerations. Thin films (20–120 nm) are desirable in order to maintain process simplicity during the steps that remove the antireflection coating. For a thin film to attenuate normally incident light on two passes (incident and upon reflection) through the film to 10% or less, the following condition must be met:

$$\rho e^{-2t\frac{4\pi\kappa}{\lambda}} \leq 0.1, \tag{4.34}$$

where t is the thickness of the film, κ is the imaginary part of the refractive index, ρ is the reflectance between the underlying substrate and the antireflection coating, and λ is the wavelength of the light. For $\rho = 0.7$ and $t = 70$ nm, this implies that

$$\frac{\kappa}{\lambda} > 0.0011 \text{ nm}^{-1}. \tag{4.35}$$

On the other hand, if κ becomes too large, then the thin film itself becomes too reflective. The reflectance between semi-infinite thick layers of material is plotted in Fig. 4.17. From Fig. 4.17 and Eq. (4.35) one obtains, for $\lambda = 248$ nm,

$$0.3 \leq \kappa \leq 1.2, \tag{4.36}$$

in order to maintain reflectance back into the photoresist below 10%. The upper bound on κ depends upon the degree to which the real parts of the refractive indices of the resist and antireflection coating are matched. Antireflection coatings are more effective when the real part of their index of refraction nearly equals

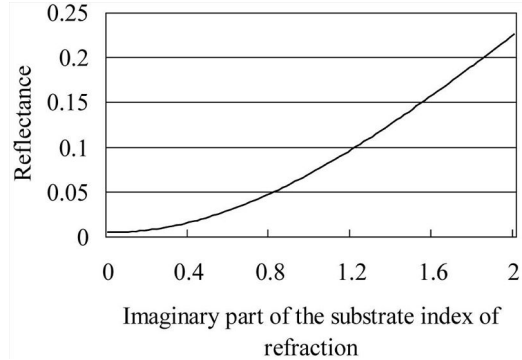


Figure 4.17 Reflectance at the interface between semi-infinite materials, calculated as the square of Eq. (4.24). One material is nonabsorbing and has an index of refraction equal to 1.74, while the real part of the index of refraction of the other material equals 2.00.

that of the resist. Also, note that the lower bound for the optimum absorption by the antireflection coating scales with the wavelength according to Eq. (4.35). Not surprisingly, most materials in use as antireflection coatings have values of κ that fall in the optimum range. For TiN, a good inorganic ARC for i-line lithography, $n = 2.01 - 1.11i$, while $n = 1.90 - 0.41i$ for BARLi, an organic i-line ARC.

Bottom antireflection coatings address all of the lithographic problems associated with reflective substrates: linewidth variations over topography (assuming the process results in adequate coverage of the antireflection coating over the steps), standing waves within the resist film, and the problem of notching, which will be discussed shortly. However, there are disadvantages as well. First, the antireflection coating must be deposited or coated. This additional processing step adds costs and potential for defects. (It may be argued that the deposition does not represent an additional step, because an adhesion promotion step is eliminated by the use of a spin-coated organic ARC. However, the ARC materials and equipment are usually much more expensive than those required for conventional adhesion promotion.) Second, the antireflection coating usually must be etched. Some spin-coated ARCs will develop out,[51] but the develop step depends critically on bake temperatures. The processing of such materials can be difficult when there is topography and accompanying variations in ARC thickness. Developing is not anisotropic, and the resulting undercut limits the use of ARCs that dissolve in developer for extremely small geometries. In general, the ARC must be anisotropically etched out of the exposed areas. Finally, the ARC usually must ultimately be removed from the unetched areas, since it is rarely a film that is part of the process architecture, other than to assist the lithography. (TiN on aluminum is a notable exception to this, where TiN is used for hillock suppression[52] and as a barrier metal.) Organic ARCs can usually be removed by the same process used to remove remaining photoresist, but inorganic ARCs require a special etch that must be compatible with the other exposed films. In spite of the expense of additional processing, the advantages for lithography are so great that antireflection coatings can often be justified.

There is at least one disadvantage to a nonreflective substrate. This can be appreciated from Fig. 4.16. Calculated in this figure are the light-intensity averages connecting the midpoints between swing-curve minima and maxima. These average curves represent, to a degree, the effective exposure following a post-exposure bake. Comparing the average curves to the light absorption for resist on a nonreflecting substrate, one sees that reflected light does give some advantage in offsetting the bulk light-absorption effect. This effect is strongest for absorbing resists where the standing-wave effect is already somewhat suppressed.

Finally, a top antireflection coating[53,54] can be used by making use of an approach that is very common in optics. Consider the situation where resist directly covers an aluminum substrate. In such a circumstance, very little light is transmitted into the substrate, and the reflectance is nearly the complement of the absorption (Fig. 4.18), which is the quantity relevant to lithography. The reflectance from the resist into air is given by the expression[55]

$$R = \frac{\rho_1^2 + \rho_1\rho_2 \cos 2\delta + \rho_2^2}{1 + 2\rho_1\rho_2 \cos 2\delta + \rho_1^2\rho_2^2}, \tag{4.37}$$

where ρ_1 is the reflectivity between the air and the resist [see Eq. (4.30)], and ρ_2 is the reflectivity between the resist and aluminum, at normal incidence. The thickness and refractive index of the resist are d and n, respectively:

$$\delta = \frac{2\pi}{\lambda} nd. \tag{4.38}$$

The oscillatory behavior of the reflectance (and absorption) results from the factor $\rho_1\rho_2 \cos 2\delta$. This factor is suppressed by a bottom antireflection coating ($\rho_2 \rightarrow 0$) or a top antireflection coating ($\rho_1 \rightarrow 0$). An ideal top antireflection

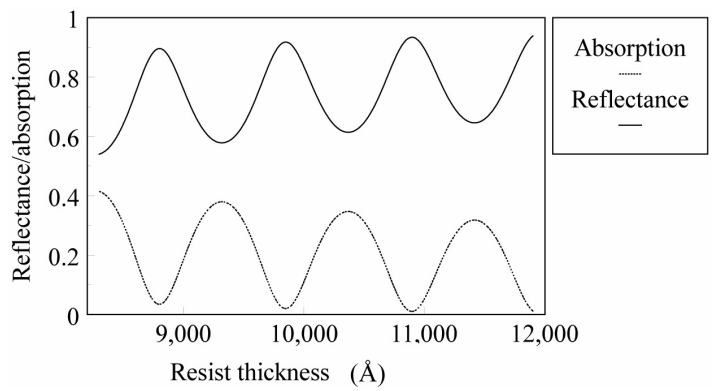


Figure 4.18 Absorption and reflectance for the i-line resist on aluminum (at the beginning of exposure). Note that absorption and reflectance are complementary.

coating is a film coated on top of the resist that has the following properties:[54]

$$\text{refractive index} = n_{arc} = \sqrt{n_{resist}}, \tag{4.39}$$

$$d = \frac{(k+1)\,\lambda}{4n_{arc}}, \tag{4.40}$$

where k is a nonnegative integer. Top antireflection coatings address situations where variations in the resist's absorption of light are modulated by changes in the thickness of nonabsorbing or moderately absorbing films. The resist itself can be one of these films. Because top antireflection coatings do not eliminate reflections from the substrate, they are less effective for highly reflective substrates and situations where the reflectivity of the substrate is highly variable. This occurs when metal interconnects have already been formed and surrounded by insulating oxides. It is also difficult to create materials that satisfy Eq. (4.39), so most top antireflective coatings are less than optimal. However, top antireflection coatings are usually water soluble, and are therefore easily integrated into semiconductor manufacturing processes, a highly desirable characteristic.

One way to minimize the impact of the standing-wave effect is to use a source with multiple wavelengths. Minima from one wavelength can be canceled out by maxima from another. This works well in terms of reducing variations in the integrated intensity, but a standing wave still exists at the interface between resist and a metallic substrate, as will now be explained.

Reflected light from a highly reflective surface maintains most of the magnitude of the incident light but undergoes a phase reversal. [In Eq. (4.30), $\rho_{i,j} \rightarrow -1$, as $n_j \gg n_i$, where n_j is the refractive index of the reflective layer and n_i is the index of refraction of the photoresist.] For a perfectly reflective surface, 100% of the light is reflected and the phase reversal is 180 deg, while light reflected from actual surfaces, which are less-than-perfect reflectors, have magnitudes and phases according to the degree to which the substrates approximate perfect reflectors. All wavelengths of interest to optical lithography are reflected from metals, silicon, and many silicides, so all wavelengths undergo a phase reversal at the resist-reflector interface. For distances on the order of several quarter waves, most wavelengths are still relatively in phase, so the standing waves interfere constructively near the resist-reflector interface. Further away from the interface, the standing waves for different wavelengths will go out of phase.[56]

For reflective substrates with considerable topography, one of the most egregious problems that result from high reflectivity is that of notching. The problem occurs when there is topography as depicted in Fig. 4.19. Light is reflected from edges and slopes into regions that are intended to be unexposed. This results in notching of resist lines, as shown in Fig. 4.20. It is useful to calculate the distance over which scattered light can be effective in producing notches. Consider the characteristic curve shown in Fig. 2.16. Note that there is negligible resist loss for exposure doses less than some fraction f of E_0. Values for f are larger for higher-contrast resist, with typical values of 0.7. Wafers typically receive doses E_e

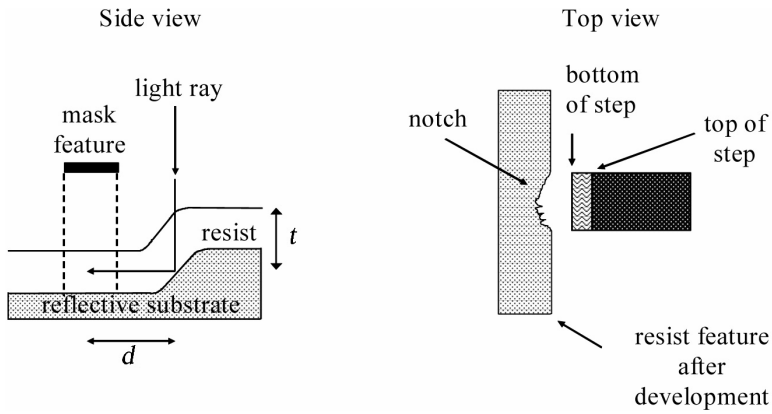


Figure 4.19 Situation for reflective notching. Light reflects from substrate topography into areas in which exposure is not desired.

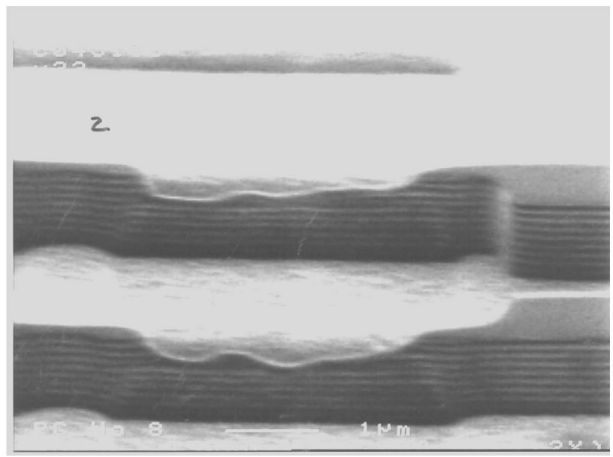


Figure 4.20 Example of reflective notching for i-line resist on polysilicon. In the area of the topography the resist is notched. Standing waves resulting from substrate reflectivity are also evident.

that are about twice E_0. Notching does not occur as long as the scattered dose E_s meets the following condition:

$$E_s \leq fE_0 \leq 0.7E_0 = 0.35E_e. \tag{4.41}$$

The scattered dose at a distance d from a step is given by

$$E_s = E_e \exp[-\alpha(d + t)], \tag{4.42}$$

where α is the absorption coefficient for the resist and t is the resist thickness. From the above two equations, there is no notching so long as

$$\rho e^{-\alpha(d+t)} \leq 0.35, \tag{4.43}$$

where the reflectance from the substrate is ρ. One obvious way to reduce the distance d over which notching may occur is to increase the absorption α. This can be done by adding a dye into the resist that does not bleach with exposure, and such an approach has been used extensively.[57] The benefit of having dye in the resist was already discussed as a method for reducing the effects of standing waves. There are tradeoffs between absorption to suppress the effects of reflective substrates and image quality. Highly dyed resists have less vertical resist sidewalls, so there is optimum dye loading for resists.[58]

Values for ρ can be as high as 0.9 for specular aluminum, and a typical value for α is 0.6 μm^{-1}. For these parameters, no notching occurs for $d + t \geq 1.5$ μm. From these considerations, one can see that the utility of dyed resists is small for deep submicron lithography. Moreover, as discussed in the section on photoresists, optical absorption reduces the performance of the resist.

Thus far the discussion of antireflection coatings has assumed that the light is normally incident or nearly normally incident onto the resist. While this is an appropriate approximation to make for imaging at low numerical aperture, it is not suitable for situations involving very high numerical apertures. To understand the complications arising from high angles of incidence, consider the configuration depicted in Fig. 4.21. The effectiveness of a single-layer bottom antireflection coating is shown in Fig. 4.22. At low angles of incidence the antireflection coating is effective, producing nearly the same low reflectance for both S- and P-polarization. The antireflective coating becomes less effective at larger angles of incidence, and there are substantial differences in behavior between the two types of polarization. Except when imaging simple gratings using S- or P-polarized light, the inability to produce low reflectance over a large range of angles of incidence is a problem when using lenses with very high numerical apertures, since light will be incident across a wide range of angles. For good process control, low reflectance is required at all angles of incidence. As it turns out, this requires complex antireflection coatings. There are insufficient degrees of freedom to design

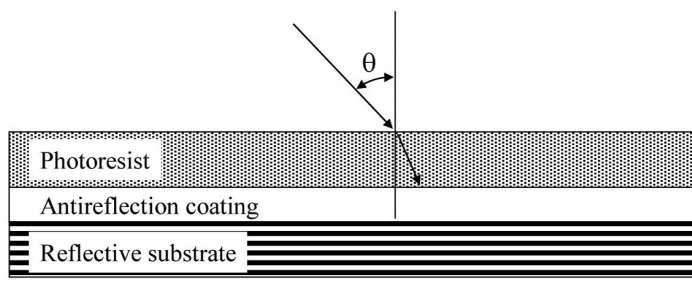


Figure 4.21 Geometry for light incident on resist.

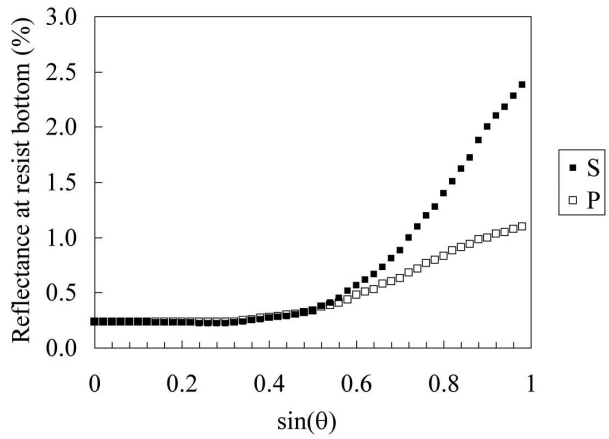


Figure 4.22 Calculated reflectance of ArF light from a bottom antireflection coating into resist, all of which is on a silicon substrate.[8] The index of refraction and thickness of the antireflection coating are $1.83–0.21i$ and 135 nm, respectively. The index of refraction of the photoresist is assumed to be 1.8. In some publications, S-polarization is referred to as transverse electric polarization, while P-polarized light is referred to as transverse magnetic light.

a single-layer antireflection coating that works across a wide range of angles of incidence. Solutions include antireflection coatings with more than one layer or coatings with indices of refraction that vary through the depth of the film.

4.7 Development

For diazonaphthoquinone resists, pattern profiles can be calculated if the development rate R is known as a function of the parameter M (discussed in Section 4.3), and the rate is usually measured directly for particular resist and developer systems. Values for M in the resist are simulated using optical models, with suitable modifications as a consequence of diffusion due to post-exposure bakes. For chemically amplified resists, M can be taken as the fraction of deprotection after post-exposure bake. Development rates are quite nonlinear and are usually fit to particular functional forms in order to facilitate calculations. Various functional forms have been proposed over time. Dill and coworkers proposed the development rate function[59]

$$\text{Rate} = e^{E_1+E_2M+E_3M^2}. \tag{4.44}$$

Other rate functions have since been adopted, notably the "Mack" model:[60]

$$\text{Rate} = r_{max}\frac{(a+1)(1-M)^n}{a+(1-M)^n} + r_{min}, \tag{4.45}$$

where r_{max}, r_{min}, n, and a are parameters of the model. By first calculating $M(x, y, z)$ using imaging models, Eq. (4.44), Eq. (4.45), or a similar equation can

then be coupled with a suitable numerical algorithm to complete the simulation of resist profiles.[61–63]

In this chapter, the major components of pattern formation in resists were discussed in terms of basic physics and chemistry. In the next two chapters, the tools used for exposing wafers and the means for ensuring overlay of successive patterns are discussed in detail. Descriptions are given of specific subsystems—such as lenses and illuminators—that determine some of the parameters that have been discussed thus far.

Problems

4.1 Assume that the index of refraction of the resist is 1.7. For P-polarized light, what numerical aperture is required to produce light rays that have orthogonal polarization in the resist?

4.2 Referring to Fig. 4.4, the reflectance for S-polarized light is nearly zero for an angle of incidence of ~60 deg. For what numerical aperture would the most oblique rays have this angle of incidence?

4.3 Using the Dill parameters of Table 4.3, show that the light intensity that reaches the bottom of a 1-μm-thick film of AZ 1470 resist is 55% and 33% of the incident intensity at the g line and i line, respectively, at the very beginning of exposure. Show that these numbers are 97% and 80% for a fully bleached 1-μm-thick film of AZ 1470 resist.

4.4 Why are post-exposure bakes used for diazonaphthoquinone resists, even though such bakes are not required to complete the photochemical conversion of photoactive compounds?

4.5 Suppose that proper exposure of a DNQ resist requires $M = 0.1$ in the middle of large exposed features. Use Eq. (4.21) to show that an exposure time of 0.15 sec is needed if the value for the Dill parameter A of the resist = 0.8 μm^{-1}, $B \approx 0$, and $C = 0.01$ cm^2/mJ, and the light intensity is 2000 mW/cm^2.

4.6 Why do bottom antireflection coatings need to absorb light?

References

1. A. Sommerfeld, *Optics*, Academic Press, New York (1954).
2. W. Oldham, S. Nandgaonkar, A. Neureuther, and M. O'Toole, "A general simulator for VLSI lithography and etching processes: part I–application to projection lithography," *IEEE Trans. Electr. Dev.* **ED-26**, 717–722 (1980).
3. C. A. Mack, "Comparison of scalar and vector modeling of image formation in photoresist," *Proc. SPIE* **2440**, 381–94 (1995).

4. A. Wong and A. Neureuther, "Examination of polarization and edge effects in photolithographic masks using three-dimensional rigorous simulation," *Proc. SPIE* **2197**, 521–527 (1994).

5. C. A. Mack, "PROLITH: a comprehensive optical lithography model," *Proc. SPIE* **538**, 207–220 (1985).

6. A. R. Neureuther, C. Hsu, T. Horie, J. Cheng, P. Chien, T. James, V. Lee, M. Li, R. Shen, R. Yang, and V. Wien, "Lithography simulation using the Internet," *Microlithog. World*, 18–20 (Winter, 1999).

7. D. G. Flagello and A. E. Rosenbluth, "Vector diffraction analysis of phase-mask imaging in photoresist films," *Proc. SPIE* **1927**, 395–412 (1993).

8. T. Brunner, N. Seong, W. D. Hinsberg, J. A. Hoffnagle, F. A. Houle, and M. I. Sanchez, "High NA lithographic imagery at Brewster's angle," *Proc. SPIE* **4691**, 1–10 (2002).

9. E. Hecht and A. Zajac, *Optics*, Addison-Wesley, Reading, Massachusetts (1974).

10. J. E. Gortych and D. Williamson, "Effects of higher-order aberrations on the process window," *Proc. SPIE* **1463**, 368–381 (1991).

11. M. S. Yeung, "Measurement of wave-front aberrations in high-resolution optical lithographic systems from printed photoresist patterns," *IEEE Trans. Semicond. Manuf.* **13**(1), 24–33 (2000).

12. D. G. Flagello, J. de Klerk, G. Davies, R. Rogoff, B. Geh, M. Arnz, Y. Wegmann, and M. Kraemer, "Towards a comprehensive control of full-field image quality in optical photolithography," *Proc. SPIE* **3051**, 672–686 (1997).

13. The image intensity contours were provided by Dr. Bernd Geh of Carl Zeiss.

14. D. G. Flagello and B. Geh, "The influence of lens aberrations in lithography," *Microlithog. World*, 11–16 (Winter, 1998).

15. R. Kim and H. J. Levinson, "Application of contrast enhancement layer to 193 nm lithography," *J. Vac. Sci. Technol.* **B25**(6), 2466–2470 (2007).

16. T. Brunner and R. R. Allen, "In-situ measurement of an image during lithographic exposure," *IEEE Electron. Device Lett.* **EDL-6**(7), 329–331 (1985).

17. T. Brunner and R. R. Allen, "In-situ resolution and overlay measurement on a stepper," *Proc. SPIE* **565**, 6–13 (1985).

18. A. K. Pfau, R. Hsu, and W. G. Oldham, "A two-dimensional high-resolution stepper image monitor," *Proc. SPIE* **1674**, 182–192 (1992).

19. R. A. Budd, D. B. Dove, J. L. Staples, R. M. Martino, R. A. Ferguson, and J. T. Weed, "Development and application of a new tool for lithographic evaluation, the stepper equivalent aerial image measurement system, AIMS," *IBM J. Res. Dev.* **41**(12), 119–127 (1997).

20. A. R. Neureuther, P. K. Jain, and W. G. Oldham, "Factors affecting linewidth control including multiple wavelength exposure and chromatic aberrations," *Proc. SPIE* **275**, 48–53 (1981).
21. W. H. Arnold and H. J. Levinson, "High-resolution optical lithography using an optimized single-layer photoresist process," *Proc. Kodak Microelectron. Sem.*, 80–92 (1983).
22. R. W. McCleary, P. J. Tompkins, M. D. Dunn, K. F. Walsh, J. F. Conway, and R. P. Mueller, "Performance of a KrF excimer laser stepper," *Proc. SPIE* **922**, 396–399 (1988).
23. F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of positive photoresist," *IEEE Trans. Electron. Device* **ED-22**(7), 445–452 (1975).
24. Shipley Microelectronics Product Catalog, Shipley Company, Newton, Massachusetts (1990).
25. M. Exterkamp, W. Wong, H. Damar, A. R. Neureuther, C. H. Ting, and W. G. Oldham, "Resist characterization: procedures, parameters, and profiles," *Proc. SPIE* **334**, 182–187 (1982).
26. Tokyo Ohka, Private Communication (2000).
27. C. A. Mack, "Photoresist process optimization," *Proc. KTI Microelectron. Sem.*, 153–167 (1987).
28. H. J. Levinson, and W. H. Arnold, "Optical lithography," Chapter 1 of *Handbook of Microlithography*, *Micromachining and Microfabrication*, P. Rai-Choudhury, Ed., SPIE Press, Bellingham, Washington (1997). These equations were derived by Bill Arnold.
29. C. A. Mack, "Analytical expression for the standing-wave intensity in photoresist," *Appl. Optic.* **25**, 1958–1961 (1986).
30. E. J. Walker, "Reduction of photoresist standing-wave effects by post-exposure bake," *IEEE Trans. Electron. Device* **ED-22**, 464–466 (1975).
31. T. Batchelder and J. Piatt, "Bake effects in positive photoresist," *Sol. State Technol.*, 211–217 (August, 1983).
32. J. Crank, *The Mathematics of Diffusion*, 2nd ed., Oxford Univ. Press, Oxford (1975).
33. D. A. Bernard, "Simulation of post-exposure bake effects on photolithographic performance of a resist film," *Philips J. Res.* **42**, 566–582 (1987).
34. P. Trefonas, B. K. Daniels, M. J. Eller, and A. Zampini, "Examination of the mechanism of the post exposure bake effect," *Proc. SPIE* **920**, 203–211 (1988).
35. J. Crank and G. S. Park, *Diffusion in Polymers*, Academic Press, London (1968).
36. K. Phan, M. Templeton, and E. Sum, "A systematic approach for i-line manufacturing resist optimization," *Proc. SPIE* **1087**, 279–289 (1989).

37. D. Matiut, A. Erdmann, B. Tollkühn, and A. Semmler, "New models for the simulation of post-exposure bake of chemically amplified resists," *Proc. SPIE* **5039**, 1132–1142 (2003).

38. S. V. Postnikov, M. D. Stewart, H. V. Tran, M. A. Nierode, D. R. Madeiros, T. Cao, J. Byers, S. E. Webber, and C. G. Willson, "Study of resolution limits due to intrinsic bias in chemically amplified photoresists," *J. Vac. Sci. Technol. B* **17**(6), 3335–3338 (1999).

39. B. Tollkühn, A. Erdmann, J. Lammers, C. Nölscher, and A. Semmler, "Do we need complex resist models for predictive simulation of lithographic process performance," *Proc. SPIE* **5376**, 983–994 (2004).

40. J. Nakamura, H. Ban, and A. Tanaka, "Influence of acid diffusion on the lithographic performance of chemically amplified resists," *Jpn. J. Appl. Phys.* **31**, 4294–4300 (1992).

41. I. I. Bol, "High-resolution optical lithography using dyed single-layer resist," *Proc. Kodak Microelectron. Sem.*, 19–22 (1984).

42. C. A. Mack, "Dispelling the myths about dyed photoresist," *Sol. State Technol.*, 125–130 (January, 1988).

43. K. Harrison and C. Takemoto, "The use of antireflection coatings for photoresist linewidth control," *Proc. Kodak Microelectron. Sem.*, 107–111 (1983).

44. H. Van den Berg and J. van Staden, "Antireflection coatings on metal layers for photolithographic purposes," *J. Appl. Phys.* **50**(3), 1212–1214 (1979).

45. W. H. Arnold, M. Farnaam, and J. Sliwa, "Titanium nitride as an antireflection coating on highly reflective layers for photolithography," U.S. Patent No. 4,820,611 (1989).

46. C. Nölscher, L. Mader, and M. Schneegans, "High contrast single layer resists and antireflection layers: an alternative to multilayer resist techniques," *Proc. SPIE* **1086**, 242–250 (1989).

47. T. Ogawa, H. Nakano, T. Gocho, and T. Tsumori, "SiOxNy:H, high performance antireflection layer for the current and future optical lithography," *Proc. SPIE* **2197**, 722–732 (1994).

48. T. Brewer, R. Carlson, and J. Arnold, "The reduction of the standing-wave effects in positive photoresists," *J. Appl. Photogr. Eng.* **7**(6), 184–186 (1981).

49. J. Lamb and M. G. Moss, "Expanding photolithography process latitude with organic AR coatings," *Sol. State Technol.*, 79–83 (September, 1993).

50. S. Kaplan, "Linewidth control over topography using a spin-on AR coating," *Proc. KTI Microelectron. Sem.*, 307–314 (1990).

51. B. Martin, A. N. Odell, and J. E. Lamb III, "Improved bake latitude organic antireflection coatings for high resolution metalization lithography," *Proc. SPIE* **1086**, 543–554 (1989).

52. M. Rocke and M. Schneegans, "Titanium nitride for antireflection control and hillock suppression on aluminum silicon metalization," *J. Vac. Sci. Technol. B* **6**(4), 1113–1115 (1988).

53. T. Brunner, "Optimization of optical properties of resist processes," *Proc. SPIE* **1466**, 297–308 (1991).

54. T. Tanaka, N. Hasegawa, H. Shiraishi, and S. Okazaki, "A new photolithography technique with antireflection coating on resist: ARCOR," *J. Electrochem. Soc.* **137**, 3900–3905 (1990).

55. O. M. Heavens, *Optical Properties of Thin Solid Films*, Dover, New York (1955).

56. J. LaRue and C. Ting, "Single and dual wavelength exposure of photoresist," *Proc. SPIE* **275**, 17–22 (1981).

57. A. V. Brown and W. H. Arnold, "Optimization of resist optical density for high resolution lithography on reflective surfaces," *Proc. SPIE* **539**, 259–266 (1985).

58. B. K. Daniels and P. Trefonas, "Optimum dye concentration to balance reflective notching against wall angle in positive rhotoresist," *J. Electrochem. Soc.* **135**(9), 2319–2322 (1988).

59. F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of positive rhotoresist," *IEEE Trans. Electron. Device* **ED-22**(7), 445–4520 (1975).

60. C. A. Mack, "Development of positive photoresists," *J. Electrochem. Soc.* **134**(1), 148–152 (1987).

61. R. E. Jewett, P. I. Hagouel, A. R. Neureuther, and T. van Duzer, "Line-profile resist development simulation techniques," *Polymer Eng. Sci.* **11**(6), 381–384 (1977).

62. J. A. Sethian, "Fast marching level set methods for three-dimensional photolithography development," *Proc. SPIE* **2726**, 262–272 (1996).

63. I. Karafyllidis, P. I. Hagouel, A. Thanailakis, and A. R. Neureuther, "An efficient photoresist development simulator based on cellular automata with experimental verification," *IEEE Trans. Semicond. Manuf.* **13**(1), 61–75 (2000).

Chapter 5
Wafer Steppers

Wafer steppers, introduced briefly in Chapter 1, are discussed further in this chapter, paying particular attention to the key subsystems of the modern reduction stepper, such as light sources, illuminators, reduction lenses, and the wafer stage. Alignment systems will be discussed in more detail in Chapter 6. In all of these discussions, the viewpoint will be that of the user.

5.1 Overview

Prior to the advent of wafer steppers, circuit patterns were transferred from masks to wafers by contact or proximity printing, or by using full-wafer scanners. In contact printing, a mask that had the pattern for all chips on the wafer was brought into contact with a resist-coated wafer. The mask was illuminated, thereby exposing the resist under the clear features on the mask. This method of exposure was used in the earliest days of the semiconductor electronics industry, but the mechanical contact caused defects on both the mask and wafer, reducing productivity. Proximity printing, where masks and wafers were brought close to each other, but not into contact, was one approach to reducing the problem of defects that arises with contact printing. Unfortunately, resolution was poor when the gap between the mask and the wafer was a practical size, as a consequence of diffraction. The first workable solution to this problem was the full-wafer scanner, which also used a mask that contained the patterns of all chips that were transferred 1:1 to the wafer.[1] More detail on tools for patterning wafers before the introduction of wafer steppers is presented in Section 5.9.

The most common method for making the masks used for contact, proximity, or full-wafer scanning lithography involved a photorepeater.[2] The photorepeater had a stage on which an unpatterned mask could be placed and moved precisely (Fig. 5.1). A reduction lens was used to image the pattern of a single chip onto the resist-coated mask blank, exposing the mask one chip at a time and using the precise stage to move the mask between exposures. In order to distinguish between the mask being made with multiple chips and the master object containing the pattern of only one chip, the single-chip mask was called a reticle. Eventually it occurred to someone to eliminate the intermediate mask and exposure tool and essentially use a photorepeater to expose the wafers directly. Thus, the wafer stepper was born.

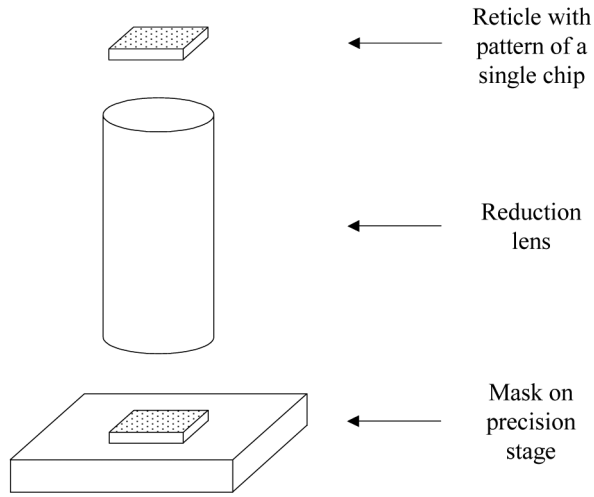


Figure 5.1 Geometry of the photorepeater.

A cutaway view of a modern wafer stepper is shown in Fig. 5.2 (see Color Plates). This view shows all the major subsystems: reduction lens and illuminator, excimer laser light source, wafer stage, reticle stage, wafer cassettes, and operator workstation. In this particular figure, wafers are being taken from cassettes. In the configuration more typical of high-performance lithography, resist-processing equipment is interfaced directly to the exposure tool. Resist-coated wafers are taken from the resist-processing equipment [or the input cassette, standard mechanical interface (SMIF) pod, or front-opening unified pod (FOUP)] and placed on a prealignment station, where the wafers are oriented with respect to the notch (or flat) and centered mechanically. The wafers are then transferred onto a very flat

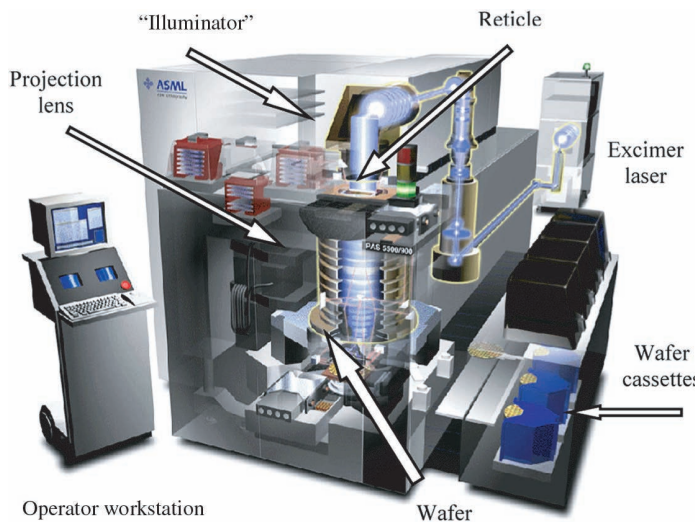


Figure 5.2 ASML PAS5500 wafer stepper (see Color Plates).

vacuum-exposure chuck that sits on a stage whose position can be controlled with extreme precision. The wafer stage is discussed in more detail later in this chapter.

Once the wafer is placed on the exposure chuck, it is aligned by automatic systems that detect wafer targets optically and move the stage in small increments to correct the wafer position with respect to the ideal image field. Alignment systems enable the overlay of new patterns to circuit patterns that already exist on the wafer. Prior to exposing each field, the wafer is positioned in the vertical axis by an autofocus system, which in modern steppers also includes the capability to pivot the vacuum chuck that holds the wafer during exposure in order to reduce any net tilt in the wafer surface due to chuck or wafer flatness errors.[3,4] Autofocus systems are discussed in more detail later in this chapter. Once the wafer is properly positioned and brought into focus, a shutter in the illumination system is opened, and the resist is exposed in the first exposure field. After exposing the first field, the shutter is closed and the wafer is moved into position to expose the next field. This process is repeated until the entire wafer is exposed. This repetitive process led to the "step-and-repeat" designation for this tool, or "stepper" for short. The exposed wafer is then moved to an output cassette (or SMIF pod or FOUP) or back to interfaced resist tracks for additional processing.

The wafer stepper was introduced commercially in the late 1970s by the GCA Corporation of North Andover, Massachusetts, based on a photorepeater that they were already producing.[5] GCA's first system, the DSW4800, had been preceded by systems designed and built by several semiconductor companies, including Philips, Thomson CSF, and IBM,[6–8] who built steppers for their own use and did not sell them commercially. (The Philips stepper technology was eventually commercialized by ASM Lithography.) The DSW in the name of GCA's stepper referred to Direct Step on the Wafer, an allusion to the stepper's origin in mask making. The GCA DSW4800 stepper, which handled 3-, 4-, or 5-in. wafers, was equipped with a 10× reduction, 0.28-NA, g-line lens that could pattern a maximum square field of 10 × 10 mm. The lens was supplied by Carl Zeiss of Oberkochen, Germany. The stepper could achieve overlay of ±0.7 μm and a resolution of 1.25-μm lines and spaces over 1.5-μm depth-of-focus. Its list price was about $300,000.

Following GCA, several companies began to produce wafer steppers and sell them commercially. The major stepper suppliers today for fine-line patterning are Nikon, Canon, and ASM Lithography (ASML). The GCA operations were absorbed by Ultratech Stepper. Ultratech has a significant share of the market for bump and packaging applications. Modern steppers have considerably greater capability than the original GCA DSW4800, but the operating principles are essentially the same. Leading-edge systems also come in the step-and-scan configuration, as well as in the step-and-repeat format. The defining characteristics of these two configurations were discussed in Chapter 1. The characteristics of contemporary step-and-scan systems are listed in Table 5.1, in comparison to the DSW4800. Modern systems provide much greater overlay and imaging capability than GCA's first machine. There has also been a substantial improvement in the productivity of exposure tools over time, a topic that is discussed further in Chapter 11.

Table 5.1 Commercially available step-and-scan 193-nm systems, compared to the first commercially available wafer stepper. The relationship between the number of shots and throughput is discussed further in Chapter 11.

	GCA	Nikon	ASML
Model	DSW4800	NSR-S310F	XT:1450G
M:1	10×	4×	4×
Wavelength (nm)	436	193	193
Maximum lens NA	0.28	0.92	0.93
Field size (mm)	10×10	26×33	26×33
Overlay, machine-to-itself (nm)	500	7	6
Wafer diameter	3, 4, 5 in.	300 mm	300 mm
Throughput (wafers per hour)	20 (4-in. wafers)	174 at 76 fields per 300-mm wafer	145 at 125 fields per 300-mm wafer

5.2 Light Sources

As a practical matter, it has proven possible to design lenses that image near the diffraction limit only over narrow bandwidths in wavelength. This is a consequence of the phenomenon of dispersion, where the index of refraction of glass changes with the wavelength of the light. This optical property of glass leads to the well-known phenomenon whereby white light is changed by a prism into a spectrum of separated colors (Fig. 5.3).[9] However, there are limits to the range of wavelengths over which lenses can be color corrected while maintaining a practical size and reasonable cost. Consequently, high-resolution stepper lenses image only over a fairly narrow range of wavelengths. Even over the bandwidth of light produced by a line-narrowed laser it is necessary for lens designers to include corrections for the varying optical properties of glass materials over the bandwidth of light. Moreover, intense light sources are needed for cost-effective stepper throughput, so exposure systems for optical lithography have operated and will continue to operate at wavelengths for which there are intense sources of illumination over narrow bandwidths. This has led to a short list of possible light sources for use in lithography. It should be noted that there are some benefits in this restriction to

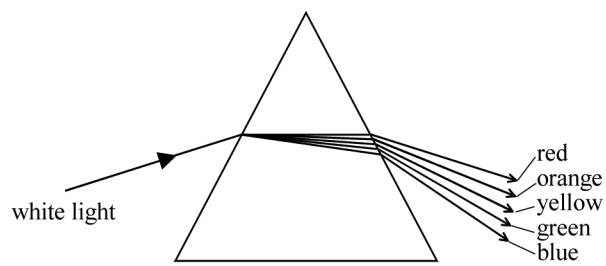


Figure 5.3 White light is broken into a spectrum of color by a prism. Typically, the refractive index of glass increases as wavelengths get shorter, increasing refraction for shorter-wavelength light. In lenses, this causes imaging to be wavelength dependent, and compensation for this effect is required in lenses for semiconductor lithography in order to achieve high resolution.

particular standard wavelengths, since additional technology that depends upon the wavelength is required. For example, resist manufacturers need to develop resists that perform optimally at the specific wavelengths used,[10] and pellicles (discussed in Chapter 7) need to be optimized for the specific wavelengths that are used. The limited number of wavelengths at which lithography is practiced enables R&D resources to be well focused.

Mercury-arc lamps and excimer lasers have been the sources of actinic light for nearly all projection photolithography. (Radiation that can induce chemical reactions in photoresists is termed actinic.) The lines of the mercury spectrum and excimer lasing wavelengths used in lithography are listed in Table 5.2. The mercury-arc lamp has three intense spectral lines in the blue and ultraviolet portions of the electromagnetic spectrum (Table 5.2), along with some continuum emission in between these spectral lines. The first commercially available wafer stepper, the GCA DSW4800, operated at the mercury g line, as did the first steppers built by Nikon, Canon, ASML, and TRE (TRE was an early supplier of wafer steppers. They later changed their name to ASET, but discontinued operations in the early 1990s.) These were based upon the same concepts as the GCA DSW4800. The mercury g line is blue light (λ = 436 nm) in the visible part of the spectrum. As mentioned previously (and to be discussed in more detail in Section 5.4), it has proven possible to design stepper lenses that meet the extreme resolution and field size requirements of microlithography only over a very narrow range of wavelengths. For mercury arc lamp–based systems, this has been over the bandwidths of the arc lamps, on the order of 4–6 nm. These bandwidths are much larger than the natural bandwidths of mercury atomic emissions, because of the collision (pressure) and thermal (Doppler) broadening[11] that can be considerable in a high-pressure arc lamp operating at temperatures approaching 2000 °C.

Some systems built by Ultratech image over somewhat broader ranges of wavelengths (390–450 nm), but these have resolutions limited to 0.75 μm or larger. This use of multiple wavelengths has significant advantages in terms of reducing standing-wave effects,[12] and the Ultratech steppers have been used effectively for lithography at ≥1.0-μm feature sizes. Unfortunately, the broadband Ultratech lens design prints the reticle at a 1:1 ratio.[13] There is no reduction, and reticle quality has largely prevented the use of Ultratech steppers for critical applications in deep submicron lithography. Besides the lenses from Ultratech, there were a few other lenses that imaged at both the g and h line,[14] but these had limited

Table 5.2 Light sources and wavelengths used in optical lithography.

Class of light source	Specific type	Wavelength (nm)	Location in the electromagnetic spectrum
Mercury-arc lamps	g line	436	visible
	h line	405	visible
	i line	365	mid-ultraviolet
	DUV	240–255	DUV
Excimer lasers	KrF	248	DUV
	ArF	193	DUV

acceptance. While the mercury h line was used on a few wafer steppers,[15] most stepper manufacturers made a transition to i-line lithography in the late 1980s as the need arose to print submicron features, while maintaining depths-of-focus ≥1.0 μm.[16] I-line lithography dominated leading-edge lithography until the advent of deep-ultraviolet (DUV) lithography in the mid-1990s.

There is a strong band of DUV emission (λ = 240–255 nm) from mercury-xenon arc lamps, and these were used on early DUV exposure tools, such as the Micrascan I and Micrascan II from SVGL. (SVGL was acquired by ASML in 2001.) Most DUV systems today use excimer lasers as light sources, as did higher-resolution versions of the Micrascan platform. The bandwidth requirements for these light sources are usually much less than 1.0 picometer (pm). Excimer lasers are considered in more detail, following a discussion of arc lamps.

A mercury-arc lamp is illustrated in Fig. 5.4. A fused silica bulb is filled through a tip with a small quantity of mercury and argon or mercury and xenon. After filling, the bulb is sealed. Operation is initiated by applying a high-frequency high voltage (>10 kV) across the two electrodes, ionizing the inert gas. The resulting discharge causes the mercury to evaporate, and the mercury begins to contribute to the discharge. The plasma, being electrically conducting, cannot support the high voltage, so the voltage drops, and steady lamp output is maintained by operating the lamp at constant current at relatively low dc voltages (50–150 V). High voltages are needed only to ignite the plasma. Condensation of the mercury to the cooler walls of the bulb near the electrodes is inhibited by reflecting coatings. Pressure inside the bulb can exceed 30 atm during operation,[17] and catastrophic failure is always a concern. The electrodes are made of refractory metals, such as tungsten, in order to withstand the internal temperatures that can be as high as 2000 °C. Thorium coatings are often used to reduce the electrode work functions and provide electrons to the plasma more easily.

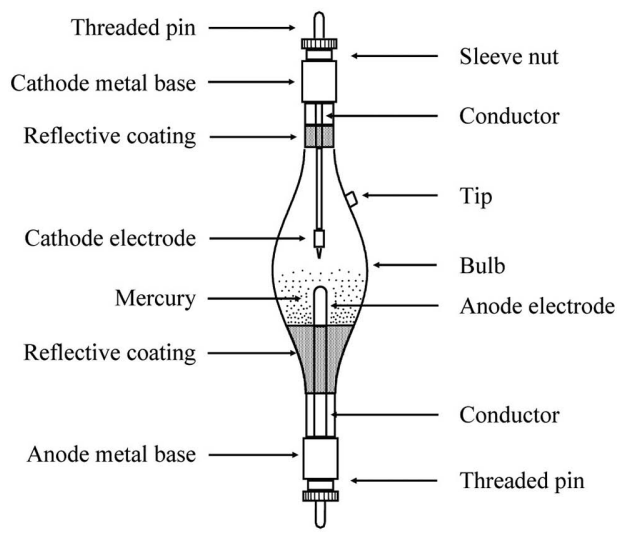


Figure 5.4 Mercury-arc lamp.

During operation, electrode material is gradually deposited onto the insides of the bulb, reducing light output. Lamps are usually replaced after several hundred hours of operation in order to maintain high stepper throughput. The probability of a lamp explosion also increases with time, and the prevention of such explosions is another reason why lamps are replaced at the end of their rated lives. The two most common mechanisms of catastrophic failure are degradation of glass-metal seals or fracture of the glass bulb.[18]

Less than 1% of the electrical power supplied to a mercury-arc lamp is converted to actinic light. The rest of the energy is nonactinic light or heat that is removed by air exhaust and cooling water. The air exhaust also serves a safety function, by exhausting mercury vapor in the event of a lamp explosion. Improved stability of light output has also been obtained by monitoring and controlling the external lamp temperature, which will exceed 700 °C.[19] As a practical matter, the lamp base temperature, typically between 150 °C and 200 °C, is more easily measured than the bulb temperature. Temperature control is obtained by adjusting the air exhaust. Catastrophic failures can also be avoided by monitors of the voltage and current to the lamp.[20] Mercury is a toxic substance, and thorium is radioactive, so both materials present handling and disposal issues. Because thorium is an alpha emitter, it represents no safety hazard to users, as long as the glass bulb is intact. Users of mercury-arc lamps need to establish procedures for handling lamp explosions and for the disposal of used arc lamps.

Another light source applied to lithography is the excimer laser. Excimer lasers are much larger and more complicated than arc lamps. Because they are so large, and cleanroom floor space is so expensive, excimer lasers are usually placed outside of the cleanroom (Fig. 5.5). The intrinsic directionality of the laser light enables the lasers to be placed up to 25 m away from the stepper, with the light being delivered by a series of mirrors and lenses without a significant loss of light energy.[21]

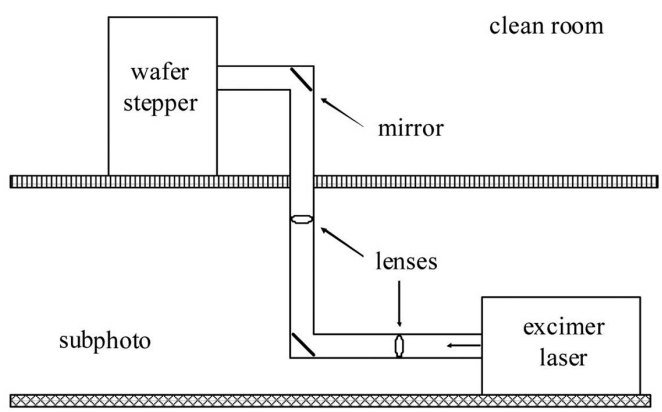


Figure 5.5 Configuration for excimer laser light sources, where the lasers are placed far from the stepper.

In KrF excimer lasers, excited dimers are created by placing a strong electric field across a gas mixture containing Kr, F_2, and Ne.[22] Early excimer lasers required voltages >20 kV, and many components failed from high-voltage breakdown.[23] A high voltage is used to produce an electrical discharge, which, in turn, drives the reactions which ultimately result in lasing (see Table 5.3).[24] Reliability has improved through laser designs that require lower voltages, in the range of 12–15 kV, to produce the electrical discharge, and these lower voltages reduce the stress on all electrical components.[25] Excimer lasers produce light in pulses, at rates up to several kilohertz. There has been a transition from thyratron-based discharging electronics to solid-state electronics, and this has also contributed to improvements in laser reliability. Excimer lasers have matured considerably since their first use on commercially available wafer steppers in 1988[26] to the point where they are sufficiently reliable for industrial use. This improvement in excimer light sources has played a critical role in bringing DUV lithography to production worthiness. To appreciate the degree of reliability required for use in manufacturing, consider that modern KrF and ArF excimer lasers are capable of pulsing at 6 kHz. With a duty factor of only 10%, this represents 18.9 billion pulses a year.

The unnarrowed fluorescence spectrum for the KrF emission is shown in Fig. 5.6.[27] There is sufficient gain over only part of this spectrum for lasing, about 400 pm. Freely running ArF lasers have similarly broad bandwidths, with full-width half-maxima of about 450 pm. These bandwidths are much too large for use with all-refractive optics and require narrowing for catadioptric lenses as well.

Excimer lasers consist of several subsystems, shown in Fig. 5.7. A high repetition rate is desirable for these pulsed light systems, and excimer-laser suppliers have improved the available rates from 200 Hz[28,29] to 4 kHz,[30,31] and now to 6 kHz.[32,33] The higher rates allow for high doses in short times without requiring high-peak light intensities, and this reduces damage to optical elements. Measurements for stepper self-metrology take place no faster than permitted by the excimer-laser frequency, so stepper set-up time is reduced with high-repetition-rate lasers.

After several years of improvement, typical (time-averaged) water-plane intensities for DUV steppers are now comparable to those achieved on i-line

Table 5.3 The key reactions in a KrF excimer laser.[24]

$F_2 + e^- \rightarrow F^- + F$	Negative fluorine production
$Kr + e^- \rightarrow Kr^* + e^-$ $Kr^* + e^- \rightarrow Kr^+ + 2e^-$	Two-step positive krypton production
$Kr^+ + F^- + Ne \rightarrow KrF^* + Ne$	Excimer formation
$KrF^* \rightarrow Kr + F + h\nu$	Spontaneous emission
$KrF^* + h\nu \rightarrow Kr + F + 2h\nu$	Stimulated emission
$F + F + Ne \rightarrow F_2 + Ne$	Recombination

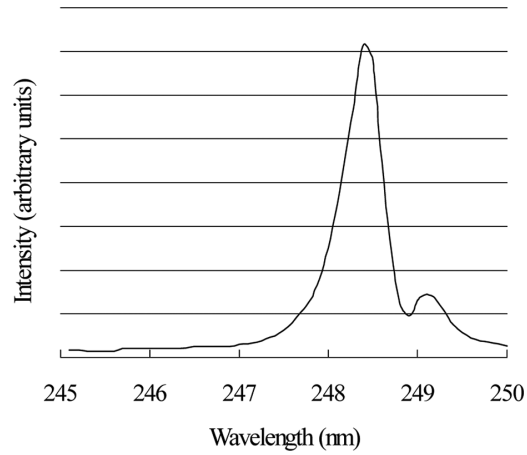


Figure 5.6 KrF* fluorescence spectrum.

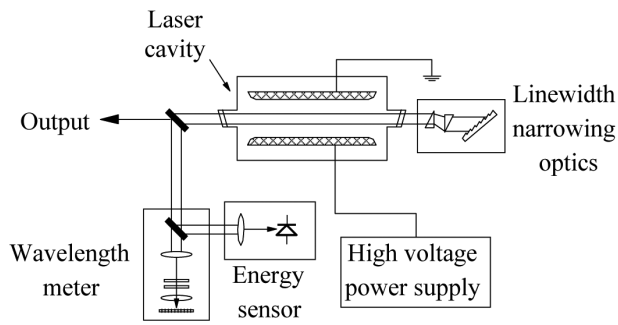


Figure 5.7 Schematic of a single-chamber excimer laser, showing key subcomponents.

steppers (2000 mW/cm^2 or more). Further improvements for DUV systems are limited by glass damage that occurs with higher peak power and difficulties in increasing the excimer-laser repetition rate beyond 6 kHz. A high level of atomic fluorine in the discharge region of an excimer laser can cause arcing and other instabilities, so the by-products from one pulse (see Table 5.3) must be removed from between the electrodes and replenished with fresh gas before another pulse can be fired. The repetition rate is limited by the ability to exchange the gas between the electrodes. Faster exchange rates place significantly greater requirements on fans and motors, with attendant concerns for reliability degradation. In principle, glass damage can be reduced by "stretching" the pulses from their unstretched length of 25–30 nsec,[34] reducing the peak energy while maintaining the total-integrated energy.[35] Another advantage of temporally longer pulses is a reduction in bandwidth.[36] An example of a stretched pulse is shown in Fig. 5.8.

Because laser-pulse intensity does not evolve symmetrically in time, temporal pulse length needs definition. The most commonly used definition is the integral-

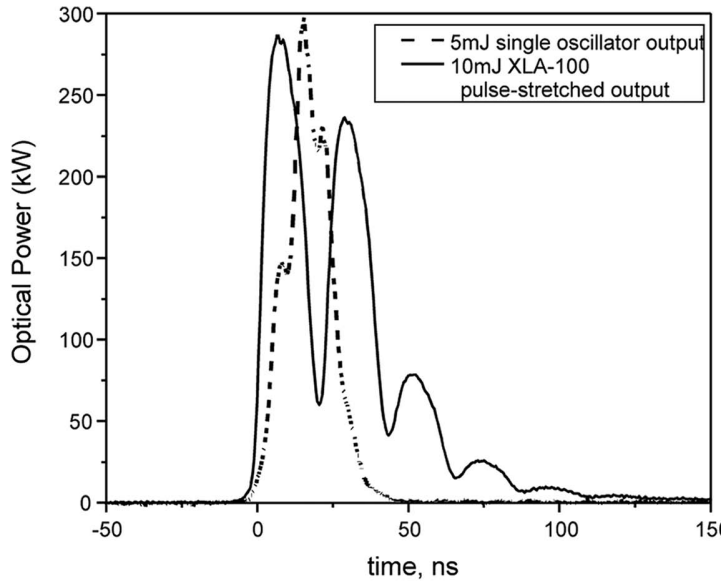


Figure 5.8 Excimer-laser power versus time for a normal and stretched pulse from a Cymer ArF excimer laser.[37]

squared pulse width:[38]

$$\tau = \frac{\left[\int I(t)\,dt\right]^2}{\int I^2(t)\,dt}. \tag{5.1}$$

This definition is used because this is the pulse duration that is most relevant to the issue of glass damage.[38]

The requirements for the laser gases cause the cost of installation of excimer laser steppers to increase relative to costs for arc-lamp systems (Table 5.4), particularly since the fluorine gases have safety requirements that must be addressed. An innovative solid source for fluorine addresses this safety issue.[39] Since fluorine leads to the etching of silicon dioxide, the windows of the excimer laser are typically made of calcium fluoride, a material that is discussed in more detail in a later section. Laser gases must be very pure, as impurities will cause degradation of laser performance. Since fluorine is a primary excimer laser gas, special materials must be used for handling this very chemically reactive gas.

Table 5.4 Installation and consumables costs for the illumination systems of arc lamp and excimer laser–based wafer steppers.

	Arc lamps	**KrF excimer lasers**
Availability	99.5%	97%
Cost for consumables	$40,000/year	$100,000/year
Installation costs	$1000–$10,000 per stepper	$100,000 per stepper

Beam-delivery systems involve another expense associated with excimer-laser steppers that does not exist for arc-lamp systems, where the illumination source is built into the stepper. The purpose of the beam-delivery system is to provide an optical path from the excimer laser, which is usually located outside of the cleanroom, to the wafer stepper. Typically, the beam-delivery system is a set of tubes, containing lenses and mirrors, through which the DUV light traverses the distance between laser and stepper (Fig. 5.5). Performance of the imaging optics requires alignment of the input beam to within 1 deg, positioned to within 150 μm,[40] while the laser may be separated from the stepper by as much as 25 m. This places tight constraints on tolerable levels of vibration throughout an extended part of the fabricated beam line. Maintaining the 150-μm placement of a beam delivered from 10 m away requires 3 arcsec of angular control. A robust way to ensure good alignment of the laser beam is to employ active beam sensing and control of the mirrors in the beam-delivery unit.[41]

The beam-delivery system must contain optics to address the natural beam divergence of an excimer laser. A laser beam of a cross section 1×2 cm will diverge 5 mrad and 2 mrad, respectively. At a distance of 25 m, the beam spreads to a cross section in excess of 13×6 cm without optics to refocus the beam. The beam-delivery system also typically requires mirrors to bend the beam around obstructions between the laser and the stepper. While not as sophisticated as the stepper's projection optics, the lenses in the beam-delivery system require antireflection coatings to maintain light intensity at the actinic wavelength and also at an optical wavelength, usually 632.8 nm (HeNe laser), for aligning the beam-delivery optics. To avoid photochemically deposited contaminants on the optical elements, the beam-delivery system is enclosed and often purged with nitrogen. Enclosure is also required to provide a Class-1 laser system.[42] The lenses and mirrors in beam-delivery units encounter high fluencies, so the lens and mirror coatings need to be resistant to damage by high-intensity light.[43] The frequency at which components need to be replaced in order to maintain good performance will depend upon optical-component quality and the level of contamination within the beam-delivery unit.

Excimer lasers today are installed and maintained as Class-1 laser systems. This was a critical advance for production environments, where protecting operators from stray DUV light formerly required elaborate shielding walls and curtains during laser installation and maintenance. Stepper manufacturers have responded to this standard of engineering by designing beam-delivery systems that can be aligned while maintaining the systems in Class-1 configurations.

Excimer laser utilization has also improved during the past few years, over and above the improvements resulting from reductions in unscheduled downtime. The gas inside the lasers requires periodic replacement, since small amounts of impurities in the gas reduce lasing efficiency and can deposit coatings on windows. Early excimer lasers for lithography required complete gas refills every 100,000 pulses, while today's lasers require such refills only every two-*billion* pulses.[44] Much of this improvement resulted from the replacement of organic materials,

such as Teflon® and Viton®, with ceramic and pure metal materials for insulators and seals, and improved methods for cleaning parts.

The step-and-scan configuration places more stringent requirements on the excimer lasers than do step-and-repeat machines. There is a direct relationship between the maximum scan speed and the number of pulses required to achieve a specified dose:

$$W_s = V_m \frac{n}{f}, \tag{5.2}$$

where W_s is the slit width, V_m is the maximum wafer-stage scan speed, n is the minimum number of pulses required to achieve the specified dose, and f is the laser repetition rate.[45,46] The minimum number of pulses used to achieve the dose is also related to dose control. The standard variation of the dose σ_D is related to the pulse-to-pulse variation σ_{P-P} by:

$$\sigma_D = \frac{\sigma_{P-P}}{\sqrt{n}}. \tag{5.3}$$

If σ_{P-P} is large, n cannot be too small or there is inadequate dose control. Typical pulse-to-pulse repeatability is about 2.5% (3σ). Thus, laser pulse-to-pulse variation is directly related to the potential throughput of step-and-scan systems. For example, with 2.5% (3σ) pulse-to-pulse energy variation, a dose tolerance of 0.25%, and a slit width of 8 mm, the wafer-stage scan speed of the exposure tool with a 4-kHz laser will be limited to 320 mm/sec by the pulse-to-pulse stability. A missed pulse, in which the excimer laser simply fails to fire, seriously degrades dose uniformity across the scan, while step-and-repeat systems keep the shutter open for an additional pulse and compensate completely. Thus, greater laser reliability is required for step-and-scan systems than for step-and-repeat systems.

Freely running KrF lasers have a natural bandwidth of approximately 300 pm [full-width half-maximum (FWHM)], which is too wide for high-resolution wafer steppers. All-refractive optics require bandwidths of <1.0 pm, and even catadioptric, moderate-NA systems require bandwidths of <100 pm.[47] (Lens types are discussed in Section 5.4.) The requirements of catadioptric lenses are met easily, while the line narrowing demanded by all-refractive lenses has presented a challenge to laser designers. For refractive systems there is a need to be able to vary the center wavelength in a controlled manner, and this wavelength is tunable on KrF lasers over a range of up to 400 pm, at least in principle.[48] In practice, the wavelength only needs to be tuned over <1 pm. As the numerical aperture increases, the bandwidth must be reduced. (The reason for this is discussed in Section 5.4.) Consequently, the 0.6-NA Micrascan III required the use of an excimer laser, narrowed to $\Delta\lambda < 100$ pm,[49] in contrast to earlier models of Micrascans which had broadband arc-lamp sources of actinic light. However, the Micrascan III, with its catadioptric lens, still tolerated a bandwidth much larger than the $\Delta\lambda < 0.8$ pm required for refractive-lens systems. Several methods

have been used to narrow the bandwidth, principally gratings[50] and Fabry-Perot etalons (discussed below), or combinations of both. Prisms are sufficient to narrow the bandwidth for application for moderate-NA catadioptric systems,[51] while extremely high-NA catadioptric lenses still have very tight bandwidth requirements (<0.5 pm FWHM). The different options for narrowing laser bandwidths are illustrated in Fig. 5.9.

Etalons are based upon the transmission properties of light through a transparent plane-parallel plate. If the reflectance from an individual surface of the plate is R, then the transmitted intensity through the plate I_t, normalized to the incident light intensity I_i, is given by[52]

$$\frac{I_t}{I_i} = \frac{1}{1 + F \sin^2 \dfrac{\delta}{2}}, \tag{5.4}$$

where

$$\delta = \frac{4\pi}{\lambda} nt \cos\theta \tag{5.5}$$

and

$$F = \frac{4R}{(1-R)^2}. \tag{5.6}$$

In these equations, t is the thickness of the parallel plate, θ is the angle of incidence, n is the refractive index of the plate, and λ is the wavelength of the light (in vacuum). The resulting transmitted light intensity is shown in Fig. 5.10. Because of their high resolving power, etalons are used to measure wavelength and bandwidth with the intensity reduced sufficiently to avoid damage. All-refractive optics require very good control of the wavelength. Variations in the wavelength centerline can cause shifts in focus,[53,54] distortion,[55,56] and other aberrations. One

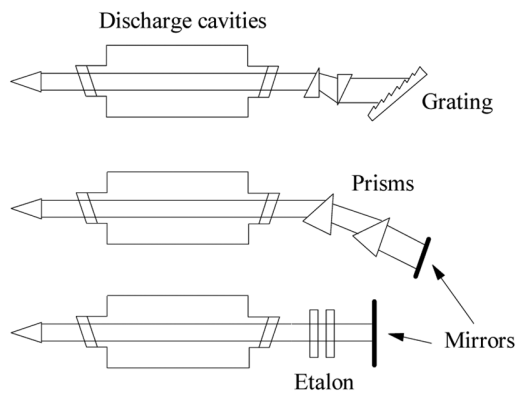


Figure 5.9 Different types of bandwidth-narrowing optics.

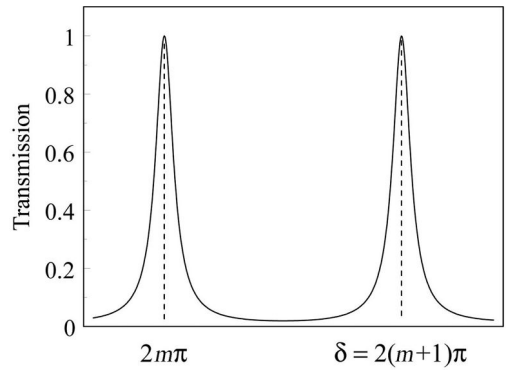


Figure 5.10 Transmitted light intensity versus δ through an etalon. δ is given by Eq. (5.5).

of the problems with early excimer lasers was maintaining absolute wavelength calibration. The gratings and Fabry-Perot etalons used to narrow the bandwidth require a level of mechanical stability to maintain a fixed wavelength that is not achievable. Modern KrF lasers use an internal iron reference, with several wavelengths in the region of 248.3 nm, to establish absolute calibration.[57] Single-isotope mercury sources have also been used.[58] For ArF lasers, there is a carbon spectral line at 193.0905 nm that is used to establish absolute wavelength calibration.[59]

Bandwidth specifications for excimer lasers for lithographic applications were originally specified in terms of full-width half-maximum. Over time it became recognized that image degradation could be caused by wavelengths on the "tails" of the bandwidth distribution. It was also found that linewidth variations were found to correlate better to a parameter referred to as E95, which is the bandwidth in which 95% of the light energy is contained, than to the FWHM metric.[60] As a consequence, bandwidth specifications for the latest generations of excimer lasers for lithographic applications are given in terms of E95.

Modern exposure tools impose conflicting requirements on lasers. Improved imaging requires illumination with narrower bandwidths, for the reasons discussed earlier in this section. There is also strong economic pressure for lasers to have high-energy output in order to achieve high exposure-tool productivity (see Chapter 11). It is very difficult to obtain high-laser-energy output at very narrow wavelengths, particularly for ArF lasers, which are intrinsically less efficient than KrF lasers. One approach to addressing both requirements is the technique of injection locking.[61,62] In this approach a laser pulse is generated in a seed laser, or master oscillator. This pulse then travels through a second laser chamber, where the power of the signal is amplified. The critical beam characteristics, such as bandwidth, are determined largely by the master oscillator, while power is generated by the power amplifier. Injection-locking technology has existed for many years, but its application to lithography has been inhibited by the high spatial coherence of the light produced by traditional injection-locking systems.[63]

Cymer introduced a two-chamber laser system, referred to as master oscillator power amplifier (MOPA), which produces light with less spatial coherence than typical injection-locking lasers. With MOPA, the first chamber (the master oscillator) is used to create a laser pulse with a narrow bandwidth. This pulse is then injected into a second chamber, the power amplifier, which amplifies the signal (see Fig. 5.11). There is an active discharge in the amplifier chamber that must be synchronized with the pulse generated in the master oscillator. Cymer's MOPA differs from traditional injection locking by not having resonator optics as part of the amplifier chamber, and this allows for reduced levels of spatial coherence. Gigaphoton has also introduced a two-chamber laser. Even though the system from Gigaphoton does have resonator optics in the power amplifier stage, it produces light with spatial coherence that is low enough for lithographic applications.[64]

Lasers often produce nonuniform illumination, because of the phenomenon of speckle,[63] associated with high degrees of coherence. (See Appendix A for a brief discussion of coherence.) Because of their high gain, excimer lasers naturally have multimode outputs, with resulting low spatial coherence,[65] and therefore lower amounts of time-averaged speckle than produced by many other types of lasers.[66] Consequently, excimer lasers are well suited as sources of light for photolithography, as lasers go. However, the very stringent requirements of lithography have resulted in speckle being manifest in resist patterns, particularly line-edge roughness.[67] The impact of speckle can be reduced by increasing the number of pulses per exposure so that the speckle noise within individual pulses is moderated by averaging with other pulses. Increasing the temporal pulse length also serves to reduce the impact of speckle.

Dual-chamber lasers achieve their objective of narrower bandwidth and higher output, as shown in Table 5.5. The bandwidth is reduced appreciably on the dual-chamber lasers relative to single-chamber lasers, while the output power is more than tripled. Improvements in excimer-laser technology are expected to continue, and it remains to be seen what the limits will ultimately be for bandwidth, output

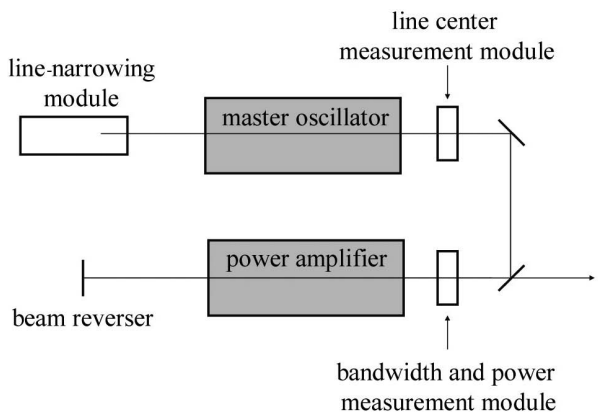


Figure 5.11 MOPA configuration.

Table 5.5 Comparison of dual-chamber ArF lasers with a representative single-chamber ArF laser. The dual-chamber lasers are available with two options for power output. MOPO is the acronym for master oscillator power oscillator.

	Cymer Nanolith 7000 (single chamber)	**Cymer XLR600ix**[69] **(MOPA)**	**Gigaphoton GT62A**[70,71] **(MOPO)**
Pulse energy	5 mJ	10 mJ/15 mJ switchable	10 mJ/15 mJ switchable
Power	20W	60 W/90 W	60W/90 W
Repetition rate	4 kHz	6 kHz	6 kHz
Spectral bandwidth (FWHM)	≤0.5 pm	not specified	not specified
Spectral Bandwidth (E95%)	≤1.3 pm	≤0.35 pm	≤0.35 pm

energy, and pulse duration. For more detail on excimer lasers for lithographic applications, refer to the excellent review paper by Das and Sandstrom.[68]

5.3 Illumination Systems

The stepper's illumination system performs several functions, which include spectral filtering (particularly for arc-lamp systems), establishing a specified level of coherence, and perhaps generating complex forms of illumination. In addition, the illumination system must control the exposure dose and provide uniform light. These functions of the illuminator are discussed in this section.

With arc-lamp light sources, it is necessary to filter out the unwanted wavelengths, including those in the infrared that lead to excessive heating. This filtration usually occurs in one of the first stages of the illumination system. The illumination of the reticle must be very uniform (≤±1%) over a large area. For step-and-repeat systems this can be up to a 130 mm × 130 mm square in size, while it is much smaller for step-and-scan systems, on the order of 104 mm × ~40 mm. The need to provide uniform illumination over a smaller area is another advantage of step-and-scan relative to step-and-repeat. Illumination uniformity is related to linewidth control across the exposure field. Exposure latitude is usually given as the amount the exposure dose can be changed without changing linewidths by more than ±10% from the target value. The typical exposure latitude of current processes is ±5%, but has been decreasing over time, placing tighter requirements on illumination uniformity. Good illumination uniformity is accomplished by a number of approaches, such as the use of fly's eyes[72] and rods.[73] A fly's eye arrangement consists of an array of lenslets, each of which images part of the light source over the entire field to be illuminated, thereby averaging the amount of light over the field (Fig. 5.12). The averaging effects of the optical systems in illuminators also help to reduce the effects of speckle that were discussed in the prior section. Additional methods for reducing illumination nonuniformity due to speckle and other causes include deflection of the laser beam with a vibrating mirror within the illuminator.[74]

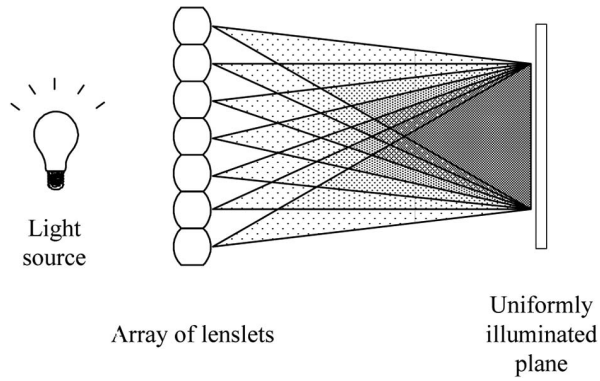


Figure 5.12 Schematic of a fly's eye array for producing uniform light intensity in an illuminator.

With uniform illumination over the entire slit, good dose control on step-and-scan systems requires the same number of pulses for every exposure. To avoid a dose quantization problem, the illumination is usually nonuniform in the direction of the scan on step-and-scan systems, a trapezoidal shape being typical (Fig. 5.13).

The illuminator must also control the dose. This requires real-time measurement of the exposure. A typical method involves the use of a beamsplitter to pick off a small fraction ($\leq 1\%$) of the light, and the intensity is measured (Fig. 5.14). In a step-and-repeat system, the light is integrated from the time the shutter is opened, and the shutter is closed when the correct exposure dose is achieved. For a step-and-scan system, the feedback is used to control the scanning speed. Constant dose on the wafer is maintained by decreasing the scanning speed of both the reticle and wafer by the same percentage to compensate for decreases in laser- or lamp-light output.

The use of a light-intensity monitor that is integral to the illuminator requires calibration of this internal intensity monitor to the light intensity that is actually achieved on the wafer. The light at the wafer plane is affected by transmission of the lens and those portions of the illuminator that are between the light monitor and the reticle. Most steppers have a light-intensity detector on the wafer stage, enabling periodic checks of the dose-control system. The calibration of the detector

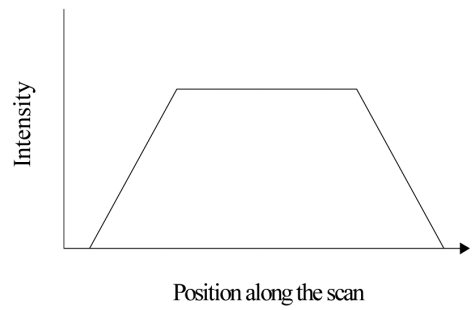


Figure 5.13 Trapezoidal illumination in the direction of the scan.

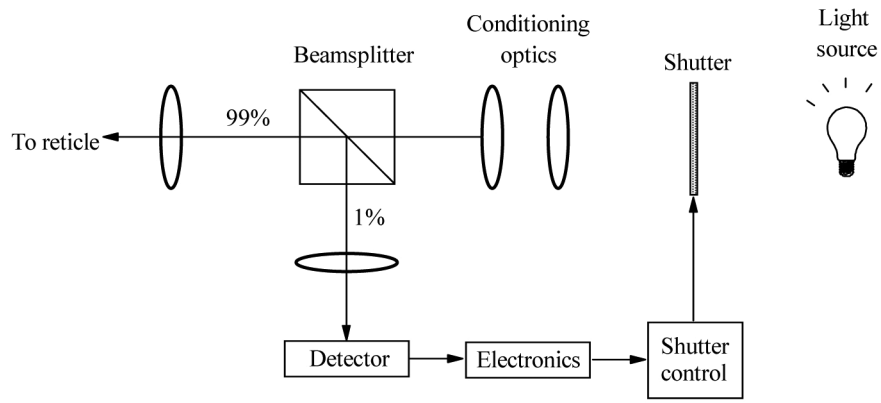


Figure 5.14 Part of the light from the source is measured in order to control the exposure dose. The particular control system shown in this figure applies to step-and-repeat systems. In step-and-scan systems, the feedback loop involves the scan speed.

on the stage requires the use of an external energy monitor, and these are available from a number of commercial suppliers. The National Institute of Standards and Technology (NIST) provides absolute calibration capability.[75]

The illuminator also establishes the partial coherence (σ) of the light for conventional illumination and generates various forms of off-axis illumination as well. (Off-axis illumination is discussed further in Chapter 8.) The subject of coherence is discussed briefly in Appendix A. On modern wafer steppers, σ is a user-controllable parameter.[76,77] The geometry that controls σ is shown in Fig. A.3 of Appendix A. One simple way to modulate σ is to use an aperture that regulates the light incident on the reticle. However, apertures reduce the total light intensity, thereby diminishing exposure-tool productivity. Modern illuminators have been designed to redirect the light collected from the light source to allow for variable partial coherence without significant reductions in light intensity when low values of σ are selected.[78] For example, the angle of the cone of light that illuminates the reticle can be modulated by the use of zoom optics, as shown in Fig. 5.15. Similarly, light can be directed into an annulus by using a conical shaped lens element, known as an axicon.[79] A cross section of an axicon is illustrated in Fig. 5.16. As shown in Fig. A.3, modifying the cone angle of the illumination changes the partial coherence of the light. In addition to using refraction to redirect light within the illuminator, diffractive optical elements (DOE) can also be used, and they are useful for generating complex illumination shapes, such as will be discussed in Chapter 8.

The effect of partial coherence on imaging is shown in Fig. 5.17. Complete coherence leads to oscillations in the light-intensity profile that result from the interference of light waves, but the edges of the profiles are steeper than those obtained with less coherent light. Sharper profiles are obtained with small values of sigma ($\sigma \leq 0.4$) and are beneficial for patterning features such as contacts[82] that need the best optical profiles possible. On the other hand, it has been found that

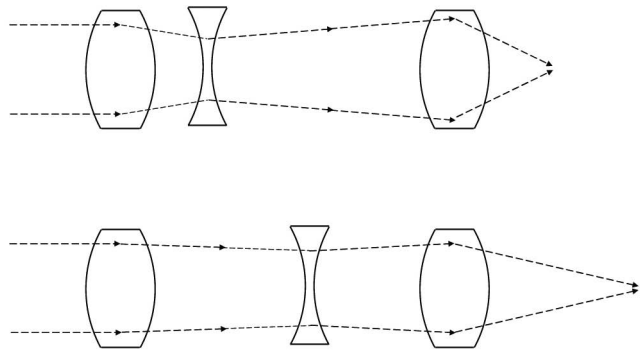


Figure 5.15 Illustration of zoom optics,[80] which can be used to modify the partial coherence of the illumination.[81]

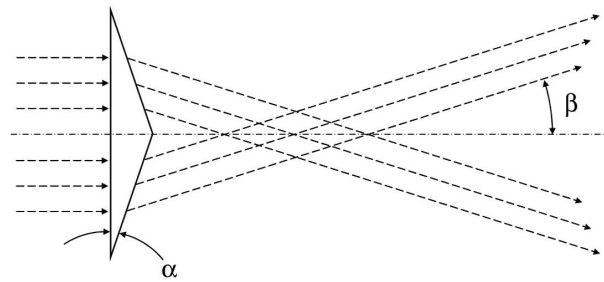


Figure 5.16 Illustration of the cross section of an axicon. $\beta = \arcsin[n\sin(\alpha)] - \alpha$, where n is the index of refraction of the glass comprising the conical lens.

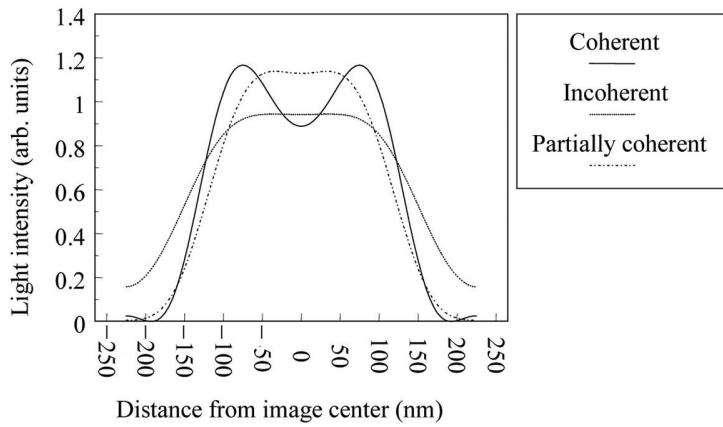


Figure 5.17 Variation of light-intensity profiles with partial coherence. The images were for 300-nm nominal features, calculated using PROLITH (version 1.5), for an aberration-free 0.6-NA lens, and imaging at a wavelength of 248 nm.

better depth-of-focus of line and space patterns is obtained with higher values of sigma ($\sigma \geq 0.6$),[83] which are often used on gate and metal layers.

As might be expected from Fig. 5.17, linewidths are modulated by the partial coherence. This effect is shown in more detail in Fig. 5.18, where linewidth is given as a function of partial coherence. As can be concluded from the data shown in Fig. 5.18, not only must illumination intensity be uniform over the exposure field, so must partial coherence[84] in order to maintain uniformity of linewidths across the exposure field.

As is discussed in detail in Chapter 8, the depth-of-focus can be enhanced in certain circumstances by illuminating the reticle with light from a ring or annulus of light brought in at oblique angles to the reticle. As with respect to partial coherence, illuminators are designed to produce off-axis illumination without significantly reducing the light intensity, by using optical elements such as axicons.[78,85]

In another implementation of off-axis illumination, quadrupole, the reticle is illuminated with four separate beams.[86] For quadrupole illumination to work properly, all quads of the illumination must have the same the light intensity. This is a more stringent requirement than having uniform light intensity overall.[87]

5.4 Reduction Lenses

On a wafer stepper, *N*-times demagnified images of the IC patterns are projected onto the wafer surface by the microlithographic lens, mounted only a few millimeters above the wafer surface. Values of *N* typically range between 1 and 5, with 4× reduction being the most common, although 5× was the dominant reduction prior to the advent of step-and-scan technology. The first commercially available step-and-repeat systems employed 10× reduction. Reduction optics have a benefit with respect to reducing sensitivities to defects on the reticles,[88–90] reticle linewidth variations, and registration errors, with greater such advantages

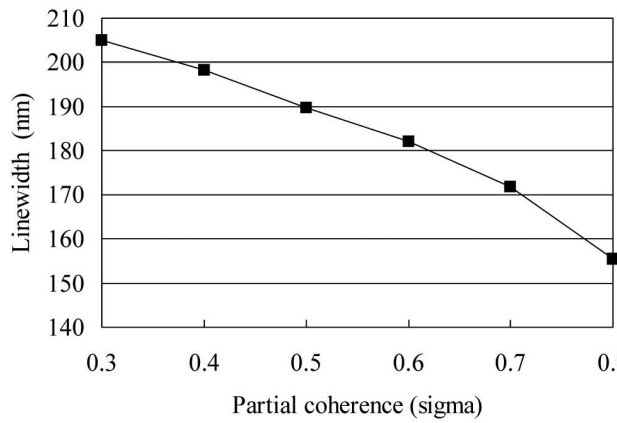


Figure 5.18 Linewidth as a function of sigma, given by calculated aerial images for isolated 180-nm lines, with optical parameters of 0.7 NA and a wavelength of 248 nm.

accompanying larger values of N. However, the largest field on the wafer equals the biggest quality area on the reticle demagnified by N. To sustain field sizes adequate to handle expected chip sizes (see Table 5.6) while maintaining practically sized reticles, and to provide adequate stepper throughput, the lens-reduction factor N has been kept to modest values, such as 5× and 4×. The factors that determine the optimum lens-reduction factor are given in Table 5.7.

Most stepper lenses with numerical apertures of ≤0.93 are refractive, meaning that the optical elements are all made of transmitting glass. Microlithographic reduction-lens design has generally followed the "double-Gaussian" form[92] first described by Glatzel[93] of Carl Zeiss. An example of a modern stepper lens design is shown in Fig. 5.19. Early stepper lenses weighed about 10 lbs. and resembled camera lenses. Indeed, the stepper is a type of camera, in which a picture of the reticle is taken and printed on a film comprised of photoresist. It is not surprising

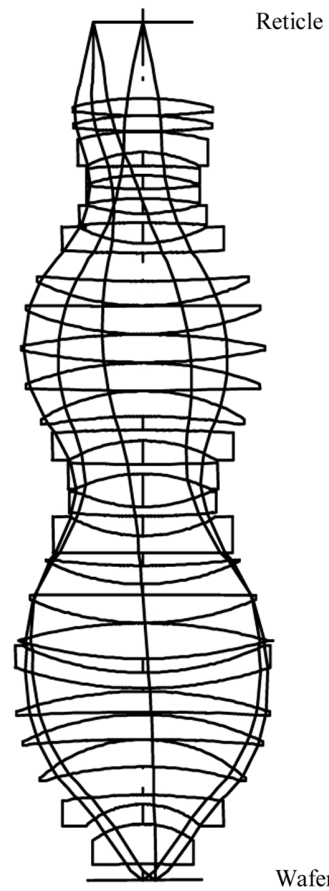


Figure 5.19 The lens drawing shows one embodiment of the invention (DE19855108, filed Nov. 30, 1998) that the lens designers from Carl Zeiss generated within the framework of a design and feasibility study related to modern optical-projection systems for microlithography at 248 nm with high numerical aperture (Courtesy of Winfried Kaiser of Carl Zeiss).

Table 5.6 Expected chip sizes at the start of production, from the 2008 International Technology Roadmap for Semiconductors.* Applications-specific ICs are expected to fill available field sizes. Memory chips tend to have 2:1 aspect ratios, while microprocessor aspect ratios are closer to 1:1. For many years exposure field size requirements were driven by DRAM chip sizes, but logic now determines how big field sizes need to be.

	2007	2010	2013	2016
Half pitch (nm)	68	45	32	22.5
DRAM (mm^2)	107	93	93	93
Microprocessor (mm^2)	652	563	560	557

* http://public.itrs.net

Table 5.7 Factors that determine the optimum lens-reduction factor.[91]

Reasons that compel a large lens-reduction factor	Reasons that may compel a small lens-reduction factor
CD and overlay control	Large die size
Reticle inspection and repair	230-mm reticle cost and reality
Small features on reticles for OPC	Stepper throughput
Reticle defects	Lens size and cost
	Potential for multiple dies on reticles

that the three largest manufacturers of stepper lenses—Nikon, Canon, and Carl Zeiss—have their roots in making cameras. The modern refractive lens has 20–30 glass elements held firmly in a steel cylindrical jacket. The lens may be a meter in length and weigh 500 kg or more. Photographs of KrF lenses are shown in Fig. 5.20.

Refractive lenses typically provide good imaging over a narrow band of wavelengths. Image formation by lenses relies on the phenomenon of refraction, where rays of light are bent at the interface between materials with different optical properties. Referring to Fig. 5.21, light rays that have an angle of incidence θ_1 will refract to the angle θ_2 according to Snell's Law:[52]

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}, \tag{5.7}$$

where n_i is the refractive index of material i. In lenses, one of these materials is usually air and the other is glass.

With refractive lenses, imaging changes if either the index of refraction of air or the index of refraction of the glass changes. These refractive indices change with wavelength—the phenomenon of dispersion. As a consequence, the imaging of the lens changes as the wavelength of light changes, leading to practical limitations on the bandwidths over which diffraction-limited lenses can be designed and built. Designing a refractive lens with nearly diffraction-limited imaging over a wide range of wavelengths is very difficult, and such lenses, if they could be designed, would be larger, more complex, and more expensive than narrow-bandwidth lenses.

(a) (b)

Figure 5.20 Two KrF lenses: (a) 0.82-NA KrF lens from Nikon, and (b) 0.8-NA lens from Canon.

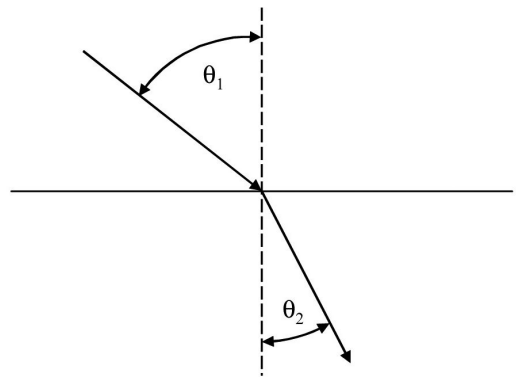


Figure 5.21 Refraction of light at an interface. The medium above the line has an index of refraction of n_1, and the medium below the line has an index of refraction of n_2.

The bandwidths over which lenses operate tend to decrease along with wavelengths and feature sizes. Consequently, the lenses used in microlithography, particularly those used for imaging features of ≤65 nm, operate over narrow bandwidths.

Lenses can be made to operate over wider bandwidths when there is more than one optical material available, because the lens designer has an additional degree of freedom to use in designing lenses. G-line and i-line lenses can operate over the bandwidths produced by mercury-arc lamps, or with slight narrowing, typically 4–6 nm (FWHM). For many years, fused silica was the only available material with sufficient transparency and quality for making 248-nm lenses, and such lenses have needed highly narrowed light sources, typically <1 pm (FWHM). This is a much smaller bandwidth than needs to be supported at g-line and i-line wavelengths, where there are several different glass materials that can be used for making chromatic corrections in the lenses. To meet the needs of ArF lithography, crystalline CaF_2 with adequate quality and in pieces of sufficient size

has become available. This material can work down to 157 nm and somewhat shorter wavelengths. Lenses for 193-nm lithography typically use both fused silica and a small quantity of CaF_2. In spite of their shorter wavelengths and higher resolution, lenses for 193-nm lithography have approximately the same bandwidth requirements as the single-material 248-nm lenses of comparable NAs, because a second lens material is used.

For good overall lens transmission, the materials used for the lens elements must be extremely transparent. The transmission properties of various types of optical glass are shown in Fig. 5.22.[94–96] Today, fused silica and calcium fluoride are the only suitable materials available in sufficient quantity for use at wavelengths below 350 nm, and these materials, too, have their limits. Lenses for 193-nm lithography usually include CaF_2 as an optical material[97] because of its excellent transparency at 193 nm and its resistance to optical damage (a problem that is discussed shortly). BaF_2 was being studied as a possible material for 157-nm lithography and 193-nm liquid immersion lithography (to be discussed in Chapter 10).

For 248-nm and 193-nm lithography, it is practical to have lenses that operate over very narrow ranges of wavelength (<1 pm), because adequate intensity from excimer lasers can be obtained even over such narrow bandwidths. On the other hand, i-line lenses need to operate over much wider bandwidths of between 4 and 6 nm. At shorter wavelengths, ingredients that change the glass's index of refraction also absorb the light. In stepper lenses this absorption is small and does not have a major impact on lens transmissions, but there are consequences. Lenses containing absorbing materials heat up and expand during exposure, causing defocus,[98] changes in magnification, and other aberrations. Focus is the parameter affected most by lens heating, with some measurable impact on magnification. Changes to the other aberrations are typically small, but are nonnegligible for very low-k_1 lithography. This heating often occurs with i-line lenses, and a similar problem occurs for 193-nm lenses, because fused silica becomes slightly absorbing

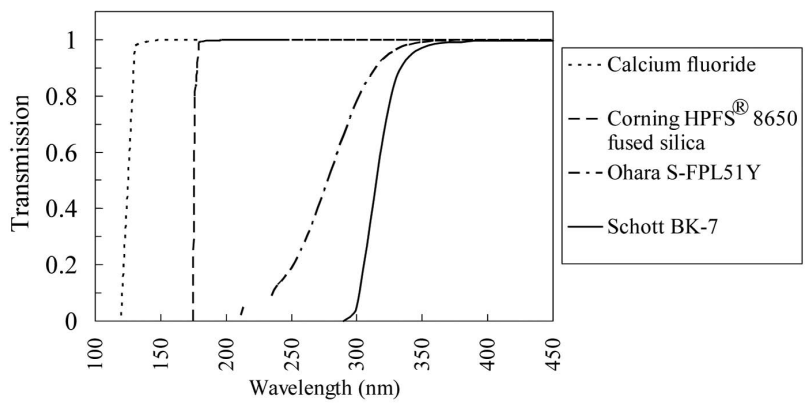


Figure 5.22 The transmission of 10-mm-thick samples of materials used in the lenses of wafer steppers. Surface losses are removed. The Ohara and Schott materials are suitable for i-line lenses, but are too absorbing at DUV wavelengths.

at that wavelength (unlike what happens at 248 nm, where fused silica is extremely transparent).

Stepper software is required to compensate for the effects of lens heating.[99] Manipulators are used to adjust lens-element positions actively to correct for changes induced by lens heating. The amount of light transmitted into the lens is determined by the patterns on the masks. Less light is transmitted through a contact mask than a typical gate-layer mask. The light that actually enters the lens is also affected by feature sizes, since much of the light can be diffracted out of the entrance pupil of the projection optics. Diffraction may also cause nonuniform lens heating, reflecting the nonuniformity of the mask's diffraction pattern. Just as the lens will heat up during exposure, it will cool when the stepper is idle, and the stepper must correct for the cooling as well.

Transparency is a necessary but insufficient requirement for a material to be appropriate for the optics used in microlithography. Many crystalline materials exhibit birefringence, where the index of refraction is different along various crystal axes, and such materials are generally not suitable for use in high-resolution optical systems. For example, SiO_2 in its purely crystalline form, quartz, is not easily incorporated into lenses, while amorphous fused silica is a standard optical material because the crystalline form of SiO_2 has dissimilar behavior among different optical axes. Other materials, such as MgF_2 and LiF have adequate transparency at 193-nm and 157-nm wavelengths, but are too birefringent for use in high-resolution lenses. CaF_2 has negligible intrinsic birefringence at long wavelengths, but at wavelengths as short as 193 nm, this birefringence can no longer be ignored [100]. Fortunately, the birefringence in CaF_2 can largely be corrected by using different [111] oriented CaF_2 lens elements with their crystallographic axes rotated 60 deg relative to each other, and also by including pairs of elements[100] of oriented material (the two elements in the pair rotated 45 deg with respect to each other).[101] Barium fluoride has also been considered as a material for stepper lenses, but it has a much higher level of intrinsic birefringence than CaF_2.[102] Indeed, the intrinsic birefringence of BaF_2 at a wavelength of 193 nm is larger than the birefringence of CaF_2 at a wavelength of 157 nm.

Low levels of birefringence that can affect imaging at levels relevant for microlithography are induced by mechanical stresses into materials that are not birefringent in their perfect crystalline or amorphous state.[103] Such stresses are created if there are large thermal inhomogeneities in the furnaces in which the CaF_2 crystals or boules of fused silica are grown. Residual stress birefringence is one factor that limits the yield of CaF_2 suitable for use in stepper lenses.

The transition to 193-nm lithography necessitated the addition of another lens material, or the use of mirrors as well as glass lenses in the design, in order to maintain practical laser bandwidths. A lens with both reflective and refractive elements is referred to as catadioptric. In addition to fused silica, 193-nm lenses often use CaF_2, which has significantly different physical properties than fused silica, such as being much softer. Because CaF_2 has long been used in television camera lenses, lens makers have had some experience in the past working with this material, albeit not at the critical levels required for the lenses used in

microlithography. Nevertheless, prior experience was sufficient to enable CaF_2 to be used in the first generation of wide-field 193-nm lenses.

In addition to providing a material with an index of refraction different from that of fused silica, CaF_2 is resistant to damage by 193-nm light. Glass damage is proportional to λ^{-4}, so light-induced damage issues are much greater for 193-nm systems than for 248-nm ones.[104] Damage is a concern because excimer lasers are used as light sources. Under high-intensity levels of light, fused silica will compact (become more dense) and become more absorbing. Decompaction, where the glass becomes less dense, can also occur. Compaction or decompaction, if such occurs in the projection optics, will degrade the image quality, and induced absorption will reduce light intensity at the wafer, thereby degrading stepper throughput.

Under exposure to 193-nm laser radiation, the density of fused silica changes, with a resulting change in the material's index of refraction. There appear to be two different mechanisms for change. One is compaction, where the fused silica becomes more dense with increasing exposure,[105] while the other is exactly opposite, rarefaction or decompaction. The change in the index of refraction Δn [in parts per billion (ppb)] can be described approximately by the following expresson:[105,106]

$$\Delta n = k_1 (NI)^a + \left(k_2 \frac{NI^2}{\tau} \right)^b , \tag{5.8}$$

where τ is the pulse length (nsec), N is the number of pulses, I is the energy density (J/cm^2) per pulse, and k_1, k_2, a, and b are fitting parameters that vary from sample to sample. Parameters obtained by measuring the change in refractive index over the course of 40 billion pulses is given in Table 5.8. The first term on the right hand side of Eq. (5.8) represents rarefaction, so k_1 is negative (or zero), while the second term represents compaction. The two processes of compaction and decompaction compensate each other to some extent. The degree to which this occurs is dependent upon material preparation, and the fused silica which is used for ArF stepper optics has a balance between the compaction and rarefaction mechanisms. Lens lifetimes can be estimated by solving Problem 5.7.

Densification and/or rarefaction can change the refractive index and imaging properties of the optical components at a level significant for lithographic optical

Table 5.8 Parameters for Eq. (5.8), for changes to the index of refraction at a wavelength of 632.8 nm, which tracks the index of refraction at 193 nm.[107]

	Sample A	**Sample B**	**Sample C**
k_1 **(J · cm^{-2})**	−0.051	0.00	−0.109
a (unitless)	1.1	n/a	0.83
k_2 **(cm^4 · ns · J^{-2})**	12.1	5.24	2.82
b (unitless)	0.98	1.0	1.0

systems. After extensive use, compaction and/or rarefaction can render 193-nm optics unsuitable for high-resolution lithography. The details of the manufacturing process for the fused silica material can influence the rate of compaction, and grades of fused silica adequate for use in 193-nm projection optics have been developed. In the illuminators of 193-nm systems, where high light intensities occur, appreciable amounts of CaF_2 are needed.

As can be seen in Eq. (5.8), the compaction and rarefaction terms have different functional dependencies. Optics lifetime can be maximized by optimizing peak power and pulse duration. Since the intensity differs at every element in the optical path, careful studies must be done for each lens design. Since light distributions within lenses depends on the diffraction patterns resulting from particular combinations of mask layouts and illumination, lens-lifetime studies must also take into account how the exposure tools will be used.

In general, some light reflects from the interface between two optically different materials. Light propagating through a semifinite medium with an index of refraction n_1 and normally incident onto a planar interface with another semifinite medium that has an index of refraction n_2 has a fraction of light energy reflected away from the interface:

$$R = \left| \frac{n_2 - n_1}{n_2 + n_1} \right|^2 . \tag{5.9}$$

The index of refraction of air ≈ 1, while the indices of refraction for the glass materials used in lithography, at the wavelengths of interest, are ~1.5. Consequently, approximately 4% of the light is reflected at each interface. For a lens with 30 optical elements (60 interfaces) the overall transmission *loss* due to these reflections would be over 90% with such simple air-glass interfaces. To avoid such a large loss of transmission, with the resulting diminution of stepper throughput and productivity, antireflection coatings are applied to the surfaces of each optical element. The physics of such coatings is identical to that of the top antireflection coatings used in resist processing and discussed in Chapter 4. As seen from Eq. (4.40), these coatings perform optimally at certain wavelengths, and this is another reason why refractive lenses should be used over a narrow band of wavelengths. On the other hand, associated with the use of exposure systems with monochromatic illumination, there is the susceptibility of the process to thin-film optical effects, a topic that was discussed in detail in Chapter 4. It has proven better to avoid compromising the projection optics and to find solutions to the problems created by thin-film optical effects.

Because of high fluences, optical coatings need to be developed that are resistant to laser-induced damage. This is particularly true for antireflection coatings in the illuminator where fluences are high, and it also applies to high-reflectance coatings.[43] While fluences are lower in the projection optics, hence the potential for damage is less, the sensitivity to damage is greater for the stepper lens, causing repair costs to be much higher than for illuminators and lasers. In all cases, coatings

must be designed to be resilient under extended exposure to high-intensity DUV light.

One example of a problem caused by damaged or nonoptimized antireflection coatings is flare.[108] This term refers to light scattered in the optical system that reaches the image plane. When there are reflections at the glass-air interface, multiple reflections occur, leading to a background of light (Fig. 5.23). Flare also arises from scattering, which occurs at surfaces or at defects in the glass material itself.

Flare occurs when the surfaces of lens elements are rough. Obtaining smooth surfaces is particularly difficult for CaF_2 lens elements because the hardness of CaF_2 depends upon crystallographic orientation. Consequently, material removal is anisotropic across curved surfaces. Special techniques have been developed for polishing CaF_2.[109]

Scattering from lens surfaces can result from contamination. The high-intensity DUV light can induce photochemical reactions that deposit material on illuminated lens surfaces.[110] If the deposited material coalesces, light is scattered by the resulting particles. One way to avoid this contamination is to purge the lenses with extremely pure gases. Levels of impurities need to be well below 1 ppb.[111]

Flare leads to light in parts of the pattern that correspond to opaque chrome on the mask.[112] This can occur over a range of distances.[113] For example, long-range flare leads to a fairly uniform background of light throughout the imaged area, which leads to reduced image contrast. If a uniform background of scattered light exists, it will be added to both the maximum and minimum light in an image; thus, the image contrast [Eq. (2.9)] becomes

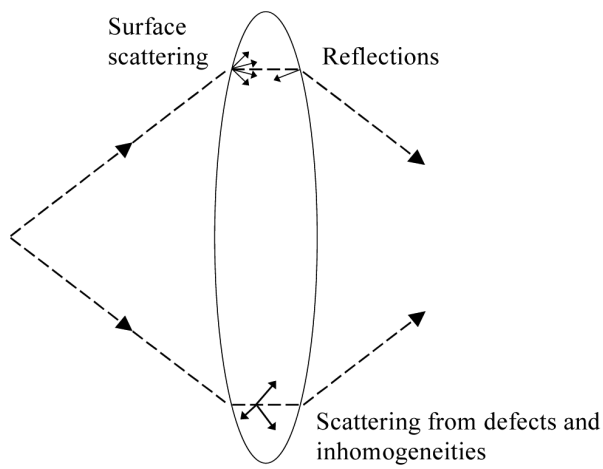


Figure 5.23 Flare resulting from reflections and scattering within lens elements.

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{5.10}$$

$$\xrightarrow{\text{with flare level } \delta} \frac{(I_{max} + \delta) - (I_{min} + \delta)}{I_{max} + \delta + I_{min} + \delta} \tag{5.11}$$

$$= \frac{I_{max} - I_{min}}{I_{max} + I_{min} + 2\delta}. \tag{5.12}$$

As can be seen, flare reduces image contrast.

Scattering produces long-range flare as light passes through the reticle and lens and ultimately creates a fairly uniform background of light across the exposure area. By its very nature, long-range flare is a function of the amount of clear area on the mask as well as the propensity of the lens to scatter light. Light can also be scattered over shorter distances, and correctly accounting for the impact of flare on images requires that medium- and short-range flare also be treated properly. This can be accomplished by measuring image contrast over a range of spatial frequencies.[113]

According to Eq. (2.8), smaller features are resolved by decreasing the wavelength or increasing the numerical aperture of the lenses. The NA can be increased by increasing the diameter of the lens or decreasing the focal length (Fig. 2.11). There are practical limits to both of these factors. The focal length of the lens is limited by the practical necessity of moving wafers underneath it without hitting it. The distance between the bottom of the lens and the wafer surface is often referred to as the *working distance* of the lens. Another reason to maintain a reasonably large working distance is the problem of outgassing from the photoresist during exposure. During exposure, volatile constituents of the resist can outgas and contaminate the bottom lens element. This contamination can lead to nonuniform intensity and scattered light. A sufficiently long working distance and substantial air flow between the lens and the wafer can be used to minimize lens contamination. Typical working distances are on the order of several millimeters.

Reflective optics are advantageous because broadband illumination can be used with them. In practice, high-NA, fully reflective optics are difficult to design and build because optical elements interfere with each other. As can be seen in Fig. 5.24, enlarging the mirrors results in some light rays being blocked by other mirrors. The lens shown in Fig. 5.24 is for an EUV lens, designed for operation at a very short wavelength, 13.4 nm. (EUV lithography is discussed in Chapter 12.) Because of the very short wavelength involved, high resolution is achieved with a very small numerical aperture (0.14). At wavelengths of 193 nm or longer, much higher numerical apertures ($\geq$0.8) are needed to achieve the desired resolution, and the problem of mirrors interfering with each other arises. Consequently, DUV microlithography lenses that employ mirrors also contain some refractive elements. Lenses containing both refractive and reflective elements are called catadioptric. For many years, catadioptric lenses were used primarily by two suppliers of wafer steppers; Ultratech Stepper and SVG Lithography (now a part of ASML). With extremely high NAs associated with immersion lithography (see Chapter 10), Nikon, Canon, and ASML have also begun to use catadioptric lenses routinely.

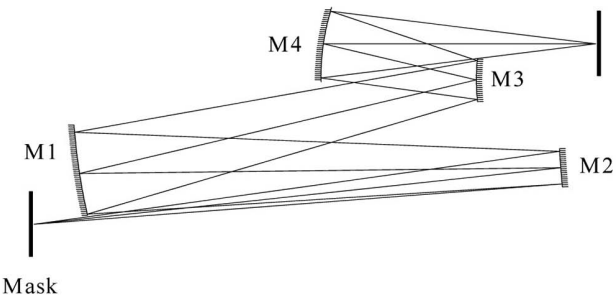


Figure 5.24 Four-mirror design with $NA = 0.14$ and $\lambda = 13.4$ nm, designed for use in EUV lithography.

Ultratech steppers use what is known as the Wynne–Dyson design for lenses.[114] These comprise a family of catadioptric 1:1 lenses capable of imaging over fairly broad ranges of wavelength. Early Ultratech steppers used the part of the mercury spectrum that included both the g- and h-lines, providing the user with the capability to minimize standing-wave effects. A more recent model, the Titan III, uses illumination over the spectrum from 390 to 450 nm, and provides a resolution of 0.75 μm.[115] Unfortunately, it has proven difficult to extend this type of lens from 1:1 to reduction formats.

In addition to being the first commercially available step-and-scan system, the Micrascan I used a catadioptric lens[116] that did provide image reduction (4:1). However, the Micrascan I used ring-field optics that were difficult to fabricate and maintain, and extension to higher numerical aperture was problematic. (Ring-field optics are discussed in Section 5.9.) To address the problem of aligning the optical elements, an innovative lens design[117–119] was introduced with the Micrascan II and extended to the Micrascan III (Fig. 5.25).[120] This lens incorporated a beamsplitter that addressed the problem of aligning the optical elements when curved mirrors were involved, by leading to a lens configuration that had only two orthogonal axes. This greatly simplified mechanical alignment of the lens elements, thereby permitting the extension of catadioptric lenses to higher NAs (0.50 for Micrascan II and 0.60 for Micrascan III, compared to 0.357 on Micrascan I). In the Micrascan II and III, polarized light from the reticle passes through several refractive lens elements and is reflected by a flat folding mirror before entering the polarizing beamsplitter. The light is reflected from the beamsplitter to a curved mirror, returns through the beamsplitter, and is finally imaged onto the wafer. The use of the aspheric mirror element allows a designed bandwidth of 4 nm for the Micrascan II, as compared to the ≤1-pm bandwidth for an all-fused silica reduction stepper lens at 248 nm. Thus the Micrascan II used a mercury-arc lamp for the illumination source, as did the Micrascan I. A mercury-arc lamp, however, would be far too weak to use with an all-refractive lens design, which requires that most of the light be filtered away. As a corollary to being able to image well at a wider range of wavelengths, catadioptric lenses also have greater stability with respect to temperature and pressure changes than is possible with all-refractive lens systems.

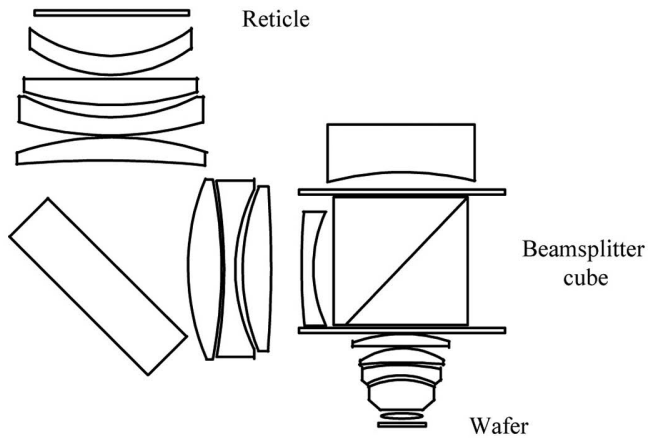


Figure 5.25 Schematic of a 4×, $NA = 0.60$, $\lambda = 248$-nm catadioptric lens used in the Micrascan III, based on the patented design of Williamson.[119]

The different intrinsic stabilities between catadioptric and refractive lenses are listed in Table 5.9.[120] Note that it is possible to compensate for focus and magnification shifts using software, so the greater sensitivity of refractive lenses to temperature and pressure increases the accuracy with which corrections need to be made, but it has not proven to be fundamentally limiting. Nevertheless, there is considerable value to having reduced intrinsic sensitivies, since there will inevitably be some level of compensation error.

With higher NAs, even the catadioptric lenses cannot support the broad bandwidth of the arc lamp's DUV emission, and the Micrascan IIIs, as well as all-refractive lens systems, used a KrF excimer laser for operation at 248 nm. The 0.6-NA version of the Williamson lens used in the Micrascan III had a bandwidth of 100 pm,[121] wide enough to allow the use of a moderately narrowed KrF laser, potentially boosting the light intensity in the wafer plane and obviating the sensitive line-narrowing optics in the laser. For shorter wavelengths and larger NAs, narrower bandwidths are required. More recent examples of catadioptric lenses without the involvement of a beamsplitter are shown in Fig. 5.26.

Table 5.9 Comparison of sensitivities of catadioptric and refractive 0.6-NA DUV lenses to changes in wavelength, temperature, and pressure.[120]

	Refractive lens	**Catadioptric lens**
Focus		
Temperature	−5.01 μm/°C	−0.04 μm/°C
Pressure	0.243 μm/mm Hg	<0.01 μm/mm Hg
Wavelength	0.24 μm/pm	<0.01 μm/pm
Magnification		
Temperature	16.5 ppm/°C	0.2 ppm/°C
Pressure	−0.8 ppm/mm Hg	<0.1 ppm/mm Hg
Wavelength	−0.8 ppm/pm	<0.1 ppm/pm

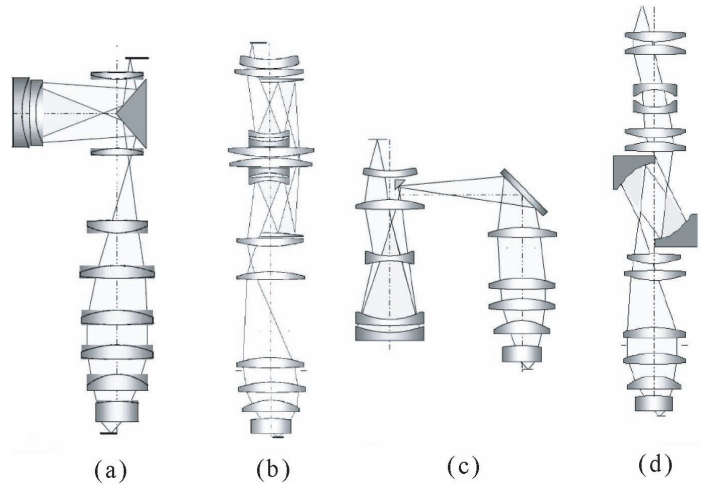


Figure 5.26 Lens designs considered by Nikon for $NA \geq 1.3$, designed to operate at a wavelength of 193 nm.[122] (a) Two-axis and three-mirror type, (b) uniaxis and four-mirror type, (c) three-axis and three-mirror type, and (d) uniaxis and two-aspheric-mirror type. Nikon selected design (a) for their S610C scanner. How numerical apertures >1 are achieved is discussed in Chapter 10.

Lens performance can fall short of the diffraction limit. Deviations from diffraction-limited performance—aberrations—were discussed in the previous chapter and can arise from a number of sources. Lens designs may be inadequate, and even when the lenses are designed nearly to perfection, fabrication will be imperfect. Important issues in lens fabrication are the purity, homogeneity, and spectral transmission of glass materials; the precision to which spherical and aspheric surfaces can be ground and polished; and the centration and spacing of elements.[123]

To begin, optical materials must be uniform in terms of refractive index and transmission. To meet these requirements, the lens materials, when being grown in a furnace, must be subjected to very gradual changes in temperature; otherwise, thermal gradients lead to variations in density. Induced stress also leads to birefringence. The necessary quality levels have resulted in prices for optical glass that have reached thousands of dollars per kilogram, while crystalline CaF_2 is much more expensive.

Once the individual lens elements are produced, they must be mounted properly. Critical machining and assembly are required to ensure the correct spacing between lens elements and to make certain that they share, as closely as possible, a common optical axis. Securing the elements to the lens housing, to ensure that they do not shift, often requires adhesives. It is important that the curing of these adhesives does not cause the lens elements to shift and does not induce stress.[124] Adhesives for lithography lenses must also have low levels of outgassing and must be stable to stray UV and DUV light.

For the lowest-aberration lenses, all sources of stress that can distort optical elements must be eliminated. Mounts for the lens elements have been carefully designed to minimize deformation.[125] A careful balance must be maintained while securing the lens elements so they do not move in an undesired way or induce undue stress.

Lens quality is assured through extensive metrology. Optical materials are measured for homogeneity in refractive index down to the parts-per-million level, and birefringence is measured as well. Transmission must also be uniform and greater than minimum required values. After polishing, the surfaces of individual lens elements are carefully measured and repolished until they meet specifications. Many lens manufacturers use phase-measuring interferometry (PMI) to determine the form of the emerging wavefront from the assembled lens, information that can then be compared to the diffraction-limited wavefront in order to measure the residual aberrations. The PMI was developed by Bruning and coworkers at AT&T Bell Laboratories.[126]

The need for reduced aberrations motivated stepper manufacturers to introduce the step-and-scan configuration in spite of its greater mechanical complexity. Because only a small part of the reticle is imaged at any one instant with a step-and-scan, a smaller part of the lens field is utilized. The part of the lens with the smallest aberrations, least distortion, best imaging, and tightest linewidth control is selected for use. The move to step-and-scan provides improvements in overlay (as a consequence of reduced lens-placement errors and improved ability to adjust for reticle registration errors) and better linewidth control, without requiring significant improvements in lens technology. In addition to being able to select the part of the field with fewest aberrations, there are improvements in across-field uniformity that result from the averaging effect of scanning. The placement of an object on the wafer is the result of the lens-placement errors averaged over the part of the lens through which it is imaged. Since lenses are judged by the worst placement errors, averaging always has an ameliorating effect. Step-and-scan also provides improved compensation for reticle errors, because magnification is adjusted independently in the directions parallel and perpendicular to the scan,[127,128] but other issues can arise when doing this, as will be discussed in Section 5.7. Similarly, critical dimension uniformity is paramount in lenses, and this is always improved by averaging. These are all general properties of step-and-scan, independent of exposure wavelength, but step-and-scan has appeared first on DUV systems for critical layer applications, although i-line step-and-scan systems are also available.[129]

This use of a restricted field in step-and-scan has important implications. Since only a rectangular strip of the lens field is used, the elements are rotated with respect to each other, or "clocked," to find the best image performance. The whole lens, in fact, can be clocked. This tends to make final assembly of the lens go more quickly and results in better performance. In addition, many aberrations are proportional to powers of the field diameter and grow steeply near the field edges, which is avoided when using the restricted field. Overall, the lens performance

is improved for both CD control and image placement when a restricted field is scanned rather than using the full image field.

Even though lenses are fabricated with extremely low levels of aberrations, it has proven necessary to have active controls on the lenses to maintain good imaging. As noted previously, lens heating and changes in barometric pressure require active compensation of focus and magnification. As imaging requirements have gotten tighter, it has proven necessary to compensate actively for other aberrations, such as spherical aberration.[130] This compensation is accomplished in several ways. In state-of-the-art scanner lenses, the position of a number of lens elements is controlled actively using actuators.[131] To address the problem of asymmetric lens heating due to the nonuniform diffraction patterns, Nikon has incorporated infrared lasers to heat the lenses in those portions of the lens not heated by the actinic light (Fig. 5.27).

Step-and-scan also provides cost-effective very large field sizes. Consider a lens with a 31.1-mm diameter. The largest square field that is imaged by such a lens is 22×22 mm. A very wide, but short, field can also be imaged (Fig. 5.28). In step-and-repeat systems, such fields are not very useful, since most integrated circuits do not fit in such a field. On the other hand, such a short imaging field can be used on step-and-scan machines, where the height of the printed field is obtained by scanning. Thus, the same 31.1-mm diameter lens that could produce only a 22×22-mm maximum square on a step-and-repeat system could print a 30-mm-wide field using an 8-mm-tall slit. Slit heights on commercially available systems are shown in Table 5.10.

The height of the field is determined by the scanning stages, lens reduction, and the size of the reticle. The stage technology exists for scanning up to 50 mm

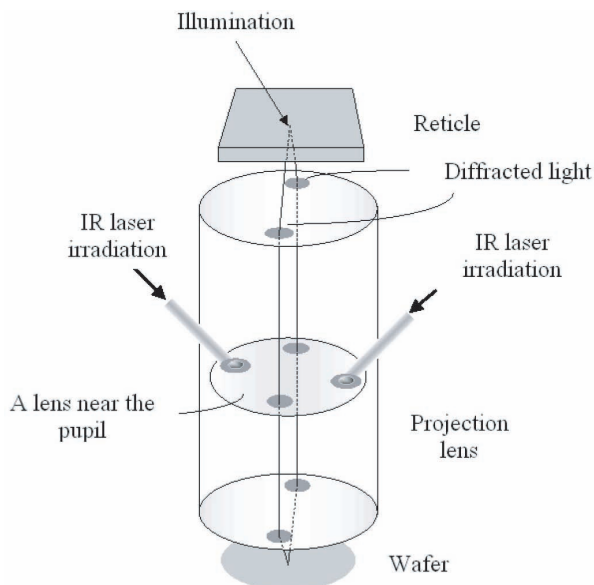


Figure 5.27 Illustration of Nikon's infrared aberration control system.[132]

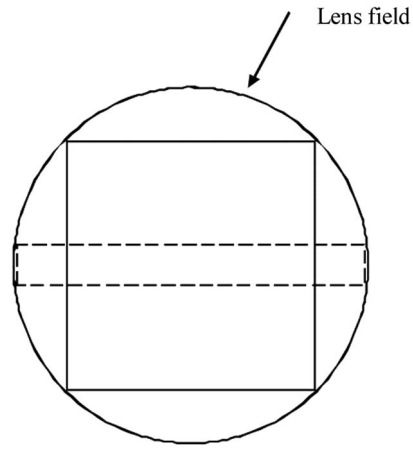


Figure 5.28 Lens field utilization for a lens with a circular image field. For step-and-repeat systems, the large square area would be used, while the wide and short area of the lens field would be used for step-and-scan machines.

Table 5.10 Slit heights of representative step-and-scan machines. All of these are KrF lenses, except for the Nikon NSR-307E, which is an ArF lens.

Stepper manufacturer	Model	Slit height (mm)
SVGL	Micrascan III[133]	5
Nikon	NSR-S204B	8
Nikon	NSR-307E	8
ASML	5500/500[134]	8
Canon	ES6[135]	7

on the wafer, and is considered to be extensible beyond that. With 4× reduction ratios, a 50-mm height requires reticles larger than 200 mm on a side. Field heights of 50 mm are not used in practice, because sufficiently large reticles are not readily available. Today's reticles are 152 × 152 mm, which results in a practical maximum field height of 33 mm for tools with 4× lens reduction. Field widths on step-and-scan systems are typically 25 or 26 mm. Expected chip sizes over the next half-decade were shown in Table 5.6. All of these chips can be accommodated in the 26 × 33-mm fields currently available.

In order to pattern large-area dies, a die-stitching method was reported by workers at Hitachi for making a 17.3 × 8.2-mm 16-Mbit DRAM with a first-generation i-line stepper with a 10-mm^2 lens field,[136] before wide-field lenses appeared. Stitching allows the use of well-corrected small-field lenses to image large die sizes. Stitching, however, is not an attractive option for volume manufacturing because of the need for two masks at every layer, overlay errors that occur at the stitching boundary, and sharply reduced throughput due to reticle exchanges needed for every wafer, among other reasons. However, stitching enables the fabrication of high-bit-count memories in early stages of development, without requiring expensive wide-field lenses.

For several device generations, lens manufacturers have been able to improve both NA and field size while maintaining diffraction-limited performance, as is illustrated in Tables 5.1 and 5.11. Figure 5.29 shows the steady progression of lens pixel count (the maximum exposure area divided by the area of a square that is one minimum feature size on a side) for 248-nm lenses from five manufacturers (Zeiss, Nikon, Canon, Tropel, and SVGL).

Increases in the pixel count, whether resulting from increases in numerical aperture or field size, generally require larger lenses. This was seen in Fig. 2.11. Over time, higher resolution and larger field sizes have been required, leading to increases in lens size. The trend in lens weight is shown in Fig. 5.30. To follow the pace required by the IC industry, lens manufacturers have had to use more glass elements and elements with ever greater diameters (~300-mm maximum). Wide-field i-line lenses were once considered expensive (~$500,000), but leading-edge KrF lenses are several times more expensive due to the cost of the optical-quality fused silica glass and more stringent requirements for surface figure and finish, as well as coating uniformity. ArF lenses are even more expensive than KrF ones. The fraction of stepper cost represented by the lens has been increasing, and today represents one-third or more of the total system cost. Also, as the size of the lens increases, the stepper body must add more dynamic compensation in order to hold the lens in a vibration-free environment. The cost of microlithographic lenses

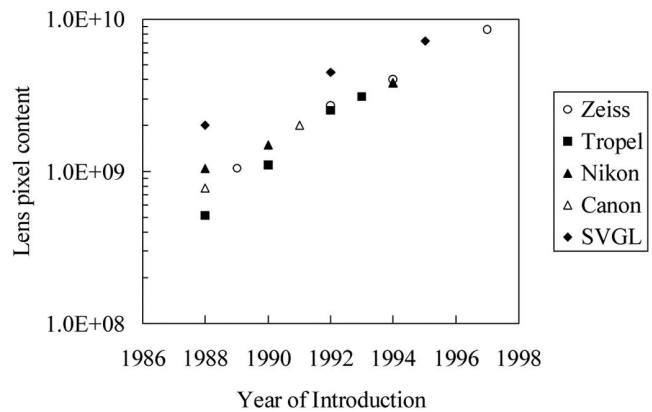


Figure 5.29 Progression of lens pixel counts versus time for 248-nm lenses.

Table 5.11 Microlithographic lenses for IC production.

Minimum features (μm)	NA	Wavelength	k_1	Depth-of-focus* (μm)	Pixel count
1.00	0.38	g line	0.80	3.02	2.4×10^8
0.35	0.60	i line	0.50	1.01	2.0×10^9
0.25	0.60	KrF	0.60	0.69	1.3×10^{10}
0.18	0.70	KrF	0.51	0.51	2.6×10^{10}
0.13	0.75	ArF	0.50	0.34	4.9×10^{10}
0.09	0.85	ArF	0.40	0.27	1.0×10^{11}

* $k_2 = 1.0$.

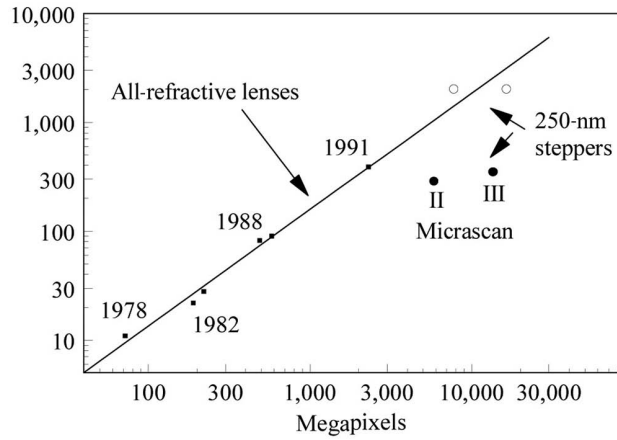


Figure 5.30 Progression of lens weight (lbs) as a function of pixel count.

continues to increase sharply, even with such changes in paradigm as step-and-scan and catadioptric designs. However, without these innovations, optical lithography might well have stopped being a cost-effective manufacturing technology long ago.

Two lenses are clearly off the trend line for refractive lenses, as shown in Fig. 5.30. These are the catadioptric Micrascan II and III lenses. In addition to other advantages, the catadioptric lenses are capable of being built with less optical glass than all-refractive lenses. These catadioptric lenses contain only about a dozen elements, compared to 20 or more for the all-refractive stepper lens, which makes them both less massive and less costly to build.

With all of the advantages of catadioptric lenses it may appear puzzling that all stepper manufacturers did not make the transition to such optics a long time ago. This transition was delayed for several reasons. First, there is considerable value in continuing to use technology in which one has the greatest expertise, and most makers of lenses for use in photolithography have substantial expertise in designing and building refractive lenses. Also, while the overall amount of glass is reduced considerably with the catadioptric lens design shown in Fig. 5.25, relative to all-refractive lenses, the beamsplitter requires a substantial quantity of glass. This means that problems induced by optical glass inhomogeneity and damage from excimer light sources are not entirely eliminated with that particular catadioptric design. Beamsplitter technology involves special expertise, particularly related to coatings. Finally, the particular design shown in Fig. 5.25 was patented by SVGL, and their competitors needed either to license this design or to find new approaches to catadioptric lenses. The latter has now happened.

The transition to catadioptric lenses gained momentum when 157-nm lithography was being considered. There was only one optical material, CaF_2, available in adequate size and quality for making lenses, and initial lenses for 157-nm lithography were constructed using CaF_2 as the only refractive material.[137–140] In this case, catadioptric lenses appeared to be a practical necessity. A similar situation occurred with immersion lithography, which will be discussed in more

detail in Chapter 10, where catadioptric designs provided the only practical solutions for $NA > 1.1$. Having gained experience with catadioptric lenses for immersion lithographic applications, ASML introduced a 0.93-NA KrF exposure system, the model XT:1000H, which incorporates a catadioptric lens to reduce the number of elements in the lens and to lower costs.[141]

Aspheric lens elements have also been introduced into microlithographic lenses, and these have enabled a reduction in the size and weight of stepper lenses. When a concave surface is moved against a convex surface in a random fashion, spherical surfaces are naturally produced by abrasion (see Fig. 5.31). This forms the basis for standard lens-fabrication methods, which produce lens elements with surfaces that are spherical.[142] Such spherical elements necessarily introduce aberrations that must be compensated by additional elements. The introduction of aspherical elements enabled the number and size of lens elements to be reduced (see Fig. 5.32 in Color Plates). However, it is necessary to use polishing tools that cover only a small area of the lens element in order to create aspheric surfaces, and this can easily lead to surface roughness over short distances, even though the general figure of the lens element is close to the desired shape. This short-distance roughness can lead to flare. Aspheric lens technology has many potential benefits, particularly reducing the number of lens elements, but extra care is required to maintain low flare with lenses that incorporate a large number of aspheric elements.

Finally, it should be pointed out that resolution is independent of lens reduction, a factor that does not appear in any of the equations for resolution. In properly designed microscopes there is a natural relationship between resolution and magnification, so the two tend to be associated. The historical transition from full-wafer scanning systems to higher-resolution steppers involved an increase in the reduction ratio of the optics, but this increase was not required to achieve the higher resolution. In lithographic equipment, resolution and magnification are independent quantities. Very high resolution 1× magnification systems have been built.[143]

5.5 Autofocus Systems

As discussed in Chapter 2, depths-of-focus are very small, typically less than ±0.1 μm, often much less. It is essential that the photoresist film be placed at

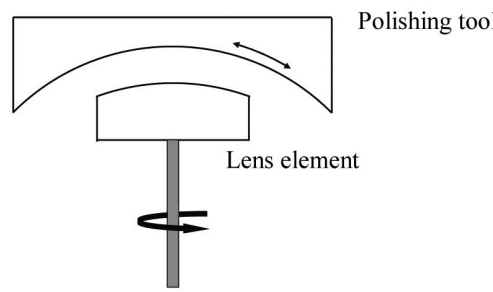


Figure 5.31 Polishing of lens elements, leading to spherical surfaces.

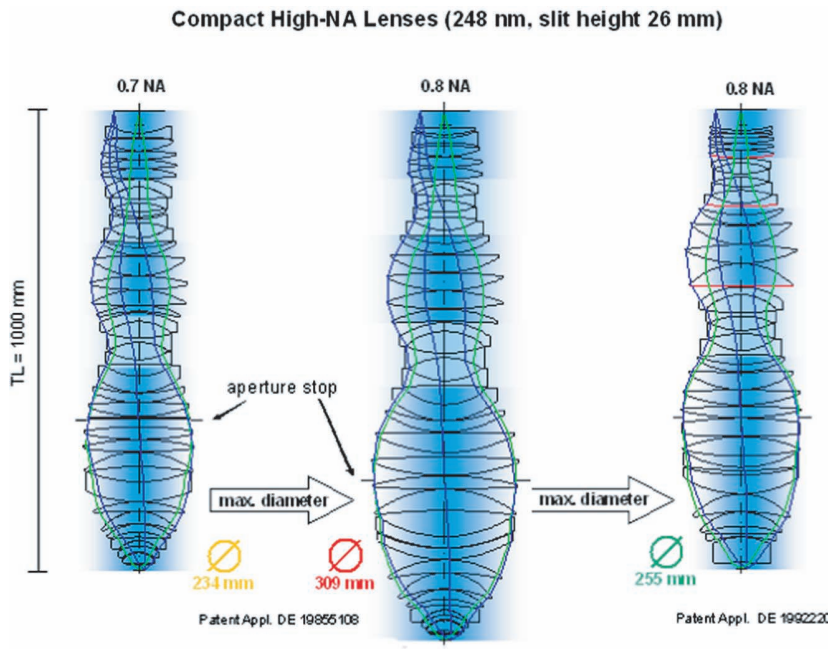


Figure 5.32 The benefit of aspherical lens elements (see Color Plates).

the correct distance from the lens, where good imaging occurs. Because wafer thickness varies by considerably more than the depth-of-focus, direct detection of the top surface of the wafer is a practical necessity, in conjunction with servo mechanisms and appropriate feedback systems. The separation between the lens and the wafer is measured prior to each exposure or during each scan, and the wafer height is adjusted with the objective of maintaining a constant lens-to-wafer distance. (On some older steppers it was the lens height that was changed.)

Because of concerns over defects, the detection of the top surface must be performed without mechanical contact with the wafer. Several methods have been devised for detecting the wafer surface, and the basic concepts behind these methods are now discussed. Three methods for focusing wafers are used on steppers: optical, capacitance, and pressure.

The optical method, illustrated in Fig. 5.33(a), is the technique used most commonly for focusing.[144–147] In this approach, light with a glancing angle of incidence is focused onto the substrate. The reflected light hits the detector at different positions, depending upon the vertical position of the substrate. In Fig. 5.33(a), light reflecting off a wafer at the position indicated by the solid line hits the detector at a different position than light that reflects from a wafer at the position indicated by the dashed line. Detectors that distinguish the position at which the light strikes the detector can be used to measure focus. The degree to which the optical method is insensitive to substrate film type and topography is critically dependent upon the system design.[148] For example, the optical system may detect the strongest reflecting surface, and the position of the resist film will be

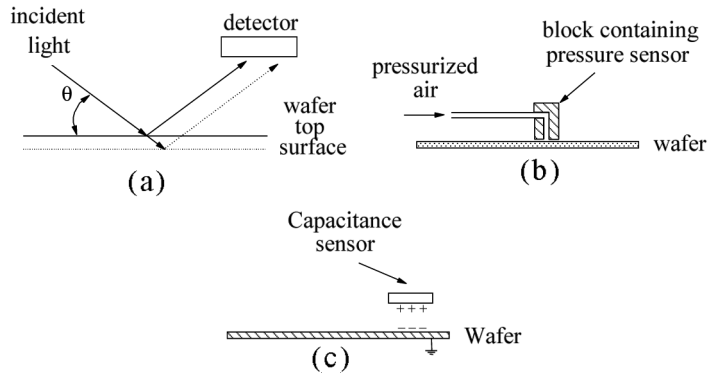


Figure 5.33 Schematic representations of different methods for focusing exposure tools. (a) Optical focusing, (b) pressure sensors for focusing, and (c) capacitance gauges for focusing.

found at a different distance from the lens if it is sitting on top of a thick oxide film rather than a bare silicon substrate. Process engineers must empirically determine if their combination of steppers and substrates creates such a sensitivity. Substrates comprised of metal and thick oxide films, as typically enountered in the back end of wafer processing, are most commonly susceptible to focus errors.[149] Since metal films are highly reflective, this is not too surprising. Minimizing the angle θ and using multiple wavelengths provides the least sensitivity to films on the wafer surface.[150,151] Autofocus errors are also reduced when the plane containing the incident and reflected autofocus beams is at 45 deg relative to substrate features.[152]

Two other methods of locating the wafer surface are also shown in Fig. 5.33. In one method, pressurized air is forced into a block that contains a pressure sensor.[153] The pressure in the block depends upon the gap between the block and the wafer surface. By mounting such blocks on the sides of lenses, the separations between lenses and wafers are measured. This method is sensitive only to the top surface of the resist film and is independent of substrate composition. It does require calibration to the ambient barometric pressure, and this type of focus sensor and special effort is required to measure the height of the wafer surface directly below the lens during exposure, since the sensor would interfere with the imaging.[154] Such air gauge systems have been used as the primary autofocus systems on exposure tools, as well as for calibrating optical focus measurement systems.[155] In one assessment, focus errors on product wafers were reduced by 20–30 nm (3σ) when the optical focus system was calibrated with an air gauge.[156]

Capacitance sensors have also been used to measure the lens-to-wafer distance.[157] A capacitor is created where one plate is the wafer [Fig. 5.33(c)]. Capacitance C is given by

$$C = \frac{\varepsilon A}{d}, \tag{5.13}$$

where A is the area of the capacitor, d is the separation between the plates, and ε is the dielectric constant of the material between the plates. The separation between the lens and wafer is determined by a measurement of capacitance because of its dependence on d. Like the optical method, this technique is sensitive to the composition of the films on the silicon wafers because of the dependence of the capacitance on ε.

Since it is necessary to have the entire exposed area at the correct distance from the lens, the tilt of the wafer surface must also be measured and corrected. This can be accomplished in a number of ways. The focus systems described above can be used to measure exposure fields at multiple positions, providing the data required to level the exposure field. In another approach, parallel beams of light are reflected from the surface of the wafer. Their angle of reflection is a measure of the local surface tilt.

5.6 The Wafer Stage

The final key stepper subsystem discussed in this chapter is the wafer stage. The ability to move wafers significant distances (up to several hundred millimeters) with extreme precision is central to the stepper concept. The position of the wafer stage is measured with great accuracy using a modern variant of the Michelson interferometer, shown schematically in Fig. 5.34. In this type of interferometer, light is generated by a HeNe laser. By placing the laser cavity in an axial magnetic field, lasing occurs at two wavelengths, separated in frequency by a few MHz about the central frequency of 4.736×10^{14} Hz. Moreover, the light is circularly polarized and has opposite polarization at the different wavelengths. A quarter-wave plate converts the circularly polarized beams into orthogonal, linearly polarized beams. The maximum frequency difference that can be achieved practically using a magnetic field is 4 MHz. Alternatively, acousto-optic modulators can be used to achieve a frequency split as large as 20 MHz.[158] The resulting laser light passes through a polarization-sensitive beamsplitter. Light with one polarization

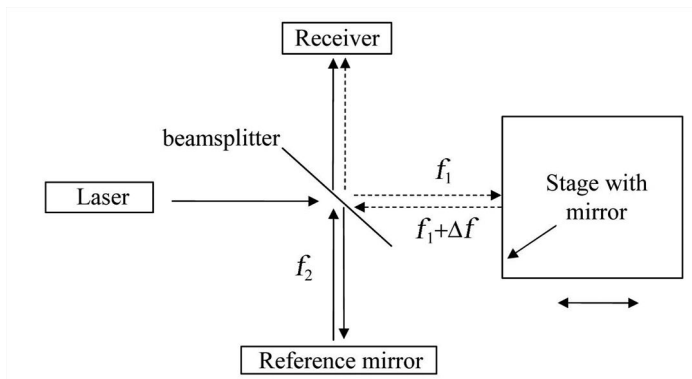


Figure 5.34 The Michelson laser interferometer used to measure the X position of the stage.

and frequency f_1 is reflected off a mirror mounted on the wafer stage, while the other beam with frequency f_2 is reflected to a fixed reference mirror. Because the wafer stage may be moving, the frequency f_1 will be Doppler shifted by an amount Δf, given by[159]

$$\Delta f = 2f_1\frac{v}{c} = 2\frac{v}{\lambda_1}, \tag{5.14}$$

where v is the velocity of the stage, c is the speed of light, and λ_1 is the wavelength of the light incident on the moving mirror.

After reflections, the two beams of light are recombined at a heterodyne receiver capable of measuring the difference frequency $f_2 - f_1 \pm \Delta f$. Maximum measurable velocity is set by the requirement that $\Delta f < f_2 - f_1$. Comparing this to a reference measurement of $f_2 - f_1$ enables a determination of the stage velocity through Eq. (5.14). By integration of the velocity, the stage position is determined.[160–162] Systems today typically use interferometers that potentially measure position with $\lambda/1024 \approx 0.6$-nm precision.

Although the interferometer is capable of measuring a position that is a small fraction of the wavelength of the laser light, the stage precision has always fallen short of such capability. The best current stages can step a Cartesian grid accurately to about 5 nm and are repeatable to about 2 nm (3σ), far less precise than $\lambda/1024 = 0.6$ nm. A simple way to measure stage precision is to use pattern structures to measure overlay on the wafer. Typical types of such structures are discussed in Chapter 9. The first part of the structure is stepped across the wafer, and then the second part of the structure is stepped, so that the two structures are concentric if the stepping is perfectly precise. The discrepancy between interferometer precision and actual stage performance is a consequence of changes in the air through which the interferometer beams travel that effectively changes the wavelength of the light.

The effective wavelength of the light in air can vary because of environmental changes (Table 5.12). Since the density of air changes nonuniformly along the path of interferometer beams, the phase of the light is shifted, resulting in measurement errors. These fluctuations in density result primarily from localized sources of heat or motion. Interferometer errors due to temperature- or pressure-induced fluctuations in the optical path length of the air surrounding the wafer stage are reduced by shielding the path from air currents and thermal gradients. Stepper designers go to great lengths to maintain temperature uniformity throughout the environmental enclosure of the steppers, in order to avoid air turbulence induced

Table 5.12 Changes in environmental parameters that will result in a one part-per-million change in the wavelength of light in air.[163]

Air temperature	1 °C
Pressure	2.5-mm Hg
Relative humidity	80%

by thermal gradients. All sources of heat that can be remotely operated are removed from the enclosure. Heat-generating parts that must be in the enclosure, such as stage motors, are covered with temperature-controlled jackets.[164] However, since the stage must be able to move rapidly and freely within a large area, it is impossible to shield the entire path length. Lis has described an air turbulence–compensated interferometer that could potentially reduce the error due to environmental fluctuations,[165] but requirements appear to have exceeded the improvements that such a system could bring.

In order to meet the very tight overlay requirements of sub-40-nm half-pitch processes, it is necessary to fundamentally address the problem of air turbulence and reduce the length of the path through air in which the light of the metrology system travels. The latest generation of exposure tools makes use of encoders to decrease such path lengths substantially (to $\leq$2 mm), thereby reducing turbulence control and compensation requirements. Encoders make use of gratings, from which light beams are diffracted, as illustrated in Fig. 5.35, where the grating is placed on the wafer stage. Alternatively, the grating could be above the stage, with the laser and detection electronics incorporated into the stage itself.

With encoders, the stage position is measured by having the $+1^{st}$- and -1^{st}-order diffracted beams recombined.[166,167] As the stage moves the resulting fringe pattern also shifts (see Fig. 5.36). This fringe pattern can be measured with precisions of 1 nm or better.[168] Unfortunately, it is not yet possible to produce gratings that are very accurate over long distances, so exposure tools that use encoders must still have interferometers for calibrating the encoders and ensuring orthogonality between the X and Y stage motions. As an historical note, an early commercially available wafer stepper from a company called Optimetrix used an encoder to measure stage position, but did not also have a laser interferometer. Lack of stage accuracy contributed to Optimetrix's inability to gain substantial market share.

Wafer stages need to move in more than one direction, so there are interferometer beams to measure motion in the Y axis as well as the X axis. More

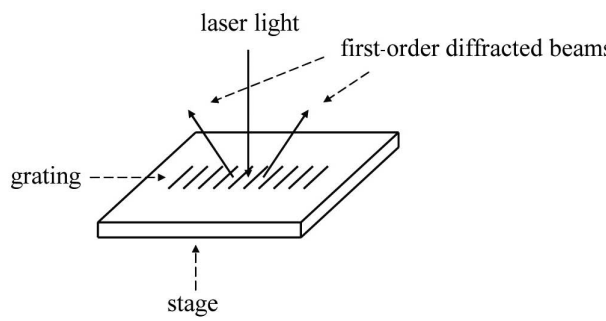


Figure 5.35 Illustration of the encoder concept used to measure stage position. Above the grating is a module that contains the laser, as well as optics and electronics for detecting the diffracted beams. The effects of air turbulence are minimized by making the distance between this module and the grating short. Two-dimensional gratings can be used to provide positional information in both the X and Y directions.

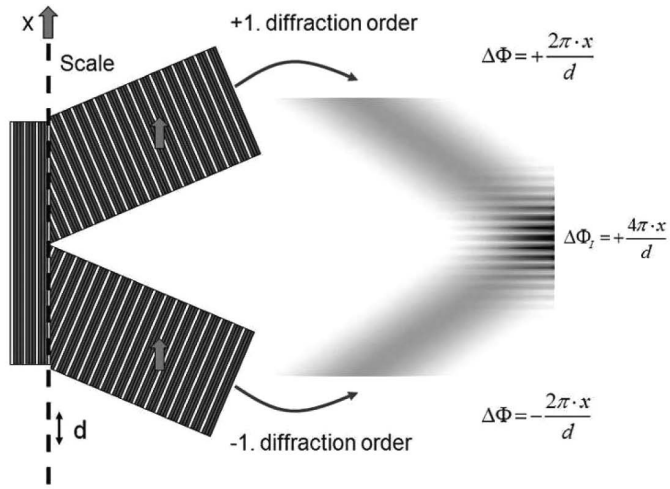


Figure 5.36 Illustration of phase differences between the $+1^{st}$- and -1^{st}-order diffracted beams as the stage moves. These phase differences lead to changes in the fringe pattern produced by recombining the diffracted beams, and these changes are used to measure stage position.[169]

than one beam in the *X* or *Y* axis enables measurement of stage yaw or rotation[170] (Fig. 5.37). This is important because stages can rotate, as well as undergo rectilinear motion. Modern stages have as many as six or more interferometer beams to control rotation about the *X*, *Y*, and *Z* axes. This is important to minimize the impact of *Abbe errors* that result from stage rotations, as illustrated in Fig. 5.38. A single interferometer beam measures the distance traveled by the point on the stage mirror where the beam is reflected, not how far the wafer center has moved. These can be different because of stage rotation.

Not only must the stage position be measurable to high precision, it must be controllable. This is accomplished by means of precision motors, bearings,

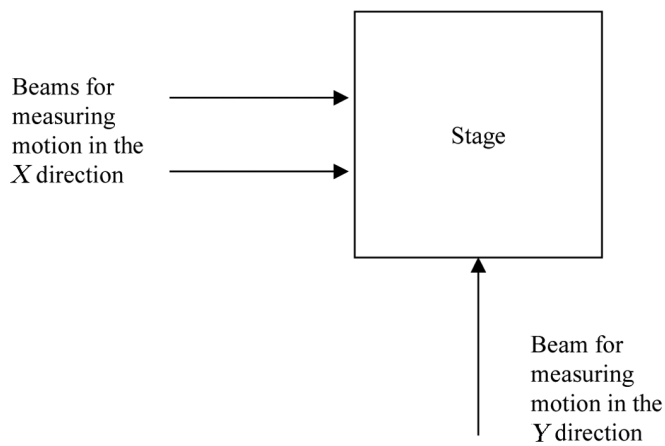


Figure 5.37 Multiple beams for measuring yaw.

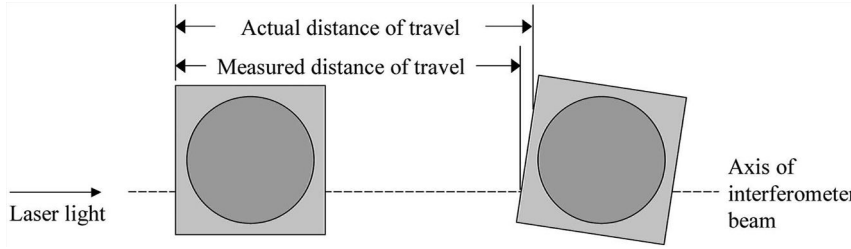


Figure 5.38 Illustration of Abbe errors.[163]

and feedback loops. In order to provide smooth motion, the wafer stage must slide on some type of low-friction bearing. The original GCA system had Rulon feet that slid on highly polished metal surfaces. Oil was applied periodically to maintain smooth motion. Other stages have incorporated roller, air,[171] or magnetic bearings.[172,173] Air bearings, riding on ultraflat granite or steel surfaces, are commonly used.[174]

Even though near-perfect stage precision might be obtained, stage motion still may not be accurate for a number of reasons. One reason might be cosine error, illustrated in Fig. 5.39. In this situation, the stage travel is not parallel to the interferometer beam. The result is an error proportional to the cosine of the angle between the line of stage travel and the interferometer beam, i.e., the path of the interferometer beams is $2d/\cos\theta$ rather than $2d$, where d is the distance that the stage actually travels. Cosine errors effectively change the scale by which distance is measured by a factor of $\cos\theta$. Systematic position errors due to deviations from ideal flatness of the stage mirrors, at least relative to a reference, can be characterized and compensated for using software.

In order to achieve high stepper throughput, the stages must be capable of moving at high velocity while maintaining good control of position. Acceleration and jerk (the third time derivative of position) capability must also be high (see Table 5.13). This is difficult to achieve because stages have become massive in order to accommodate large wafers, be robust against vibrations, and remain stiff, yet have the light mass needed for high acceleration. Since the imaging performance of the stepper can be severely affected by vibrations,[175] the rapid movements of the wafer stage must be effectively decoupled from the optical

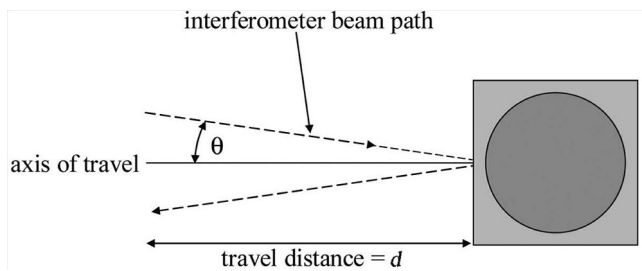


Figure 5.39 Illustration of cosine error.

Table 5.13 Capability of high performance stages. These values are those available in the open literature, but they likely do not represent today's state-of-the-art for acceleration and jerk, which are probably much larger. During operations that do not require great precision, such as wafer loading and unloading, the wafer stage speed is usually greater than the maximum speed during exposure scanning.[176]

Wafer stage speed (scanning)[177]	700 mm/sec
Acceleration[176]	0.5G = 4.9 m/sec^2
Jerk[178]	500 m/sec^3

column through pneumatic isolation and active damping mechanisms. Since the reticle stage must move 4× faster than the wafer stage (in systems with 4× lens reduction), there are even greater challenges for accelerating and scanning the reticle. The difficulties involved with scanning the reticle become greater with increased lens reduction, providing additional motivations for maintaining moderate values for lens reduction.

Vibrational analysis is typically accomplished by finite-element analysis in the design of new wafer-stepper bodies to find and minimize natural resonances at the low frequencies (0 to 200 Hz) that can blur projected images. High-frequency vibrations are naturally suppressed due to the large mass of wafer steppers. Vibrations need to be controlled in order to maintain good imaging and overlay.

With the advent of scanning reticle and wafer stages, control of vibrations assumes an even larger role. The mechanical problem of scanning a reticle stage at four or five times the velocity of a wafer stage moving in the opposite direction, while keeping the positions of the two stages synchronized to an error of less than 10 nm, requires the highest degree of active vibration control available. Scanner throughput is ultimately limited by the stepper's overlay error at the highest scan rates.[179] The problem of measuring the positions of two stages moving in opposite directions at speeds that are several hundreds of millimeters per second and synchronizing their positions to within a few nanometers so that overlay rules are met is a significant challenge to modern interferometry, digital computing, and mechanical control systems.

One of the potential problems of scanning is the shift in the center of mass of the exposure tool during the scan. This is particularly significant when the reticle stage is scanned, since it moves four times as fast and as far as the wafer stage (assuming 4× lens reduction). With a large shift in the center of mass, forces generated during the reticle scanning can be transmitted mechanically to the wafer stage, and from the wafer stage back to the reticle stage. These mechanical interactions can reduce scanning accuracy significantly. To reduce these errors, balance masses are used. While the reticle is being scanned in one direction, the balance mass is scanned in the opposite direction, thereby maintaining a fairly fixed center of mass.[180] A similar balance mass can be used for the wafer stage.

5.7 Scanning

In a step-and-scan exposure tool, it is necessary that the reticle stage scanning be synchronized with the wafer stage scanning. The optical image of a feature k may be misplaced because of imperfect scanning or synchronization between the reticle and wafer stages. If the placement error in the x direction at time t_i is $\Delta x(t_i)$, then the resulting feature in resist will be misplaced by an amount

$$\overline{\Delta x_k} = \sum_i \Delta x_k\,(t_i)\,, \tag{5.15}$$

where the sum is over all times during which the particular feature is being imaged through the slit of lens. For a scanner with an excimer laser light source, each time t_i will correspond to a laser pulse. For scanners using arc-lamp light sources there are analogous equations involving integrals. It will be assumed for the remainder of the discussion in this section that the light source is pulsed. Since not all position errors $\Delta x(t_i)$ will be identical there will also be variations in the positions of the optical images of a specific feature k about its average position:

$$\sigma_k = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(\Delta x_k\,(t_i) - \overline{\Delta x_k}\right)^2}, \tag{5.16}$$

where there are N pulses during the scan. There are usually specifications for the quantities given in Eqs. (5.15) and (5.16) for a particular model of scanner, and statistics of the sets of values $\left\{\overline{\Delta x_k}\right\}$ and $\{\sigma_k\}$. The specified tool *moving average* (MA) is usually given as the mean $+3\sigma$ for the set $\left\{\overline{\Delta x_k}\right\}$, while the *moving standard deviation* (MSD) is the mean $+3\sigma$ for the set $\{\sigma_k\}$.

A nonzero MA implies that there will be pattern placement errors, contributing to imperfect overlay (discussed in the next chapter), while the MSD will cause images to be blurred (see Fig. 5.40). This blurring has been called image "fading."[134] For advanced exposure tools, $MA < 1$, while $MSD < 3$ nm, and perhaps even below 2 nm.[181–183]

There are also systematic errors that can cause image fading. For example, fading occurs if there is a mismatch between the lens reduction and the scanning speed. For an explanation of this, suppose the reticle is moving at a speed v_R and the wafer is moving at a speed v_W. During a scan of time t, the aerial image moves a distance $v_R t/N$ at the wafer plane, where N is the lens reduction. During this same period of time, a point on the wafer moves a distance $v_W t$; therefore the amount of blur caused during the scan is

$$\left(\frac{v_R}{N} - v_W\right)t. \tag{5.17}$$

As shown in the next chapter, the quantity in brackets is nonzero when there is an asymmetric magnification or skew error. Thus, there is systematic image fading when there is an asymmetric overlay error.

Image fading leads to a number of consequences.[184] Random scanning errors (caused by MSD) lead to general image blur and degradation of the image log slope. Line-end shortening can also result, as well as a change in the bias between isolated and dense features (Fig. 5.41). Image fading due to asymmetric magnification errors will affect lines that are perpendicular to the scan, while not having an effect on lines that are parallel to the scan. As one can see from

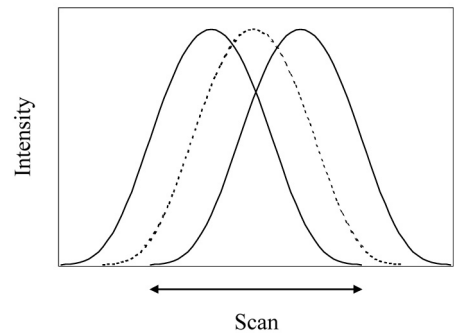


Figure 5.40 Image fading from imperfect synchronization between the reticle and wafer-stage scanning that leads to the aerial image not staying at a fixed position on the wafer during the scan. The result is image blurring.

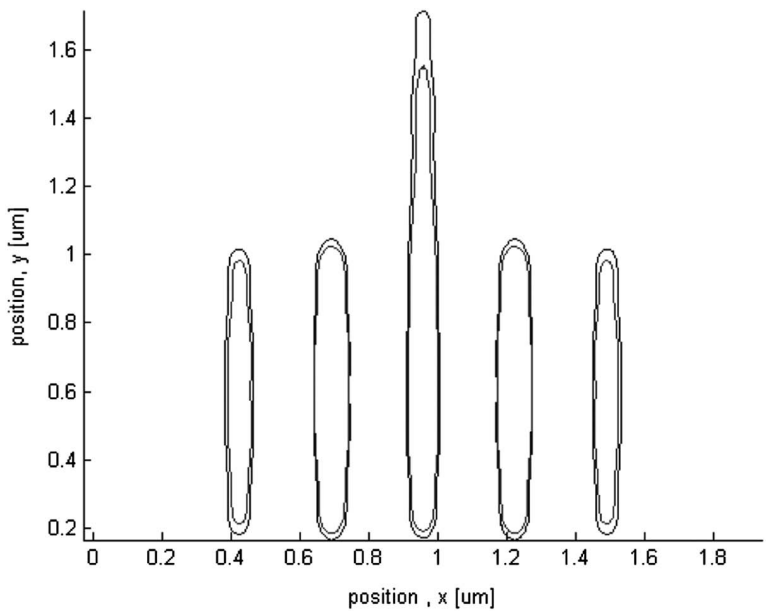


Figure 5.41 Simulated image contours for 100-nm lines and spaces, with no random stage synchronization errors and with MSD in both the X and Y directions, having random Gaussian distribution, with $\sigma = 25$ nm. The parameters for the imaging are a wavelength of 193 nm, $NA = 0.8$, and quadrupole illumination. Note that line shortening is affected more than linewidths, and lines are shortened more with fading.

Fig. 5.40, image fading will cause an increase in linewidth. Since asymmetric magnification errors affect only lines parallel to the scan, such errors lead to a bias in linewidth between horizontal and vertical lines.[185] Thus, one should be cautious when applying asymmetric overlay corrections, as might be done to correct for such errors as reticles. Because linewidth bias varies quadratically with fading (Fig. 5.42), small asymmetric field terms can be employed without affecting linewidths significantly.

5.8 Dual-Stage Exposure Tools

The transition from 200-mm wafers to 300-mm wafers brought the prospect of a significant decrease in exposure throughput, as measured in wafers per hour, since many more scans of the wafer are required for larger wafers. (For further details on exposure-tool throughput, see Chapter 11.) This also occurred at a time when the projection optics of the exposure systems were becoming an increasing fraction of the system cost. That meant that there was an increasing economic penalty for the time when the lens was not being used, i.e., when the system was aligning wafers or mechanically moving wafers on and off the chuck, instead of exposing. One solution to this problem was to have two stages.[186]

In a dual-stage exposure tool, one stage is used for loading and unloading wafers, and for metrology, such as alignment and focusing. Once these operations are completed, the stage is moved under the lens, where the wafer is then exposed. While this exposure is taking place, a new wafer is loaded on the other stage (after unloading a wafer that may already be on that stage), and the process is repeated. In an alternative implementation, only the wafer chucks are swapped, rather than the complete stages. With either approach, dual-stage exposure tools have the potential for very high throughput.

In addition to providing greater throughput, the dual-stage systems have several other potential advantages. One of these is that extensive wafer metrology

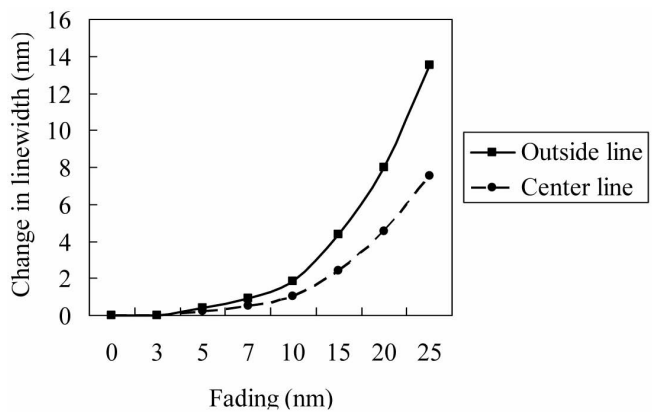


Figure 5.42 Change in linewidth as a function of image fading for lines in a 5-bar structure. The measure for the abscissa is the amount of fading across the scan due to asymmetric magnification [Eq. (5.17)].[184]

(alignment and focus) can be performed without loss of throughput, at least up to a point, because the measurements can be performed without detracting from wafer exposure time, thus compromising productivity. In the first dual-stage exposure tool that was introduced, the Twinscan platform from ASML, the wafer topography is mapped in detail. A z-axis interferometer is then used during exposure to control the focusing. This has the potential for much better focus control.[186]

Nikon has introduced exposure tools that also have two stages, but only one of them is a wafer stage—the other is for metrology. In order to maintain optimum tool control it is necessary for the exposure tools to perform self-calibration periodically. Beginning with their model S609, Nikon has used a separate "tandem stage" that performs this scanner self-metrology during the times when wafers are being loaded onto or unloaded from the wafer stage.[187] An illustration of a tandem stage is shown in Fig. 5.43. Regardless of the dual-stage approach adopted, significant improvements in scanner throughput have been achieved by both Nikon and ASML.

Dual-stage exposure tools are obviously much more complex than their single-stage counterparts. With such complexity comes the potential for reduced reliability and greater tool purchase price. Nevertheless, because of their potential for higher throughput, such systems can be cost competitive. The relationship between throughput and the lithography costs will be discussed further in Chapter 11.

5.9 Lithography Exposure Tools before Steppers

The first type of lithography used for the fabrication of integrated circuits was contact printing. This was a simple technique, in which the mask was brought into direct contact with the resist, as illustrated in Fig. 5.44. Contact printers were very simple and inexpensive tools, and good resolution could be achieved, because the Fresnel-diffracted light just below the mask produced features that were very close to the sizes of the features on the mask (see Fig. 5.45). However, defects were a very serious problem with contact printing. Dirt and particles on the mask would

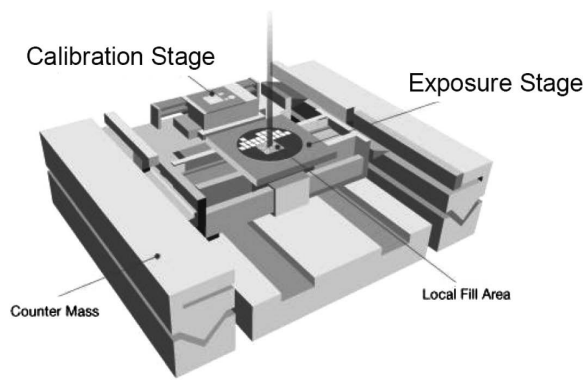


Figure 5.43 Illustration of Nikon's tandem stage.[187]

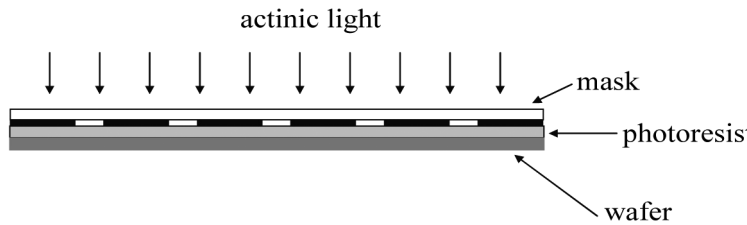


Figure 5.44 Illustration of contact printing.

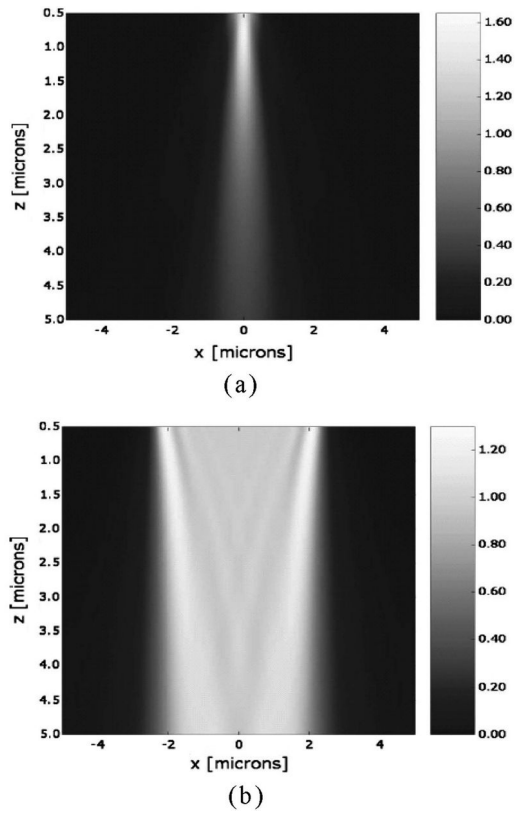


Figure 5.45 Calculated diffraction pattern just below a mask surface. Partially coherent light with the wavelengths of 365 nm, 405 nm, and 436 nm, with a cone angle of incidence of 8 deg was assumed. (a) 1-μm mask feature and (b) 5-μm mask feature.[188]

cause defects on the wafers. After each wafer printing the mask would pick up additional particles from the wafer, or get scratched, and the mask would then need to be replaced after being used to expose only a few wafers, in order to maintain yield.

In order to reduce mask costs and improve yields, proximity printing was introduced, where the mask was positioned a few tens of microns above the resist surface (Fig. 5.46). By not coming directly into contact with the wafer, the

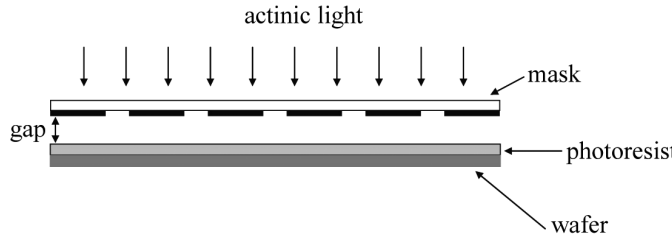


Figure 5.46 Illustration of proximity printing.

problems of wafer defects and mask degradation were both addressed. However, having a gap between the resist and the mask reduced resolution, as can be seen in Fig. 5.45. As the distance from the mask is increased, the light distribution broadens due to diffraction, and edge acuity degrades. A useful expression that estimates the resolution of proximity printing as a function of the gap is[189]

$$\text{resolution} = k_P \sqrt{\lambda\left(g + \frac{T}{2}\right)}, \tag{5.18}$$

where λ is the wavelength of the light, g is the gap between the mask and wafer, T is the resist thickness, and k_P is a parameter analogous to k_1 in optical-projection lithography. Resolution versus the size of the gap is shown in Fig. 5.47.

A schematic of a commercially available proximity and contact printer is shown in Fig. 5.48. Such systems can provide both proximity and contact printing. Automatic wafer handling, alignment capability, and gap control are also available on contemporary proximity printers.

Because of the defects associated with contact printing and the resolution limits and difficulties of maintaining linewidth control with proximity printing, optical

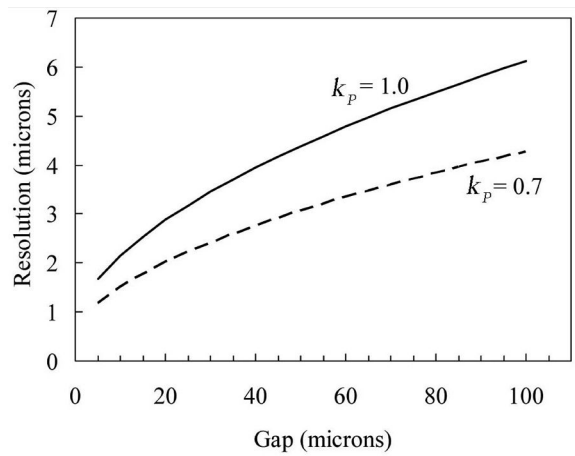


Figure 5.47 Resolution versus gap for proximity printing, assuming a wavelength of 0.365 μm.

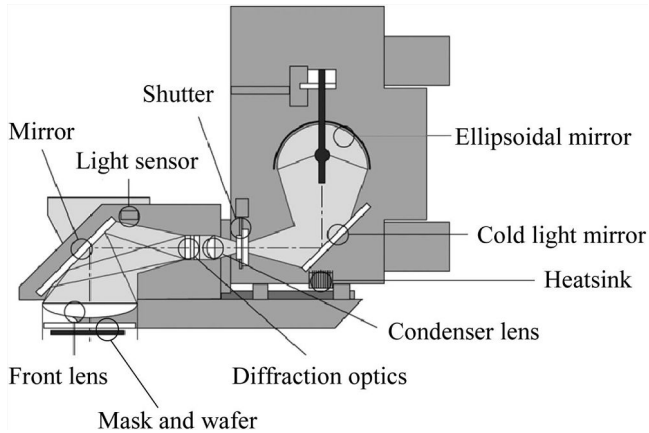


Figure 5.48 Exposure units of a SUSS MicroTec Mask Aligner. For this system, resolution is specified as <3 μm for proximity printing with a gap of 20 μm and <0.8 μm with vacuum-assisted contact printing.[190]

projection was pursued as an alternative to address both of these problems. The result was the Micralign series of full-wafer scanners, built by Perkin-Elmer (see Fig. 5.49).[191] These scanners differed from contemporary step-and-scan systems in several fundamental ways. First, the entire wafer was exposed during a single scan. Since the wafers were only 3 in. (76.2 mm) or smaller at that time, the required imaging area was much smaller than would be necessary today to cover an entire wafer. Second, the imaging was 1×. By engineering the optics properly, this enabled the use of existing proximity-printing masks in the Micraligns.

Another key ingredient in the Micralign design was an all-reflecting, ring-field lens.[192] The use of mirrors to form images enabled the use of broadband light sources, reducing issues associated with standing waves. Ring-field optics,

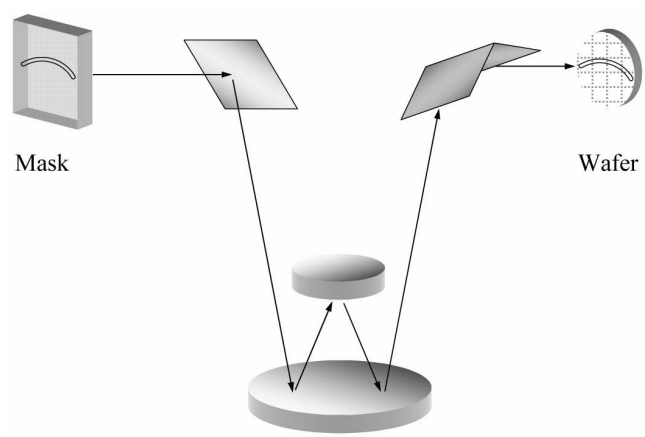


Figure 5.49 Schematic of the optics of a Perkin-Elmer scanner. The mask and wafer are scanned together.

in which the imaging field is a curved slit, enabled a significant reduction in aberrations without introducing a large number of optical elements. The numerical aperture of the Micraligns was 0.167. Although this seems like a low numerical aperture by contemporary standards, with a wavelength of 365 nm it had a Rayleigh resolution of 1.3 μm and a practical resolution of ~2 μm, which was quite adequate at the time the Micralign was introduced. Since the wafer and mask were scanning together with the same mechanical mechanism, the specifications for the mechanical scanning mechanism were based on the requirements of linewidth uniformity, not overlay, and were much looser than the requirements of wafer steppers. Later, Canon also introduced full-wafer scanners using similar optics.[193] In the late 1970s and early 1980s, full-wafer scanners were the dominant tools of choice for producing leading-edge devices.

Over time the Micralign systems went through a series of improvements. As wafer sizes increased from 76.2 mm to 150 mm, the Micralign field sizes were increased accordingly in the Models 500 and 600. Imaging could be changed slightly from 1:1 printing to enable improvement in overlay by providing capability for compensation of wafer expansion and contraction during processing. The wavelength range was extended to DUV wavelengths, enabling printing ~1 μm.

The full-wafer scanners had a number of advantages over early wafer steppers.[194] The scanner's broadband illumination resulted in significantly less difficulty with standing waves and associated issues, particularly when patterning aluminum metal layers. Throughput was also significantly higher on the scanners. Over time, the scanners lost all advantages to steppers. Advances in scanner optics were required just to provide capability for larger wafers, leaving little potential for increasing numerical aperture and resolution. For submicron lithography on larger wafers, the smaller imaging fields of the wafer steppers became a necessity. By the mid-1980s there was wide-spread use of wafer steppers for leading-edge lithography. Perkin-Elmer sold their exposure tool business to Silicon Valley Group (SVG), and SVG was later acquired by ASML.

Contact printing, proximity printing, and full-wafer scanners are still in use for making micromechanical devices, packaging, and components of various types, even if not used for producing leading-edge integrated circuits. Suppliers of proximity printers include EV Group of St. Florian am Inn, Austria, and SUSS MicroTec of Germany. The advantages and disadvantages of the various exposure methods are summarized in the following table (Table 5.14).

Table 5.14 Advantages and disadvantages of various exposure methods.

System	Advantages	Disadvantages
Contact printer	Simple and inexpensive	Defects
Proximity printer	Simple and inexpensive	Resolution limited
Full-wafer scanner	Moderate cost and high throughput	Resolution limited
Step-and-repeat	High resolution	Cost

Problems

5.1 Suppose that 0.5% (3σ) dose control is required from excimer laser pulse-to-pulse stability. If the single pulse-to-pulse stability of the laser is 2.5% (3σ), show that the maximum scan speed consistent with the dose control requirement is 960 mm/sec for a scanner with a slit height of 4 mm.

5.2 For a laser interferometer with a maximum Zeeman frequency split of $f_2 - f_1 = 4$ MHz, show that the maximum stage velocity that can be measured is 1266 mm/sec.

5.3 If the heterodyne receiver in a laser interferometer is capable of measuring frequency changes as small as 10 Hz, show that the the slowest measurable velocity that such a system can measure is 3.16×10^{-3} mm/sec.

5.4 Suppose that the projection optics of a stepper consists of 25 refractive elements, and with the application of antireflection coatings the reflectance from each element surface is 0.5%. Show that the total transmission of the projection lens is 78%.

5.5 Suppose that fields are scanned at a length of 30 mm. Show that 0.1 ppm of asymmetric magnification (in ppm) will lead to 3 nm of error across the scan. From Fig. 5.38, is this expected to lead to significant linewidth errors?

5.6 Why are lenses that are designed to image well over a wide bandwidth less sensitive to changes in barometric pressure than narrowband optics?

5.7 A typical ArF resist has 30 mJ/cm^2 sensitivity, and about 100 pulses are required to expose the resist. This means that the energy density per pulse is 0.3 mJ/cm^2. This is also approximately the energy density in the bottom element of the lens, which is close to the wafer. Imaging will be degraded at a significant level if the index of refraction in the final lens element is $\Delta n = 1000$ ppb or greater. With this energy density and a pulse duration of $\tau = 45$ nsec, show that a lens made from Sample B of Table 5.8 will be degraded after 95.4 billion pulses. With 12.5 billion pulses per year in manufacturing, what is the lifetime of a lens made from material B? What is the lifetime if resists with sensitivity of 50 mJ/cm^2 are used? Should the user request a laser with pulses stretched to 100 nsec?

References

1. A. Offner, "New concepts in projection mask aligners," *Opt. Eng.* **14**(2), 130–132 (1975).
2. F. Klosterman, "A step-and-repeat camera for photomasks," *Philips Tech. Rev.* **30**(3), 57–69 (1969).
3. M. van den Brink, B. Katz, and S. Wittekoek, "New 0.54 aperture i-line wafer stepper with field-by-field leveling combined with global alignment," *Proc. SPIE* **1463**, 709–724 (1991).

4. K. Suwa and K. Ushida, "Optical stepper with a high numerical aperture i-line lens and a field-by-field leveling system," *Proc. SPIE* **922**, 270–276 (1988).

5. J. Roussel, "Step-and-repeat wafer imaging," *Solid State Tech.*, 67–71 (May, 1978).

6. A. Bouwer, G. Bouwhuis, H. van Heek, and S. Wittekoek, "The silicon repeater," *Philips Tech. Rev.* **37**(11/12), 330–333 (1977); see also S. Wittekoek, "Optical aspects of the silicon repeater," *Philips Tech. Rev.* **41**, 268 (1983/84).

7. M. Lacombat, A. Gerard, G. Dubroeucq, and M. Chartier, "Photorepetition et projection directe," *Rev. Tech. Thomson-CSF* **9**(2), 337 (1977); see also Hans Binder and Michel Lacombat, "Step-and-repeat projection printing for VLSI circuit fabrication," *IEEE Trans. Electron. Dev*. **ED-26**(4), 698–704 (April, 1979).

8. J. Wilcyncski, "Optical step-and-repeat camera with dark field automatic alignment," *J. Vac. Sci. Tech.* **16**, 1929–1933 (1979).

9. W. J. Smith, *Modern Optical Engineering*, 2nd ed., McGraw Hill, Boston (1990).

10. H. Sewell and T. Friedman, "High-resolution imagery: the matching of optical and resist systems in the mid-UV," *Proc. SPIE* **922**, 328–334 (1988).

11. R. Loudon, *The Quantum Theory of Light*, 2nd ed., Clarendon Press, Oxford (1983).

12. W. H. Arnold and H. J. Levinson, "High resolution optical lithography using an optimized single layer photoresist process," *Proc. Kodak Microelectron. Sem.*, 80–82 (1983).

13. R. Hershel, "Optics in the model 900 projection stepper," *Proc. SPIE* **221**, 39–43 (1980).

14. A. R. Neureuther, P. K. Jain, and W. G. Oldham, "Factors affecting linewidth control including multiple wavelength exposure and chromatic aberrations," *Proc. SPIE* **275**, 48–53 (1981).

15. I. Friedman, A. Offner, and H. Sewell, "High resolution imagery: the matching of optical and resist systems," *Proc. KTI Microelectron. Sem.*, 239–250 (1987).

16. W. H. Arnold, A. Minvielle, K. Phan, B. Singh, and M. Templeton, "0.5 micron photolithography using high numerial aperture i-line wafer steppers," *Proc. SPIE* **1264**, 127–142 (1990).

17. T. C. Retzer and G. W. Gerung, "New developments in short arc lamps," *Illumin. Eng.* **51**, 745–752 (1956).

18. W. E. Thouret, "Tensile and thermal stresses in the envelope of high brightness high pressure discharge lamps," *Illumin. Eng.* **55**, 295–305 (1960).

19. G. Gear, "Reliability and stability of mercury-arc lamps used in wafer steppers," *Proc. Kodak Microelectron. Sem.*, 104 (1985).

20. D. W. Peters, "Improvements in linesize control resulting from exposure lamp temperature stabilization and optimization," *Proc. Kodak Microelectron. Sem.*, 137–149 (1986).

21. D. Cote, K. Andresen, D. Cronin, H. Harrold, M. Himel, J. Kane, J. Lyons, L. Markoya, C. Mason, D. McCafferty, M. McCarthy, G. O'Connor, H. Sewell, and D. Williamson, "Micrascan III-performance of a third generation, catadioptric step-and-scan lithographic tool," *Proc. SPIE* **3051**, 806–816 (1997).

22. R. Sze, "Rare-gas halide avalanche discharge lasers," *IEEE J. Quant. Electron.* **QE-15**, 1338–1347 (1979).

23. H. J. Levinson, "GCA ALS excimer stepper experience in the ASTC," *Proc. Sematech DUV Lithog. Workshop* (1991).

24. *Excimer Lasers*, 2nd ed., in *Topics in Applied Physics Series*, C. K. Rhodes (Ed.), Vol. 30, Springer Verlag (1984).

25. W. Partlo, R. Sandstrom, I. Fomenkov, and P. Das, "A low cost of ownership KrF excimer laser using a novel pulse power and chamber configuration," *Proc. SPIE* **2440**, 90–95 (1995).

26. R. W. McCleary, P. J. Tompkins, M. D. Dunn, K. F. Walsh, J. F. Conway, and R. P. Mueller, "Performance of a KrF excimer laser stepper," *Proc. SPIE* **922**, 396–399 (1988).

27. B. Rückle, P. Lokai, H. Rosenkranz, B. Kikolaus, H. J. Kahlert, B. Burghardt, D. Basting, and W. Mückenheim, "Computerized wavelength stabilized 248.4-nm excimer laser for stepper," *Proc. SPIE* **922**, 450–460 (1988).

28. R. Morton, I. Fomenkov, W. Partlo, P. Das, and R. Sandstrom, "Design considerations and performance of 1kHz KrF excimer lasers for DUV lithography," *Proc. SPIE* **2726**, 900–909 (1996).

29. J. Kleinschmidt, R. Pätzel, O. Heist, U. Stamm, and D. Basting, "Extremely narrow bandwidth, high repetition rate laser for high NA step-and-scan tools," *Proc. SPIE* **2726**, 821–830 (1996).

30. I. Bragin, V. Berger, R. Paetzel, U. Stamm, A. Targsdorf, J. Kleinschmidt, and D. Basting, "Prospects for very high repetition rate lasers for microlithography," *Proc. SPIE* **4000**, 1445–1451 (2000).

31. J. M. Hueber, H. Besaucele, P. Das, R. Elis, A. Ershov, V. Fleurov, D. Gaidarenko, T. Hormann, P. Melcher, W. Partlo, B. Kikolaus, S. Smith, and K. Webb, "Performance of very high repetition rate ArF lasers," *Proc. SPIE* **4000**, 1418–1423 (2000).

32. V. Fleurov, S. Rokitski, R. Bergstedt, H. Ye, K. O'Brien, R. Jacques, F. Trintchouk, E. Figueroa, T. Cacouris, D. Brown, and W. Partlo, "XLT

600i: recirculating ring ArF light source for double patterning immersion lithography," *Proc. SPIE* **6924**, 69241R (2008).

33. T. Kumazaki, T. Suzuki, S. Tanaka, R. Nohdomi, M. Yoshino, S. Matsumoto, Y. Kawasuji, H. Umeda, H. Nagano, K. Kakizaki, H. Nakarai, T. Tatsunaga, J. Fujimoto, and H. Mizoguchi, "Reliable high power injection locked 6 kHz 60W laser for ArF immersion lithography," *Proc. SPIE* **6924**, 69242R (2008).
34. U. K. Sengupta, Private Communication (2000).
35. T. Hofmann, B. Johanson, and P. Das, "Prospects for long pulse operation of ArF lasers for 193-nm microlithography," *Proc. SPIE* **4000**, 511–518 (2000).
36. T. Enami, T. Hori, T. Ohta, H. Mizoguchi, and T. Okada, "Dynamics of output spectra within a single laser pulse from a line-narrowed excimer Laser," *Jpn. J. Appl. Phys.* **39**(1), 86–90 (2000).
37. V. B. Fleurov, D. J. Colon III, D. J. W. Brown, P. O'Keeffe, H. Besaucele, A. I. Ershov, F. Trinchouk, T. Ishihara, P. Zambon, R. Rafac, and A. Lukashev, "Dual-chamber ultra line-narrowed excimer light source for 193-nm lithography," *Proc. SPIE* **5040**, 1694–1703 (2003).
38. T. P. Duffey, T. Embree, T. Ishihara, R. Morton, W. N. Partlo, T. Watson, and R. Sandstrom, "ArF lasers for production of semiconductor devices with CD < 0.15 μm," *Proc. SPIE* **3334**, 1014–1020 (1998).
39. R. Pätzel, J. Kleinschmidt, U. Rebban, J. Franklin, and H. Endert, "KrF excimer laser with repetition rates of 1 kHz for DUV lithography," *Proc. SPIE* **2440**, 101–105 (1995).
40. T. Lizotte, O. Ohar, T. O'Keefe, and C. Kelly, "Stable beam delivery expands excimer laser use," *Laser Focus World*, 163–169 (February, 1997).
41. J. Vipperman, P. Das, J. Viatella, K. Webb, L. Lublin, D. Warkentin, and A. I. Ershov, "High-performance beam-delivery unit for next-generation ArF scanner systens," *Semicond. Manuf.* **5**(4), 104–110 (2004).
42. ANSI Standard Z136 1-1993, "Safe use of lasers," *American National Standards Institute*, New York (1993).
43. M. Case, "193-nm coatings resist excimer-laser damage," *Laser Focus World*, 93–96 (April 2004).
44. H. Tsushima, M. Yoshino, T. Ohta, T. Kumazaki, H. Watanabe, S. Matsumoto, H. Nakarai, H. Umeda, Y. Kawasuji, T. Suzuki, S. Tanaka, A. Kurosu, T. Matsunaga, J. Fujimoto, and H. Mizoguchi, "Reliability report of high power injection lock laser light source for double exposure and double patterning ArF immersion lithography," *Proc. SPIE* **7274**, 72743L (2009).
45. K. Suzuki, S. Wakamoto, and K. Nishi, "KrF step-and-scan exposure system using higher NA projection lens," *Proc. SPIE* **2726**, 767–773 (1996).
46. M. van den Brink, H. Jaspar, S. Slonaker, P. Wijnhoven, and F. Klaassen, "Step-and-scan and step-and-repeat, a technology comparison," *Proc. SPIE* **2726**, 734–753 (1996).

47. H. Sewell, "Advancing optical lithography using catadioptric projection optics and step-and-scan," *Proc. SPIE* **2726**, 707–720 (1996).

48. P. Oesterlin, P. Lokai, B Burghardt, W. Muckenheim, H. J. Kahlert, and D. Basting, "Advanced lithography excimer system: absolute wavelength adjustment, five watt average power," *Proc. SPIE* **1088**, 404–406 (1989).

49. H. Sewell, "Advancing optical lithography using catadioptric projection optics and step-and Scan," *Proc. SPIE* **2726**, 707–720 (1996).

50. O. Semprez, "Excimer lasers for future lithography light sources," *Solid State Tech.*, 255–260 (July, 2000).

51. H. J. Kahlert, U. Rebhan, P. Lokai, and D. Basting, "Comparison of 248-nm line narrowing resonator optics for DUV lithography lasers," *Proc. SPIE* **1463**, 604–609 (1991).

52. M. Born and E. Wolf, *Principles of Optics*, 7th ed., Cambridge Univ. Press, New York (1999).

53. P. Yan, O. Qian, J. Langston, and P. Leon, "Effect of chromatic aberration in excimer laser lithography," *Proc. SPIE* **1674**, 316–327 (1992).

54. A. Kroyen, N. Farrar, J. Bendik, O. Semprez, C. Rowan, and C. Mack, "Modeling the effects of excimer laser bandwidths on lithographic performance," *Proc. SPIE* **4000**, 658–664 (2000).

55. S. K. Jones, E. S. Capsuto, B. W. Dudley, C. R. Peters, and G. C. Escher, "Wavelength tuning for optimization of deep UV excimer laser performance," *Proc. SPIE* **1674**, 499–508 (1992).

56. M. E. Preil and W. H. Arnold, "Aerial image formation with a KrF excimer laser stepper," *Polymer Eng. Sci.* **32**(21), 1583–1588 (1992).

57. R. K. Brimacombe, T. J. McKee, E. D. Mortimer, B. Norris, J. Reid, and T. A. Znotins, "Performance characteristics of a narrow band industrial excimer laser," *Proc. SPIE* **1088**, 416–422 (1989).

58. H. Mizoguchi, O. Wakabayashi, and N. Itoh, "Narrow band KrF excimer laser for mass production of ULSI ICs," *Proc. SPIE* **1674**, 532–542 (1992).

59. R. Sandstrom, "Argon-fluoride excimer laser source for sub-0.25 μm optical lithography," *Proc. SPIE* **1463**, 610–616 (1991).

60. A. Kroyan, I. Lalovic, and N. Farrar, "Effects of 95% Integral vs. FWHM Bandwidth Specifications on Lithographic Imaging," *Proc. SPIE* **4346**, 1244–1253 (2001).

61. V. N. Ishchenko, S. A. Kochubei, and A. M. Razhev, "High-power efficient vacuum ultraviolet F_2 laser excited by an electric discharge," *Sov. J. Quant. Electron.* **16**(5), 707–709 (1986).

62. J. Goldhar and J. R. Murray, "Injection-locked narrow band KrF discharge laser using an unstable resonator cavity," *Optic. Lett.* **1**, 199 (1977).

63. K. Jain, *Excimer Laser Lithography*, SPIE Press, Bellingham, Washington (1990).

64. O. Wakabayashi, T. Ariga, T. Kumazaki, K. Sasano, T. Watanabe, T. Yabu, T. Hori, K. Kakizaki, A. Sumiani, and H. Mizoguchi, "Beam quality of a new-type MOPO laser system for VUV laser lithography," *Proc. SPIE* **5377**, 1772–1780 (2004).

65. J. Hecht, *The Laser Guidebook*, McGraw Hill, New York (1986).

66. Y. Lin and J. Buck, "Numerical modeling of the excimer beam," *Proc. SPIE* **3677**, 700–710 (1999).

67. O. Kritsun, I. Lalovic, S. Rotkitski, B. Partlo, B. La Fontaine, N. Farrar, and H. Levinson, "Improving lithography pattern fidelity and line-edge roughness by reducing laser speckle," *J. Vac. Sci. Tech. B* **26**(6), 2145–2150 (2008).

68. P. Das and R. L. Sandstrom, "Advances in excimer laser technology for Sub-0.25-μm lithography," *Proc. IEEE* **90**, 1637–1652 (2002).

69. S. Rokitski, V. B. Fleurov, R Bergstedt, H. Ye, R. Rafac, R. Jacques, F. Trintchouk, T. Ishihara, R. Rao, T. Cacouris, D. Brown, and William Partlo, "Enabling high volume manufacturing of double patterning immersion lithography with the XLR 600ix ArF light source," *Proc. SPIE* **7274**, 72743O-1 (2009).

70. K. Kakizaki, J. Fujimoto, T. Yamazaki, T. Suzuki, T. Matsunaga, Y. Kawasuji, Y. Watanabe, M. Kaminishi, T. Inoue, H. Mizoguchi, T. Kumazaki, A. Tatsuya, T. Watanabe, T. Yabu, K. Sasano, T. Hori, O. Wakabayashi, and A. Sumitani, "Development of high power ArF/F_2 platform for VUV microlithography," *Proc. SPIE* **5377**, 1805–1814 (2004).

71. K. Wakana, H. Tsushima, S. Matsumoto, M. Yoshino, T. Kumazaki, H. Watanabe, T. Ohta, S. Tanaka, T. Suzuki, H. Nakarai, Y. Kawasuji, A. Kurosu, T. Matsunaga, J. Fujimoto, and H. Mizoguchi, "Optical performance of laser light source for ArF immersion double patterning lithography tool," *Proc. SPIE* **7274**, 72743J (2009).

72. R. Voelkel, H. P. Herzig, P. Nussbaum, R. Daendiliker, and W. B. Hugle, "Microlens array imaging system for photolithography," *Opt. Eng.* **35**(11), 3323–3330 (1996).

73. J. van Schoot, F. Bornebroek, M. Suddendorf, M. Mulder, J. van der Spek, J. Stoeten, and A. Hunter, "0.7-NA DUV step & scan system for 150-nm imaging with improved overlay," *Proc. SPIE* **3679**, 448–463 (1999).

74. Y. Ichihara, S. Kawata, Il. Hikima, M. Hamatani, Y. Kudoh, and A. Tanimoto, "Illumination system of an excimer laser stepper," *Proc. SPIE* **1138**, 137–143 (1989).

75. R. W. Leonhardt and T. R. Scott, "Deep-UV excimer laser measurements at NIST," *Proc. SPIE* **2439**, 448–459 (1995).

76. M. Nei and H. Kawai, "Reduction projection type exposure device: NSR2005i9C," *Denshi Zairyo*, 80–84 (March, 1993).

77. J. Mulkens, J. Stoeldraijer, R. Rogoff, J. de Klerk, P. Jenkins, J. Wangler, G. Kalmback, W. Rupp, and E. Piper, "High-throughput wafer stepper with adjustable illumination modes," *Sol. State Tech.*, 193–199 (July, 1996).

78. R. Rogoff, G. Davies, J. Mulkens, J. de Klerk, P. van Oorschot, G. Kalmbach, J. Wangler, and W. Rupp, "Photolithography using the AERIAL™ illuminator in a variable NA wafer stepper," *Proc. SPIE* **2726**, 54–70 (1996).

79. H. H. McLeod, "The axicon: a new type of optical element," *J. Opt. Soc. Am.* **44**, 592–597 (1954).

80. W. J. Smith, *Modern Optical Engineering*, 2nd ed., McGraw-Hill, Boston (1990).

81. Y. Kudo, "Illuminating optical system and exposure appartus utilizing the same," U.S. Patent No. 5,237,367 (1993).

82. W. L. Krisa, C. M. Garza, and R. D. Bennett, "Contact performance with an attenuated phase-shift reticle and variable partial coherence," *Proc. SPIE* **2440**, 524–531 (1995).

83. M. Op de Beeck, K. Ronse, K. Gandehari, P. Jaenen, H. Botermans, J. Finders, J. Lilygren, D. Baker, G. Vandenberghe, P. De Bisschop, M. Maenhoudt, and L. Van den hove, "NA/sigma optimization strategies for an advanced DUV stepper applied to 0.25 μm and sub-0.25 μm critical levels," *Proc. SPIE* **3051**, 320–332 (1997).

84. Y. Borodovsky, "Impact of local partial coherence variation on exposure tool performance," *Proc. SPIE* **2440**, 750–770 (1995).

85. T. Ogawa, M. Uematsu, F. Uesawa, M. Kimura, H. Shimizu, and T. Oda, "Sub-quarter-micron optical lithography with practical super resolution technique," *Proc. SPIE* **2440**, 772–783 (1995).

86. M. Noguchi, M. Muraki, Y. Iwasaki, and A. Suzuki, "Sub-half-micron lithography system with phase-shifting effect," *Proc. SPIE* **1674**, 92–104 (1992).

87. J. Kirk and C. Progler, "Pinholes and pupil fills," *Microlithography World*, 25–34 (Autumn, 1997).

88. H. J. Levinson, described in Pieter Burggraaf, "5X reticle problems and 1 μm design rules," *Semicond. Int.* **6**(12), 22–24 (1983).

89. J. N. Wiley, "Effect of stepper resolution on the printability of sub-micron 5× reticle defects," *Proc. SPIE* **1088**, 58–73 (1989).

90. G. Arthur, B. Martin, F. Goodall, and I. Loader, "Printability of sub-micron 5× reticle defects at g-line, i-line and DUV exposure wavelengths," *Proc. SPIE* **2196**, 241–252 (1994).

91. H. J. Levinson, M. E Preil, P. Ackmann, and B. Rericha, "The factors that determine the optimum lens reduction factor," *Proc. SPIE* **3677**, 468–478 (1999).

92. J. Braat, "Quality of microlithographic projection lenses," *Proc. SPIE* **811**, 22–30 (1987).

93. E. Glatzel, "New lenses for microlithography," *Proc. SPIE* **237**, 310–320 (1980).

94. *Ohara i-line Glasses*, Ohara Corporation, Kanagawa, Japan (June, 1993).

95. *Schott Catalog of Optical Glass No. 10000 on Floppy Disk*, SCHOTT Glass Technologies, Duryea, Pennsylvania (1992).

96. "HPFS® 8650 fused silica for ArF applications," Corning Inc., Canton, New York (2006).

97. T. Matsuyama, Y. Shibazaki, Y. Ohmura, and T. Suzuki, "High NA and low residual aberration projection lens for DUV scanner," *Proc. SPIE* **4691**, 687–695 (2002).

98. T. A. Brunner, S. Cheng, and A. E. Norton, "A stepper image monitor for precise setup and characterization," *Proc. SPIE* **922**, 327–377 (1988).

99. P. van Oorschot, B. Koek, J. van der Spek, E. Stuiver, H. Franken, H. Bother, and R. Garreis, "Performance of an i-line step-and-scan system for sub-0.25-μm mix and match applications," *Proc. SPIE* **3334**, 423–436 (1998).

100. J. H. Burnett, Z. H. Levine, and E. L. Shirley, "Intrinsic birefringence in calcium fluoride and barium fluoride," *Phys. Rev*. **B64** 241102(R), 1–4 (2001).

101. J. H. Burnett, Z. H. Levine, and E. L. Shirley, "Hidden in plain sight: calcium fluoride's intrinsic birefringence," *Photon. Spectra* **35**(12), 88–92 (December, 2001).

102. J. H. Burnett, Z. H. Levine, E. L. Shirley, and J. H. Bruning, "Symmetry of spatial-dispersion-induced birefringence and its implications for CaF_2 ultraviolet optics," *J. Microlith., Microfab., Microsyst.* **1**(3), 213–224 (2002).

103. B. Wang, "Birefringence in fused silica and CaF_2 for lithography," *Sol. State Tech.*, 77–82 (February, 2000).

104. R. E. Schenker, F. Piao, and W. G. Oldham, "Material limitations to 193-nm lithographic system lifetimes," *Proc. SPIE* **2726**, 698–706 (1996).

105. W. G. Oldham and R. E. Schenker, "193-nm lithographic system lifetimes as limited by UV compaction," *Sol. State Tech.* **40**(4), 95–102 (April, 1997).

106. J. M. Algots, R. Sandstrom, W. Partlo, P. Maroevic, E. Eva, M. Gerhard, R. Lindner, and F. Stietz, "Compaction and rarefaction of fused silica with 193-nm excimer laser exposure," *Proc. SPIE* **5040**, 1639–1650 (2003).

107. J. M. Algots, R. Sandstrom, W. Partlo, K. Takahashi, H. Ishii, and Y. Hasegawa, "Verification of the compaction and rarefaction models for

fused silica with 40 billion pulses of 193-nm excimer laser exposure and their effects on projection lens imaging performance," *Proc. SPIE* **5377**, 1815–1827 (2004).

108. P. Bousquet, F. Flory, and P. Roche, "Scattering from multilayer thin films: theory and experiment," *J. Opt. Soc. Am.* **71**(9), 1115–1123 (1981).

109. D. Golini, M. DeMarco, W. Kordonski, and J. Bruning, "MRF polishes calcium fluoride to high quality," *Laser Focus World*, S5–S13 (July, 2001).

110. R. R. Kunz, V. Lieberman, and D. K. Downs, "Photo-induced organic contamination of lithographic optics," *Microlithography World*, 2–8 (Winter, 2000).

111. S. Barzaghi, A. Plenga, G. Vergani, and S. Guadagnuolo, "Purged gas purification for contamination control of DUV stepper lenses," *Sol. State Tech.* **44**(9), 14–99 (September, 2001).

112. J. P. Kirk, "Scattered light in photolithographic lenses," *Proc. SPIE* **2197**, 566–572 (1994).

113. B. La Fontaine, M. Dusa, A. Acheta, C. Chen, A. Bourov, H. Levinson, L. C. Litt, M. Mulder, R. Seltmann, and J. Van Braagh, "Analysis of flare and its impact on low-k_1 KrF and ArF lithography," *Proc. SPIE* **4691**, 44–56 (2002).

114. R. Hershel, "Optics in the model 900 projection stepper," *Proc. SPIE* **221**, 39–43 (1980).

115. Ultratech Stepper, San Jose, California.

116. J. D. Buckley and C. Karatzas, "Step and scan: a systems overview of a new lithography tool," *Proc. SPIE* **1088**, 424–433 (1989).

117. D. Williamson, "Optical reduction system," U.S. Patent No. 4,953,960 (1990).

118. M. Barrick, D. Bommarito, K. Holland, K. C. Norris, B. Patterson, Y. Takamori, J. Vigil, and T. Wiltshire, "Performance of a 0.5 NA broadband DUV step-and-scan system," *Proc. SPIE* **1927**, 595–607 (1993).

119. D. Williamson, Inventor, SVG Lithography Systems, Inc., Assignee, "Catadioptric optical reduction system with high numerical aperture," U.S. Patent No. 5,537,260 (1996).

120. H. Sewell, "Advancing optical lithography using catadioptric projection optics and step-and scan," *Proc. SPIE* **2726**, 707–720 (1996).

121. D. Williamson, J. McClay, K. Andresen, G. Gallitin, M. Himel, J. Ivaldi, C. Mason, A. McCullough, C. Otis, J. Shamaly, and C. Tomczyk, "Micrascan III, a 0.25-μm resolution step-and-scan system," *Proc. SPIE* **2726**, 780–2726 (1996).

122. T. Matsuyama, Y. Ohmura, Y. Fujishima, and T. Koyama, "Catadioptric Projection Lens for 1.3 NA scanner," *Proc. SPIE* **6520**, 652021 (2007).

123. P. R. Yoder, *Mounting Lenses in Optical Instruments*, SPIE Press, Bellingham, Washington (1995).

124. D. C. McCarthy, "Optical adhesives provide a cure for any problem," *Photon. Spectra* **34**(9), 117–122 (September, 2000).

125. Y. Sudoh and T. Kanda, "A new lens barrel structure utilized on the FPA-6000 AS4 and its contribution to the lens performance," *Proc. SPIE* **5040**, 1657–1664 (2003).

126. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surfaces and lenses," *Appl. Optic.* **13**(11), 2693–2703 (November, 1974).

127. R. Rogoff, S. S. Hong, D. Schramm, and G. Espin, "Reticle specific compensations to meet production overlay requirements for 64 Mb and beyond," *Proc. SPIE* **807**, 806–813 (1993).

128. H. J. Levinson, M. Preil, and P. Lord, "Minimization of total overlay errors using an advanced optimization scheme," *Proc. SPIE* **3051**, 362–373 (1997).

129. J. P. Kuijten, T. Harris, L. van der Heijden, D. Witko, J. Cossins, J. Foster, and D. Ritchie, "Advanced mix and match using a high NA i-line scanner," *Proc. SPIE* **4000**, 843–856 (2000).

130. J. de Klerk, C. Wagner, R. Droste, L. Levasier, L. Jorritsma, E. van Setten, H. Kattouw, J. Jaccobs, and T. Heil, "Performance of a 1.35 NA ArF immersion lithography system for 40 nm applications," *Proc. SPIE* **6520**, 65201Y (2007).

131. T. Nakashima, Y. Ohmura, T. Ogata, Y. Uehara, H. Nishinaga, and T. Matsuyama, "Thermal aberration control in projection lens," *Proc. SPIE* **6924**, 69241V (2008).

132. T. Matsuyama, Y. Ohmura, T. Nakashima, Y. Uehara, T. Ogata, H. Nishinaga, H. Ikezawa, T. Toki, S. Rokitski, and J. Bonafede, "An Intelligent Imaging System for ArF Scanner," *Proc. SPIE* **6924**, 69241S (2008).

133. D. Williamson, J. McClay, K. Andresen, G. Gallatin, M. Himel, J. Ivaldi, C. Mason, A. McCullough, C. Otis, J. Shamaly, and C. Tomczyk, "Micrascan III, 0.25-μm resolution step-and-scan system," *Proc. SPIE* **2726**, 780–786 (1996).

134. G. de Zwart, M. van den Brink, R. George, D. Satriasuptra, J. Baslemans, H. Butler, J. van Schoot, and J. de Klerk, "Performance of a step-and-scan system for DUV lithography," *Proc. SPIE* **3051**, 817–835 (1997).

135. R. Ebinuma, K. Iwamoto, H. Takeishi, H. Itoh, M. Inoue, K. Takahashi, and M. Ohta, "Imaging performance of a scanning exposure system," *Proc. SPIE* **3334**, 437–447 (1998); Sugimori, Private Communication.

136. Y. Kawamoto, S. Kimura, N. Hasegawa, A. Hiraiwa, T. Kure, T. Nishida, M. Aoki, H. Sunami, and K. Itoh, "A half-micron technology for an experimental

16-Mbit DRAM using i-line stepper," *VLSI Symposium*, Paper III-1, 17–18 (1988).

137. J. Webb, "All-calcium fluoride system using 157-nm light," *Laser Focus World*, 87–92 (September, 2000).
138. H. Nogawa, H. Hata, and M. Kohno, "System design of a 157-nm scanner," *Proc. SPIE* **4691**, 602–612 (2002).
139. N. Shiraishi, S. Owa, Y. Omura, T. Aoki, Y. Matsumoto, J. Nishikawa, and I. Tanaka, "Progress of Nikon's F_2 exposure tool development," *Proc. SPIE* **4691**, 594–601 (2002).
140. J. Mulkens, T. Fahey, J. McClay, J. Stoeldraijer, P. Wong, M. Brunotte, and B. Mecking, "157-nm technology: where are we today?" *Proc. SPIE* **4691**, 613–625 (2002).
141. F. Bornebroek, M. de Wit, W. de Boeji, G. Dicker, J. Hong, and A. Serebryakov, "Cost-effective shrink of semi-critical layers using the TWINSCAN XT:1000H NA 0.93 KrF Scanner," *Proc. SPIE* **7274**, 72743I (2009).
142. T. Miller, "Aspherics come of age," *Photon. Spectra*, 76–81 (February 2004).
143. G. Owen, R. F. W. Pease, D. A. Markle, and A. Grenville, "1/8 μm optical lithography," *J. Vac. Sci. Tech. B* **10**(6), 3032–3036 (1993).
144. K. Suwa, K. Nakazawa, and S. Yoshida, "10:1 step-and-repeat projection system," *Proc. Kodak Microelectron. Sem.*, 61–66 (1981).
145. A. Suzuki, S. Yabu, and M. Ookubo, "Intelligent optical system for a new stepper," *Proc. SPIE* **772**, 58–65 (1987).
146. K. Suwa and K. Ushida, "The optical stepper with a high numerical aperture i-line lens and a field-by-field leveling system," *Proc. SPIE* **922**, 270–276 (1988).
147. M. A. van den Brink, B. A. Katz, and S. Wittekoek, "New 0.54 aperture i-line wafer stepper with field-by-field leveling combined with global alignment," *Proc. SPIE* **1463**, 709–724 (1991).
148. T. O. Herndon, C. E. Woodward, K. H. Konkle, and J. I. Raffel, "Photocomposition and DSW autofocus correction for wafer-scale lithography," *Proc. Kodak Microelectron. Sem.*, 118–123 (1983).
149. B. La Fontaine, J. Hauschild, M. Dusa, A. Acheta, E. Apelgren, M. Boonman, J. Krist, A. Khathuria, H. Levinson, A. Fumar-Pici, and M. Pieters, "Study of the influence of substrate topography on the focusing performance of advanced lithography scanners," *Proc. SPIE* **5040**, 570–581 (2003).
150. J. E. van den Werf, "Optical focus and level sensor for wafer steppers," *J. Vac. Sci. Tech.* **10**(2), 735–740 (1992).
151. M. A. van den Brink, J. M. D. Stoeldraijer, and H. F. D. Linders, "Overlay and field-by-field leveling in wafer steppers using an advanced metrology system," *Proc. SPIE* **1673**, 330–344 (1992).

152. T. Tojo, M. Tabata, Y. Ishibashi, H. Suzuki, and S. Takahasi, "The effect of intensity distribution in the reflected beam on the detection error of monochromatic optical autofocus systems," *J. Vac. Sci. Tech. B* **8**(3), 456–462 (1990).

153. T. H. Newman, "Air-gauge nozzle probe structure for microlithographic image focusing," U.S. Patent No. 6,029,361 (2000).

154. T. Thomas, E. Christenson, B. Holstrom, and E. Mino Jr., "On-axis air gage focus system," U.S. Patent No. 5,087,927 (1992).

155. G. Zhang, S. DeMoor, S. Jessen, Q. He, W. Yan, S. Chevacharoenkul, V. Vellanki, P. Reynolds, J. Ganeshan, J. Hauschild, and M. Pieters, "Across wafer focus mapping and its applications in advanced technology nodes," *Proc. SPIE* **6154**, 61540N-1 (2006).

156. F. Kalhenberg, R. Seltmann, B. La Fontaine, R. Wirtz, A. Kisteman, R. N. M. Vaneer, and M. Pieters, "Best focus determination: bridging the gap between optical and physical topography," *Proc. SPIE* **6520**, 65200Z (2007).

157. J. J. Henry, "Capacitance gauge for measuring small distances," U.S. Patent No. 3,716,782 (1973).

158. S. J. A. G. Cosijns, "Displacement laser interferometry with sub-nanometer uncertainty," Technische Universiteit Eindhoven (2004).

159. E. A. Hylleraas, *Mathematical and Theoretical Physics*, Vol. II, Wiley Interscience, New York (1970).

160. J. N. Dukes and G. B. Gordan, "A two-hundred-foot yardstick with graduations every micro-inch," *Hewlett-Packard J.* **21**(12), 2–8 (1970).

161. H. de Lang and G. Bouwhuis, "Displacement measurement with a laser interferometer," *Philips Tech. Rev.* **30**, 160–165 (1969).

162. A. Rude and M. Ward, "Laser transducer system for high accuracy machine positioning," *Hewlett Packard J.*, 2–7 (February, 1976).

163. C. R. Steinmetz, "Displacement measurement repeatability in tens of nanometers with laser interferometry," *Proc. SPIE* **921**, 406–420 (1988).

164. N. Deguchi and S. Uzawa, "150-nm generation lithography equipment," *Proc. SPIE* **3679**, 464–472 (1999).

165. S. A. Lis, "An air-turbulence compensated interferometer," *Proc. SPIE* **2440**, 891–901 (1995).

166. D. Crespo, J. Alonso, and E. Bernabeu, "Reflection optical encoders as three-grating Moire systems," *Appl. Optic.* **39**, 3805–3813 (2000).

167. C. Liu, H. Huang, and H. Lee, "Five-degrees-of-freedom diffractive laser encoder," *Appl. Optic.* **48**, 2767–2777 (2009).

168. C. Kao, S. Lu, and M. Lu, "High resolution planar encoder by retro-reflection," *Rev. Sci. Instr.* **76**, 085110 (2005).

169. This figure was provided by Dr. Wolfgang Holzapfel and Callimici Christian of Heidenhein GmbH, a supplier of high-precision encoders.

170. S. Wittekoek, H. Linders, H. Stover, G. Johnson, D. Gallagher, and R. Fergusson, "Precision wafer stepper alignment and metrology using diffraction gratings and laser interferometry," *Proc. SPIE* **565**, 22–31 (1985).

171. M. A. van den Brink, S. Wittekoek, H. F. D. Linders, F. J. van Hout, and R. A. George, "Performance of a wafer stepper with automatic intra-die registration correction," *Proc. SPIE* **772**, 100–117 (1987).

172. M. Williams, P. Faill, P. M. Bischoff, S. P. Tracy, and B. Arling, "Six degrees of freedom mag-lev stage development," *Proc. SPIE* **3051**, 856–867 (1997).

173. P. T. Konkola, "Magnetic bearing stages for electron beam lithography," Master of Science Thesis, Massachusetts Institute of Technology (1998).

174. S. Wittekoek, H. Linders, H. Stover, G. Johnson, D. Gallagher, and R. Fergusson, "Precision wafer stepper alignment and metrology using diffraction gratings and laser interferometry," *Proc. SPIE* **565**, 22–31 (1985).

175. B. J. Lin, "Vibration tolerance in optical imaging," *Proc. SPIE* **1088**, 106–114 (1989).

176. D. Cote, A. Ahouse, D. Galburt, H. Harrold, J. Kreuzer, M. Nelson, M. Oskotsky, G. O'Connor, H. Sewell, D. Williamson, J. Zimmerman, and R. Zimmerman, "Advances in 193-nm lithography tools," *Proc. SPIE* **4000**, 542–550 (2000).

177. Y. Shibazaki, H. Kohno, and M. Hamatani, "An innovative platform for high-throughput, high-accuracy lithography using a single wafer stage," *Proc. SPIE* **7274**, 72741I (2009).

178. J. Mulkens, Private Communication.

179. J. Bischoff, W. Henke, J. van der Werf, and P. Dirksen, "Simulations on step and scan optical lithography," *Proc. SPIE* **2197**, 953–964 (1994).

180. B. Slujik, T. Castenmiller, R. du Croo de Jongh, H. Jasper, T. Modderman, L. Levasier, E. Loopstra, G. Savenije, M. Boonman, and H. Cox, "Performance results of a new generation of 300-mm lithography systems," *Proc. SPIE* **4346**, 544–557 (2001).

181. J. de Klerk, L. Jorritsma, E. van Setten, R. Droste, R. du Croo de Jongh, S. Hansen, D. Smith, M. van den Kerkhof, F. van de Mast, P. Graeupner, T. Rohe, and K. Kornitzer, "Performance of a high NA, dual stage 193-nm TWINSCAN step and scan system for 80-nm applications," *Proc. SPIE* **5040**, 822–840 (2003).

182. I. Fujita, F. Sakai, and S. Uzawa, "Next generation scanner for sub-100 nm lithography," *Proc. SPIE* **5040**, 811–821 (2003).

183. K. Hirano, Y. Shibazaki, M. Hamatani, J. Ishikawa, and Y. Iriuchijima, "Latest results from the Nikon NSR-S620 double patterning immersion scanner," *Proc. SPIE* **7520**, 75200Z (2009).

184. I. Lalovic, A. Kroyan, J. Kye, H. Lu, and H. J. Levinson, "Image-blur tolerances for 65-nm and 45-nm node IC manufacturing," *Proc. SPIE* **5040**, 1570–1580 (2003).

185. H. J. Levinson, "Equipment for DUV lithography," *Proc. Semicon. West*, (1997).

186. M. E. Boonman, M. Dusa, and S. Miller, "Performance advantages of a dual stage system," *Proc. SPIE* **5377**, 742–757 (2004).

187. M. Okumura, J. Ishikawa, M. Hamatani, and M. Nei, "Mass production level ArF immersion exposure tool," *Proc. SPIE* **6154**, 61541U (2006).

188. B. Meliorisz and A. Erdmann, "Simulation of mask proximity printing," *J. Micro/Nanolithog. MEMS MOEMS* **6**(2), 023006 (2007).

189. L. F. Thompson and M. J. Bowden, "The lithographic process—the pohysics," in *Introduction to Microlithography: Theory, Materials and Processing*, L. F. Thompson, C. G. Willson, and M. J. Bowden, Eds., American Chemical Society, Washington, DC (1983).

190. This figure was provided courtesy of SUSS MicoTec.

191. D. P. Burbank, "The near impossibility of making a microchip," *Invention and Technology*, 44–51 (Fall, 1999).

192. A. Offner, "Unit power imaging catoptric anastigmat," U.S. Patent No. 3,748,015 (1973).

193. J. Brunning, "Optical lithography . . . 40 years and holding," *Proc. SPIE* **6520**, 652004 (2007).

194. H. O. Madsen, R. O. Rice, and G. E. Southers, "Is a stepper really better than a scanner? A 1:1 comparison," *Proc. Kodak Microlectron. Sem. Interface '84*, 4–11 (1984).

Chapter 6
Overlay

Integrated circuits are built through a sequence of patterning steps. If the transistor shown in Fig. 6.1 is going to be contacted electrically so that it functions properly, the contact to the gate must physically connect to the gate and not short electrically to the source or drain. This requires that the pattern created at the contact-masking step be placed correctly on top of the pre-existing transistor structures. As one can see from Fig. 6.1, this placement does not need to be perfect, but it must be within certain tolerances. Transistors will function so long as the contacts have sufficient overlap with the appropriate parts of the transistor and do not contact the parts of the transistor from which they are supposed to be electrically isolated.

The lateral positioning between layers comprising integrated circuits is called *overlay* (see Fig. 6.2), which is defined precisely in SEMI Standard P18-92 as follows:[1]

> Overlay—A vector quantity defined at every point on the wafer. It is the difference $\vec{O}$ between the vector position $\vec{P}_1$ of a substrate geometry, and the vector position of the corresponding point $\vec{P}_1$ in an overlaying pattern, which may consist of photoresist:

$$\vec{O} = \vec{P}_2 - \vec{P}_1. \tag{6.1}$$

As feature sizes shrink, so do the overlay tolerances between layers (see Table 6.1). While overlay is a vector quantity, the quantities in Table 6.1 are, by convention, the maximum tolerable magnitudes for overlay in either the X or Y direction. A related quantity, registration, is defined similarly to overlay:[1]

> Registration—A vector quantity defined at every point on the wafer. It is the difference $\vec{R}$ between the vector position $\vec{P}_1$ of a substrate geometry, and the vector position of the corresponding point $\vec{P}_0$ in a reference grid:

$$\vec{R} = \vec{P}_1 - \vec{P}_0. \tag{6.2}$$

Note that overlay is a relative quantity, while registration is an error compared to an absolute standard $\vec{P}_0$.

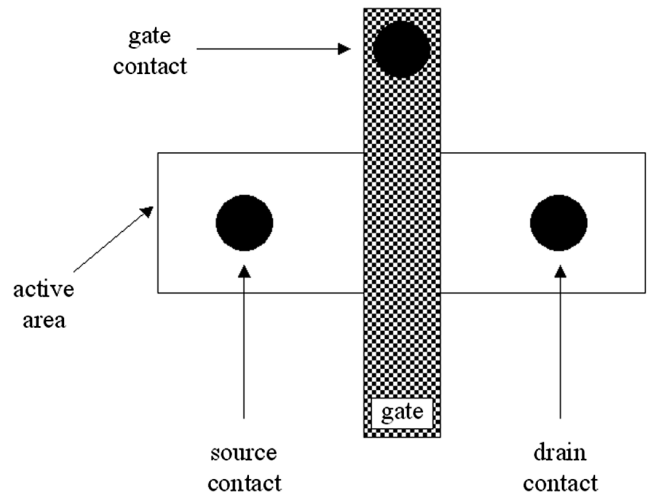


Figure 6.1 A contacted transistor. For the transistor to function properly, the contacts must overlay the appropriate features.

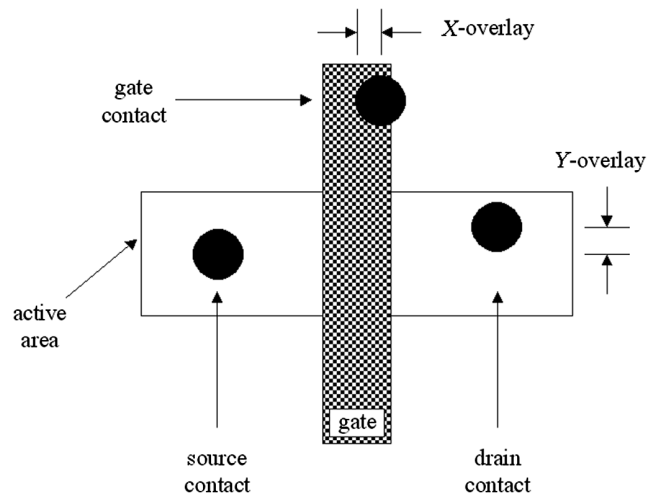


Figure 6.2 Examples of overlay errors.

Table 6.1 DRAM overlay requirements from the 2008 International Technology Roadmap for Semiconductors. Logic overlay requirements are ~25% less stringent. As will be discussed in Chapter 8, overlay requirements are much tighter for double patterning processes.

Node(DRAM half pitch)	Year	Overlay requirement (nm)
68	2007	13.6
45	2010	9.0
32	2013	6.4
22.5	2016	4.5

Overlay is achieved through the process of *alignment*, which is discussed in this chapter. There are many types of alignment systems, and several of these are described as well. In addition, sources of overlay errors and techniques for analyzing these errors are discussed. Further information on controlling overlay in manufacturing operations is available in *Lithography Process Control*.[2]

6.1 Alignment Systems

Wafers on laser stages can be moved with great precision, as described in the previous chapter. For this capability to be useful, it is necessary to relate positions in the wafer's coordinate system to positions in the stage's coordinate system.[3] The ultimate goal of the alignment process is to place the wafer under the lens so that the patterns being exposed will overlay patterns already existing on the wafer. To be able to accomplish this using the extraordinary precision of the laser stage it is necessary to determine the positions of patterns on the wafer in terms of what is already known, which is the position of the stage. Chips and other special patterns are placed on the wafer at positions specified in terms of the wafer's coordinate system. The goal of wafer alignment is to determine those chip positions in stage coordinates.

It is easier to understand how the positions of chips known in the wafer's coordinate system can be related to stage coordinates by first assuming that both the wafer's and stage's coordinate systems are orthonormal and rectilinear; that is, the coordinate systems have X and Y linear axes that are perpendicular to each other. It is further assumed that distances measured along the X axis are the same as those measured along the Y axis in both the wafer and stage coordinate systems, but the wafer and stage coordinate systems may measure distances differently. This means the "ruler" for the wafer coordinate system may be different from that of the wafer stage. These assumptions are only approximate for actual situations encountered in wafer processing (which will be discussed shortly), but it is useful to consider this simple case first. With these initial assumptions, the position of a point (X_W, Y_W) in the wafer coordinate system is related to the position (X_S, Y_S) of that point in the stage coordinate system:

$$X_S = S(\cos\theta\, X_W - \sin\theta\, Y_W) + O_X, \tag{6.3}$$

and

$$Y_S = S(\cos\theta\, Y_W + \sin\theta\, X_W) + O_Y, \tag{6.4}$$

where (O_x, O_y) is the position of the origin of the wafer coordinate system in the stage coordinate system, while θ is the angle of rotation between the two coordinate systems (Fig. 6.3). The measurement scales of the wafer in relation to the stage coordinate systems is represented by S. Thus, the wafer and stage coordinate systems are related by four parameters: S, θ, O_X, and O_Y. These parameters can be determined by measuring the wafer stage positions of two points on the wafer.

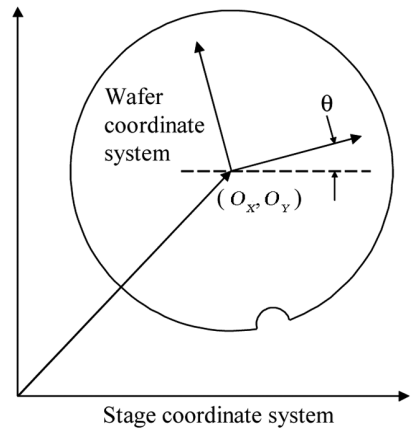


Figure 6.3 Transformation between the stage and wafer coordinate systems.

Issues that arise from having stages where the X and Y motions are not orthogonal or do not have the same scale are discussed later in this chapter.

Locating points on the wafer with nanometer precision is accomplished by the stepper's alignment system. Specially designed structures called *alignment marks* or *alignment targets* are placed on the wafer. The positions at which they are placed are specified in the coordinate system of the wafer. The stepper's *alignment system* is designed to recognize the alignment targets optically and determine their locations in the coordinate system of the stage. The key objective is to place the wafer under the projection optics so that new exposures overlay pre-existing ones, and this is accomplished by using the laser stage as a very accurate ruler. The details of the alignment process are discussed in this chapter, along with the methods used to characterize overlay errors.

Before the wafer is placed on the stage it is positioned mechanically. The stepper's *prealignment system* is used to locate the wafer's center and orient the wafer, using the wafer's edges and notch (or flat) as mechanical references. The centered and oriented wafer is then placed on the stage. Representative values for the placement accuracy of prealigned wafers are shown in Table 6.2. These values are typical of mechanical tolerances. Wafer prealignment enables the alignment targets to be brought within the alignment system's field of view, which is usually at most a few hundred microns in diameter. Some alignment systems have such small fields of view that an optical alignment, intermediate between the coarse mechanical prealignment and the fine wafer alignment, is necessary.

Table 6.2 Typical values for prealignment accuracy.

	Typical repeatability values (3σ)
Translation	10–100 μm
Rotation	0.1–1 mrad

The prealignment or the intermediate optical alignment must ensure that the fine alignment system captures the correct alignment target, which can be a problem when multiple alignment marks have been generated at several prior process layers. Because there may be limited silicon real estate in which alignment targets can be placed, the targets will need to be placed as close together as possible. Mark separation must be consistent with prealignment or intermediate-alignment capability to avoid having the fine-alignment system capture the wrong target. Some alignment targets consist of periodic structures, such as gratings, and misalignment can occur when the prealignment or intermediate-alignment error is more than half of the periodicity of the alignment mark.

Alignment marks are placed on the wafer at the first masking step. The marks may be contained within each exposure field, typically within scribe lines, or they may be placed in areas near the wafer edge (Fig. 6.4). The information of ultimate importance consists of the locations of the product dies. The relationship between the positions of the alignment marks and the product dies is established by the design of the mask that is first used to pattern the alignment marks. Thus, knowledge of the locations on the wafer of the alignment marks coupled with the design data enables the determination of the product die positions. Within an exposure field, there may be one or several alignment marks depending upon the stepper's design and the user's alignment strategy.

It is also important at the first masking step that the X axis of the reticle is aligned parallel to the X axis of the wafer stage. The consequence of failing to do this is shown in Fig. 6.5. The reticle is oriented properly by placing alignment marks on the reticle. Generally, these reticle alignment marks are outside of the exposure field.

At each subsequent masking step, the stepper's alignment system is used to determine the position of some of the alignment marks on the wafer. It is also possible to place new alignment marks on the wafer with each new exposure, and

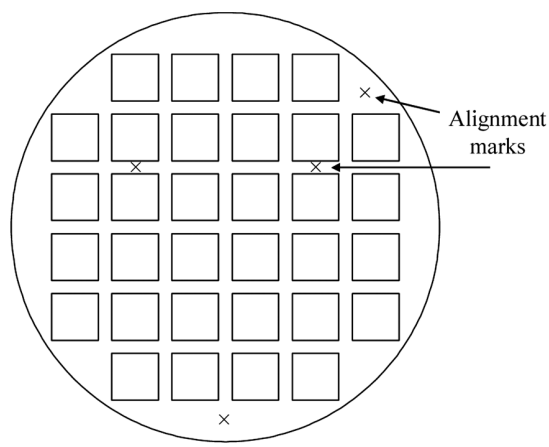


Figure 6.4 Alignment marks can be placed within exposure fields, typically in the scribe lines between product die, or outside of typical exposure fields.

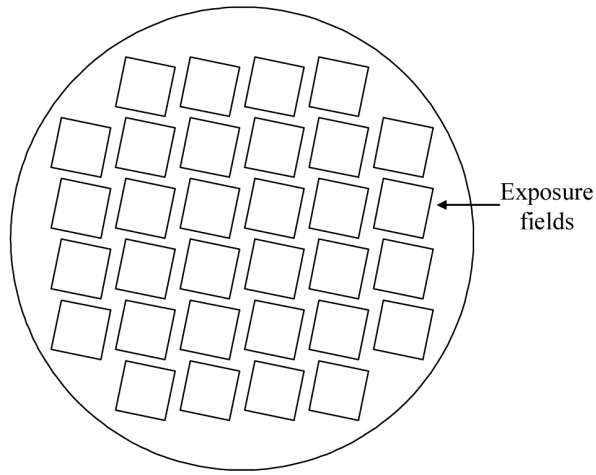


Figure 6.5 The pattern that results when the reticle is rotated with respect to the wafer stage.

align subsequent layers back to the new alignment targets. Alignment schemes are discussed in more detail later in this chapter.

After the positions of the alignment marks are determined, the next step toward achieving layer-to-layer overlay is to align the X axis of the pattern on the wafer with the X axis of the stage, similar to the requirement for the reticle (Fig. 6.3). This requires a wafer rotation that is typically accomplished by rotating the chuck on which the wafer is mounted. The wafer can then be moved, using the laser interferometer to precisely measure the wafer's movements in X and Y, and to position the wafer correctly under the lens, so that the patterns of new exposures will overlay pre-existing patterns. Wafer stage rotation is measured using multiple beams in the interferometer (Fig. 5.37).

After the existing wafer patterns are made parallel to the stage's X axis, the wafer is translated using the interferometrically controlled stage. The amount of translation between alignment and exposure is dependent upon the designed relationship of the alignment targets and the exposure fields. By aligning to two marks, four pieces of positional information are acquired: the X and Y positions of both wafer alignment marks. This provides information for the required rotation and translations. However, there is one additional piece of information acquired during the alignment of two marks that is essential for achieving good overlay. The separation between the marks provides a measure of the expansion or contraction of the wafer between masking steps. Wafer expansion or contraction up to several parts per million is quite normal during the processing of silicon wafers, a consequence of the deposition and removal of thin films from the wafer.[4–7] For example, wafer contractions of ~3 ppm of 100-mm wafers were reported for wafers prepared for LOCOS oxide isolation, as a consequence of silicon nitride depositions. Similarly, oxide depositions cause wafers to expand by similar amounts. The importance of correcting for these expansions or contractions is the subject of Problem 6.1.

With two alignment marks, it is possible to correct for wafer rotation, X and Y translation, and isotropic expansion; i.e., it is possible to correct for expansion that is the same in the X and Y directions. With three or more alignment marks, it is possible for the stepper to correct for different expansion in the two axes if the alignment marks are not colinear with a single axis. Moreover, by choosing multiple alignment marks that are not all colinear on a single axis, it is possible to correct for rotation of the patterns on the wafer with respect to both axes of the stage. Such may be necessary when a stepper, either the one used for the pre-existing pattern or the one used for patterning the overlaying exposure, has X and Y stages that do not move orthogonally, i.e., at 90 deg to each other.

6.1.1 Classification of alignment systems

There are many different types of alignment systems and approaches to alignment. Alignment systems can be classified in a number of ways as listed in Table 6.3. Every alignment method can be classified within all three categories of alignment method given in Table 6.3. For example, an alignment system might be a dark-field, off-axis system, and also use enhanced global alignment.

6.1.2 Optical methods for alignment and wafer-to-reticle referencing

One type of classification is the optical method used to locate the alignment marks. Most commonly, the wafers are imaged through microscope optics using bright-field or dark-field methods that parallel common modes of imaging in microscopes. With the diffraction method, beams from a diffraction grating are used to locate positions on the wafer.

The original GCA 4800DSW had an off-axis, white-light, bright-field microscope that could inspect two alignment marks on the wafer (global alignment) at high magnification with a binocular viewing head. In later versions, the images of the alignment marks were displayed on a black-and-white TV monitor. A pair of alignment marks consisting of crosses (+) separated by three inches on the wafer were viewed and moved manually by the joystick control of the stage to fall within fiducial marks inside the microscope's optics. The reticle was aligned to the optical column by a separate alignment microscope and manipulator system. Once the

Table 6.3 Classification of alignment methods.

Alignment classification	Types
Optical method	Bright field Dark field Diffraction
Referencing of wafer to reticle	On axis or through-the-lens (TTL) Off axis
Number of alignment sites	Global Enhanced global Die by die

wafer was aligned to the fiducial marks, it was moved a known distance of several centimeters to the appropriate location for exposure. This was done using the laser interferometrically controlled stage to ensure precise motion control. If the reticle and wafer were perfectly aligned and if all the distances were as designed, then the wafer was well aligned. The overlay capability of this system was approximately ± 0.7 μm.

There were a number of inadequacies with the alignment system in the first steppers. Alignment was accomplished manually, both in terms of having the operator visually ascertain when the wafer alignment marks were positioned properly relative to the fiducials, and in having the operator move the wafer manually using a joystick. The reticle was also aligned manually and positioned using mechanical manipulators. Today's wafer steppers use automatic forms of optical pattern recognition, and wafer motion is controlled using automatic feedback between the pattern recognition system and the precise wafer stage.

Because the wafer was not aligned directly to the reticle in the GCA DSW4800, periodic calibrations involving test wafers were required to ensure that the individual wafer and reticle alignment provided the required overlay.[8] Moreover, the *baseline*—the distance between the align position and the expose position—in this stepper was unstable with respect to temperature and barometric pressure, and varied widely in time. This baseline error severely limited both the overlay performance and the productivity of this stepper since frequent calibrations were required to maintain adequate overlay.

Modern steppers that use off-axis alignment have automated self-calibration that essentially eliminates the baseline error effectively and efficiently.[9] The highly precise laser stage again plays a central role in the calibrations. In a typical implementation, an artifact containing a wafer-style alignment mark and a detector is placed on the wafer stage.[10,11] The separation between the alignment mark and the detector must be known very accurately. The detector is designed to detect marks located on the reticle and must be responsive to the exposure wavelength. By shining light at the exposure wavelength through the reticle marks and locating them with the stage-mounted detector, the position of the reticle image is determined in the coordinate system of the stage. For excimer laser steppers, the self-metrology must be synchronized with the detection of pulsed light. Upon determining the position of the reticle, the position of the wafer alignment system is identified by aligning the wafer-style mark on the artifact using the wafer alignment system. The distance between the wafer alignment system and the reticle image is measured using the laser stage, which provides an extremely accurate determination of the baseline. Because the self-calibration can be done quickly, the baseline can be updated frequently, thereby ensuring good alignment at all times and without suffering the significant loss of productivity that accompanied early off-axis alignment systems. By using light at the exposure wavelength to align the reticle, challenges related to chromatic correction for performing the reticle alignment are avoided. An excellent summary of the issues associated with off-axis alignment has been written by van den Brink, Linders, and Wittekoek.[12]

Systems have also been developed for aligning wafers directly to the reticle.[3,13–15] There are a number of challenges associated with this approach, because direct wafer-to-reticle alignment must be accomplished through the projection optics instead of a dedicated alignment microscope. In order to achieve the best imaging, the projection optics are designed to image over a very narrow range of wavelengths. Through-the-lens (TTL) alignment must necessarily avoid broadband illumination. The exposure wavelength is the simplest to use for TTL alignment since this is the wavelength at which the projection lens has minimum aberrations. However, there are disadvantages with alignment of wafers at the exposure wavelength (see Problem 6.2). Other wavelengths can be used for alignment, but they require additional optics to correct for chromatic aberrations. These must be redesigned for each new family of projection optics. Moreover, to accommodate TTL alignment at nonactinic wavelengths, it may be necessary to compromise some aspect of imaging at the exposure wavelength. For example, antireflection coatings on the lens elements may no longer be optimal at the exposure wavelength in order to avoid spurious reflections at different wavelengths used for alignment. In spite of such problems, the disadvantages associated with alignment at the exposure wavelength have been found to outweigh the benefits, and late-generation TTL alignment systems operated at nonactinic wavelengths. Monochromatic illumination used for alignment may lead to thin-film interference effects for wafer reflectance signals when faced with the normal range of thickness variation in films like nitride or oxide resist.[16,17] Asymmetries in resist coatings over the alignment marks, when viewed in monochromatic light, may also lead to alignment errors.[18] However, issues associated with thin-film optical effects have been encountered and solved with many off-axis alignment systems that use monochromatic illumination, so this problem is not unique to TTL alignment, nor is it insurmountable.

Until quite recently, the relatively low NA of the stepper lens, which necessarily serves as the objective lens of a TTL alignment system, has had lower resolution than a dedicated alignment microscope. The resolution for the purpose of alignment is further degraded, relative to the imaging resolution, if an alignment system uses a longer wavelength than is used for imaging (see Problem 6.3). This is problematic, since overlay control must be a fraction of the minimum feature size.

For many years, the steppers from ASML used a phase-grating alignment system[19] to directly align reticle and wafer images through the main stepper lens. Instead of relying on an optical-line-edge profile to produce an alignment signal, the ASML alignment system involved a phase grating on the wafer produced by etching an alignment mark into the wafer to produce alternating-phase light (Fig. 6.6). The maximum alignment signal resulted when the two phases differed by 180 deg, something that was achieved by etching the alignment mark to the appropriate depth (see Problem 6.5). Auxiliary optics compensated for optical path differences between HeNe laser light ($\lambda = 632.8$ nm) as used by the alignment system and the actinic light for which the imaging lens was designed (Fig. 6.7). By working with the diffracted signal, it was possible to filter the image to produce

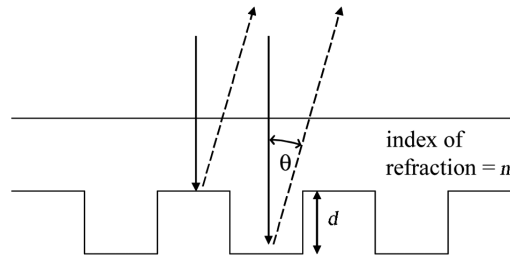


Figure 6.6 A phase-grating alignment mark. In practice, the lower parts of the mark may be filled with a different material than that which covers the higher parts, so there may be an intensity as well as a phase difference between the two sets of beams.

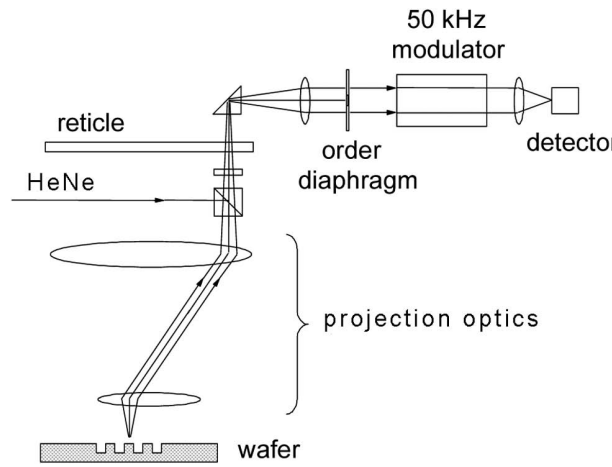


Figure 6.7 ASML's through-the-lens alignment system.

excellent signal-to-noise, with good immunity to small defects and poor linewidth-edge definition. By not relying on edge detection, the resolution requirements of the alignment optics were relaxed considerably (see Problem 6.4). The diffraction alignment method can also be used in an off-axis mode.[20]

ASML also promoted the use of the so-called *zero-level* alignment strategy in which the wafer has alignment marks etched directly into the substrate before any other processing. All subsequent masks are aligned to the zero-level marks. This alignment strategy has some obvious disadvantages, such as requiring additional exposures and wafer processing to produce the alignment marks. However, the zero-level approach has certain advantages, particularly if it is possible to maintain the integrity of the mark through all subsequent processing. This is often possible because the process used to generate the alignment marks can be optimized for the purpose of alignment rather than some overriding circuit device consideration. The ability to align to well-maintained marks is a significant advantage. Moreover, since all layers align to the same marks, certain offsets, such as those that result from placement errors of alignment marks due to lens distortion, do not affect overlay. This can be understood as follows. The absolute position of an alignment

mark is displaced from its designed position due to lens distortion and reticle registration errors. If all process layers are aligned to the same alignment marks, they all have the same registration error due to the lens and reticle errors, and this registration error produces no net overlay error. When alignment marks are generated at many different masking steps, on many different steppers, different errors are introduced at different masking steps. It is possible to correct for these errors, but the corrections can become difficult to maintain over a base of many steppers.

There is another advantage to the zero-layer alignment scheme. If every process layer has alignment variation σ to the zero level, then the layer-to-layer alignment variation is $\sqrt{2}\sigma$ for any pair of process layers. On the other hand, when alignment is sequential (Fig. 6.8), the overlay between two process layers can be larger.[21] For example, the overlay between layers D and A would be $\sqrt{3}\sigma$ with sequential alignment. There are, of course, some disadvantages to the zero-layer scheme. It may involve an extra processing step, and it does not allow for ultimate alignment accuracy of one layer to another since all alignments are secondary.

Because there is no associated baseline error, TTL alignment appears to have an inherent advantage over off-axis alignment systems. However, there are a number of problems that arise during the implementation of TTL alignment systems. As mentioned previously, it is necessary to provide for color correction since there are a number of reasons to avoid aligning at the actinic wavelength.[23] This is not a difficult optical problem to solve, but it does require new alignment optics every time new projection lenses are introduced. It also means that the coatings on the lens elements must accommodate the alignment and actinic wavelengths. As a consequence, both imaging and alignment are ultimately compromised. For these reasons, companies that, in the past, have implemented TTL alignment systems, such as ASML and SVGL, have more recently introduced off-axis alignment systems,[20] often not incorporating the TTL alignment capability at

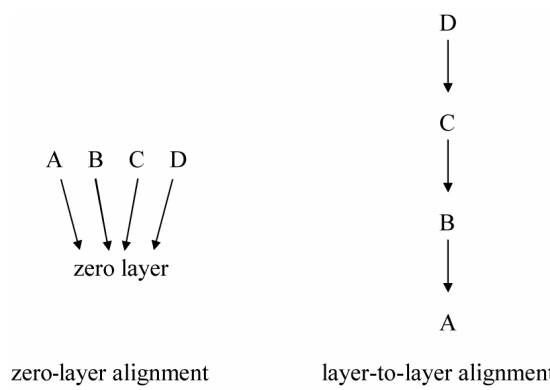


Figure 6.8 Alternative alignment trees. A, B, C, and D are different process layers. It is possible to have hybrid alignment trees.[22] For example, both layers C and D could be aligned to marks generated on layer B.

all. For their newest generations of exposure tools, ASML offers only off-axis alignment capability, similar to what is available from Nikon and Canon.

Most alignment systems employ some type of microscope optics to view alignment marks. The first steppers used bright-field imaging, but these often had significant process sensitivities. For example, resist that coats asymmetrically over alignment marks[24,25] with appreciable topography can induce shifts in the image of the alignment target.[26] As shown in Fig. 6.9, the spin coating of resist over topography does not produce symmetrical films. With nonplanar alignment marks, optical alignment systems are susceptible to thin-film interference effects that can result in low-contrast signals[27] or misalignment.[28] Because spin coating is a radial process, the misalignment goes in opposite directions on diametrical sides of the wafer and results in a scaling error. (Classes of overlay errors are discussed later in this chapter.) Resist-coating processes can be modified to reduce the impact of thin-film optical effects.[29] Optical modeling has shown that dark-field alignment systems are less sensitive to these thin-film optical effects than bright-field or diffraction alignment systems.[30] Because of reduced process sensitivity, alignment systems based on dark-field imaging were introduced in steppers made by Ultratech,[31] GCA,[32,33] and Nikon.[34]

Dark-field alignment systems have problems of their own. For example, with grainy and highly reflective metal, polysilicon, or silicide surfaces, a significant amount of light is scattered into the dark-field cone. Consequently, steppers with dark-field alignment systems also feature bright-field options for use on metal layers. The user can then select, by layer, the alignment optical type that gives the best overlay. Modeling has been used to gain further insight into the acquisition of alignment targets.[35–38] The influence of grains can be minimized by the use of low-coherence illumination (large σ) in the alignment optics,[39] or through the use of interference alignment techniques, which involve the design of the alignment system and are generally beyond the control of the user. A good reference detailing optimization of dark-field alignment on a stepper is given by Smith and Helbert. [40]

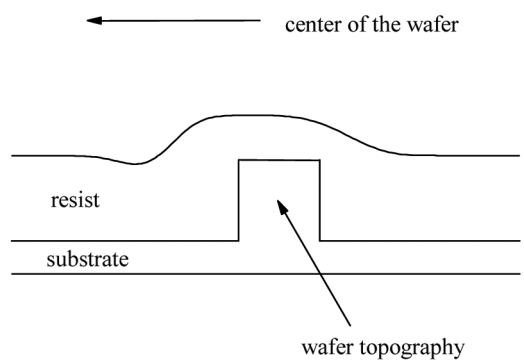


Figure 6.9 Resist films coated over topography will be asymmetric. When this topography is a wafer alignment mark that is detected optically, misalignment can result from the asymmetry.

Misalignment can happen due to poor optics in any alignment system. For example, nontelecentric lenses (to be discussed later in this chapter), in which there is a coupling between defocus and lens magnification, result in alignment variations as focus changes. Asymmetric illumination of alignment marks also produces misalignment, similar in effect to real asymmetry in the alignment marks.

Canon has developed a broadband TV alignment system that uses a high-resolution CCD camera system to align the wafer mark to the reticle mark.[41] The broadband illumination helps to reduce thin-film interference effects, while the CCD camera has extremely small effective pixel sizes. However, the broadband illumination dictates a separate optical system from the projection lens, so this is an off-axis alignment system and must rely on autocalibration to achieve tight registration.

6.1.3 Number of alignment marks

Different numbers of alignment marks per wafer can be used. The original GCA wafer stepper used the minimum number of alignment marks—two per wafer—as did a number of early machines from other stepper suppliers. This often resulted in less-than-optimal overlay because of noise in the alignment signal. One method that was attempted in order to improve overlay was die-by-die alignment, where wafers were aligned to each field just prior to exposure.[15] This had some advantages over competing systems of the time, because these die-by-die alignment systems did not have significant baseline errors. Particular implementations of die-by-die alignment achieved a certain level of success, but it was ultimately a misguided effort because in the most common circumstances, where the poor overlay was caused by noise in the alignment signal, overlay could not be improved by this approach. In fact, die-by-die alignment actually made overlay worse when overlay errors were dominated by alignment signal noise. Moreover, throughput was decreased when compared to two-point global alignment, adding insult to injury. With good stage interferometry and control, the long distances that the wafer stage needs to travel between alignment and exposure is not a problem. Die-by-die alignment is fundamentally advantageous only when the wafer-stage control is worse than alignment noise. It has proven straightforward to build highly precise wafer stages where the operating conditions can be controlled, while wafer alignment targets vary with the process and are therefore difficult to anticipate. The baseline problem has also been overcome by suitable automatic calibration methods.

An alternative to both two-point global alignment and die-by-die alignment is *enhanced global alignment* (EGA), introduced by Nikon.[42] With EGA, usually between five and ten alignment marks are acquired, but then the rest of the procedure is similar to that of two-point global alignment. Enhanced global alignment has the advantage of acquiring more data for the alignment than two-point global alignment, enabling the effect of noise to be reduced by averaging.

The positions of the alignment targets are measured in the coordinates of the stage (X_S, Y_S). On the other hand, the alignment targets are placed at positions

(X_W, Y_W) in the coordinate system of the wafer. In terms of the stage coordinates, the positions of the product die can be determined by assuming that θ in Eqs. (6.3) and (6.4) is small, and the stage and wafer coordinate systems are then linearly related:

$$X_S = S_X X_W - \theta_X Y_W + O_X, \tag{6.5}$$

and

$$Y_S = S_Y Y_W + \theta_Y X_W + O_Y. \tag{6.6}$$

Typically, the coefficients are computed by a least-squares fit to the measured alignment data.[43,44] These equations differ from Eqs. (6.3) and (6.4) in that the coefficients in the two equations are decoupled. This is done so that actual deviations from the ideal situation, such as having nonorthogonal stages or different scales in the X and Y directions, can be dealt with. The two pairs of equations are equivalent when $S_x = S_y = S$, $\theta_x = \theta_y = \theta$, and a small value of θ is assumed. The parameters in Eqs. (6.5) and (6.6) have similar significance for the alignment operation as the factors in Eqs. (6.3) and (6.4). For example, the coefficients θ_X and θ_Y represent the amount the wafer needs to be rotated, bringing the wafer's X and Y axes parallel to the stage X and Y axes, respectively, in the absence of scaling corrections. It is possible for $\theta_x \neq \theta_y$, which occurs when the stage X and Y axes are not perpendicular for all steppers, and similarly, $S_x \neq S_y$ when the distance scales are different between the X and Y axes. The terms O_x and O_y still give the distance between the center of the stage and wafer coordinate systems. With the coefficients in Eqs. (6.5) and (6.6), the positions of the product die and exposure fields can be known in the coordinates of the stage, enabling the exposure fields to be moved into proper position for exposure.

Enhanced global alignment has proven to be extraordinarily effective and has been adopted on commercial steppers today as a standard alignment scheme. Enhanced global alignment gives better overlay than die-by-die alignment whenever the precision of the laser stage is greater than the precision of the stepper's ability to capture alignment signals. In particular, it has reduced sensitivity to degraded individual alignment targets.

6.2 Overlay Models

Lithographers need to characterize overlay errors in order to make improvements. Overlay errors can be considered in terms of a hierarchy (Fig. 6.10). The most fundamental errors are those that occur when only a single stepper and ideal substrates are used, and where the latter provide high signal-to-noise alignment signals. This basic set of overlay errors is well described by overlay models, which will be discussed in detail shortly. When more than a single stepper is used, an additional set of overlay errors is introduced, referred to as matching errors. Finally, there are process-specific contributions to overlay that can result

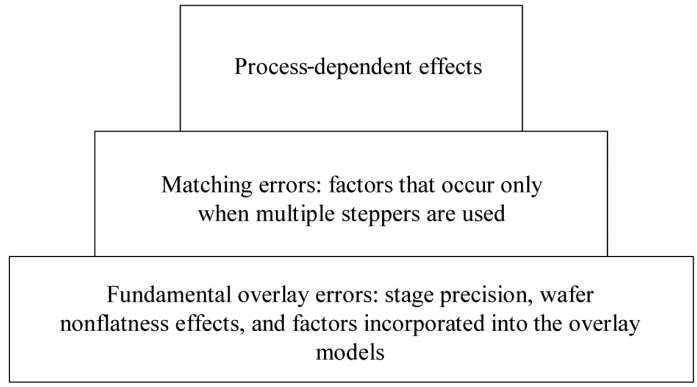


Figure 6.10 Hierarchy of overlay errors. The wafer stage and linear errors comprise the foundation.

in nonideal alignment targets. A full accounting of overlay errors in terms of a hierarchy is useful for assessing and improving overlay,[45] and each of these issues is discussed in the remainder of this chapter.

The control of overlay requires the use of mathematical models because such models are employed directly in the software of wafer steppers when aligning wafers and transforming wafer positions to stage coordinates (as described in the previous section). There is a necessary interplay between the model parameters used to evaluate overlay errors and the model parameters of the stepper's internal software. The overlay models conform to the physics of wafer steppers, either step-and-repeat or step-and-scan, as appropriate. There are two categories of overlay errors: intrafield, which involve the variation of overlay errors within exposure fields, and interfield, which are the errors that vary from exposure field to exposure field across wafers. Both categories are discussed in this section.

Interfield errors can be understood by considering the overlay error at equivalent points in each exposure field. Conventionally, this is the center of the exposure field. Let (X, Y) be the coordinate of a point on the wafer corresponding to the center of some exposure field. The center of the wafer is the most convenient location for the origin of the coordinate system, and is placed there for the following discussion. Overlay errors vary across each wafer, so one can consider the overlay errors at the exposure field whose center is located at position (X, Y) to be given by

$$\Delta X = f_X(X, Y), \tag{6.7}$$

and

$$\Delta Y = f_Y(X, Y), \tag{6.8}$$

where ΔX is the overlay error in the X direction and ΔY is the error in the Y direction. The overlay error ΔX can be expressed in a MacLaurin series of the

function f_X:

$$\Delta X = T_X + e_x X - R_X Y + \text{higher order terms.} \tag{6.9}$$

The sign prior to the $R_X Y$ term is negative to put these in the same form as the alignment equations [Eqs. (6.3) and (6.5)]. There is a similar expression for ΔY:

$$\Delta Y = T_Y + e_y Y + R_Y X + \text{higher order terms.} \tag{6.10}$$

If only linear terms are retained, then Eqs. (6.9) and (6.10) become

$$\Delta X = T_X + e_x X - R_X Y + e_x \tag{6.11}$$

and

$$\Delta Y = T_Y + e_y Y + R_Y X + e_y, \tag{6.12}$$

where e_x and e_y are the residual errors that do not conform to the model. The most common source of residual error is the imprecision of the stepper's stage, which is on the order of 5 nm or less (3σ) for the current generation of wafer steppers.[46] Another source of nonmodeled error arises when multiple steppers are used, which is the issue of matching that is discussed in the next section.

The linear terms in Eqs. (6.11) and (6.12) have physical meanings. Each layer can be considered as patterns printed on a rectangular or near-rectangular grid. The parameters T_X and T_Y represent translation errors in the X and Y directions, respectively, and indicate an overall shift of the printed grid relative to the grid of the substrate layer. The factors E_X and E_Y are scale errors, which represent the errors made by the stepper in compensating for wafer expansion or contraction. Scale errors are dimensionless and are usually expressed in parts per million. The coefficients R_X and R_Y are rotation factors. When $R_X = R_Y$, one grid is rotated relative to the other, and this accounts for the sign convention chosen in Eq. (6.11). If the angle between the axes of the grids is not equal, then $R_X \neq R_Y$. This latter error is referred to as an orthogonality error since it represents the situation in which the grid for at least one of the layers (substrate or overlaying) has nonorthogonal axes. In the presence of orthogonality errors, one can still talk of a grid rotation error, given by

$$\frac{R_X + R_Y}{2}. \tag{6.13}$$

The rotation prefactors are sines of angles, but since

$$\theta \approx \sin\theta, \tag{6.14}$$

for small θ, the rotations are usually expressed in radian measure, typically in units of microradians. This linear model for characterizing grid overlay errors [Eqs. (6.11) and (6.12)] was introduced by Perloff,[44] and it is still widely used today.

Stage precision is typically measured by stepping a pattern on a wafer and then printing a second pattern on top of the first. These exposures are performed without removing the wafer from the chuck or removing the reticle between exposures. The reticle used for this testing typically has both parts of a structure that can be used for measuring overlay between two layers. (Such structures will be discussed in more detail in Chapter 9.) The two parts of the structure are offset on the mask, so the wafer is shifted when it is placed under the lens for the second exposure. Thus, stage precision means that a wafer can be moved from one position to another many millimeters away and then returned to the original position to within 5 nm or better (3σ). This is an extraordinary level of mechanical control.

Nonlinear overlay errors can arise. While these are typically smaller than the linear ones, they may not be insignificant. One source—matching—is discussed in the next section. Another observed example of nonlinearity was caused by the heating of wafers during i-line exposures resulting in nonlinear errors along with a wafer scaling–error component.[47] A wafer may also change temperature during the alignment and exposure operations if the wafer is at a temperature different from that inside the stepper environmental enclosure when the wafer is first moved into the stepper and is not allowed to come to thermal equilibrium prior to processing[48] (see Problem 6.6). Wafers can experience nonlinear plastic distortions, often caused by rapid thermal processing,[49] which are not corrected by alignment methods that account only for linear errors. Nonlinear errors are not corrected by the software of most steppers, which use a linear model, though there have been proposals for incorporating nonlinear effects.[50] Some of the errors that are not included in the model of Eqs. (6.11) and (6.12) are discussed in the next section.

Overlay errors may also vary across each exposure field. These types of errors are referred to as intrafield errors, examples of which are illustrated in Fig. 6.11. Consider, for example, a magnification (or reduction) error. The lens reduction typically has nominal values of 4:1 or 5:1, but the magnification deviates from these nominal values by some small amount. When there is wafer expansion or contraction between masking steps it is necessary for the stepper to be programmed to measure this change and compensate for it, not only in grid terms, but in the size of each exposure field as well. Errors in this correction result in magnification errors. The other primary intrafield errors are shown in Fig. 6.11.

Intrafield overlay models for wafer steppers were introduced by MacMillan and Ryden in 1982.[51] Since then, the most significant change has been the introduction of models for step-and-scan systems. The models for step-and-scan systems include parameters for asymmetric magnification and skew, which are not relevant for step-and-repeat systems.

Suppose the same stepper is used for printing two layers, one overlaying the other. The overlay error at any point on the wafer is the sum of the grid errors, defined previously as the overlay errors at the center of the exposure field, and

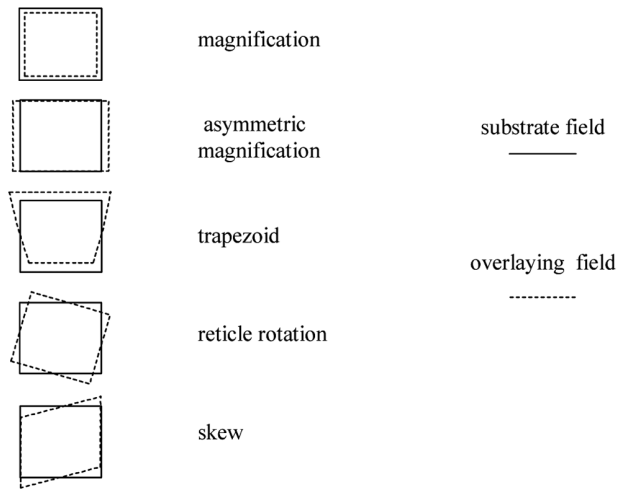


Figure 6.11 Intrafield overlay errors deviate from these nominal values by some small amount.

overlay errors that vary across each exposure field. Let (x, y) be the coordinates of a point on the wafer, relative to the center of the exposure field in which it is contained (Fig. 6.12). For a step-and-repeat system, the intrafield overlay errors are modeled as[51]

$$\delta x = t_x + mx - ry + T_x xy + T_y x^2 + e_x, \tag{6.15}$$

and

$$\delta y = t_y + my + rx + T_y xy + T_x y^2 + e_y. \tag{6.16}$$

In these equations, the parameter m represents a magnification error. It is dimensionless and is usually expressed in parts per million. The magnification error can be interpreted as follows. The error in overlay increases linearly with distance from the center of the exposure field. The rate at which this error increases with distance from the center of the exposure field is the magnification error coefficient m. For step-and-repeat systems the magnification error is the same in x and y directions. The coefficient r represents reticle rotation. As with wafer rotation, this is usually expressed in microradians. Reticle-rotation errors also increase linearly with the distance from the center of the exposure field, and the error coefficient is the same for errors in x and y for step-and-repeat systems. The terms e_x and e_y are residual errors that do not conform to the model. The factors T_x and T_y are trapezoid error coefficients, and are discussed next.

Trapezoid errors result from magnification that varies across the exposure field and results from the phenomenon of perspective, where more distant objects appear smaller in size (Fig. 6.13). When reticles are not perpendicular to the optical axis, some portions of the exposure field are further away from the lens than others. For

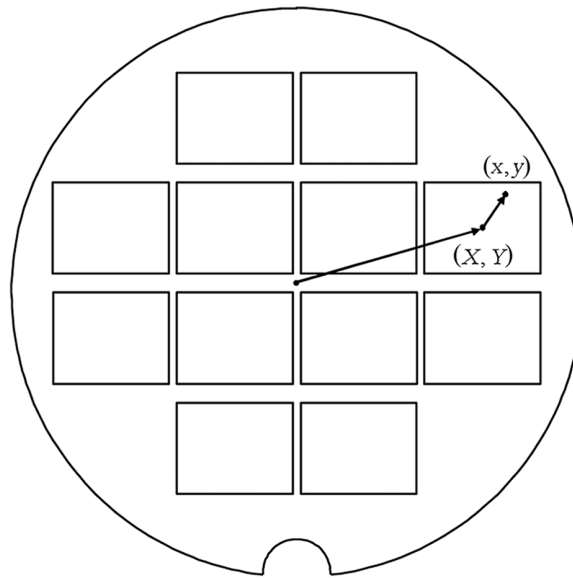


Figure 6.12 Intrafield overlay is measured relative to the center of each exposure field.

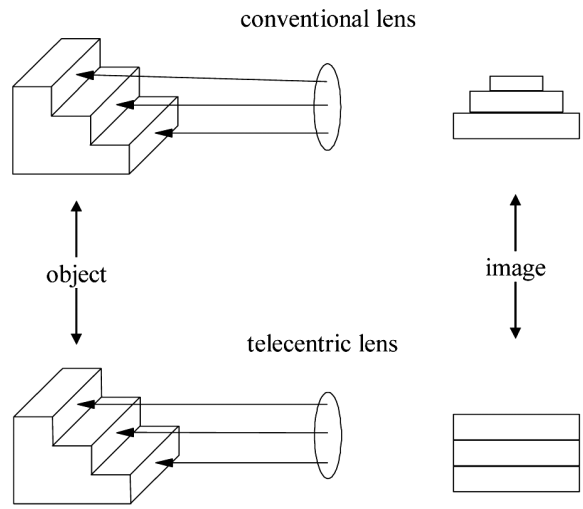


Figure 6.13 The imaging of telecentric and nontelecentric lenses.

normal lenses this results in a variation in magnification across the exposure field—that is, trapezoid errors. Telecentric lenses have magnification independent of the distance between the object and lens over a limited range of separations. Stepper lenses have always been telecentric on the wafer side (magnification independent of wafer focus), but newer generations of lenses are also telecentric on the reticle side. These double-telecentric lenses are free of trapezoid errors.

Trapezoid errors introduce an anomaly into the overlay models. If all intrafield errors are symmetric about the origin, then the intrafield errors average to zero (by definition of being relative to the center of the exposure field) up to measurement

noise, at least when considering overlay where the same exposure tool is used for both layers. As one can see in Fig. 6.11, trapezoid errors are not symmetric with respect to the center of the exposure field. The introduction of trapezoid to the model requires the inclusion of translation terms t_x and t_y to adjust for the center offset of the trapezoid errors.

Grid (interfield) models are identical for step-and-repeat and step-and-scan systems, but the appropriate intrafield models differ. For step-and-scan systems, the intrafield model is

$$\delta x = m_x x - r_x y + e_x, \tag{6.17}$$

and

$$\delta y = m_y y + r_y x + e_y. \tag{6.18}$$

The step-and-scan model is similar to the step-and-repeat model for double-telecentric lenses, with some differences. Most notably, the magnification coefficients (m_x and m_y) and rotation coefficients (r_x and r_y) in Eqs. (6.17) and (6.18) are independent in the two equations and are not coupled, as in Eqs. (6.15) and (6.16).

The magnification for step-and-scan systems can be different in the x and y directions because the magnification in the x direction has a different origin than the magnification in the y direction. The magnification in the direction of the scan is determined by the relative speeds of the reticle and wafer stages. Ideally, the reticle stage should move N times faster than the wafer stage, where $N : 1$ is the reduction ratio of the lens (typically, $N = 4$.) When the reticle stage speed is faster than N times the wafer stage speed, the printed field is shorter in the scan direction. This can be understood as follows. Suppose the scanning occurs for a time t, during which the entire reticle field is scanned. The length of the field scanned on the reticle, in the scan direction, is given by

$$L_R = v_R t, \tag{6.19}$$

where v_R is the speed at which the reticle is scanned. Similarly, the size of the field on the wafer, in the scan direction, is given by

$$L_W = v_W t, \tag{6.20}$$

where v_W is the scanning speed of the wafer. The reduction in field size, in the scan direction, from reticle to wafer is

$$\frac{L_R}{L_W} = \frac{v_R}{v_W}. \tag{6.21}$$

If $v_R > Nv_W$, then the field reduction in the scan direction is larger than N. The magnification in the direction perpendicular to the scan is determined only by the lens-reduction factor ($\sim N$:1), since the image size in that direction is defined entirely by the reduction optics.

For step-and-scan systems, intrafield rotation issues are more complicated than for step-and-repeat systems. For step-and-repeat exposure tools, intrafield rotation involves the relative orientations of the wafer stage, reticle, and existing patterns. Step-and-scan systems have an additional component, the reticle scanning stage. Pure reticle rotation occurs when the scans of the reticle and wafer stages are parallel, but the reticle is rotated relative to both. The origin of field skew is left to the reader to derive in Problem 6.7.

For any point on the wafer, the overlay error is the sum of the interfield and intrafield errors. In the X direction, the total overlay error is $O_{X,x} = \Delta X + \delta x$, and the overlay error in the Y direction is $O_{Y,y} = \Delta Y + \delta y$, at the point $(X + x, Y + y)$ (see Fig. 6.12). For step-and-scan systems:

$$O_{X,x} = T_X + E_X X - R_X Y + m_x x - r_x y + \rho_{X,x}, \tag{6.22}$$

and

$$O_{Y,y} = T_Y + E_Y Y + R_Y X + m_y y + r_y x + \rho_{Y,y}, \tag{6.23}$$

where $\rho_{X,x}$ and $\rho_{Y,y}$ are total residual errors. There are similar equations for step-and-repeat systems.

Classifying errors is important for controlling overlay, because intrafield and interfield overlay errors generally arise from different causes. For example, reticle-rotation errors involve the alignment of the reticle, while wafer-rotation issues involve separate alignments. In order to classify overlay errors the coefficients in Eqs. (6.22) and (6.23) need to be determined. The most common method for extracting these coefficients from measured data $(O_{X,x}, O_{Y,y})$ is the method of least squares.[52,53] In this technique, more data are needed than there are parameters. For Eqs. (6.22) and (6.23), measurements need to be made at two points per exposure field or more, in order to model intrafield effects. (Each measurement site provides two measurements, one in X and one in Y.) The least-squares method determines the set of coefficients $\{T_X, T_Y, m_x, \ldots\}$, which minimizes the sum of squares of the residual errors:

$$\sum_{X,x,Y,y} \left(\rho_{X,x}^2 + \rho_{Y,y}^2\right), \tag{6.24}$$

where the sum is extended over all measured points. Computational speed and insensitivity to noise are among the reasons why the least-squares method is the mathematical technique most commonly used. Coefficients obtained from the least-squares method do not represent local minima. The choice of mathematical

methods used to determine the model coefficients is further discussed later in this chapter.

Example: Misregistration of patterns on reticles can produce overlay errors. An engineer wants to compensate for reticle registration errors measured on a product reticle. Errors on the reticle that fit the model of Eqs. (6.15) and (6.16) were corrected on step-and-scan systems. Constants were added to each of these equations as well, to correct for translation shifts. The reticle errors, as measured on the product reticle, are given in Table 6.4. The model coefficients are given in Table 6.5. As can be seen, a substantial fraction of the reticle error is eliminated by applying magnification, rotation, and translation offsets. For example, the registration error at the position $(-69.5, 0)$ is given by

$$\delta x = X\text{-translation} + m_x x - r_x y + e_x, \tag{6.25}$$

$$= -10.6 + 0.963 \times 69.5 - 0.329 \times 0.0 + e_x, \tag{6.26}$$

$$= 56.3 + e_x. \tag{6.27}$$

The measured error is 63 nm, so the model was able to account for most of the registration error at that point, and most of the error was therefore correctable.

Table 6.4 Reticle registration errors. The first two columns are the positions where registration is measured. The third and fourth columns are the registration errors, as measured, and the last columns are the residual errors, after the modeled part of the errors is subtracted from the raw data.

Measurement positions		**Measured registration errors**		**Residual registration errors:** step-and-scan model	
x **(mm)**	y **(mm)**	δx **(nm)**	δy **(nm)**	e_x **(nm)**	e_y **(nm)**
−69.5	0.0	63.0	0.0	6.7	−9.7
69.5	0.0	−63.0	0.0	14.5	−12.9
−30.2	51.0	17.0	−15.0	15.3	−2.2
30.2	51.0	−14.0	1.0	42.4	12.4
30.2	−51.0	−21.0	30.0	1.9	−5.4
−30.2	−51.0	48.0	29.0	12.8	−5.0
−29.9	50.8	−41.0	−16.0	−42.5	−3.3
29.9	50.8	−82.0	−7.0	−25.9	4.3
29.9	−50.8	−61.0	39.0	−38.3	3.7
−29.9	−50.8	48.0	52.0	13.1	18.1

Table 6.5 The model coefficients, determined by least squares, for the reticle misregistration data in Table 6.4.

x **translation**	**−10.6 nm**	y **translation**	**11.3 nm**
m_x	−0.963 ppm	m_y	0.023 ppm
r_x	0.329 μrad	r_y	−0.459 μrad

The step-and-scan models considered to this point have been purely linear. From Eqs. (6.9) and (6.10) it is readily seen that the interfield terms can be expanded to include nonlinear terms, and a similar approach may be taken to characterize intrafield overlay errors. A nonlinear model was used in the early days of stepper lithography to describe mix-and-match overlay between a stepper and a Perkin-Elmer scanner.[54] With occasional exceptions,[55] only linear terms were considered sufficient for controlling overlay until approximately 2001, at which point interest in nonlinear models began to increase. Part of this increased interest was the result of tightened requirements for overlay, beyond the typical 0.7× reduction node-to-node. With shrinking, everything did not scale at the same pace. As a consequence, overlay went from being ~1/3 of the half pitch to ~20% of the half pitch (see Table 6.1). At the same time the difficulty of improving overlay also increased, and it became necessary to find improvement wherever possible. The introduction of 300-mm wafers appears to have been accompanied by a higher level of nonlinear wafer distortion induced by processes such as film deposition and thermal annealing.[56] After years of concentrated efforts to minimize linear errors, the nonlinear components and residuals became a larger fraction of the total overlay errors.[57]

The need to address nonlinear overlay errors requires a substantial increase in the number of wafer alignments and overlay measurements. If one can legitimately assume purely linear and isotropic errors, then only two across-wafer and two within-field alignments are needed, and only a few more measurements are required if parameters such as field magnification require asymmetric compensation. On the other hand, a much larger number of measurements is needed when there are substantial nonlinear contributions to the overlay error, potentially reducing scanner throughput as well as increasing metrology costs. If the nonlinear signature is reproducible wafer-to-wafer, at least within a single lot of wafers, then a large number of alignment and measurement sites can be used on one wafer to determine this signature, which can then be applied to other wafers on which fewer alignments are made. Nonlinear contributions are also a significant contributor when two layers, between which overlay is a concern, are exposed on a different scanner. This problem that arises from using more than one scanner is the topic of the next section.

6.3 Matching

A set of nonrandom overlay errors not included in the overlay models are found in situations where more than one stepper is used. These errors are referred to as matching errors since they refer to the degree on which the pattern placement produced on one stepper matches that of other steppers. There can be grid and intrafield matching errors. Grid matching errors arise from absolute stepping errors. While stepper stages are extremely repeatable, their stepping may deviate at each stage position, on average, from a perfect grid. When a single stepper is used for both the substrate and overlaying layers, these average deviations cancel out. A different situation arises when different steppers are used for the two layers. Some

differences in the stepping of different stages are correctable. For example, scale errors can arise when the beams of the interferometer are not exactly perpendicular to the mirrors on the stage (Fig. 6.14). The laser beam travels a distance of $2h$, while the stage travels a distance d. The system is designed on the assumption that $h = d$, while actually

$$h = \frac{d}{\cos\theta}. \tag{6.28}$$

This indicates a scaling error proportional to $\cos\theta$, which is why these errors are designated as *cosine errors*—being a linear error, it can easily be corrected.

It is also possible that the X and Y steppings are not truly perpendicular on one machine, resulting in a grid that is not perfectly orthogonal. This is correctable through software if the magnitude of the error is known. Stability of orthogonality is important, so stage mirrors are usually made from thermally stable materials, such as Zerodur, and it is preferable to have the X and Y stage mirrors fabricated from a single block of material to maximize stability.[58]

Nonlinear grid-registration errors originate in the nonflatness of the stage mirrors. Consider the stage mirrors drawn in Fig. 6.15. The branch of the interferometer that measures the stage's position in the Y direction measures apparent movement in the Y direction as the stage is moved from left to right because of the mirror's nonflatness. When a single stepper is used, the resulting stepping errors, relative to a perfect grid, do not lead to overlay errors because these stage-stepping errors occur for all layers and cancel each out. Because mirror nonflatness arises from mirror polishing, the nonflatness generally varies from stepper to stepper, and stage mismatch occurs. For modern wafer steppers, the stage mirrors are flat to approximately $\lambda/40$, or about 16 nm, where $\lambda = 633$ nm is the wavelength of light of the stage's interferometer.

Grid-registration errors can be corrected using software.[59] Typically, reference-matching wafers are used for matching all steppers in a facility to the same positions. Look-up tables can be used for these corrections, or the grid models can

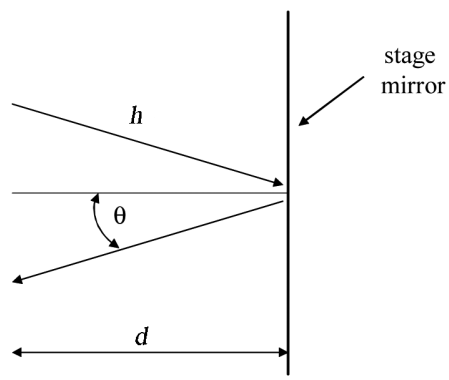


Figure 6.14 Cosine error in a stage interferometer.

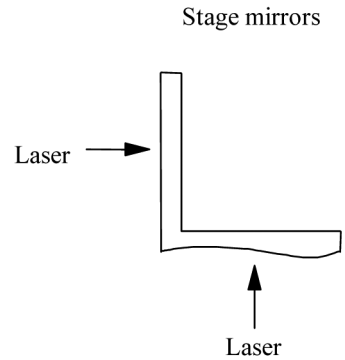


Figure 6.15 Stage mirror nonflatness that leads to nonlinear stage-matching errors.

be expanded to higher orders. The second-order interfield term is *wafer bow*.[60] Matching grids to less than 5 nm in not uncommon. Laser interferometers for wafer steppers usually have more than one beam for each direction in order to measure stage rotation during stepping. Mirror nonflatness can then lead to intrafield rotation as fields are stepped across wafers[55] that are not captured by models that treat the reticle rotation the same in every field. This is eliminated by proper accounting of mirror nonflatness.

Lenses do not always place geometries where they belong. For example, suppose the pattern on the reticle consists of a perfect rectangular grid (Fig. 6.16). Because of the lens-placement errors, the vertices of the pattern are slightly moved. Placement errors run from 200 nm on older lenses to <10 nm on contemporary machines. These errors may be inherent in the design or result from fabrication imperfections. The design contribution is referred to as lens distortion and has a known functional form for static exposures:

$$\Delta r = D_3 r^3 + D_5 r^5, \tag{6.29}$$

$$\Delta x = D_3 x r^2 + D_5 x r^4, \tag{6.30}$$

$$\Delta y = D_3 y r^2 + D_5 y r^4, \tag{6.31}$$

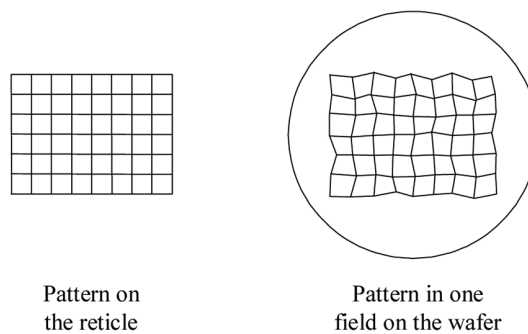


Figure 6.16 Lens-placement errors.

where r is the radial distance from the center of the lens field,

$$r = \sqrt{x^2 + y^2}. \tag{6.32}$$

Because of the magnitude of the exponents of terms in Eq. (6.29), D_3 and D_5 are referred to as third- and fifth-order distortion, respectively. Systematic distortion in scanning systems results from the averaging of Eqs. (6.30) and (6.31) across the slit during the scan. This leads to image blurring, so it is important that distortion be low in the lenses used for scanning lithography.

Distortion tends to be fairly constant within a particular lens family. However, there are strong economic reasons to minimize the numbers of very high-performance steppers, and to mix-and-match steppers of different types.[61] Consequently, lenses of more than one type are usually used to fabricate semiconductor devices. Older generations of projection optics had nonzero levels of distortion in the design, where the magnitude of Δx and Δy, due to distortion alone, could be as large as several tens of nanometers, sometimes up to 200 nm. In such situations good matching could be achieved only within given lens families, particularly when different lens designs might have distortions with opposite signs for D_3 and D_5. The designs of modern lenses have negligible levels of radial third- and fifth-order distortion. For these newer lenses, placement errors are due primarily to manufacturing imperfections, which tend to produce intrafield errors that are not systematic, as in the case of third- and fifth-order lens distortion. Polishing variations across the surfaces of lenses and mirrors and inhomogeneity in optical materials can lead to intrafield registration errors. These tend to have random patterns.

While distortion is extremely stable for i-line systems, distortion varies with very small shifts—less than a picometer—in wavelength for DUV systems.[62,63] Consequently, additional controls are needed on DUV machines, the most critical being wavelength control. Modern KrF excimer lasers, the light source for 248-nm exposure systems, contain internal wavelength references that provide good wavelength control.[64] As with anything, this calibration system can malfunction or drift, and there is always the potential for change in third- and fifth-order distortion as a consequence of a small shift in wavelength. Also, changes in barometric pressure change the wavelength of light in the air, and excimer steppers must adjust for this correctly, or third- and fifth-order distortion may be introduced.

For step-and-scan systems the intrafield placement errors also have a scanning component. In some cases, lenses have improved to the point that intrafield placement errors are dominated by the stage scanning.[65] Part of this reduced contribution from the lens is due to the scanning itself, because the static lens-placement errors are averaged during the scan. If the lens-placement error at static field position (x, y) is $\vec{\varepsilon}(x, y)$, then the placement error $\vec{e}(x)$ of the scanned image at

slit position x is[66]

$$\vec{e}(x) = \frac{1}{W}\int_{\frac{-W}{2}}^{\frac{W}{2}} \vec{\varepsilon}(x,y)dy, \tag{6.33}$$

which is illustrated in Fig. 6.17. Note that there is no variation (at a given slit position of x) of the intrafield placement errors in the y direction from the static lens-placement errors if the scanning is perfect. Such errors in the y direction must necessarily result from the scanning.

There are random and systematic contributions to the scanning errors. In this case random means more than failure to conform to some intrafield error model. These random errors vary from field-to-field and are independent of measurement error. These are average errors during the scan, so they lead to overlay errors but not image fading, which results from variations during the scan. This random component does not exist in step-and-repeat systems, so there is more potential for overlay errors with scanners, even for a stepper to itself. Moreover, the stage that scans the reticle can malfunction, another control problem unique to step-and-scan systems.

The intrafield overlay contains some theoretically correctable contributions that can be fit to the overlay model [Eqs. (6.22) and (6.23)]. When using two different exposure tools, one for the first layer and the other tool for the second layer, there are additional components that arise from differences between the two lenses that do not conform to the model. The matching of the lenses is the set of residual overlay vectors remaining after subtracting out the correctable contributions. These sets of vectors are dependent upon the criterion used for determining the model parameters. The least-squares method is used most commonly, but it is not the only reasonable criterion. For example, since devices are usually designed to yield so long as overlay is less than a particular value and fail when the value is exceeded, yield can be maximized when the worst overlay errors are minimized.[67] This criterion consists of the minimization of

$$\max_{x,y}\left(\left|\rho_x\right|,\left|\rho_y\right|\right), \tag{6.34}$$

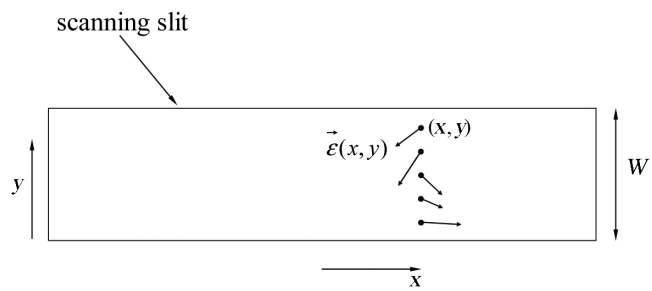


Figure 6.17 Lens-placement errors in the scanning slit.

where ρ_x and ρ_y are the residual errors in Eqs. (6.22) and (6.23). The minimization of Eq. (6.34) differs from the least-squares criterion, which consists of the minimization of the quantity in Eq. (6.24) and typically leads to different model coefficients. We say that lenses are matched when correctable parameters are set to minimize the selected criterion, either Eq. (6.24), (6.34), or another that may be chosen.

The least-squares method has two key advantages. The first is computational simplicity, since model coefficients can be found through simple matrix calculations. The least-squares method also provides transitivity,[68] that is, if Stepper A is matched to Stepper B, and Stepper B is matched to Stepper C, then Stepper A is also matched to Stepper C. Minimization of maximum error does not share this characteristic.[69] This is a practical problem when matching many steppers within a large fabricator.

When only a single stepper is used, the intrafield overlay model accounts for nearly all of the intrafield overlay errors, and the overlay errors throughout the entire field are therefore inferred from measurements at only a few points within the exposure fields. It is quite a different situation when more than one stepper is used. Matching methodology must account for overlay in the areas of the exposure field that are not accessible when measuring product wafers. Specialized structures are normally used for measuring overlay, and these are usually placed in the scribe lanes between product dies. This limits the amount of overlay data collected on product wafers, since most of the exposure field is occupied by product devices, not overlay-measurement structures. This limitation on the number of points within each exposure field at which overlay can be sampled has significant consequences since overlay is measured in areas where the product is not placed.

Suppose overlay is measured on several fields of the wafer, and also within each field at four sites with the same intrafield positions. One can fit the acquired data to the overlay model [Eqs. (6.22) and (6.23)] by minimizing the appropriate metric (least squares, minimum error, etc.). This approach may not optimize the overlay over the entire exposure field, particularly the parts of the field in which the product is located and overlay is not measured. This is demonstrated in the following way.[69] Between two steppers the overlay is measured on a 12×11 grid with 1.95-mm spacings in the x and y directions within the exposure field. The resulting lens matching is shown in Fig. 6.18. In a *gedanken* (thought) experiment, the overlay measurements are considered at only four points among these 132 sites, in a pattern that resembles a typical overlay-measurement plan. Several subsets of four points are considered in Fig. 6.19. For each set of four sites, the overlay model coefficients are recalculated and plotted in Fig. 6.20. As one can see, the resulting model coefficients vary significantly among sampling plans. Recall that the baseline set of coefficients was the one that optimized the overlay over the entire exposure field, not just at the four corner points. By measuring at only four points and adjusting the stepper to minimize overlay at just the measured points, overlay is moved away from the optimum, overall. This problem occurs when different steppers are used for the overlaying of patterns.

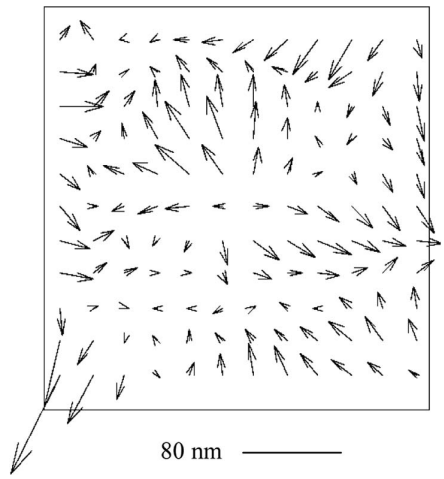


Figure 6.18 Matching between two particular step-and-repeat and step-and-scan systems.

1	3	5	8	10		13	9	8	5	3	1
2	6	12	15	18			18	15	12	6	2
4	11	17	20	22			22	20	17	11	4
7	14	19	23	25			25	23	19	14	7
9	16	21	24	26			26	24	21	16	10
13											13
10	16	21	24	26			26	24	21	16	9
7	14	19	23	25			25	23	19	14	7
4	11	17	20	22			22	20	17	11	4
2	6	12	15	18			18	15	12	6	2
1	3	5	8	9	13		10	8	5	3	1

Figure 6.19 Various sampling plans for measuring overlay. In Sampling Plan #1, measurements are taken in the four extreme corner measurements sites.

In earlier discussions it was mentioned that the production of good devices requires that critical parameters remain within specifications at all locations within the dies. Lens mismatch may result in particular points at which overlay is very poor. Overall overlay may be very good, but the die at the field location where the mismatch occurs yields poorly. An example of this is shown in Fig. 6.18, where the bottom left corner of the field has a point with particularly bad overlay. Overall, the overlay is good, but the die located in that corner consistently has poor overlay. With one to eight die per field being typical, a single bad site within the exposure field can degrade the yield significantly. The concept that good yield requires parameters at all points of a die to be within specifications was first introduced in the context of overlay.[67]

To complicate matters further, it has been observed that lens-placement errors vary with illumination conditions, such as partial coherence[71–73] or off-axis

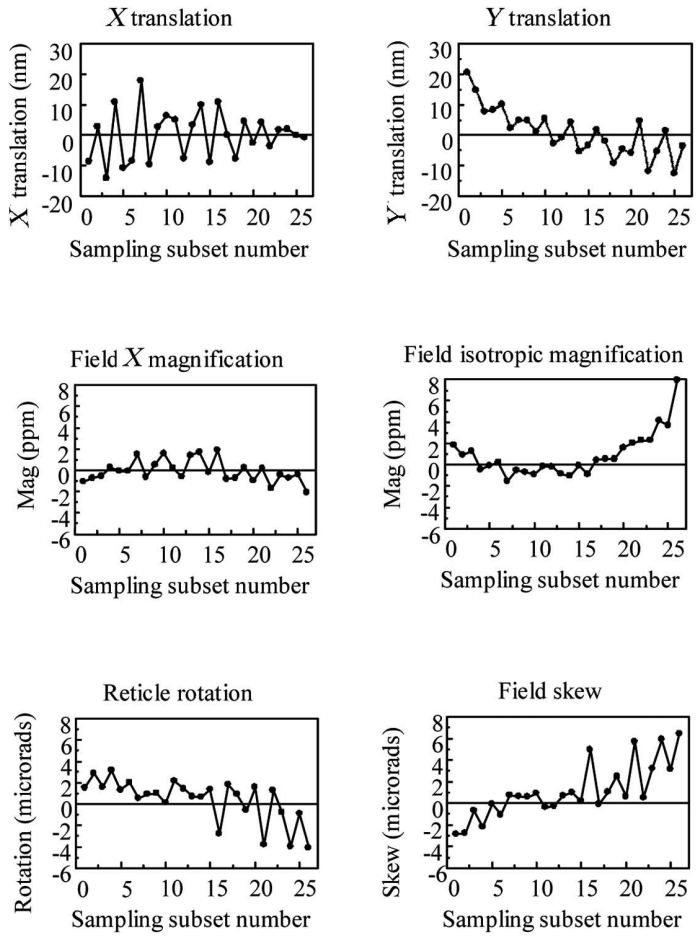


Figure 6.20 Model parameters for different sampling plans of the single data set shown in Fig. 6.18.

illumination.[74–76] For purposes of maximizing depth-of-focus or exposure latitude, different illumination conditions are often used at different layers. For example, large values for the partial coherence are used for gate and metal layers, at least when binary masks are used. Small values of partial coherence are used for contact layers and phase-shifting masks. Consequently, there are intrafield errors, even when the same stepper is used for all layers, if different illumination conditions are used for different layers. The overlay errors between a first layer printed with standard illumination ($NA = 0.63, \sigma = 0.65$) and a second layer exposed using the same numerical aperture, but a partial coherence of $\sigma = 0.3$, are shown in Fig. 6.21. Even though both layers are printed using the same lens, there is considerable intrafield error. Each vector is the overlay error at the point located at the tail of the vector. It should also be noted that the use of a single stepper does not guarantee the nonexistence of grid and intrafield matching errors, because drift, database corruption, and stepper malfunction are possible. However, such events are rare,

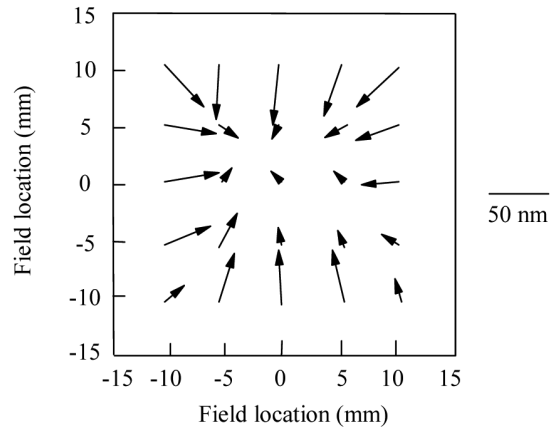


Figure 6.21 The overlay errors between a first layer printed with standard illumination ($NA = 0.63$, $\sigma = 0.65$) and a second layer exposed using the same numerical aperture, but a partial coherence of $\sigma = 0.3$.[69]

and the matching errors discussed here occur infrequently when single steppers are used with fixed operating parameters.

The changes in lens-placement errors caused by variations in illumination conditions can be understood by considering the imaging of a simple diffraction grating, discussed in Chapter 2. With various angles of illumination incidence numerical apertures, and the pitches of the pattern, the light of a grating pattern projected through a lens passes through different parts of the lens. Many aberrations are a result of variations in the polishing of optical surfaces and inhomogeneity in glass materials. Light rays going through one part of the lens have different errors than rays going through different parts of the lens. As a consequence, aberrations vary for light passing through different parts of the lens. For a given feature, the aberrations of its image depends on the particular imperfections in those parts of the lens through which the light from that feature passes. Hence, aberrations vary with pitch, numerical aperture, and the illumination conditions. The light from patterns other than gratings is also distributed throughout the optics with dependence upon feature size and proximity to other features. Since light is diffracted in fairly specific directions with highly coherent illumination and is more spread out with less coherent light, variations in distortion for different pitches and geometries are smaller with reduced coherence (larger values of σ).

To address these issues, at least partly, capability for measuring overlay within product die has been developed.[77] Specialized overlay-measurement marks are still required, but they have been made small enough ($<5\ \mu m \times 5\ \mu m$) that a quantity of them can be placed within product die areas without interfering too greatly with the circuit design. Use of such targets can reduce the impact of incomplete sampling.

As feature sizes shrink, these subtle issues generate overlay errors of significance. For example, certain aberrations, such as coma, result in a feature-size dependency for intrafield registration. This is shown in Fig. 6.22, where simulated

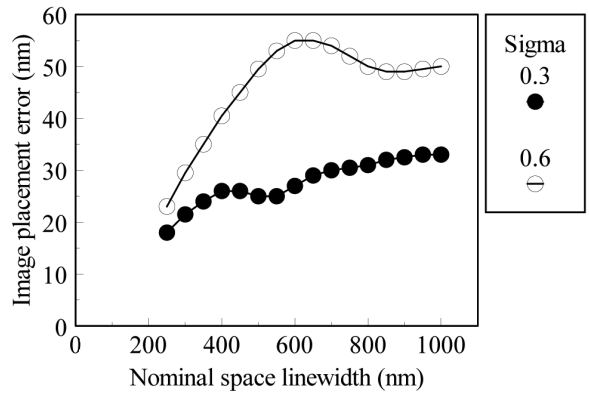


Figure 6.22 Overlay errors from coma.[78]

image-placement shifts of isolated clear lines are plotted against space width, where the lens has $a_7 = 0.035$ waves of coma.[78] In the presence of coma, overlay-measurement structures with large features measure different overlay than actually occurs for critical fine-linewidth features in the product.

This implies that overlay measurements do not represent the overlay of critical features in the circuits. For example, using overlay-measurement structures of different widths, effective field magnification is found in one instance to differ by more than 1 ppm between 2.0-μm and 0.2-μm features.[79] For a 20-mm × 20-mm field, the 0.2-μm features that might be found in the circuit are placed ±20 nm differently at the edges of the field, relative to the positions indicated by the 2.0-μm features that are conventionally used for overlay measurement. Modern lenses have lower levels of aberrations than assumed for the calculations of Fig. 6.22, but image-placement errors of several nanometers still exist between large and small features. Measuring overlay directly using small features is a problem. Overlay is typically measured using optical tools (Chapter 9) that can measure features reliably only if their size is greater than about 0.25 μm. Optical tools are used for measuring overlay in spite of this limitation because they have throughput and cost advantages over measurement tools, such as scanning-electron microscopes, that are capable of measuring smaller features. Subtle issues such as these need to be addressed when nanometers represent a significant fraction of the total overlay budget.

In order to enhance productivity, high-throughput steppers with large exposure fields are often used for noncritical layers, while smaller-field steppers are used for the critical layers. The mixing of these different types of steppers often results in nonconcentric exposure fields (Fig. 6.23). Control of overlay requires extensions of the models discussed thus far, which have assumed that the centers of exposure fields sit approximately on top of each other.[80,81] Overlay needs to be measured at the corners of the critical fields in order to identify overlay errors associated with intrafield errors that are associated with the smaller field. Mathematical models describing field rotation and magnification must take into account the nonconcentric nature of the fields.

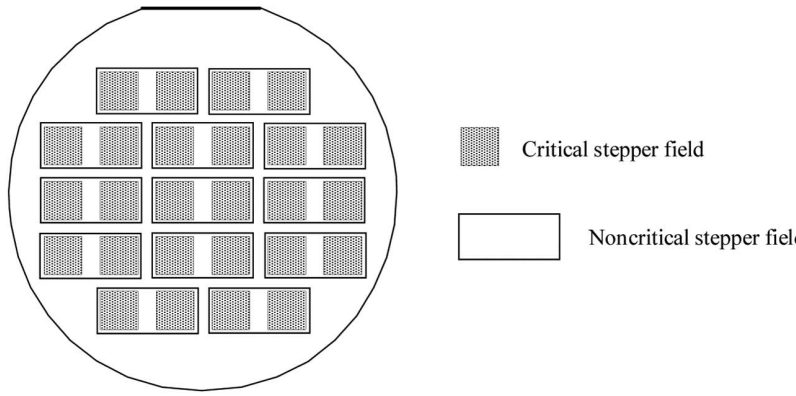


Figure 6.23 Nonconcentric matching.

Misregistration of geometries on reticles results in overlay errors on wafers. This misregistration can have random components, as well as other contributions that vary systematically across the reticles. An example of a systematic variation is one that varies linearly across the reticle. Such linearly varying misregistration will appear as a magnification error. It is possible to correct for systematically varying reticle errors corresponding to adjustable intrafield parameters.[70,82] For example, reticle misregistration resulting from orthogonality errors of the reticle beam writer corresponds to field skew. Such an error is correctable on a step-and-scan system, but not a step-and-repeat machine. Correction of reticle errors was described in an example shown in the previous section.

6.4 Process-Dependent Overlay Effects

The quality of the overlay is very dependent upon the ability of the stepper's alignment system to acquire alignment targets accurately. There is an interplay between the alignment signal and the nature of the alignment targets, which depends upon the overall process for making semiconductors, including film depositions, resist coatings, etches, and polishes. Consequently, lithography engineers need to optimize alignment targets. This optimization includes the dimensions of targets, film stacks, etch depths, resist coatings, and polishing processes.

Sputter deposition of metals can cause asymmetries in overlay targets. Metal ions are partially shadowed by alignment-mark topography, leading to apparent shifts in the position of the alignment mark (Fig. 6.24). Because sputtering geometries are usually radially symmetric, the resulting overlay errors often appear as wafer-scaling errors. The overlay-measurement structures are affected in the same way as the alignment targets, and the overlay errors are apparent only after etching the metal. The asymmetry is affected by the geometries of substrate features, and it is different for large structures used to measure alignment relative to critical features in the product. If the asymmetries are consistent within a lot and lot-to-lot, then overlay offsets may be determined off line and used for correction.

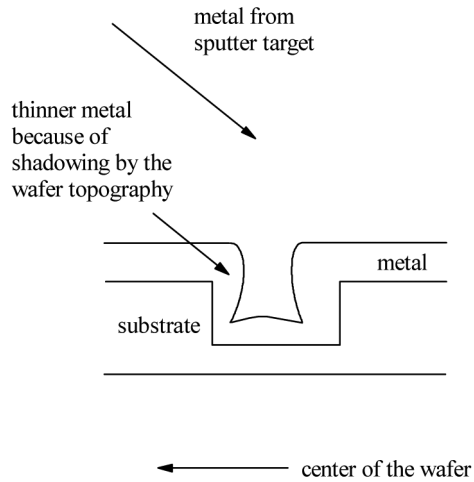


Figure 6.24 Metal deposited asymmetrically over alignment targets or overlay-measurement marks cause shifts in overlay.

Chemical-mechanical polish (CMP) represents a particularly difficult challenge for overlay.[83,84] The purpose of CMP is to produce highly planarized surfaces. While this helps lithography by reducing thin-film effects and increasing depth-of-focus, it makes overlay difficult because it reduces alignment-target contrast. Consider the situation depicted in Fig. 6.25(a). Alignment targets are invisible when planarization is complete and wafers are covered completely by metal films. The polishing process is adjusted in order to produce some degree of topography [Fig. 6.25(b)],[85] but the polishing often produces asymmetries [Fig. 6.25(c)], particularly when the width of the pattern is large. Alignment target deformation occurs from all CMP processes, such as for shallow trench isolation, and not just those processes involving metal layers.

There are several things that can be done to improve overlay in the context of CMP. Alignment mark dimensions can be varied or segmented in order to

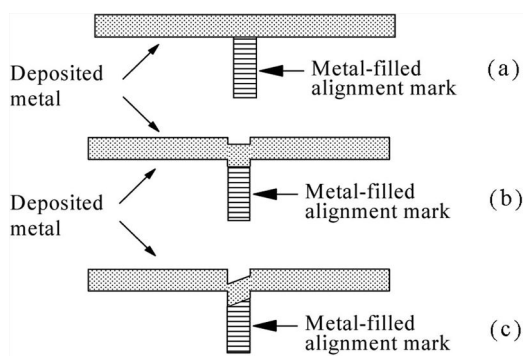


Figure 6.25 Effects of chemical-mechanical polishing on alignment marks: (a) alignment target polished flat, (b) symmetrically recessed alignment target, and (c) asymmetrically recessed alignment target.

determine the best mark. Polishing engineers should participate in any programs to improve overlay, since modifications of the polish process can often reduce overlay errors considerably, particularly when such errors result from highly variable polish processes. For example, the use of harder polishing pads may lead to reduced overlay errors. Stepper manufacturers have also responded with modifications to their alignment systems, using modified illumination[86] and new algorithms for interpreting alignment signals.[87]

Problems

6.1 Acceptable values for wafer scaling can be assessed by comparing the effect on overlay of a 0.01-ppm error in correcting for wafer expansion to the overlay requirements in Table 6.1. Across a 300-mm wafer, show that a range of 3-nm overlay error can result from an 0.1-ppm wafer-scaling error. Is this a significant error for the 32-nm node?

6.2 Give two reasons why alignment systems should be designed to function at wavelengths other than the ones used for patterning the resist. (Hint: Are there any particular resist processes that are particularly problematic for actinic-wavelength-alignment systems?)

6.3 Consider a 0.7-NA 193-nm exposure tool used to pattern 130-nm features. Show that the diffraction-limited resolution of a through-the-lens alignment system on this tool, operating at the HeNe laser wavelength of 632.8 nm, is 168 nm. Is this adequate for achieving overlay control that is one-third the minimum feature size?

6.4 The alignment mark for the ASML alignment system is a phase grating consisting of approximately equal lines and spaces of 8.0 μm width. If a HeNe laser is used for alignment and is directed normal to the plane of the grating, show that the minimum-NA lens required to capture the ±first-order diffraction beams is 0.04 [use Eq. (2.1)]. What minimum NA is required for third-order? Fifth-order? Seventh-order? Is this NA requirement too large for microlithographic-projection optics for through-the-lens alignment?

6.5 Show that the depth required to achieve a 180-deg phase difference for the first-order diffraction beams from the phase grating shown in Fig. 6.6 is

$$\frac{\lambda}{2n}\frac{\cos\theta}{1+\cos\theta}.$$

Show that $\cos\theta \approx 1$ is a suitable approximation for first-order alignment on the ASML alignment system. How much should the depth be adjusted for optimized seventh-order alignment relative to first-order alignment?

6.6 If silicon wafers have a 2.6-ppm/°C coefficient of thermal expansion, show that a ~0.01 °C change in temperature will cause a 10-nm change in the distance between two points on opposite sides of a 300-mm wafer. Is ±0.01 °C wafer temperature control adequate for the 45-nm node?

6.7 In a step-and-scan system, the reticle must be parallel with the reticle stage, and the reticle and wafer stages must be parallel to each other. Show that intrafield skew errors arise when the reticle is parallel to the wafer stage, but the reticle and wafer stages do not scan in the same direction.

6.8 Show that the range of displacement errors across a 300-mm wafer due to a 0.01 arcsec rotation error is 14.5 nm.

6.9 Suppose a lens has third-order distortion [Eqs. (6.29)–(6.31)]. Show that the image-placement error caused by this distortion error in the x direction at position (x, y) in the exposure field is given by:

$$\Delta x = D_3 x^3 + D_3 x \frac{W^2}{12},$$

where $(0, 0)$ is the center of the exposure field. Show that the image-placement error due to third-order distortion in the y direction is zero everywhere.

References

1. Semiconductor Equipment and Materials International, San Jose, California.
2. H. J. Levinson, *Lithography Process Control*, SPIE Press, Bellingham, Washington (1999).
3. M. A. van den Brink, S. Wittekoek, H. F. D. Linders, F. J. van Hout, and R. A. George, "Performance of a wafer stepper with automatic intra-die registration correction," *Proc. SPIE* **772**, 100–117 (1987).
4. V. R. Nagaswami, "In-plane distortions in silicon wafers induced by a sub-micron CMOS process," *Microelectron. Eng.* **9**, 457–461 (1989).
5. L. D. Yau, "Process-induced distortion in silicon wafers," *IEEE Trans. Electron Dev.* **ED-26**, 1299–1305 (1979).
6. J. D. Cuthbert, "Characterization of in-plane wafer distortions by double polysilicon NMOS processing," *Proc. Microcircuit Eng.* **190**, (1979).
7. A. Imai, N. Hasegawa, S. Okazaki, and K. Sakaguchi, "Wafer and chip deformation caused by pattern transfer," *Proc. SPIE* **2726**, 104–112 (1996).
8. W. C. Schneider, "Testing the Mann type 4800DSW wafer stepper," *Proc. SPIE* **174**, 6–14 (1979).
9. H. E. Mayer and E. W. Loebach, "Improvement of overlay and focusing accuracy of wafer step-and-repeat aligners by automatic calibration," *Proc. SPIE* **470**, 178–184 (1984).

10. S. Drazkiewicz, G. Gallatin, and J. Lyons, "Micrascan adaptive x-cross correlative independent off-axis modular (AXIOM) alignment system," *Proc. SPIE* **2726**, 886–899 (1996).

11. D. Cote, K. Andresen, D. Cronin, H. Harrold, M. Himel, J. Kane, J. Lyons, L. Markoya, C. Mason, D. McCafferty, M. McCarthy, G. O'Connor, H. Sewell, and D. Williamson, "Micrascan III–performance of a third generation, catadioptric step-and-scan lithographic tool," *Proc. SPIE* **3051**, 806–816 (1997).

12. M. van den Brink, H. Linders, and S. Wittekoek, "Direct-referencing automatic two-points reticle-to-wafer alignment using a projection column servo system," *Proc. SPIE* **633**, 60–71 (1986).

13. A. Stephanakis and H. Coleman, "Mix and match 10× reduction wafer steppers," *Proc. SPIE* **334**, 132–138 (1982).

14. S. Lee, S.-K. Yao, and M. Nuhn, "Alignment wavelength optimization for wafer stepper microscope," *Proc. SPIE* **633**, 79–84 (1986).

15. H. Stover, "Stepping into the 80s with die-by-die alignment," *Sol. State Technol.*, 112–120 (May, 1981).

16. S. Cosentino, J. Schaper, and J. Peavey, "Application of thin film reflectance calculations to linewidth measurements for HMOS circuit fabrication," *Proc. Kodak Microelectron. Sem.*, 34–39 (1983).

17. S. Kuniyoshi, T. Terasawa, T. Kurosaki, and T. Kimura, "Contrast improvement of alignment signals from resist coated patterns," *J. Vac. Sci. Technol. B* **5**(2), 555–560 (1987).

18. K. A. Chivers, "A modified photoresist spin process for a field-by-field alignment system," *Proc. Kodak Microelectron. Sem.*, 44–51 (1984).

19. G. Bouwhuis and S. Wittekoek, "Automatic alignment system for optical projection printing," *IEEE Trans. Electron. Dev.* **ED-26**(4), 723–728 (1979).

20. F. Bornebroek, J. Burghoorn, J. S. Greeneich, H. J. Megens, D. Satriasuptra, G. Simons, S. Stalnaker, and B. Koek, "Overlay performance in advanced processes," *Proc. SPIE* **4000**, 520–531 (2000).

21. R. J. Kopp and D. J. Stevens, "Overlay considerations for the selection of integrated-circuit pattern-level sequences," *Solid State Technol.*, 79–87 (July, 1980).

22. W. Wu and S. Cheng, "A simple method to determine alignment tolerance in photolithography process," *Proc. SPIE* **633**, 94–97 (1986).

23. S. Lee, S.-K. Yao, and M. Nuhn, "Alignment wavelength optimization for wafer stepper microscope," *Proc. SPIE* **633**, 79–93 (1986).

24. N. Bobroff and A. Rosenbluth, "Alignment errors from resist coating topography," *J. Vac. Sci. Technol. B* **6**(1), 403–408 (1988).

25. L. M. Manske and D. B. Graves, "Origins of asymmetry in spin-cast films over topography," *Proc. SPIE* **1463**, 414–422 (1991).

26. G. Flores and W. W. Flack, "Photoresist thin-film effects on alignment process capability," *Proc. SPIE* **1927**, 367–380 (1993).

27. S. Kuniyoshi, T. Terasawa, T. Kurosaki, and T. Kimura, "Contrast improvement of alignment signals from resist coated patterns," *J. Vac. Sci. Technol. B* **5**(2), 555–560 (1987).

28. G. Flores and W. W. Flack, "Photoresist thin-film effects on alignment process capability," *Proc. SPIE* **1927**, 367–380 (1993).

29. R. Mohondro, S. Bachman, T. Kinney, G. Meissner, and D. Peters, "High-contrast education spin-coat process effects on uniformity and overlay registration," *Proc. Olin Microlithog. Sem.*, 131–140 (1995).

30. C. P. Kirk, "Theoretical models for the optical alignment of wafer steppers," *Proc. SPIE* **772**, 134–141 (1987).

31. R. Hershel, "Autoalignment in step-and-repeat wafer printing," *Proc. SPIE* **174**, 54–62 (1979).

32. J. S. Wilczynski, "Optical step-and-repeat camera with dark-field automatic alignment," *J. Vac. Sci. Technol.* **16**(6), 1929–1933 (1979).

33. D. R. Beaulieu and P. Hellebrekers, "Dark-field technology–a practical approach to local alignment," *Proc. SPIE* **772**, 142–149 (1987).

34. S. Murakami, T. Matsuura, M. Ogawa, and M. Uehara, "Laser step alignment for a wafer stepper," *Proc. SPIE* **538**, 9–16 (1985).

35. G. M. Gallatin, J. C. Webster, E. C. Kintner, and F. Wu, "Modeling the images of alignment marks under photoresist," *Proc. SPIE* **772**, 193–201 (1987).

36. C.-M. Yuan, "Modeling of optical alignment images for semiconductor structures," *Proc. SPIE* **1088**, 392–402 (1989).

37. N. Magome and N. Shiraishi, "Laser alignment signal simulation for analysis of Al layers," *Proc. SPIE* **1088**, 238–247 (1989).

38. P. D. Coon, A. A. Aiyer, H. Chau, and H. Ooki, "High speed alignment simulator for Nikon steppers," *Proc. SPIE* **3679**, 208–218 (1999).

39. K. Ota, N. Magome, and K. Nishi, "New alignment sensors for wafer stepper," *Proc. SPIE* **1463**, 304–314 (1991).

40. C. Smith and J. Helbert, "Improving dark-field alignment through target and mapping software optimization," *Microlithog. World*, 5–9 (April-June, 1993).

41. S. Uzawa, A. Suzuki, and N. Ayata, "A new alignment system for submicron stepper," *Proc. SPIE* **1261**, 325–331 (1990).

42. S. Slonaker, S. McNamara, K. Konno, and R. Miller, "Enhanced global alignment for production optical lithography," *Proc. SPIE* **922**, 73–81 (1988).

43. S. Murakami, T. Matsura, M. Ogawa, and M. Uehara, "Laser step alignment for a wafer stepper," *Proc. SPIE* **538**, 9–16 (1985).

44. D. S. Perloff, "A four point electrical measurement technique for characterizing mask superposition errors on semiconductor wafers," *IEEE Solid State. Circ.* **SC-13**(4), 436–444 (1978).

45. N. Magome and H. Kawai, "Total overlay analysis for designing future aligner," *Proc. SPIE* **2440**, 902–912 (1995).

46. M. A. van den Brink, J. M. D. Stoeldraijer, and H. F. D. Linders, "Overlay and field-by-field leveling in wafer steppers using an advanced metrology system," *Proc. SPIE* **1673**, 330–344 (1992).

47. T. Saito, S. Sakamoto, K. Okuma, H. Fukumoto, and Y. Okuda, "Mask overlay scaling error caused by exposure energy using a stepper," *Proc. SPIE* **1926**, 440–449 (1993).

48. E. Morita, M. Kawakubo, F. C. Leung, S. J. McNamara, and J. T. Parry, "Impacts of reticle and wafer elasticity control on overall alignment management strategy," *Proc. SPIE* **3334**, 510–519 (1998).

49. G. Rivera and P. Canestrari, "Process induced wafer distortion: measurement and effect on overlay in stepper based advanced lithography," *Proc. SPIE* **807**, 806–813 (1993).

50. A. A. Ghazanfarian, X. Chen, M. A. McCord, R. Fabian, and W. Pease, "Exploiting structure of wafer distortion in global alignment," *J. Vac. Sci. Technol. B* **16**(6), 3642–3646 (1998).

51. D. MacMillan and W. D. Ryden, "Analysis of image field placement deviations of a 5× microlithographic reduction lens," *Proc. SPIE* **334**, 78–89 (1982).

52. J. D. Armitage Jr. and J. P. Kirk, "Analysis of overlay distortion patterns," *Proc. SPIE* **921**, 207–222 (1988).

53. T. E. Zavecz, "Lithographic Overlay measurement precision and calibration and their effect on pattern registration optimization," *Proc. SPIE* **1673**, 191–202 (1992).

54. W. H. Arnold, "Image placement differences between 1:1 projection aligners and 10:1 reduction wafer steppers," *Proc. SPIE* **394**, 87–98 (1983).

55. M. A. van den Brink, C. G. M. de Mol, and R. A. George, "Matching performance for multiple wafer steppers using an advanced metrology procedure," *Proc. SPIE* **921**, 180–197 (1988).

56. T. Mono, U. P. Schröder, D. Nees, K. Palitzsch, W. Köstler, J. Bruch, S. Kramp, M. Veldkamp, and R. Schuster, "Overlay considerations for 300 mm lithography," *Proc. SPIE* **5038**, 121–125 (2003).

57. S. Chang, S. DeMoor, J. Brown, C. Atkinson, and J. Roberge, "An improved method to determine optimal alignment sampling layouts," *Proc. SPIE* **5038**, 70–80 (2003).

58. M. A. van den Brink, B. A. Katz, and S. Wittekoek, "New 0.54 aperture i-line wafer stepper with field-by-field leveling combined with global alignment," *Proc. SPIE* **1463**, 709–724 (1991).

59. A. Sukegawa, S. Wakamoto, S. Nakajima, M. Kawakubo, and N. Magome, "Overlay improvement by using new framework of grid compensation for matching," *Proc. SPIE* **6152**, 61253A (2006).

60. V. Nagaswami, and W. Geerts, "Overlay control in submicron environment," *Proc. KTI Microelectron. Sem.*, 89–106 (1989).

61. J. G. Maltabes, M. C. Hakey, and A. L. Levine, "Cost/benefit analysis of mix-and-match lithography for production of half-micron devices," *Proc. SPIE* **1927**, 814–826 (1993).

62. S. K. Jones, E. S. Capsuto, B. W. Dudley, C. R. Peters, and G. C. Escher, "Wavelength tuning for optimization of deep UV excimer laser performance," *Proc. SPIE* **1674**, 499–508 (1992).

63. M. E. Preil and W. H. Arnold, "Aerial image formation with a KrF excimer laser stepper," *Polymer Eng. Sci.* **32**(21), 1583–1588 (1992).

64. R. K. Brimacombe, T. J. McKee, E. D. Mortimer, B. Norris, J. Reid, and T. A. Znotins, "Performance characteristics of a narrow band industrial excimer laser," *Proc. SPIE* **1088**, 416–422 (1989).

65. J. de Klerk, "Performance of a high NA, dual stage 193-nm TWINSCAN step-and-scan system for 80-nm applications," *Proc. SPIE* **5040**, 822–840 (2003).

66. J. Braat and P. Rennspies, "Effect of lens distortion in optical step-and-scan lithography," *Appl. Optic.* **35**(4), 690–700 (1996).

67. H. J. Levinson and R. Rice, "Overlay tolerances for VLSI using wafer steppers," *Proc. SPIE* **922**, 82–93 (1988).

68. M. A. van den Brink, C. G. M. de Mol, and J. M. D. Stoeldraijer, "Matching of multiple wafer steppers for 0.35-μm lithography, using advanced optimization schemes," *Proc. SPIE* **1926**, 188–207 (1993).

69. J. C. Pelligrini, "Comparisons of six different intrafield control paradigms in an advanced mix-and-match environment," *Proc. SPIE* **3050**, 398–406 (1997).

70. H. J. Levinson, M. E. Preil, and P. J. Lord, "Minimization of total overlay errors on product wafers using an advanced optimization scheme," *Proc. SPIE* **3051**, 362–373 (1997).

71. N. R. Farrar, "Effect of off-axis illumination on stepper overlay," *Proc. SPIE* **2439**, 273–280 (1995).

72. T. Saito, H. Watanabe, and Y. Okuda, "Effect of variable sigma aperture on lens distortion and its pattern size dependence," *Proc. SPIE* **2725**, 414–423 (1996).

73. A. M. Davis, T. Dooly, and J. R. Johnson, "Impact of level specific illumination conditions on overlay," *Proc. Olin Microlithog. Sem.*, 1–16 (1997).

74. C. S. Lee, J. S. Kim, I. B. Hur, Y. M. Ham, S. H. Choi, Y. S. Seo, and S. M. Ashkenaz, "Overlay and lens distortion in a modified illumination stepper," *Proc. SPIE* **2197**, 2–8 (1994).

75. C.-M. Lim, K.-S. Kwon, D. Yim, D.-H. Son, H.-S. Kim, and K.-H. Baik, "Analysis of nonlinear overlay errors by aperture mixing related to pattern asymmetry," *Proc. SPIE* **3051**, 106–115 (1997).

76. N. Seong, H. Kang, J. Kye, H. Cho, and J. Moon, "Pattern displacement error under off axis illumination," *Jpn. J. Appl. Phys.* **37**, 6695–6697 (1998).

77. S. Girol-Gunia, B. Schulz, N. Smith, and L. Binns, "Using in-chip overlay metrology," *Proc. SPIE* **6922**, 69220N (2008).

78. T. A. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. Res. Develop.* **41**(1/2), 57–67 (1997).

79. T. Sato, and H. Nomura, "Coma aberration measurement by relative shift of displacement with pattern dependence"' *Jpn. J. Appl. Phys.* **37**, Part 1, 3553–3557 (1998).

80. M. E. Preil, T. Manchester, and A. Minvielle, "Minimization of total overlay errors when matching non-concentric exposure fields," *Proc. SPIE* **2197**, 753–769 (1994).

81. W. W. Flack, G. E. Flores, J. C. Pellegrini, and M. Merrill, "An optimized registration model for 2:1 stepper field matching," *Proc. SPIE* **2197**, 733–752 (1994).

82. R. Rogoff, S. S. Hong, D. Schramm, and G. Espin, "Reticle specific compensations to meet production overlay requirements for 64 Mb and beyond," *Proc. SPIE* **2197**, 781–790 (1994).

83. D. L. Meunier, D. Humphrey, B. Peck, P. Feeney, J. Paquette, and J. Thibault, "Optimization of metal layer overlay and alignment targets in a chemical-mechanical polishing environment," *Proc. Olin Microlithog. Sem.*, 355–366 (1996).

84. S.-W. Hsia, G. Miyagi, and M. Brongo, "Alignment characterization of CMP tungsten process," *Proc. Olin Microlithog. Sem.*, 381–390 (1996).

85. D. L. Meunier, B. Plambeck, P. Lord, and N. Knoll, "The implementation of coherence probe microscopy in a process using chemical mechanical polishing," *Proc. OCG Microlithog. Sem.*, 155–169 (1995).

86. N. Shirishi, A. Sugaya, and D. P. Coon, "Alignment strategies for planarizing technologies," *Proc. SPIE* **3051**, 836–845 (1997).

87. T. Kanda, K. Mishima, E. Murakami, and H. Ina, "Alignment sensor corrections for tool induced shift (TIS)," *Proc. SPIE* **3051**, 846–855 (1997).

Chapter 7
Masks and Reticles

7.1 Overview

In optical-projection lithography, the patterns on wafers are reproductions of those on photomasks. The quality of the wafer patterns, as measured by linewidth control, overlay, and defects, is strongly affected by the quality of the corresponding parameters on the masks. Linewidth variations on the reticles ultimately result in linewidth variations on the wafer. Mask-registration errors contribute to overlay errors. Defects on the reticle may result in nonworking die. Consequently, masks are critical components of lithographic technology.

The impact from variations and defects on masks is particularly significant when wafer steppers are used, where the patterns of integrated circuits are formed on the wafers by the repeated imaging of reticles. This repetitive imaging process imposes stringent requirements on the reticles. Consider a reticle that contains four product dies. A single defect on the reticle capable of causing product failure reduces yield by 25%. Such a large yield loss from a single defect implies zero tolerance for defects on the reticle. Similarly, linewidth control and registration must be very good on reticles. Because reticle defects and variations are reproduced repetitively, photomask technology is an important aspect of lithography. Current and future mask requirements from the 2009 International Technology Roadmap for Semiconductors (ITRS) are shown in Table 7.1.

Photomasks are fabricated with techniques similar to those used in wafer processing. A photomask *blank*, consisting of an opaque film (usually chromium- or molybdenum-containing compounds) deposited on a glass substrate, is covered with resist. The resist is exposed according to the circuit pattern; the resist is then developed, and the exposed opaque material is etched. Photomask patterning is accomplished primarily by means of beam writers, which are tools that expose mask blanks according to suitably formatted circuit designs. Two types of beam writers—electron and optical—are used, and both are discussed in detail in this chapter.

At this early point in the discussion of photomasks, the strict difference between a mask and reticle is explained. In the early days of integrated circuits, the patterns of an integrated circuit were formed by transferring the patterns from masks to the wafers. For every layer, the transfer was 1:1 between the mask and wafer; the patterns for all of the dies on the wafer were contained on the mask. The mask

Table 7.1 Optical mask requirements. White cells mean solutions exist or are expected. Light gray cells indicate that significant work is required before solutions are found. Dark gray cells indicate that no solution may exist by the appropriate date. The phase-shifting mask (PSM) is described in more detail in Chapter 8.

Year of introduction **"Technology node"**	**2010** **45 nm**	**2012** **36 nm**	**2014** **28 nm**
DRAM/MPU/ASIC wafer minimum metal-1 half-pitch (nm)	45	36	28
MPU gate in resist (nm)	35	28	22
Contact in resist (nm)	56	44	35
Magnification	4	4	4
Mask minimum primary feature size (nm)	99	78	62
Mask subresolution feature size opaque (nm)	71	56	44
Image placement (nm, multipoint)	5.4	4.3	3.4
CD uniformity (nm, 3σ)			
Isolated lines (MPU gates, binary or attenuated phase-shifting masks)	2.0	1.7	1.3
Dense lines (DRAM half pitch, binary or attenuated phase-shifting masks)	3.4	2.7	2.1
Contact/vias	1.9	1.5	1.2
Linearity (nm)	7.2	5.7	4.5
CD mean-to-target (nm)	3.6	2.9	2.3
Defect size (nm)	36	29	23
Data-volume (GB)	825	1310	2080
Mask-design grid (nm)	1	1	1
Attenuated-PSM transmission mean deviation from target (+/− % of target)	4	4	4
Attenuated-PSM transmission uniformity (+/− % of target)	4	4	4
Attenuated-PSM phase mean deviation from 180° (+/− deg)	3	3	3
Alternating-PSM phase mean deviation from nominal phase-angle target (+/− deg)	1	1	1
Alternating-PSM phase uniformity (+/− deg)	1	1	1

Magnification: Lithography tool reduction ratio *N*:1.
Mask minimum primary feature size: Minimum printable feature after OPC application to be controlled on the mask for CD placement and defects.
Mask subresolution feature size: The minimum width of nonprinting features on the mask such as subresolution assist features.
Image placement: The maximum component deviation (x or y) of the array of image centerlines relative to a defined reference grid after removal of isotropic magnification error.
CD uniformity: The 3σ deviation of actual image sizes on a mask for a single-size and tone-critical feature. This applies to features in x and y and isolated features.
Linearity: Maximum deviation between mask "mean to target" for a range of features of the same tone and different design sizes. This includes features that are equal to the smallest subresolution assist mask feature and up to 3× the minimum wafer half pitch multiplied by the magnification.
CD mean-to-target: The maximum difference between the average of the measured feature sizes and the agreed-to feature size (design size). Applies to a single feature size and tone.
Defect size: A mask defect is any unintended mask anomaly that prints or changes a printed image size by 10% or more. The mask-defect size listed in the roadmap is the square root of the area of the smallest opaque or clear "defect" that is expected to print for the stated generation. Printable 180-deg phase defects are 70% smaller than the number shown.
Data volume: This is the expected maximum file size for uncompressed data for a single layer as presented to a pattern generator tool.
Mask-design grid: Wafer-design grid multiplied by the mask magnification.
Transmission: Ratio, expressed in percentage, of the fraction of light passing through an attenuated-PSM layer relative to the mask blank with no opaque films.
Phase: Change in optical-path length between two regions on the mask expressed in degrees. The mean value is determined by averaging phase measured for many features on the mask.
Phase uniformity: The maximum phase-error deviation of any point from the mean value.

layout was identical to the wafer layout. This approach is different from the way that wafers are patterned with wafer steppers, where the mask pattern has only part of the wafer pattern, and full wafer coverage is obtained by repetitive imaging of the mask. In the earlier 1:1 lithography, the pattern on the mask was often generated through a lithographic process in which the mask blank was exposed die-by-die on a photorepeater, a tool that was essentially a stepper that patterned masks instead of wafers. This method of mask making required an object whose pattern was imaged onto the mask by the photorepeater. To distinguish between the object whose image was repeated and that which was the mask shop's product, the object used on the photorepeater was referred to as a reticle. The inefficiency of this process, by which circuit patterns were first transferred from a reticle to a mask and then to the wafer, was recognized, and the wafer stepper was invented, whereby the circuit patterns were stepped directly onto the wafers. To be technically correct, the object imaged by a step-and-repeat system should be called a reticle, but producers of integrated circuits in wafer fabricators have always referred to their master patterns as masks, so the terms *masks* and *reticles* now tend to be used interchangeably. Regardless of whether they are called masks or reticles, their form, fabrication, and use is much the same.

7.2 Mask Blanks

For high-resolution stepper lithography, the most common size for reticles today is the so-called 6-in. (0.25-in.-thick) format, whose defining dimensions are given in Table 7.2. For most applications, the glass type of the mask substrate is fused silica, chosen primarily for its high transmission throughout the ultraviolet portion of the electromagnetic spectrum and excellent thermomechanical properties. The coefficient of thermal expansion of fused silica is only 0.5 ppm/°C, compared to 4–100 ppm/°C for other types of glass, such as borosilicate glass.[1] During the patterning of the photomask, temperature variations in the mask result in small amounts of misregistration when using fused silica, while substantial misregistration occurs when using other types of glass substrate materials. For example, across 100 mm (a typical distance across the exposed field on the reticle), a temperature rise of only 0.1 °C causes a 5-nm registration error across a fused silica plate. Detailed simulations of mask-substrate heating during patterning using 50-keV electrons have shown temperature increases up to 2 °C, with resulting misregistration errors of 30–50 nm (3σ).[2] A substantial portion of these placement

Table 7.2 Sizes of selected SEMI standard reticle formats.[8]

	Edge length		Thickness	
Nominal size	**Minimum (mm)**	**Maximum (mm)**	**Minimum (mm)**	**Maximum (mm)**
6.0 × 6.0 × 0.25 in.	151.6	152.4	6.25	6.45
6.0 × 6.0 × 0.15 in.	151.6	152.4	3.70	3.90
7.0 × 7.0 × 0.25 in.	177.0	177.8	6.25	6.45
230 × 230 × 9 mm	229.6	230.0	8.90	9.10

errors contribute to correctable magnification errors, but there are residual errors that are very significant if a substrate less thermally stable than fused silica is modeled—even with 4× lens reduction. While there is an acceptable level of registration errors with fused-silica masks seen in these simulations, substrates, such as borosilicate glass, with coefficients of thermal expansion more than ten times larger than that of fused silica, would require unrealistic temperature control to achieve adequate performance. Fortunately, fused silica has excellent transparency at DUV wavelengths, and new grades have been developed that have adequate transparency for use as reticle substrates for wavelengths down to 157 nm.[3] Thermomechanical stability and good transparency justify the use of fused silica as photomask substrates, in spite of poor electrical conductivity and associated electrostatic damage issues.

Ideal fused silica is amorphous, which constitutes another advantage of this material for mask substrates. Crystalline materials, even those with cubic symmetry, will have optical properties that depend on the orientation of the polarization vector of light relative to crystal axes,[4] a property referred to as birefringence.[5] Under stress, fused silica may also acquire a significant degree of birefringence,[6] so care must be taken when manufacturing the fused silica from which photomasks are to be fabricated.[7]

Just as images on the wafer become blurred when the imaging plane is outside the depth-of-focus, images also become blurred when the objects are outside the lens's *depth-of-field*. The depth-of-field is the distance that the object can be moved, parallel to the optical axis, while maintaining good imaging. The depth-of-field is related to the depth-of-focus by

$$\text{depth-of-field} = (\text{depth-of-focus}) \times N^2, \tag{7.1}$$

where N is the lens reduction (4, 5, etc.). Alternatively, one could say that the effective depth-of-focus is reduced by reticle nonflatness, by the amount

$$\text{depth-of-focus} = \frac{\text{reticle nonflatness}}{N^2}. \tag{7.2}$$

Depth-of-focus considerations drive specifications for reticle flatness.[9] Very flat reticles are becoming necessary (see Problem 7.2), driving up the cost of mask blanks considerably. Stress from absorber deposition also warps mask blanks that are initially quite flat. It is also becoming necessary to consider gravitational effects. During use, photomasks are supported only outside of the exposure area. Gravity causes reticles to sag in the areas of the device pattern.[10] This gravitational sag needs to be factored into overall stepper design and focus uniformity budgets.

Chromium- and molybdenum-containing compounds are the most common opaque materials used for making photomasks. Tantalum-based absorber materials have also been considered.[11] Films of chromium- and molybdenum-containing compounds[12–15] are typically sputtered onto the glass substrates to thicknesses between 500 Å and 1100 Å. To provide some context, the optical constants of

Table 7.3 Characteristics of chromium films as opaque materials for photomasks.

Wavelength (nm)	Index of refraction[7]	Reflection R from chromium glass interface	Transmission T through a 600-Å film of chromium	$(1-R)\times T\times 100\%$
436	1.79–*i*4.05	0.60	0.0009	0.036
365	1.39–*i*3.24	0.56	0.0012	0.055
248	0.85–*i*2.01	0.47	0.0022	0.119
193	0.84–*i*1.65	0.38	0.0016	0.098
157	0.68–*i*1.11	0.32	0.0048	0.310

chromium are given in Table 7.3. From these values one can see that chromium films of typical thickness are very opaque. It should be kept in mind that the absorbers on commercially produced chromium-based photomask blanks typically are not comprised of pure chromium.[16] The properties of chromium- and molybdenum-containing compounds vary among mask-blank suppliers, who can provide details for their own specific films.

It should be noted that most of the light is blocked by the chromium through absorption. Consequently, masks will heat during use.[17] Again, most of the heating effect results in correctable magnification errors with suitably designed symmetrical reticle platens. Compensating for this will require periodic reticle realignments, which provide measurements of reticle expansion or contraction, though this will usually reduce stepper productivity. Regardless, the noncorrectable registration errors may become significant for the 22-nm node and beyond.

In addition to optical properties, the stress of absorber films is important. There will be stress relief when the absorber film is etched, and this can lead to mask distortion—often nonlinear—between patterning by the beam writer and final mask use.[18] The actual distortion pattern will depend upon the mask layout. Extremes between clear and dark areas on the mask will lead to greater nonlinearities. Mask distortion following etch can clearly be reduced by using absorbers that have been deposited with low stress.

7.3 Mechanical Optical-Pattern Generators

The creation of a photomask requires that a circuit design be transformed into a physical pattern on the mask. Optical-pattern generators were the first tools to be used for automatically patterning reticles,[19] which were then used on photorepeaters to generate 1× photomasks. A mechanical optical-pattern generator projected the image of a small rectangle onto a resist-coated reticle blank. The blank was moved under projection optics on a laser-interferometrically controlled stage and was exposed according to the circuit design by repetitive exposure of the rectangle. If it was desired to have an opening in the chrome at a certain location on the mask, the pattern generator exposed the (positive) resist on the mask blank when that particular location was in the exposure position. This was achieved by dividing the design into rectangles. The simplest way to do this was to lay out the design on a square grid, and each square would become an area to be exposed or not, depending upon the design.

Consider the generation of a reticle for a chip with an area of 2 mm × 2 mm on the wafer. Suppose that this reticle was used for fabricating a 1× mask using a photorepeater with a 10× lens. If the minimum feature size for the circuit was 10 μm (on the wafer), then the photorepeater would expose a square of 100 μm × 100 μm a maximum of 40,000 times in order to pattern the 20-mm × 20-mm chip area on the 10× reticle. At the rate of two squares exposed per second, the reticle would be exposed in 5 ½ hours. As feature sizes decreased and die sizes grew, the time to expose reticles on mechanical optical-pattern generators increased geometrically. A faster method for generating reticles was needed, and this need was met by electron-beam writers.

7.4 Electron-Beam Lithography and Mask Writers

Energetic electrons can induce chemical reactions in resists, just as photons can. While some resists can operate as both optical- and electron-beam resists, the resists used in mask making are usually specialized materials, because of the need for good sensitivity, the need to maintain throughput through the (expensive) beam writers, and a primary exposure reaction that differs from that in many optical photoresists (see Section 7.6). Patterning with electrons is accomplished by focusing the electrons into beams with very narrow diameters. The technology for making such beams is very mature, having been developed initially for electron microscopy. Electrons have potential for very high resolution because electron diffraction occurs over atomic distances. However, electrons scatter in resists, limiting resolution and inducing proximity effects.

Electron beams can potentially be moved at high speed because the electrons can be deflected electromagnetically instead of relying on mechanical motion to scan the beams across reticles that are being patterned. Consequently, electron-beam mask writers were adopted as chips were designed with high levels of integration.[20] Much is known about electron-beam technology through earlier and parallel development in scanning-electron microscopy. While many electron-beam patterning systems have been introduced, the ones most widely used for mask making for many years originated in the electron-beam exposure system (EBES) developed by Bell Telephone Laboratories in Murray Hill, New Jersey.[21] This technology was licensed and found the greatest commercial success in the manufacturing electron-beam exposure system (MEBES) tools built by Etec Systems. For two decades, the MEBES systems were the primary beam writers used to make photomasks. Applied Materials acquired Etec Systems, but later greatly reduced operations, as new architectures for electron-beam writers came to provide superior performance and ultimately supplanted MEBES as the tool used for making advanced photomasks.

The general structure of electron-beam writers is shown in Fig. 7.1. A resist-coated unpatterned mask substrate is placed on the *X-Y* stage. Electron beams are scanned across the resist surface by a combination of electron optics for short-range scanning and mechanical stage motion for long-range coverage of the mask blank. Electron-beam writers are classified according to two characteristics, beam

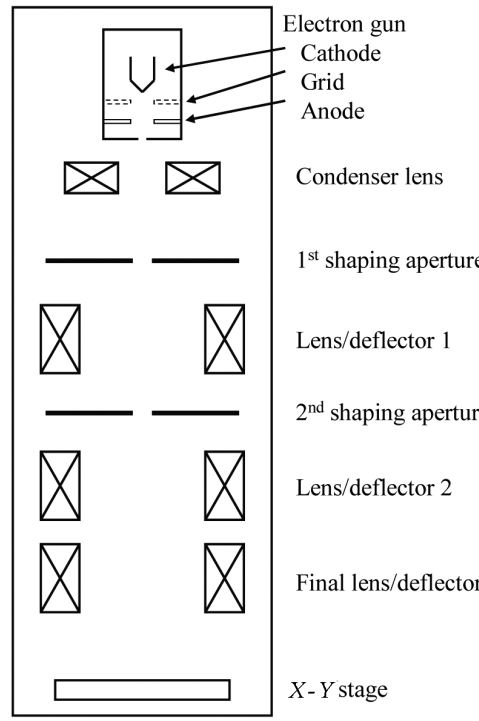


Figure 7.1 Schematic of an electron optical system for mask writing. For electron writers where throughput is a priority, the electron source is a thermionic emitter (such as lanthanum hexaboride), while systems designed for extremely high resolution employ thermal-field emission sources.[22,23]

shape and the method of scanning the beam. The two most common beam shapes are *Gaussian round beam* and *variable-shaped beam*. As the name suggests, a Gaussian round beam has circular symmetry in the writing plane, and the intensity is well approximated by a Gaussian distribution. In order to form the corners of rectangular shapes on the mask, such beams must have diameters that are much smaller than the minimum feature sizes. Shaped beams usually have rectangular shapes that allow good corner fidelity with relatively large beams, but other shapes are also possible. As might be expected, shaped-beam systems typically have higher throughput than Gaussian-beam systems, but there are methods that will be discussed later in this chapter for maintaining reasonable throughput for Gaussian-beam systems. Gaussian-beam systems are used where the highest resolution is needed, while shaped-beam writers are more typically used for the patterning of reduction masks for use in manufacturing, where resolution can be compromised to some degree in the interest of economy.

There are two basic approaches to scanning in electron-beam writers. The first is raster scanning, the principle behind television picture tubes and scanning-electron microscopes. In this approach, one part of the electron optics is set to scan the beam back-and-forth across the mask blank, while separate components turn the beam on and off. In the second type of electron-beam writers—vector systems—the beam

is directed only at those areas that actually get exposed. The distinction between raster and vector systems is illustrated in Fig. 7.2. As can be seen in the figure, a significantly smaller area is actually scanned when a vector scanning system is used for mask writing, compared to a raster scanning system. Further writing-time reductions are possible by employing a variable-shaped beam, where larger spot sizes can be used for writing large features, thereby reducing the amount of scanning even further. Vector systems currently are the primary machines used to write state-of-the-art masks for the manufacturing of integrated circuits. Examples of different electron-beam writes are given in Table 7.4.

Improvements in edge roughness are obtained by shaping the electron beam, and vector scanning systems today typically have shaped beams. A shaped beam is produced by passing the electron beam through a shaped aperture, or sets of apertures, in the electron optics.[26–29] Examples of shapes that could be produced by apertures in the Leica[30] ZBA32H+ are shown in Fig. 7.3.

While the conceptual advantages of vector scanning with shaped beams have been long understood, it has taken many years for this architecture to be adopted widely for mask making. One obstacle that needed to be overcome was the lack of robust software for converting circuit designs into data formats that could be used by the e-beam writers.[36] Because of strong interest in direct e-beam writing on wafers, for which system throughput was a major concern, vector scanning systems continued to be developed even while masks were made predominately by raster scanning tools. Additional resources became available to develop vector scanning systems further to meet the stringent challenges of 1× x-ray masks (see Chapter 13), which required very high pattern fidelity (including geometries with

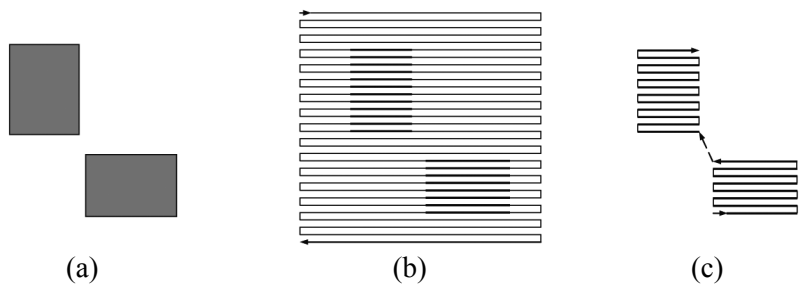


Figure 7.2 The area to be written is (a), while (b) illustrates exposure using a raster scanning system, where the beam is directed over a large area, with the beam blanked off except when directed at the dark areas. (c) The same area is scanned more quickly with a vector scanning beam writer, which scans only over the areas that are to be exposed.[24]

Table 7.4 Examples of commercially available electron-beam writers with different architectures.

	Raster scanning	**Vector scanning**
Gaussian beam	ETEC MEBES 5500	Vistec EBPG5000 plus
Shaped beam	No systems currently available[25]	NuFlare EBM-7000

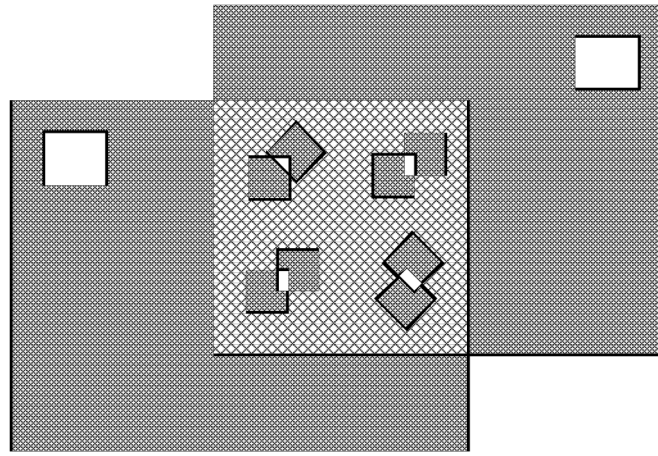


Figure 7.3 Different beam shapes that can be formed by two overlapping apertures.[31] The apertures can often be in different planes within the electron optical column.[32–35]

corners) and practical writing times.[37] Eventually vector scanning systems reached maturity, and there are several vector scanning systems available today for mask making, such as the JEOL JBX-3050MV and the NuFlare EBM-7000.[38] For high resolution, the beam voltage in these systems is 50 kV. Parameters of the EBM-7000 are given Table 7.5.

The area of a mask is patterned using a combination of electromagnetic beam deflection and mechanical scanning,[40] an example of which is shown in Fig. 7.4. Scanning can move beams over a distance of up to ~1 mm, while mechanical motion is required to cover longer distances. Scanning typically occurs simultaneously with mechanical movement to minimize periods of acceleration and deceleration that reduce throughput.

If beam writers are set up incorrectly, local linewidth errors can be produced because of errors at a stripe boundary. This is illustrated in Fig. 7.5. Geometries that lie on stripe boundaries can have dimensional errors because of pattern misplacement of separate geometries that comprise the complete feature. Errors of this type are <1.5 nm (3σ) on a NuFlare EBM-7000, mitigated in part by the use of two-pass printing, illustrated in Fig. 7.6. With two-pass printing, the extent to which patterned placement errors cause dimensional errors is lessened by reducing the degree to which geometries are split across strip boundaries.

Another example of an error at a stripe boundary is shown in Fig. 7.7. One of the challenges of scanning is ensuring that the stripes are butted against each other

Table 7.5 The key parameters of the EBM-7000.[39]

Data address unit	0.1 nm–100 nm
Stripe height	184 μm
Minimum feature size	20 nm
Pattern placement accuracy (3σ)	3.5 nm
CD uniformity (3σ)	2.0 nm
Beam voltage	50 kV

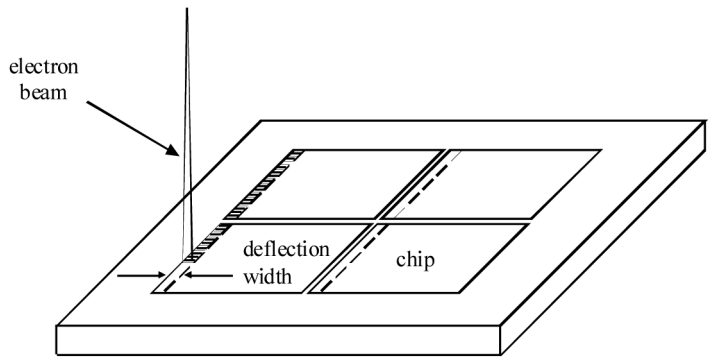


Figure 7.4 Scanning a pattern on a mask.

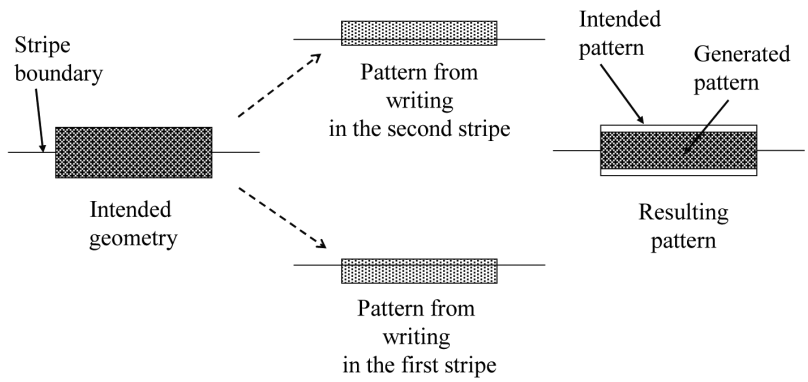


Figure 7.5 A CD error at a stripe boundary. The intended geometry is made too short.

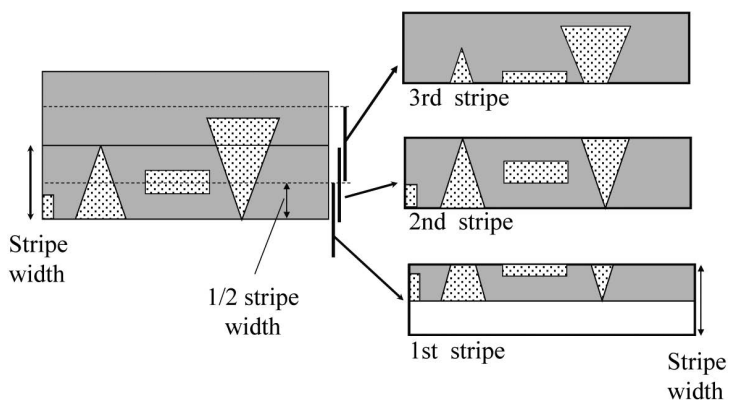


Figure 7.6 Illustration of two-pass printing on the NuFlare EBM 7000.[41]

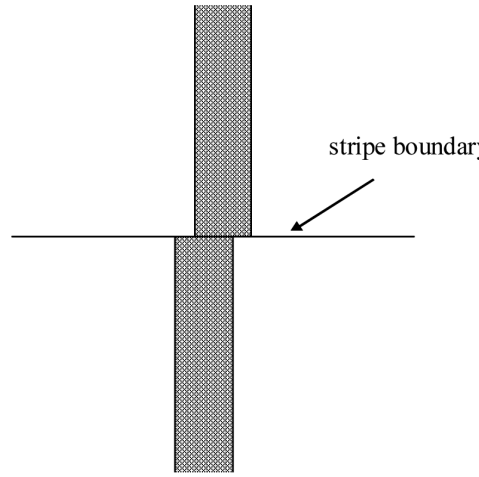


Figure 7.7 Example of a butting error.

correctly. Because some geometries extend across stripe boundaries, it is necessary to ensure that the parts written in separate stripes are connected properly. Deformed patterns at stripe boundaries, known as "butting errors," occur. An example of a butting error is shown in Fig. 7.7. As one can see from Fig. 7.6, this type of error is also reduced by multipass printing.

There are several types of electron sources used for electron-beam writers. Thermal sources, typically lanthanum hexaboride (LaB_6), are usually employed when high throughput is given priority. In such situations, a secondary concern is the intrinsic thermal energy broadening that degrades resolution. For high-resolution tools, field-effect sources, usually thermally assisted, are used, since these have lower energy spreads than LaB_6 sources. The trajectory of an electron is dependent on its energy, ultimately resulting in electron beams that often acquire a near-Gaussian intensity profile after traversing electron optics.[42] In the following discussion of raster scanning it is assumed that the beams are Gaussian.

The first electron-beam exposure system (EBES) was based upon raster scanning. The electron optics in the original EBES system was built from a modified scanning-electron microscope optical column. One part of the electron optics scanned a 0.5-μm-diameter electron beam, while another part of the system blanked the beam.[43–45] The system wrote along a scan line that was as long as 0.128 mm. The length of the scan was limited because aberrations in the electron-beam optics increased significantly as the distance from the beam's central axis increased. While the electron optics scanned the 10-keV beam in one direction, a precision mechanical stage moved the mask blank in the perpendicular direction. By this means, a stripe the length of the pattern on the mask in one direction, and up to 0.128 mm in the other direction, was written (Fig. 7.4). Following the completion of one stripe, another was written, and this continued until the entire mask was patterned. The original MEBES system evolved through a series of successive

models, MEBES II–IV,[46–48] the MEBES 4500, and the last model, MEBES 5500, released in 1999.

Electron-beam scanning lithography suffers from a fundamental productivity problem as the feature size decreases. In the most primitive implementation of raster scanning, the design needs to be divided into pixels or *addresses* whose dimensions are equal to those of the exposure beam. As the minimum feature size on the mask diminishes linearly, the spot size (area) of the exposure beam must decrease quadratically. For example, if the feature-size width decreases by a factor of two, the area of the spot size decreases by a factor of four, and, as a consequence, the number of pixels needed to expose a fixed area on the reticle increases by a factor of four. If pixels are exposed at a fixed rate, the throughput of the beam writer decreases by approximately a factor of four each time the minimum feature size decreases by a factor of two. Fortunately, the writing time of beam writers has improved, though not at a pace commensurate with decreases in feature size. When 5× steppers were first introduced, the MEBES exposure rate was 40 MHz,[49] and the wafer's minimum feature size was approximately two microns. Today's minimum feature size on the wafer is less than 50 nm, and mask features have scaled even faster, due to the widespread use of subresolution features (see Chapter 8). To maintain beam-writer productivity, it would have been necessary for the MEBES exposure rate to increase to over 64 GHz. Because the exposure rate could not increase at a rate commensurate with the increase in pixels, mask-writing time has increased, leading to higher mask-production costs, even with a transition made to vector-shaped beam architectures. The issue of lithography costs is covered in more detail in Chapter 11. Optical-beam writers, which can provide lower mask costs when mask features no smaller than 400 nm are required,[50] are discussed later in this chapter.

Before discussing writing strategies further, some definitions are needed. It is important to note that designers place circuits on grids that are different from those used to fabricate the masks. The distinction must therefore be made between the design grid and the writing grid. When writing, a particular grid point is often referred to as an *address*.

To understand the raster-writing strategies alternative to primitive raster scanning, it must be recognized that the important requirement for many masks is placement of the edges of geometries, not the minimum geometry size. For example, consider a process where the minimum pitch is 100 nm, the nominal feature size is the half pitch, and 4× reduction steppers are used. On the reticle, the minimum features are nominally $4 \times 50 = 200$ nm. However, because of the needs of optical proximity correction, it may be necessary to shrink some features on the mask by 20 nm, or 10 nm per edge, while other features remain sized at the original 200 nm. In a mask-writing scheme where the writing address equals the spot size of the beam writer, a writing grid of 10 nm would be necessary to generate this mask, even though the minimum feature size on the mask is 180 nm. With a writing grid and spot size of 10 nm, beam-writer throughput is very low.

To avoid this problem, raster writing schemes have been developed where the writing grid and spot sizes are larger than the design grid. Moreover, masks are

often patterned with multiple passes. Consider the following example. A circuit is designed on a grid, which defines the unit scale: the design-grid size is 1.0. In a writing scheme called multipass, the writing grid is a multiple (laterally) of the design grid (see Fig. 7.8). Moreover, the x and y 2σ points of the Gaussian spot are at the edges of the writing grid. By making the spot size large enough, there is overlap of the tails of the spots, which smooths out the exposure.

Illustrated in Fig. 7.8(a) is one method for moving line edges with the fine granularity of the design grid, even when writing on a grid twice that size. By exposing every other spot along part of the right edge the dose is reduced by half, and the edge is shifted to the left, relative to the edge where every spot is exposed. This enables higher throughput by using a spot size that is larger than the design grid.

The multipass gray approach, illustrated in Fig. 7.8(b), takes this concept one step further. With this technique the writing grid is four times that of the design grid.[51] The pattern area is scanned four times without requiring more writing time than if the mask was written with a writing grid equal to the design grid by using only one-quarter dose per pass. Not every pixel is exposed on each writing pass. By exposing some pixels only once, twice, or three times, the edges of features can be moved. There is also averaging that takes place with multiple passes that has the net effect of reducing linewidth, registration, and butting errors.[52]

Edge features are placed with finer granularity by reducing the number of exposures at the edge pixels, as shown in Fig. 7.9. In this figure, Gaussians are placed in a row and centered at points 0, 1, 2, and 3. Each has a standard deviation equal to two thirds. The total dose is the sum of all of the Gaussians. The three curves show the doses at the left edge of the pattern when the Gaussian centered at point 0 has 25%, 50%, 75%, or 100% of the peak magnitude of the doses at the other points. This represents one, two, three, or four exposures at point 0, while the other points are exposed on all four passes. As can be seen, the edge of the

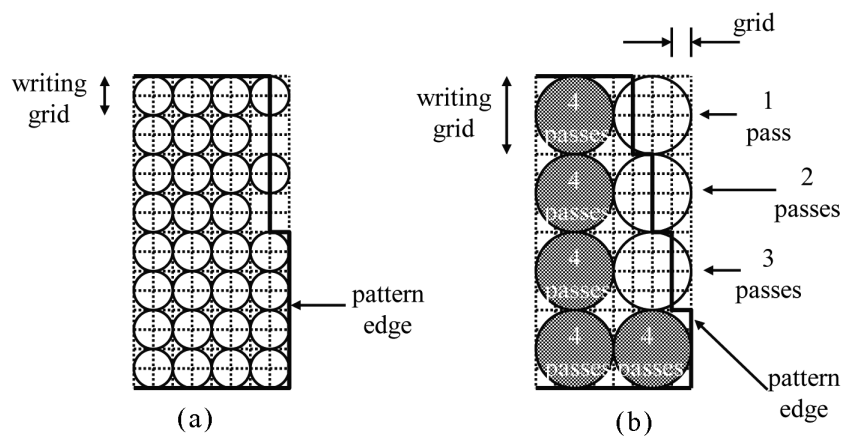


Figure 7.8 Different raster-writing schemes: (a) virtual addressing, and (b) multipass gray.

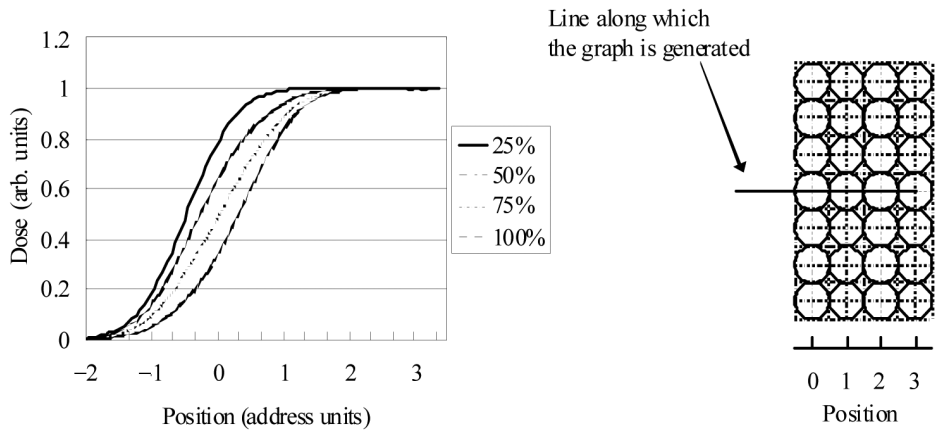


Figure 7.9 Exposure dose as a function of the dose of the Gaussian center at 0.

dose moves as the dose of the edge feature is varied. This approach can also be combined with pixel deflection for adjusting edge locations on masks.[53]

From Fig. 7.8 one might expect rough pattern edges to result from patterning schemes that involve large writing spots. However, reasonably smooth edges are actually made with large spot sizes, as shown in Fig. 7.10 (see Color Plates). In this figure, two-dimensional Gaussian spots were placed on grid points with integer coordinates, and $x \leq 0$. The resulting line edge, around a dose of 0.3, is seen to be reasonably smooth.

It was seen in Fig. 7.10 that reasonably smooth lines are generated from round Gaussian beams, but shaped beams nevertheless provide improvement, particularly

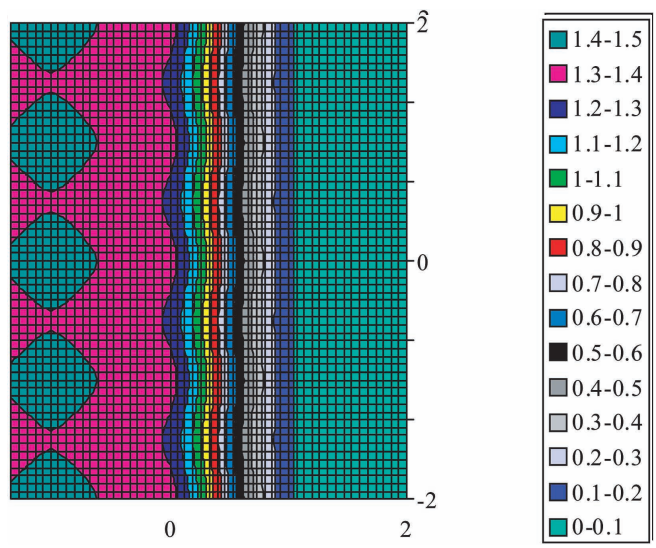


Figure 7.10 Exposure dose contours from Gaussians placed at integer coordinates (x, y), with $x \leq 0$ (see Color Plates).

with respect to sharpening corners. This was important for making 1× x-ray masks, and x-ray programs motivated the development of several vector scanning tools.

Another way to improve performance in electron-beam writers is to use higher beam energies. The original EBES system used 10-keV electrons, as did subsequent generations of MEBES machines. While the use of electrons avoids the problems of diffraction and reflection associated with optical imaging, electron lithography has a different set of issues. One problem is scattering.[54] When energetic electrons pass through matter, electrons scatter. Part of this scattering, the transfer of energy from the electrons to the resist, is an essential step in the lithographic process, whereby the solubility of the resist is altered by exposure to radiation. However, many electrons scatter into directions different from their original trajectories (Figs. 7.11 and 7.12), resulting in degraded resolution and proximity effects.[55,56] Scattering into the resist film, particularly the forward scattering that degrades resolution, is reduced by increasing the voltage of the electron beam. Consequently, newer electron-beam exposure systems have higher beam energies, with 50 keV being typical.

The scattering of electrons has been characterized as Gaussian broadening, with different amplitudes and widths for forward and backward scattering.[55] With this approach to the characterization of electron scattering, the energy distribution from a beam of electrons incident at the origin of a coordinate system is given by:

$$f(r) = \frac{1-\eta}{\pi\alpha} e^{-(r^2/\alpha^2)} + \frac{\eta}{\pi\beta} e^{-(r^2/\beta^2)}, \tag{7.3}$$

where r is the radial distance from the origin in the plane of the resist, α is the width of forward scattering (see Fig. 7.12), β is the width of the backscattering, and η is the fraction of the total scattered energy that is backscattered.

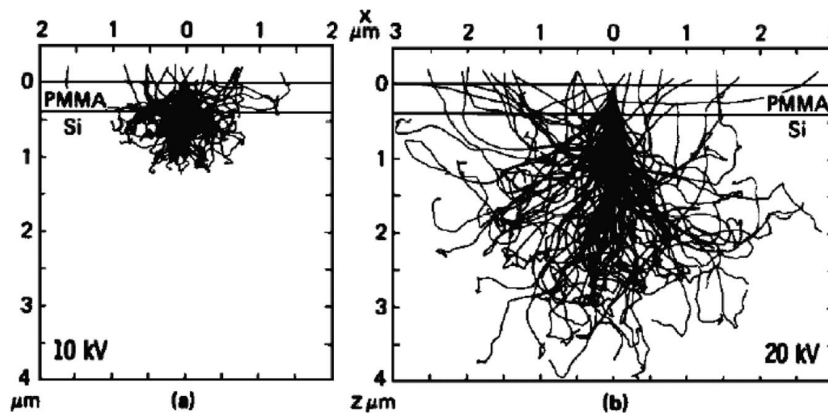


Figure 7.11 Monte Carlo simulation of electron scattering for PMMA on a silicon substrate.[56] The energies of the electrons for the two different examples are shown in the corners of the graphs.

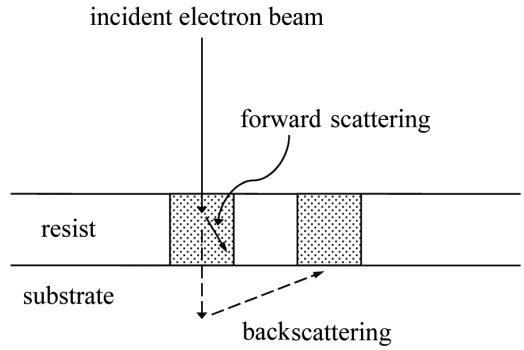


Figure 7.12 Illustration of resolution degradation and proximity effects in electron-beam lithography. The shaded areas are intended for exposure, but electrons will scatter into the regions that should be unexposed. Both forward and backscattering are shown. The dose in the "unexposed" portions of the resist depends upon the amount of e-beam exposure in adjacent areas.

The validity of this characterization has been assessed using Monte Carlo simulations,[57,58] with a representative result shown in Fig. 7.13. While there is imperfect quantitative agreement between the Monte Carlo simulations and the double-Gaussian model, there are some observable characteristics. First, the broadening attributed to forward scattering has higher peak intensity than that attributed to backscattering, and it also has a range <100 nm. On the other hand, the broadening attributed to backscattering extends over distances of several microns. Consequently, forward and backscattering have difference effects. Resolution is reduced by forward scattering, while backscattering leads to proximity effects.

Regardless of the beam energy, scattering occurs at a level of significance, and adjustments are required in order for patterns to be sized correctly, regardless of proximity. Consider the situation shown in Fig. 7.14. A seven-bar pattern is imaged using electron beams. The total exposure dose is the sum of incident and backscattered electrons. Note that the scattered dose is greater for the centerline than for lines near the edges of the seven-bar pattern. One might expect exact corrections to be very complex for random logic patterns, but fortunately, the backscattering range of 50-keV electrons is very long (~10 μm), enabling proximity corrections to be based upon average pattern densities rather than detailed layouts.

One clever approach to proximity correction is GHOST, which is based upon the idea that the background dose can be equalized by using a second exposure.[59] The pattern of the second exposure is the complement of the original, and the halfwidth of the beam used for the second pass is approximately equal to the distance over which the electrons are backscattered. The dose for the second exposure is proportional to the incident dose and the magnitude of backscattering.

The GHOST technique has the disadvantage of requiring a second exposure pass, reducing overall exposure-tool productivity. Consequently, many electron-beam writers rely on sophisticated software for proximity corrections.[60–62] In areas that receive considerable dose from electron backscattering, the beam writer

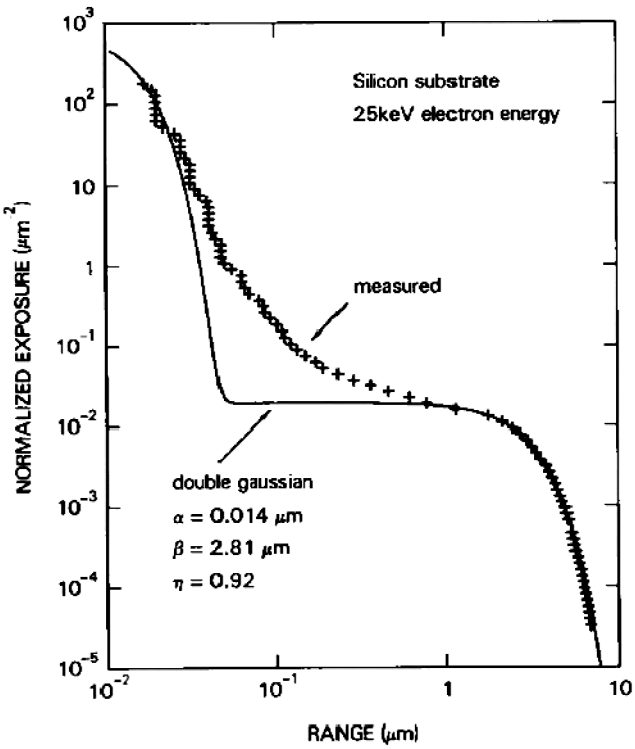


Figure 7.13 Monte Carlo simulation of electron forward and backscattering, compared to the double-Gaussian model.[57,58]

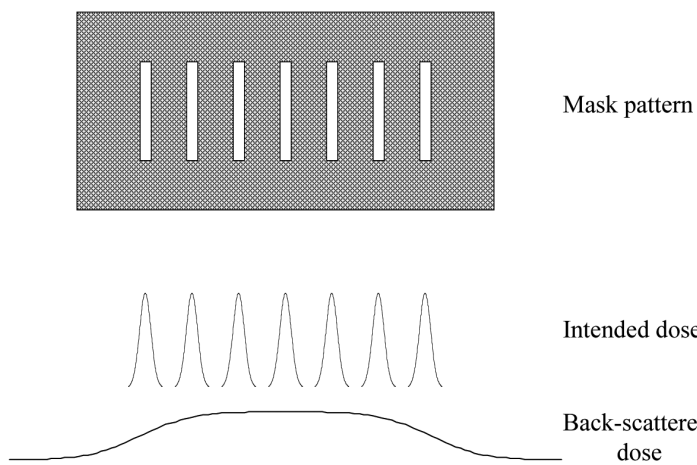


Figure 7.14 Dose in electron-beam lithography, including scattered electrons.

exposes with a reduced dose, or the edges of the directly exposed pattern are shifted, using some of the methods described previously.

Another way that electrons can scatter into unintended parts of the resist film is by reflection from the bottom of the electron-beam optical column (Fig. 7.15), a phenomenon often referred to as *fogging*.[63] Software solutions have been proposed, similar to those used to correct for more direct proximity effects.[64–66] Hardware solutions have been implemented involving various types of sufficiently deep chambers, sometimes with an electrostatic potential to reduce reflections of electrons from the bottom of the electron-beam column.[67,68]

Electron beams can impart an electrostatic charge to the substrate, particularly when a nonelectrically conducting resist is the top layer. The amount of charge will vary across the substrate according to pattern density, and this will also vary over time as the mask pattern is being written. This charging can lead to registration errors, because it will deflect the electron beam from the intended location.[70–72] A number of schemes have been proposed for mitigating the effects of substrate charging. One proposal involves the generation of a model to predict the electric fields generated by the electron beam,[73] but such models are necessarily complex, involving both the particular pattern being generated on the mask as well as the writing sequence. Another approach incorporates electrically conducting resists[74] or electrically conducting overcoats. This method is effective at reducing electrostatic charging, but overcoats often result in additional defects,[75] and the imaging properties of the electrically conducting resists are not always adequate.

Fundamental to assessing the magnitude of pattern-placement errors is equipment for measurement. Measuring absolute pattern locations over distances exceeding 100 mm, yet accurate to nanometers, is extremely challenging. Such specialized tools exist for performing such measurement, most notably the IPRO4, currently manufactured by KLA-Tencor.[76] While such tools can be calibrated quite precisely to artifacts,[77] it is very difficult to achieve absolute accuracy.[78–80] To accomplish this it is necessary to have accurate linearity across long distances and nearly perfect orthogonality between axes.

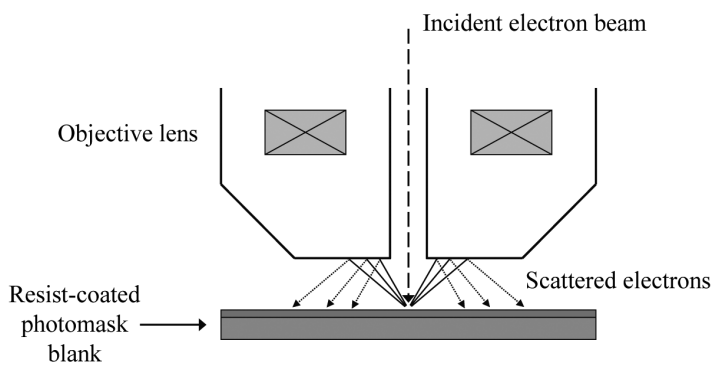


Figure 7.15 Backscattered electrons reflect from the bottom of the electron-beam column, causing a background of diffuse exposure.[69]

A beam writer requires input data in a particular format, which is typically specific to an individual model of writing tool, and this format is different from those used for circuit design and layout. What is also necessary is a standard intermediate format for the output of circuit layout tools. Every manufacturer of beam writers then only needs to be able to input the mask layout in this standard format and be able to convert the information to a form that can be used directly by the company's beam writers (see Fig. 7.16). A standard format is also useful for mask-defect inspection when comparisons are made between physical inspection of actual masks and design data. For many years the most popular intermediate format was GDSII, originally developed by the Calma division of General Electric. The legal rights to GDSII now reside with Cadence Design Systems. As circuit complexity increased, GDSII data files became very large and unwieldy, and a new format was required to reduce file sizes.[81] Such a standard is OASIS™—the Open Artwork System Interchange Standard.[82–84] OASIS has largely displaced GDSII as the intermediate data format for large, complex designs.

To facilitate layout and to control database size, design software typically involves hierarchical data structures, where sets of data are collected into cells that can be inserted repetitively. This approach is particularly useful for laying out memories, but is valuable for all types of design. For example, a NAND gate is a typical logic structure that can be used repetitively. GDSII and OASIS are hierarchical data formats, where NAND gates are made part of the layout as a whole structure, rather than through specification of each primitive shape

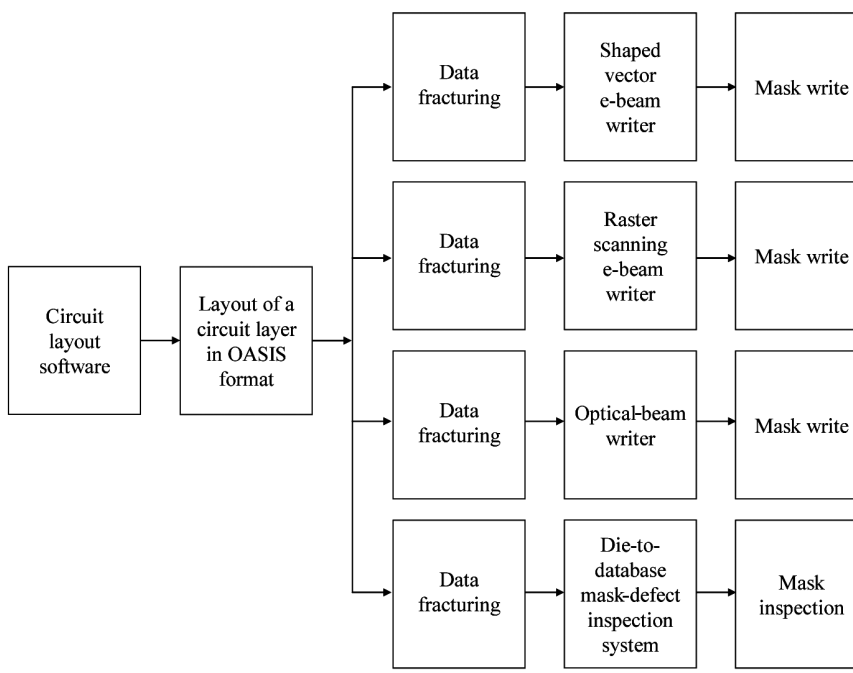


Figure 7.16 Use of an intermediate data format.[85]

comprising the structure. However, beam writers typically require data that can be streamed at very high writing rates. This requires that design or OASIS data hierarchy must first be removed, often referred to as "flattening" the data. After the hierarchy is removed, the circuit patterns need to be converted from polygons of arbitrary sizes and shapes to the primitive shapes (rectangles, triangles, etc.) consistent with the beam writer's architecture. Finally, the data needs to be broken down into the fields and subfields of the beam writer. This whole process of conversion from the hierarchical format is known as "fracturing." As a consequence of flattening and fracturing, beam-writer databases are usually much larger than the original design files.

7.5 Optical Mask Writers

Optical raster scanning systems are also widely used for patterning masks. The ALTA systems produced by Applied Materials are examples of such optical raster scanning beam writers. The architecture of an ALTA system is shown schematically in Fig. 7.17.[86] Light is generated by a laser. In the first models of optical raster scanning systems, this was an argon ion laser, operating at a wavelength of 364 nm, and improved resolution of optical-beam writers has since been achieved by using shorter wavelengths. For example, second harmonic generation of the 514-nm Ar laser line produces DUV light at a wavelength of 257 nm.[87] This light is then passed through conditioning optics and a beamsplitter. In

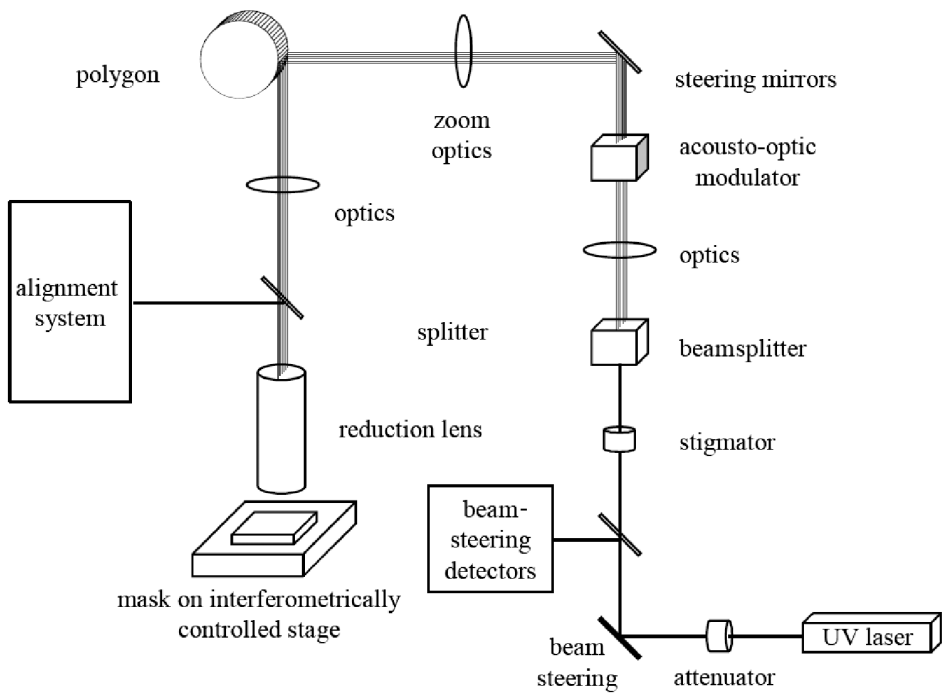


Figure 7.17 Schematic of the ALTA architecture.

the ALTA 4700, 32 separate beams are produced. By means of an acousto-optic modulator, the light intensity in each beam is set to one of 16 levels. The beams then propagate through a zoom lens that, along with a large number of gray levels, enables fine adjustment of line edges, down to a 5-nm effective address unit. Rapid scanning of the light across subfields on the mask is accomplished by the use of a rotating polygon, which is a 24-faceted mirror that can spin at rates up to 20,000 rpm.[88]

The ALTA 4700 has a 42× reduction lens with an NA of 0.9, enabling a minimum feature size of 400 nm.[89] (With a 4× stepper lens, this represents a minimum feature size on the wafer of 100 nm.) There have been proposals to extend this resolution to 200 nm.[90] Such a high numerical aperture necessitates good focus control. The mask undergoing patterning is moved on a laser-interferometrically controlled stage. There is also an alignment system on the ALTA 4700 that enables overlay of second patterns to those pre-existing on the mask, a requirement for the production of certain types of phase-shifting masks (discussed in the next chapter).

There are several advantages to optical-beam writers. With 32 beams, these systems are capable of high throughput. Beam-writer throughput has a significant impact on mask cost, since a substantial fraction of mask cost is due to the high cost of beam writers. With a high-throughput beam writer, this cost can be distributed among a larger number of masks. This helps to reduce the cost of masks with features that are not so small that they need to be fabricated using an electron-beam mask writer. The small effective address unit is useful for fine adjustment of linewidths on the mask, an important factor for implementing optical proximity corrections (also discussed in the next chapter). Optical exposures allow the use of well-characterized optical resists for mask making. This provides some advantage over electron-beam pattern generation, particularly as plasma etching of masks becomes more common.

Another approach to the optical patterning of masks is to use a micromirror array, much like those used in data projectors. Such systems are made by Micronic Laser Systems AB. The basic configuration for their Sigma7500-II system is shown in Fig. 7.18. Light is produced by a KrF excimer laser. After conditioning by illuminator optics, the light is reflected from a beamsplitter and then from a spatial light modulator (SLM).[91] The SLM is an array of 10^6 mirrors, each of which is 16×16 μm in size, which are demagnified by 200× using an 0.82-NA lens.[92] The mirrors of the SLM are deflected to produce the desired pattern on the reticle substrate. The minimum main feature size for the Sigma7500-II is specified as 220 nm.

7.6 Resists for Mask Making

Resists for optical pattern generation need to be matched with the beam writer's exposure wavelength, just as the resists used for wafer patterning need to be matched to the stepper's exposure wavelength. For many years, it has been common practice for mask makers to use mask blanks precoated with resist by the

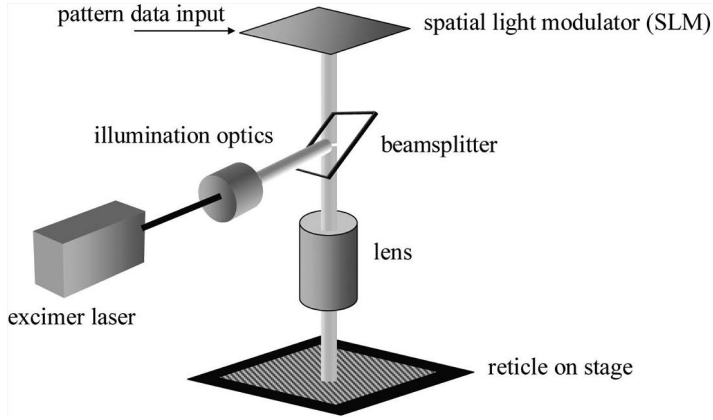


Figure 7.18 Schematic of the Micronic Sigma7500 II architecture.

blank supplier. As typical i-line resists are sufficiently stable, this is a reasonable thing to do for masks generated on optical-beam writers. An example of a resist commonly used with 364-nm exposure systems is iP3500 from Tokyo Ohka Kogyo Co., Ltd. With optical exposure, there are standing waves in the resist, reducing process control. This is improved by the use of post-exposure bakes, just as is found in wafer processing.[93]

Mask-patterning tools operating at DUV wavelengths lead to the prospect of chemically amplified resists, with all of the instabilities discussed in Chapter 3. The transition to DUV optical-patterning tools has motivated mask shops to coat the mask blanks themselves. These also require post-exposure bakes. Control of bake temperatures of the resist is more difficult with thick glass substrates relative to what is achievable with silicon wafers. Consequently, low sensitivity of the resist to the post-exposure bake temperature is an important parameter to consider when choosing a chemically amplified resist for mask making.

Electron-beam pattern generation usually requires e-beam-sensitive resists. Since the mechanism for e-beam exposure is different from optical exposure,[94] materials optimization is beneficial, although there are some resists, such as Fujifilm Electronic Material's FEP171, that work well exposed on e-beam or DUV optical mask writers. For many years, the most commonly used electron-beam resist was a positive material developed at Bell Laboratories, poly(butene-1-sulfone), usually referred to by its acronym, PBS[95] (Fig. 7.19). This material is an alternating copolymer of 1-butene and sulfur dioxide that undergoes scissioning upon exposure to energetic electrons. The lower-molecular-weight byproducts of the exposure are soluble in organic solvents. Typical developers for PBS are mixtures of pentanone and other solvents, such as methyl isoamyl ketone.[96]

PBS is a reasonably sensitive resist and can be exposed with doses on the order of 1 $\mu C/cm^2$ at 10-keV beam energies. Sensitivity is critical for high patterning-tool throughput. Even with such sensitive resists, several hours are required to pattern a single mask for a leading-edge microprocessor or memory, so significantly less-sensitive resists are problematic. PBS also has the virtue of

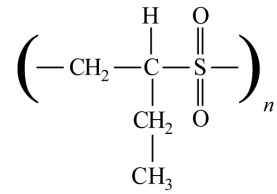


Figure 7.19 Chemical structure of PBS.

being able to hold up to the wet-etch chemistries used to etch chromium masks. For these reasons, mask makers long tolerated the otherwise poor lithographic performance of PBS.[97] However, new resists are now being used, such as FEP171 from FujiFilm Electronic Materials, SEBP9092 from Shin-Etsu, or ZEP 7000, a polymer of methylstyrene and chloromethyl acrylate from Nippon Zeon. All of the resists require somewhat higher doses (8–10 $\mu C/cm^2$) than PBS, but provide good lithographic performance and can be used when plasma-etching chromium.[98] ZEP 7000 uses an organic solvent developer (a mixture of diethyl ketone and diethyl malonate, or a mixture of methyl isoamyl ketone and ethyl malonate), which evaporates much faster than water. Because the development rate is temperature dependent, and evaporation causes nonuniform cooling, it is more difficult to control the development of resists that use organic solvents as developers. A TMAH-based aqueous developer is used with FEP171.[99]

As electrons penetrate a resist film the electrons undergo inelastic scattering, resulting in a cascade of secondary electrons.[100] Attendant energy transfers can induce chemical reactions, i.e., exposure of the resist. As discussed previously, this scattering can limit the resolution potential of electron-beam lithography. In optical lithography involving photons with 193-nm wavelengths and longer, absorption occurs between molecular energy levels. Since this is a localized event, it is typically a photoactive compound or photoacid generator that directly absorbs the light. In contrast, the high-energy electrons used to pattern masks cause ionization at the atomic level. Ionization can also be produced by secondary electrons produced by inelastic scattering of the primary electron beam. The essential chemical reactions that ultimately lead to changes in the solubility properties of the resists result from secondary reactions that originate with ionization.

In addition to positive resists, negative resists are used for mask making. With vector scanning, where scanning is limited to the features that are to be exposed, exposure times can be reduced substantially for certain patterns by using negative resists. Another advantage of negative resists is illustrated in Fig. 7.20. Although there is no significant difference between positive and negative resist processes in dimensional control of directly patterned features, there is a difference when the critical feature is a fixed tone. For example, the critical feature for gate masks is typically a line. Features need to be exposed on both sides of this line when using positive resists. As a consequence, pattern placement as well as direct linewidth control will have an impact on the dimension of this most critical feature. When line-dimensional control is critical, such as for gate masks, negative resists have

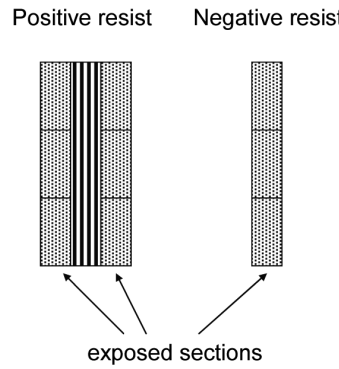


Figure 7.20 Exposures using positive and negative resists. To create a line in resist, exposures to the left and right of the line are required with positive resists, while only the line itself needs to be exposed when using negative resists.

inherent advantages.[101,102] Examples of negative chemically amplified resists are Fujifilm Electronic Material's FEN271 and Shin Etsu's SEBN1637.

Chemical amplification provides good resist sensitivity, but such resists usually have problems of delay stability and sensitivity to bakes. The resist KRS-XE, developed at IBM, appears to be relatively insensitive to PEB temperature and delays between exposure and bake.[103] This resist uses an aqueous developer and has slower etch rates in plasma environments than ZEP 7000.

One of the problems accompanying the transition to higher electron-beam voltage is heating. While the total mask temperature may not rise very far (see Problem 7.3), localized heating can be tens of degrees centigrade. This localized heating has the potential to affect resist behavior significantly. This is shown in Fig. 7.21. For ZEP 7000 resist, linewidth variations greater than 10 nm have been attributed to this temperature effect, although chemically amplified resists appear to be less sensitive to heating by the electron beam.[104] If the temperature rise becomes

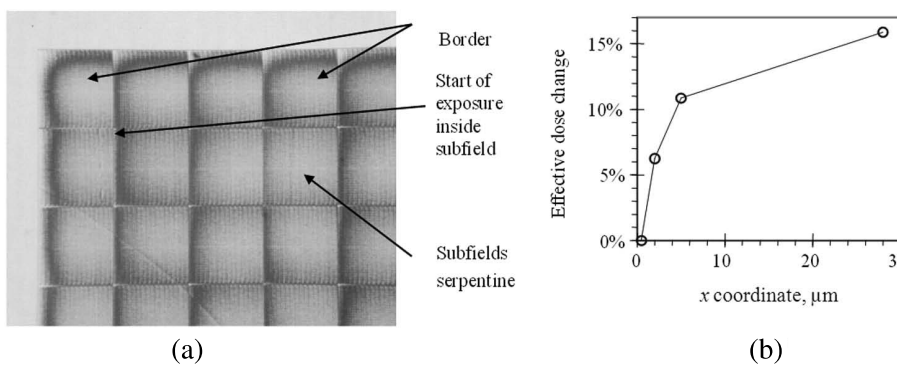


Figure 7.21 (a) Partially developed novolak-type EBR900 resist following four-pass vector scanning exposure. The effect of resist heating is apparent, although much less than in single-pass scanning. (b) Change of effective dose due to heating in four-pass vector scanning exposure.[106]

high enough, positive resist can even be converted to negative tone! Because the local temperature rise is related to the electron-beam dose, the overall effect depends upon the beam-writing strategy. In particular, vector scanning systems result in greater local temperature rises than raster scanning systems. The effect is mitigated in both cases by the use of multipass writing strategies. [105]

7.7 Etching

For many years the chromium films of photomasks have been wet etched. Typical chromium wet etchants contain ceric ammonium nitrate—$Ce(NH_4)_2(NO_3)_6$—mixed with nitric, perchloric, or acetic acid.[107] With wet etching there is appreciable undercut. Consequently, chromium thickness variations contribute to linewidth variations on the reticle, particularly when wet etching is used.

Patterns cannot be transferred with good fidelity into the chromium film with wet etches when the undercut becomes comparable in size to the features. Features smaller than 100 nm are found on the mask today, dimensions comparable to absorber thicknesses. For critical applications, chromium mask etching has moved to dry etching for the same reasons that wafer etching underwent a transition to plasma etch processes. Because CrO_xCl_y species are volatile,[108] typical chromium etches involve mixtures of Cl-containing gases (such as Cl_2 and CCl_4) and O_2,[109] with additional gases added to reduce etch-loading effects that can have an impact on isolated-dense biases.[110]

For decades chromium and chromium-containing materials were used for absorber films on photomasks. However, chromium has proven to be a difficult film to dry etch, so there has been some work on alternative opaque mask materials, most notably MoSi. MoSi has been applied extensively for the fabrication of attenuated phase-shifting masks (which are discussed in the next chapter), and more recently this material has been used more heavily for binary masks, notably with Shin Etsu's Opaque-Molybdenum-Over-Glass (OMOG) material.[111] Because MoF_6, $MoCl_4$, SiF_4, and $SiCl_4$ are volatile compounds, halogens and halogen-containing gases are used for etching MoSi.[112,113] Since SiO_2 is etched by fluorine chemistries, oxygen is sometimes added to improve selectivity between the absorber and the glass substrate.

As will be discussed in the next chapter, the SiO_2 glass substrate is sometimes etched to produce a phase shift. In this case, there is a significant knowledge base from silicon-wafer processing regarding plasma etching of SiO_2. In the case of photomask etching, compatibility with the resists used for mask making and the materials used for absorbers is necessary.

7.8 Pellicles

As mentioned at the beginning of this chapter, step-and-repeat and step-and-scan modes of wafer patterning require masks with no killer defects in order to achieve good yields. While masks are made without any defects that result in nonfunctional die, preventing particles from depositing on masks during extended mask usage is

a challenge, even in state-of-the-art cleanrooms. To avoid new printable defects, pellicles are attached to photomasks.[114] Pellicles are thin (~1 μm) polymer films stretched across a frame that is attached to the mask (Fig. 7.22). Typical frame heights are 5–10 mm; 6.35 mm is a typical value, which is the same as the thickness of the photomask blank, although lower heights are becoming more common. Particles deposited on the pelliclized photomask fall onto the pellicle or glass backside of the mask, and are therefore several millimeters away from the chrome features that are being imaged. With small depths-of-field, these particles are not going to be in focus.

With typical depths-of-focus less than 1 μm, millimeter standoffs are expected to blur the image of the particles significantly. The requirements of pellicles go beyond blurring the images of particulate defects. Pellicle standoffs must be large enough to prevent defects from reducing the light intensity of the desired mask patterns significantly. Early theoretical studies[115] showed that image intensities are not affected by amounts greater than 10% as long as the pellicle standoff is at least as large as

$$t = \frac{4Md}{NA}, \tag{7.4}$$

where M is the lens reduction, NA is the numerical aperture of the lens, and d is the diameter of the particle on the pellicle. More detailed theoretical investigations have shown that imaging is protected for particles only about one half of that given by Eq. (7.4). For example, process windows are not reduced significantly with pellicle standoff distances of 6.3 mm and particles <90 μm.[116] Usually a pellicle needs to be attached only to the chrome side of the mask, since the glass blank itself serves the same purpose as the pellicle with respect to particles.[117]

Pellicle films are usually polymers, with nitrocellulose and forms of Teflon being common. These materials must be mechanically strong when cast as thin films, be transparent, and resistant to radiation damage. Good light transmission through the pellicles is a combination of transparency and optimization of the thin-film optics.[118] Transmission through a nonabsorbing film as a function of film thickness is shown in Fig. 7.23. As can be seen, the pellicle transmission is maximized at particular thicknesses, and pellicles are fabricated at thicknesses corresponding to such maxima. In some instances, antireflection coatings are

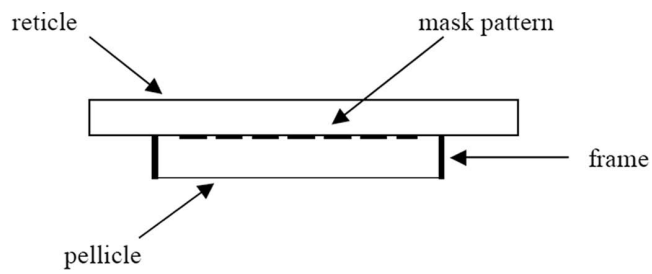


Figure 7.22 Cross-sectional view of a mask with a pellicle attached.

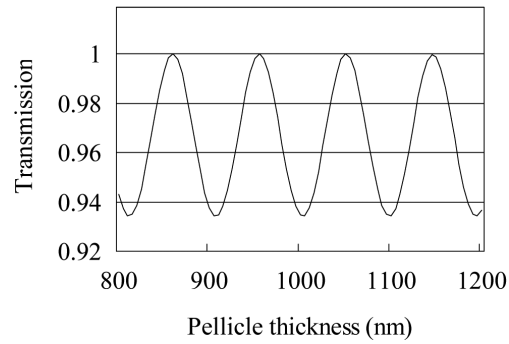


Figure 7.23 Calculated transmission through a thin Teflon AF™ film, as a function of film thickness, for normally incident 248.3-nm light. The index of refraction for the film is 1.3.

applied to the pellicle films. Transparency and resistance to radiation damage become more of a challenge as the wavelength is shortened. Transitions to shorter wavelengths have always required some level of pellicle re-engineering. It has been possible to design and fabricate pellicles that have high transmission at multiple wavelengths. For example, there are pellicles that are suitable for both i-line and KrF lithography.

Pellicle frames are usually made of anodized aluminum. Small holes are commonly drilled into the frame so the pressure of the air in the space enclosed by the mask and pellicle remains equal to the ambient air pressure.[119] These holes are very small, or involve circuitous routes, to prevent these air paths from becoming routes for particles that would defeat the purpose of the pellicle. If the surface of the frame attached to the mask is not very flat, applying the pellicle to the mask can cause the mask to distort, resulting in registration and focus errors. Induced registration errors on the order of 100 nm have been measured.[120] These distortions are often unrecognized, because reticle registration measurements are made prior to pellicle application in many mask shops. The magnitude of the distortion depends upon the compliance of the adhesive used to bond the pellicle frame to the mask. Less mask distortion occurs with flatter pellicle frames and more flexible adhesives.

Even with perfectly flat pellicles and masks, there is a potential for mask deformation induced by the pellicle during use. Pellicle frames are typically made of aluminum, which has a significantly different coefficient of thermal expansion than the fused-silica substrate of the photomasks. During exposures, the masks will absorb some light and heat up. The difference in coefficients of thermal expansion can lead to reticle deformation. For this reason, people have considered using fused silica as a material for pellicle frames.

Not only must pellicle adhesives be compliant, but also they need to be low-outgassing, to avoid material deposition on the mask surface. Because scattered light can reach the pellicle adhesives, these materials also need to be stable to exposure by the actinic light. This is a challenge that evolves with changes in wavelength.

Having a very thin pellicle reduces many of the requirements for the pellicle. This is a consequence of refraction, and can be understood by considering Fig. 7.24. If a pellicle is not perfectly flat, normally incident light refracts and is displaced by a distance of δ given by

$$\delta = t\left(1 - \frac{1}{n}\right)\sin\phi, \tag{7.5}$$

where t is the thickness of the pellicle and n is the index of refraction of the pellicle material, which is typically 1.3–1.5. For thin pellicles, the displacement of light rays by refraction is small, even for moderate amounts of pellicle nonflatness. Other aberrations can be introduced by a tilted pellicle.[121]

Even a perfect pellicle distorts the wavefront. Consider the situation shown in Fig. 7.25. Because of diffraction, light propagates from openings in the mask at a various angles ϕ. The largest relevant angle is set by the numerical aperture of the projection optics on the reticle side of the lens, which is the numerical aperture usually quoted, divided by the lens reduction. This results in moderate angles of incidence being relevant. For a lens with a specified NA (on the wafer side), there can be angles ϕ up to arcsin(NA/N), where N is the lens reduction. For a 0.93-NA 4× lens, angles can be as large as 13.4 deg. The light rays that are nonnormal to the pellicle result in wavefront errors. These are calculated as follows. Light propagating through the space of the pellicle has an optical path length of $t/\cos\phi$

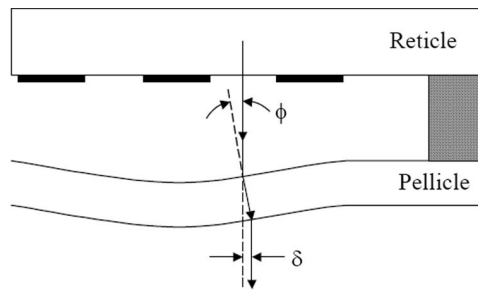


Figure 7.24 Geometry placement errors because of nonflat pellicles.

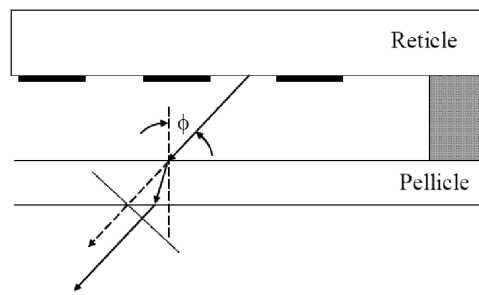


Figure 7.25 Wavefront aberrations from a pellicle.

in the absence of the pellicle. With a pellicle, the light traverses a different optical path length to reach the line perpendicular to the straight light ray. The optical path difference (OPD) is

$$\mathrm{OPD} = t\left\{\frac{-1}{\cos\phi} + \frac{n}{\cos\left[\arcsin\left(\frac{\sin\phi}{n}\right)\right]} + \left(\sin\phi\tan\phi - \frac{\sin^2\phi}{n\sqrt{1-\frac{\sin^2\phi}{n^2}}}\right)\right\}. \quad (7.6)$$

The first term is the optical path for a ray traveling straight in air. The second term is for the refracted ray in the pellicle, and the third term is the optical path of the refracted light from the pellicle to the point where it meets the line normal to the rays, shown in Fig. 7.25.

To the second order, this gives an optical path difference of

$$t\left[n - 1 + \frac{\phi^2}{2}\left(1 - \frac{1}{n}\right)\right]. \quad (7.7)$$

A constant optical path difference simply represents an overall phase shift for all rays and has no net effect on imaging, unlike variations in phase as a function of angle. From Table 4.2, we see that Eq. (7.7) represents a focus error. There are also higher-order aberrations induced by pellicles, representing spherical aberration.[122] These aberrations are minimized by the use of very thin pellicles, as seen from Eqs. (7.6) and (7.7). Aberrations are also minimized by a low value for the index of refraction n of the pellicle material. Fluoropolymers, which are useful as pellicle materials because they are transparent and durable at deep ultraviolet wavelengths, also have reasonably low indices of refraction at exposure wavelengths. For example, Teflon AF2400 has an index of refraction ≈ 1.35 at a wavelength of 193 nm.[123]

Additional complexities have accompanied the use of immersion lithography, which results in high angles of incidence on pellicles (see Problem 7.5). Shown in Fig. 7.26 is the transmission through a typical ArF pellicle as a function of the angle of incidence. At low angles, the transmission is very high and varies little with the angle of light rays incident on the pellicle. However, at numerical apertures >1.0 relevant to immersion lithography (see Chapter 10), the pellicle's transmission varies with angle of incidence, effectively inducing apodization and causing linewidths to vary (Fig. 7.27) as a function of pitch (among other consequences). There will also be differences in phase of the transmitted rays as a function of the angle of incidence.[124]

Even with pellicles, particulate defects can form on the surfaces of photomasks. Pellicles protect masks from particles, but they do not prevent gases from entering the volume between the mask and the pellicle. As noted above, pellicle frames typically have holes in them to allow for pressure equalization between the ambient air and the volume between the pellicle and mask. Although the air inside

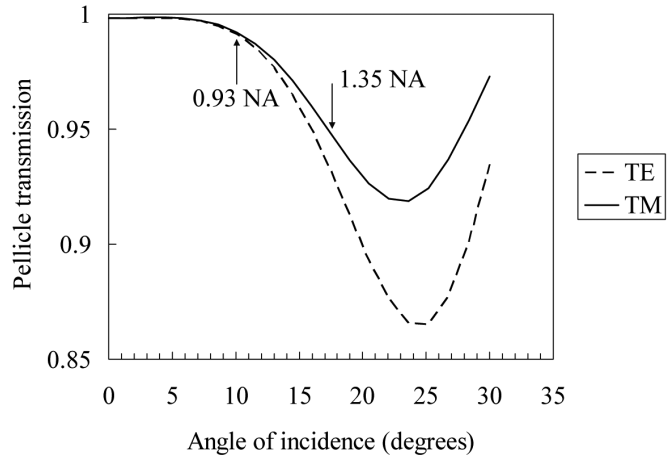


Figure 7.26 Calculated transmission of light through a pellicle as a function of the angle of incidence. For the calculations it was assumed that the pellicle thickness was 825 μm and index of refraction was 1.4.[125] Transverse electric light is perpendicular to the plane of incidence, while transverse magnetic light is parallel to the plane of incidence.

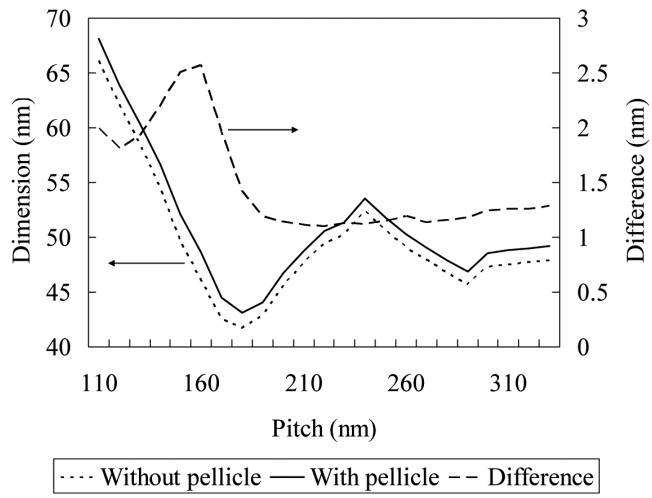


Figure 7.27 Calculated critical dimensions through pitch for 193-nm immersion lithography.[125] A 1.35-NA lens is assumed, with azimuthally polarized C-quad illumination (0.8/0.5/30 deg), for a nominally 55-nm line on a binary mask. Differences in critical dimensions over 2 nm are calculated between masks with and without a pellicle. For the calculations it was assumed that the pellicle thickness was 825 μm and index of refraction was 1.4.

steppers is very clean, very small amounts (<1 ppb) of contaminants can lead to photochemical deposition on mask surfaces. This may be enhanced because of traces of chemicals remaining on the mask surfaces from the mask fabrication process. Under intense DUV illumination, these chemicals can react and ultimately form particulates.[126,127] For example, ammonium sulfate was one material found on photomasks on which haze had grown. The ammonium could have come from ammonium hydroxide used for reticle cleaning, while the sulfur could have come from sulfuric acid (also used for reticle cleaning) or from sulfur dioxide in the ambient air. Molecular contaminants can also outgas from pellicle adhesives. A possible mechanism for the generation of ammonium sulfate is[128,129]

$$O_2 + \text{photon (193 nm)} \rightarrow 2O$$
$$SO_2 + O \rightarrow SO_3$$
$$SO_3 + H_2O \rightarrow H_2SO_4$$
$$H_2SO_4 + 2NH_3 \rightarrow (NH_4)_2SO_4.$$

Masks, being made from glass, are fragile, and need to be handled with care. In addition to the risk of mechanical damage, the use of an electrical insulator as a substrate makes photomasks susceptible to damage from electrostatic discharge (ESD).[130] For assessing the susceptibility to ESD of particular mask-handling methods, a mask (the Canary Reticle™) has been designed that has structures particularly liable to electrostatic discharge.[131] These consist of arrays of large chrome rectangles with isolated chrome lines extending from them. The tips of these isolated lines are close (1.5 μm) to adjacent large chrome areas (Fig. 7.28). These structures are arranged so that they point towards the interior of the reticle, with arrays originating from all four sides of the reticle.

Sparks from discharges cause the mask absorber material to melt and can lead to bridging across gaps on the mask. An example of this is shown in Fig. 7.29. This has been observed in a controlled experiment when the Canary Reticle was contained in a nonstatic-dissipative storage case that was subjected to a potential of several thousand volts. Inferring what will happen to actual reticles from tests involving the Canary Reticle is not clear, but this reticle can be quite useful in identifying causes of electrostatic discharge that are actually occurring. Reticle damage from ESD is a rare event, and tracking down causes is often difficult. The Canary Reticle, with its enhanced sensitivity to ESD, can expedite the identification of sources of ESD. More recently, a method has been developed to assess risk for

Figure 7.28 The basic structures of the Canary Reticle.

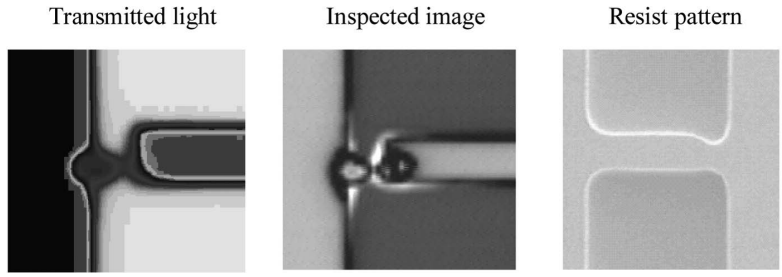


Figure 7.29 Example of mask damage from ESD.

electrostatic damage to masks that does not involve the use of something like the Canary mask, which is permanently damaged by testing.[132]

More recently, another electric field–induced damage mechanism has been identified: electric field–induced metal migration (EFM).[133–135] This occurs when voltages are too low to cause sparks, but can induce material migration (see Fig. 7.30). As features have become smaller, the gaps between geometries on masks have become very small. With very small gaps, even modest voltages can lead to large electric fields. Simulations indicate that voltages less than 100 V between structures can lead to field-induced migration. Features close to the edges of masks are particularly at risk. Masks are often constructed with rings of chrome surrounding the device patterns, so small features should not be allowed to be close to this guard ring. Keeping reticles away from the influence of external electric

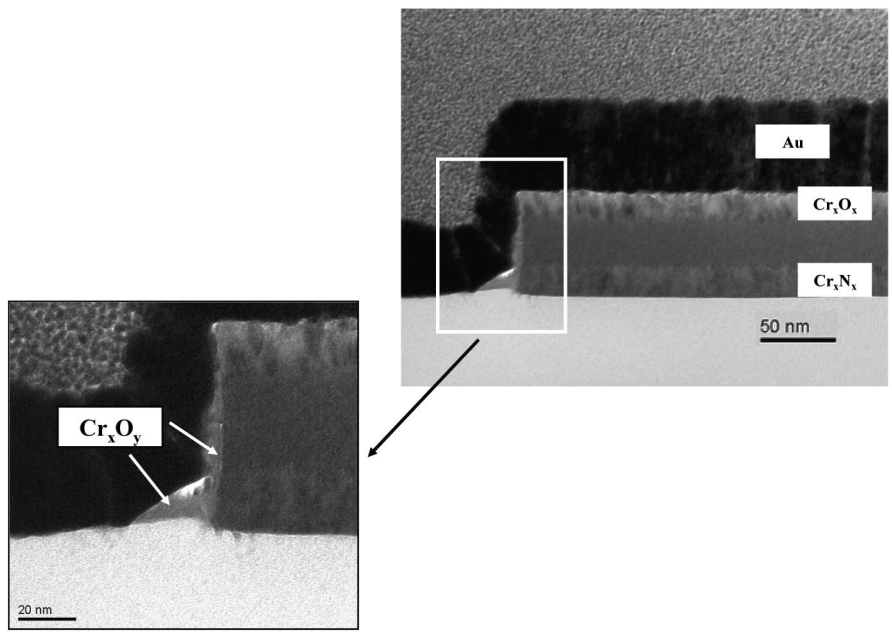


Figure 7.30 Transmission electron micrographs reveal chromium-oxide migration. Gold is used to reduce sample charging during the microscopy.

fields are also important, which can be accomplished with suitably designed reticle cassettes that contain the masks within a Faraday cage.[136] The situation is even more complex, as the amount of material that migrates appears to be related to the amount of light exposure that the mask receives.[137] Fortunately, there are test devices that can measure electric fields to assess risk.[138]

7.9 Mask-Defect Inspection and Repair

As noted earlier, there is a very low tolerance for imperfections on masks and reticles, since a single defect can cause a substantial loss of die yield. Consequently, there have been substantial engineering efforts with the purpose of designing and building inspection tools capable of finding defects on masks. Since every fabricated mask involves a substantial investment, tools have also been developed for repairing defects, once located. Defect inspection and repair are the subjects of this section.

Mask-defect inspection tools fall into three general categories: die-to-die, die-to-database, and particle inspection. The first die-to-die inspection tools were manual optical comparators,[139] which provided human operators with images that accentuated differences between two features viewed optically. In the operation of these systems two microscope objectives were focused on equivalent portions of different dies on the mask. Only small differences were seen when comparing the images of equivalent portions of two defect-free die of the same design, but there would be substantial differences if one of the die had a defect. Only a small part of the mask could be inspected at a time, equal to the field of view of the objectives. As designs increased in complexity and feature sizes became smaller, it became necessary to automate this inspection and improve the resolution of the optics.[140] This has been accomplished, and several automatic mask-defect inspection tools are commercially available.

With the introduction of ultra-large-scale integration and the transition to reduction steppers and scanners, it is possible that only a single die might reside on a reticle.[141] To inspect such reticles for defects it is necessary to compare the measured optical images of the reticle with calculations of what the image should be, given a defect-free reticle. Such tools have also been developed, with die-to-die inspection capability also combined with die-to-database capability within a single tool.[142] Examples of commercially available mask-defect inspection tools are listed in Table 7.6. In some instances it is not strictly optical images that are being compared, but auxiliary quantities, such as the amount of light transmitted through a small area on the mask.[143]

Finally, there are many systems that detect particles through light scattering. These are particularly useful for inspections after masks have been manufactured and initially qualified. In these situations the mask pattern typically remains undamaged, and acquired defects consist primarily of particles that land on the mask or pellicle during usage or handling. However, as discussed below, the assumption that the mask pattern remains undamaged is not always a good one.

Table 7.6 Examples of commercially available mask-defect inspection tools[144,145] All of these tools (except those from Lasertec) have die-to-die and die-to-database inspection capability.

Company	Model name	Inspection wavelength (nm)
KLA-Tencor	Teron 600	193
Applied Materials	Aera2	193
NuFlare	NPI-5000 Plus	199
Lasertec	MATRICS X700	213

Because the optical properties of materials vary with the wavelength of light, particularly when considering light of wavelengths in the visible versus ultraviolet portions of the electromagnetic spectrum, defect detection is strongly wavelength dependent. The most dependable defect inspection can be expected from inspection tools that use light at or near the wavelength of light used for exposures. Hence, when KrF lithography was the leading-edge optical technology, inspection tools often used 257-nm light. This was near the 248-nm wavelength of the KrF light and could be produced conveniently by frequency doubling the emission of 514-nm argon-ion continuous-wave lasers. Continuous-wave lasers are useful for inspection, because it is convenient to acquire data at rates considerably faster than the 1–6 kHz frequencies of excimer lasers. Nevertheless, some inspection tools use excimer lasers for illumination, and good inspection rates have been achieved.

Masks need to be cleaned to remove any particles that sit on the surfaces of the mask and could block the transmission of light. Such particles are usually referred to as "soft" defects. The majority of "hard" defects on masks consist of missing absorber, or absorbing material being where it should have been removed. Over time, four approaches to defect repair have been developed and implemented:

(1) Laser ablation and deposition[146–148]
(2) Focused ion-beam sputter etch and ion beam–assisted deposition[149–151]
(3) Micromachining using atomic-force microscope techniques[152–155]
(4) Electron beam–assisted chemical etching and deposition.[156]

Laser repair tools were the first to be adopted and served the semiconductor industry well for many years. However, such tools are limited by finite optical resolution, and ion-beam tools were introduced to address the needs associated with small features on masks, since ion beams can be focused to very small spots.[157] Most typically, gallium ions are used, because high-current gallium-ion sources can be made.[158] The ion energy is typically tens of keV, to provide the kinetic energy for sputtering opaque defects.[159] However, such highly energetic beams also tend to implant into the glass substrate, leaving a "stain" that results in an imperfect repair. To reduce this "stain," gases are introduced into the chamber containing the mask to assist chemically in the removal of unwanted material. In their neutral state these gases are inert, but they will produce energy in the area that is being bombarded with ions. Chemically assisted processes also enable sputtering and subsequent redeposition of material to be reduced. Ion beams can also be used to add absorber where it is missing by injecting appropriate gases into the chamber

containing the mask that is under repair. With the use of hydrocarbon-containing gases,[160] a carbon patch can be formed where the ion beam strikes the surface of the mask.

Over time, concerns with gallium staining and sputtering led to the development of electron beam–based repair tools.[161] For removing unwanted absorber, gases are introduced into the chamber containing the mask, and material is removed by means of electron beam–assisted etching. The addition of material to replace missing absorber is accomplished in an analogous way. Micromachining approaches have also been used for repairing masks, but only for material removal.[162]

Inspection tools will find defects and imperfections on masks, but the printability of these defects is an issue of practical importance, since the repair of defects is a complex and expensive process that should be avoided if unnecessary. It was recognized long ago that the printability of a defect depends upon its proximity to other features on the mask.[163] This is illustrated in Fig. 7.31. Consider a process for which the smallest feature that can be resolved on the wafer is 40 nm in size. For 4× reduction optics, this means that the smallest printable feature on the mask is 160 nm. An isolated object of half that size (80 nm) is well below the resolution of the system and will not print. On the other hand, an 80-nm size feature that is a protrusion into a narrow space will push that same space below the resolution limit, and such a defect will likely print. Since the printability of a defect is context related, and there are many possible configurations for geometries on masks, the subject of mask-defect printability is complex and has been the subject of much study.[164–169]

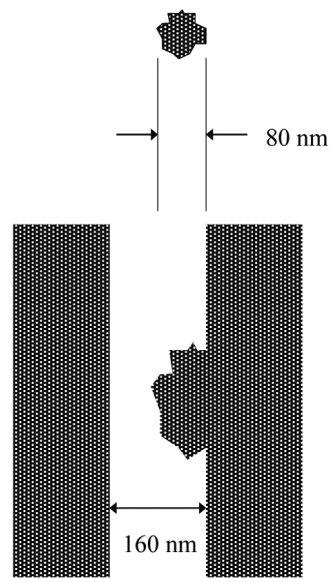


Figure 7.31 Illustration of mask defects of identical widths. The isolated feature will not print, while the defect which is a protrusion of the right line is printable.

As noted, the printability of a defect is important to know in order to decide whether or not to attempt a repair. Moreover, the repair itself may induce damage that can print. Hence, it is very useful for mask shops to have means of determining mask-defect printability, and it is greatly preferable to have the means to do this other than by using extremely expensive steppers or scanners. One way to do this is through the use of an aerial-image-microscope system, more commonly referred to by its acronym, AIMS^{TM}.[170–172] In these tools, high-resolution optics are used to measure the aerial image from a small portion of the mask. Because the field of view (~15 μm × 15 μm) is considerably smaller than the imaging field of systems used to pattern wafers, the optics of AIMS tools are much smaller and considerably less expensive than those of full-field exposure tools. Moreover, the complexity of resist processing is avoided, with the downside that subtleties involving the resist are not taken into account. An example of the information obtained from an AIMS tool is shown in Fig. 7.32. The pictures in this figure illustrate that

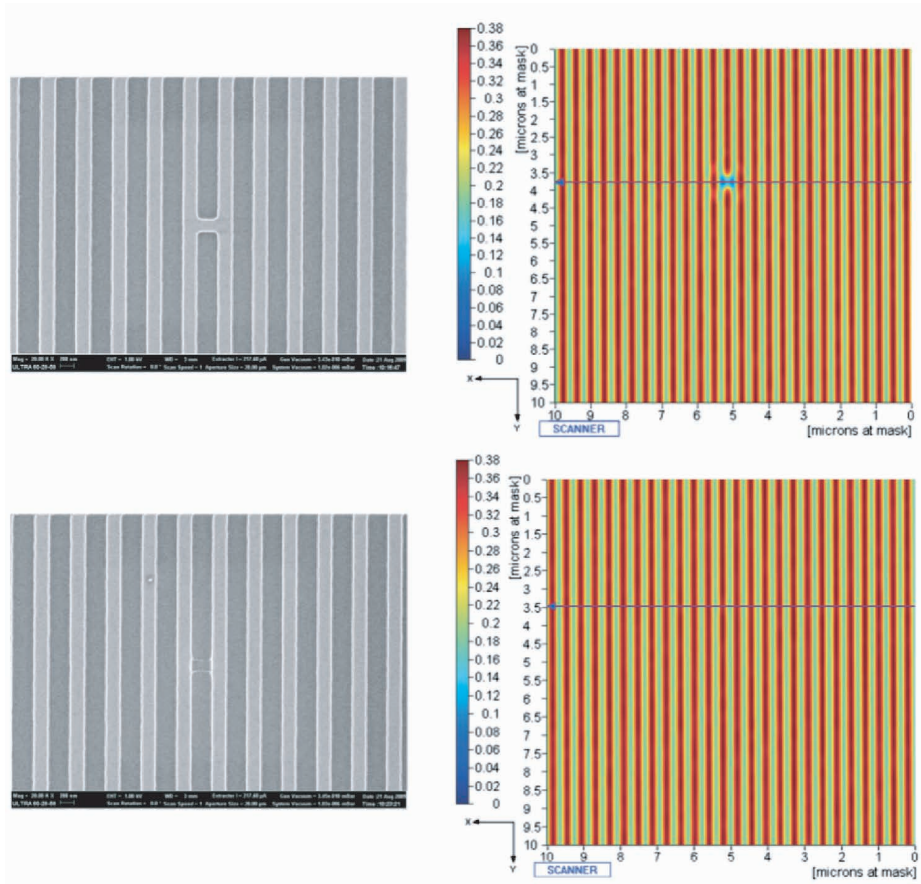


Figure 7.32 (a) A scanning-electron micrograph of a bridging defect on a mask, and (b) the measured aerial image from the mask. (c) The corresponding results after the mask was repaired (see Color Plates).

defect printability can be directly measured. AIMS measurement prior to repair provides data for determining that the repair is actually necessary, while a later AIMS measurement is used to assess the quality of the repair. Inspection tools that are based on the images, as would be produced by exposure tools, simplify mask making by combining the AIMS function into the inspection tool.[173] Even with such types of defect-inspection tools, an AIMS tool still might be used in the mask shop in order to qualify defect repairs.

As seen in this chapter, mask and reticle technology is quite complex. The need has been discussed for zero defects, good linewidth control, and good registration, because of the "tyranny of the stepper," which causes reticle imperfections to be repeated hundreds of times per wafer. In the next chapter it is seen that these requirements have become even more challenging because highly complex masks are often used to provide enhanced lithographic capabilities.

Problems

7.1 For a mask fabricated on a fused silicon substrate, show that the separation between two geometries nominally separated by 100 mm changes by 5 nm for a 0.1 °C temperature change. (The coefficient of thermal expansion for fused silica equals 0.5 ppm/°C.) For the same change in temperature, show that the separation error would be 600 nm for a mask fabricated on a borosilicate glass substrate with a coefficient of thermal expansion equal to 60 ppm/°C.

7.2 Reticle nonflatness consumes part of the depth-of-focus budget. It is desired that reticle nonflatness reduce the depth-of-focus budget by no more than 10%. For a 100-nm depth-of-focus, show that the reticle flatness must be <160 nm in order to be consistent with this 10% criterion and assuming a 4× lens. What would this be for a 6× lens?

7.3 Suppose a mask is patterned by a beam writer with a 50-keV beam energy over an area of 100 × 100 mm, using a resist with 10-$\mu C/cm^2$ sensitivity. Assume the pattern density is 50%. For a standard 6-in. mask (152 × 152 × 6.35 mm), with a heat capacity of 0.71 J/gm·K and density 2.2 gm/cc, show that the mask temperature rises by 0.11 K, assuming all electron energy is converted to heat and there is no heat conducted away from the mask substrate. (The electronic charge is 1.6×10^{-19} C and 1 eV = 1.6×10^{-19} J.) Assume that the absorber is sufficiently thin that its contribution to the thermal mass is negligible. Refer to Problem 7.1 to assess the significance of such a temperature change on registration errors.

7.4 Show that the maximum angle for a ray that passes through the top element of a 1.35-NA 4× lens is 19.7 deg.

7.5 Consider a 1.35-NA 4× lens. For a perfectly flat, 1-μm-thick pellicle, show that, due to the pellicle, the optical path of the wavefront varies by 15.3 nm. Assume that the refractive index of the pellicle = 1.35, and take the refraction by the pellicle into account by using [Eq. (7.7)].

7.6 Using Eq. (7.4), show that pellicles with 6.3-mm standoff provide protection for particles as large as 366 μm when using a 0.93-NA 4× lens for imaging.

7.7 From Eq. (7.6), show that the pellicle-induced wavefront errors → 0 as either $n \rightarrow 1$ or $t \rightarrow 0$.

References

1. R. Iscoff, "Photomask and reticle materials review," *Semicond. Int.*, 82–86 (March, 1986).
2. B. Shamoun, R. Engelstad, and D. Trost, "Assessment of thermal loading-induced distortions in optical photomasks due to e-beam multi-pass patterning," *J. Vac. Sci. Technol. B* **16**(6), 3558–3562 (1998).
3. Y. Ikuta, S. Kikogawa, T. Kawahara, H. Hishiro, N. Shimodaira, and S. Yoshizawa, "New silica glass 'AQF' for 157 nm lithography," *Proc. SPIE* **4000**, 1510–1514 (2000).
4. J. H. Burnett, Z. H. Levine, and E. L. Shirley, "Intrinsic birefringence in calcium fluoride and barium fluoride," *Phys. Rev. B* **64**, 241102 (2001).
5. M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford (1975).
6. B. B. Wang, "Residual birefringence in photomasks," *J. Microlithogr. Microfab. Microsyst.* **1**(1), 43–48 (2002).
7. R. S. Priestly, D. R. Sempolinski, and C. C. Yu, "Photolithography method, photolithography mask blanks, and method of making," U.S. Patent Application No. US 2002/0187407 A1 (2002).
8. SEMI Standards P1-92, "Specification for hard surface photomask substrates," and P34-0200, "Specification for 230 mm square photomask substrates," Semiconductor Equipment and Materials International (2007).
9. S. Inoue, M. Itoh, M. Asano, K. Okumura, T. Hagiwara, and J. Moriya, "Desirable reticle flatness from focus deviation standpoint in optical lithography," *Proc. SPIE* **4691**, 530–540 (2002).
10. A. R. Mikkelson, R. L. Englestad, E. G. Lovell, T. M. Bloomstein, and M. E. Mason, "Mechanical distortions in advanced optical reticles," *Proc. SPIE* **3676**, 744–755 (1999).
11. K. Bubke, S. Teuber, I. Hoellein, H. Becker, H. Seitz, and U. Buttgereit, "Investigation of polarization effects on new mask materials," *Proc. SPIE* **5754**, 587–598 (2005).
12. R. A. Synowicki and J. N. Hilfiker, "Spectroscopic ellipsometry applications in photomask technology," *Proc. SPIE* **7122**, 712231 (2008).
13. O. Nagarekawa, and S. Matsui, "Photo-mask blank for use in lithography including a modified chromium compound," U.S. Patent No. 4,530,891 (1985).

14. T. Faure, E. Gallagher, M. Hibbs, L. Kindt, K. Racette, R. Wistrom, A. Zweber, A. Wagner, Y. Kikuchi, T. Komizu, and S. Nemoto, "Characterization of binary and attenuated phase shift mask blanks for 32 nm mask fabrication," *Proc. SPIE* **7122**, 712209 (2008).
15. Y. Kojima, M. Shirasaki, K. Chiba, T. Tanaka, Y. Inazuki, H. Yoshikawa, S. Okazaki, K. Iwase, K. Ishikawa, and K. Ozawa, "Alternating phase-shift mask and binary mask for 45-nm node and beyond: the impact on the mask error control," *Proc. SPIE* **6607**, 66070C (2007).
16. O. Nagarekawa, and S. Matsui, "Method of manufacturing a mask blank including a modified chromium compound," U.S. Patent No. 4,657,648 (1987).
17. J. Chang, A. Abdo, B. Kim, T. Bloomstein, R. Engelstad, E. Lovell, W. Beckman, and J. Mitchell, "Thermomechanical distortions of advanced optical reticles during exposure," *Proc. SPIE* **3676**, 756–767 (1999).
18. L. K. Siewert, A. R. Mikkelson, R. L. Engelstad, E. G. Lovell, M. E. Mason, and R. S. Mackay, "Influence of stress on advanced optical reticle distortions," *Pro. SPIE* **4000**, 1594–1604 (2000).
19. G. M. Henriksen, "Reticles by automatic pattern generation," *Proc. SPIE* **100**, 86–95 (1977).
20. R. A. Geshner, "The electron beam: a better way to make semiconductor masks," *Solid State Technol.*, 69–73 (June, 1979).
21. D. R. Herriott, R. J. Collier, D. S. Alles, and J. W. Stafford, "EBES: a practical electron lithography system," *IEEE Trans. Electron. Dev.* **ED-22**, 385–392 (1975).
22. M. Gesley, F. Abboud, D. Colby, F. Raymond, and S. Watson, "Electron beam column developments for submicron- and nanolithography," *Jpn. J. Appl. Phys.* **32**, 5993–6005 (1993).
23. M. Gesley, "MEBES IV thermal-field emission tandem optics for electron beam lithography," *J. Vac. Sci. Technol. B* **9**(6), 2949–2954 (1991).
24. T. H. P. Chang, M. Hatzakis, A. D. Wilson, and A. N. Broers, "Electron-beam lithography draws a finer line," *Electronics* **50**(10), 89–98 (May, 1977).
25. The MEBES RSB platform was intended to be a raster, shaped-beam system, but it did not become available as a product. See T. Newman, I. Finkelstein, H. -M. Kao, S. Krishanswami, D. Long, R. Lozes, H. Pearce-Percy, A. Sagle, J. Verner, S. Winter, M. Gesley, and F. Abboud, "Raster shaped beam pattern generation for 70 nm photomask production," *Proc. SPIE* **4889**, 169–176 (2002); also S. A. Rishton, J. K. Varner, L. H. Veneklasen, V. Boegli, A.L. Sagle, U. Hofmann, H. Kao, and W. Wang, "Raster shaped beam pattern generation," *J. Vac. Sci. Technol. B* **17** (6), 2927–2931 (1999).
26. H. C. Pfeiffer, P. M. Ryan, and E. V. Weber, "Method and apparatus for forming a variable size electron beam," U.S. Patent No. 4,243,866 (1981).

27. R. J. Collier, and M. G. R. Thomson, "Variable-spot scanning in an electron beam exposure system," U.S. Patent No. 4,393,312 (1983).
28. M. G. R. Thomson, R. J. Collier, and D. R. Herriott, "Double-aperture method of producing variably shaped writing spots for electron lithography," *J. Vac. Sci. Technol.* **15**(3), 891–895 (1978).
29. H. Elsner, H. Hahmann, G. Dahm, and H. W. P. Koops, "Multiple beam-shaping diaphragm for efficient exposure of gratings," *J. Vac. Sci. Technol. B* **11**(6), 2373–2377 (1993).
30. The electron-beam business unit of Leica is now part of Vistec Semiconductor Systems.
31. S. Johnson, D. Loughran, P. Osborne, P. Sixt, and H.-J. Doering, "Production data from a Leica ZBA31H+ shaped e-beam mask writer located at the Photronics facility, Mancester, England," *Proc. SPIE* **3676**, 518–527 (1999).
32. M. G. R. Thomson, R. J. Collier, and D. R. Herriott, "Double-aperture of producing variably shaped writing spots for electron lithography," *J. Vac. Sci. Technol.* **15**(3), 891–895 (1978).
33. J. Yashima, K. Ohtoshi, N. Nakayamada, H. Anze, T. Katsumata, T. Iijima, R. Nishimura, S. Fukutome, N. Miyamoto, S. Wake, Y. Sakai, S. Sakamoto, S. Hara, H. Higurashi, K. Hattori, K. Saito, R. Kendall, and S. Tamamushi, "Electron-beam mask writer EBM-6000 for 45 nm HP node," *Proc. SPIE* **6607**, 660703 (2007).
34. S. Nishimura, M. Ogasawara, and T. Tojo, "Improvement of beam-adjustment accuracy by beam intensity distribution measurement on a 2nd shaping aperture in electron-beam writing systems," *Proc. SPIE* **4343**, 684–691 (2001).
35. H. Kawano, H. Ito, K. Mizuno, T. Matsuzaka, K. Kawasaki, N. Saitou, H. Ohta, and Y. Sohda, "Development of an electron-beam lithography system for high accuracy masks," *J. Vac. Sci. Technol. B* **21**(2), 823–827 (2003).
36. W. D. Grobman, "An overview of pattern data preparation for vector scan electron beam lithography," *J. Vac. Sci. Technol.* **17**(5), 1156–1163 (1980).
37. M. A. Sturans, J. G. Hartley, H. C. Pfeiffer, R. S. Dhaliwal, T. R. Groves, J. W. Pavick, R. J. Quickle, C. S. Clement, G. J. Dick, W. A. Enichen, M. S. Gordon, R. A. Kendall, C. A. Kostek, D. J. Pinckney, C. F. Robinson, J. D. Rockrohr, J. M. Safran, J. J. Senesi, and E. V. Tressler, "EL5: One tool for advanced x-ray and chrome on glass mask making," *J. Vac. Sci. Technol. B* **16**(6), 3164–3167 (1998).
38. T. Kamikubo, S. Golladay, R. Kendall, V. Katsap, K. Ohtoshi, M. Ogasawara, S. Nishimura, R. Nishimura, O. Iizuka, T. Nakayama, S. Shinakawa, T. Nishiyama, and S. Tamamushi, "New electron optics for mask writer EBM-7000 to challenge Hp 32 nm generation," *Proc. SPIE* **7122**, 71220J (2008).

39. T. Kamikubo, K. Ohtoshi, N. Nakayamada, R. Nishimura, H. Sunaoshi, K. Akeno, S. Mitsui, Y. Tachikawa, H. Inoue, S. Oogi, H. Higurashi, A. Mine, T. Ishimura, S. Tsuchiya, Y. Gomi, H. Matsui, and S. Tamamushi, "Electron beam mask writer EBM-7000 for Hp 32 nm generation," *Proc. SPIE* **7488**, 74881E (2009).
40. M. Kawano, K. Mizuno, H. Yoda, Y. Sakitani, K. Andou, and N. Saitou, "Continuous writing method for high speed electron-beam direct writing system HL-800D," *J. Vac. Sci. Technol. B* **11**(6), 2323–2326 (1993).
41. This figure was provided by Takashi Kamikubo of NuFlare.
42. E. B. Bas and G. Cremosnik, "Experimental investigation of the structure of high-power-density electron beams," *Proc. First Int. Conf. Electron Ion Beam Sci. Technol.*, 108–132 (1965).
43. L. H. Lin and H. L. Beauchamp, "High speed beam deflection and blanking for electron lithography," *J. Vac. Sci. Technol.* **10**(6), 987–990 (1973).
44. M. Gesley and P. Condran, "Electron beam blanker optics," *J. Vac. Sci. Technol. B* **8**(6), 1666–1672 (1990).
45. M. Gesley, D. Colby, F. Raymond, D. McClure, and F. Abboud, "Electrodynamics of fast beam blankers," *J. Vac. Sci. Technol. B* **11**(6), 2378–2385 (1993).
46. J. L. Freyer and K. Standiford, "Design of an accurate production e-beam system," *Solid State Technol.*, 165–170 (September, 1983).
47. J. Freyer, K. Standiford, and R. Sills, "Enhanced pattern accuracy with MEBES III," *Proc. SPIE* **471**, 8–17 (1984).
48. F. Abboud, J. Poreda, and R. L. Smith, "MEBES IV: A new generation of raster-scan electron-beam lithography system," *Proc. SPIE* **1671**, 111–125 (1992).
49. P. S. Burggraaf, "E-Beam lithography: a standard tool?" *Semicond. Int.*, 41–60 (September, 1982).
50. C. Spence, C. Tabery, R. Cantrell, L. Dahl, P. Buck, and B. Wilkinson, "A comparison of DUV wafer and reticle lithography—what is the resolution limit?" *Proc. SPIE* **4889**, 177–186 (2002).
51. F. Abboud, R. Dean, J. Doering, W. Eckes, M. Gesley, U. Hofmann, T. Mulera, R. Naber, M. Pastor, W. Phillips, J. Raphael, F. Raymond III, and C. Sauer, "Multipass gray printing for new MEBES® 4500S mask lithography system," *Proc. SPIE* **3115**, 116–124 (1997).
52. D. H. Dameron, C.-C. Fu, and R. F. W. Pease, "A multiple exposure strategy for reducing butting errors in a raster-scanned electron-beam exposure system," *J. Vac. Sci. Technol. B* **6**(1), 213–215 (1988).
53. A. Murray, F. Abboud, F. Raymond, and C. N. Berglund, "Feasibility study of new graybeam writing strategies for raster scan mask generation," *J. Vac. Sci. Technol. B* **11**(6), 2390–2396 (1993).

54. J. S. Greeneich, "Impact of electron scattering on linewidth control in electron-beam lithography," *J. Vac. Sci. Technol.* **16**(6), 1749–1753 (1979).

55. T. H. P. Chang, "Proximity effect in electron beam lithography," *J. Vac. Sci. Technol.* **12**, 1271–1275 (1975).

56. D. F. Kyser and N. S. Viswanathan, "Monte Carlo simulation of spatially distributed beams in electron-beam lithography," *J. Vac. Sci. Technol.* **12**(6), 1305–1308 (1975).

57. S. A. Rishton and D. P. Kern, "Point exposure distribution measurements for proximity correction in electron beam lithography on a sub-100 nm scale," *J. Vac. Sci. Technol. B* **5**, 135–141 (1987).

58. K. Murata, H. Kawata, K. Nagami, Y. Hirai, and Y. Mano, "Studies of energy dissipation in resist films by a Monte Carlo simulation based on the Mott cross section," *J. Vac. Sci. Technol. B* **5**, 124–128 (1987).

59. G. Owen and P. Rissman, "Proximity effect correction for electron beam lithgraphy by equalization of background dose," *J. Appl. Phys.* **54**(6), 3573–3581 (1983).

60. F. Murai, H. Yoda, S. Okazaki, N. Saitou, and Y. Sakitani, "Fast proximity effect correction method using a pattern area density map," *J. Vac. Sci. Technol. B* **10**(6), 3072–3076 (1992).

61. H. Satoh, Y. Someda, N. Saitou, K. Kawaski, H. Itoh, and K. Mizuno, "New electron beam mask writing system for 0.25 micrometer lithography," *Proc. SPIE* **2723**, 102–111 (1996).

62. H. Eisenmann, T. Waas, and H. Hartmann, "PROXECCO – Proximity effect correction by convolution," *J. Vac. Sci. Technol. B* **11**(6), 2741–2745 (1993).

63. M. Ogasawara, N. Shimomura, J. Takamatsu, S. Yoshitake, K. Ooki, N. Nakayamada, H. Okabe, T. Tojo, and T. Takigawa, "Reduction of long range fogging effect in a high acceleration voltage electron beam mask writing system," *J. Vac. Sci. Technol. B* **17**(6), 2936–2939 (1999).

64. N. Kuwahara, T. Ohfuji, N. Hayashi, C. Jackson, N. Kitano, and D. H. Hwang, "PEC-fogging effect analysis using high-performance EB simulator capable of large area mask pattern simulation," *Proc. SPIE* **4889**, 767–775 (2002).

65. S.-H. Yang, Y.-H. Choi, J.-R. Park, Y.-H. Kim, S.-W. Choi, H.-S. Yoon, and J.-M. Sohn, "Fogging effect considerations in mask process at 50KeV e-beam systems," *Proc. SPIE* **4889**, 786–791 (2002).

66. P. Hudek and D. Beyer, "Exposure optimization in high-resolution e-beam lithography," *Microelectron. Eng.* **83**, 780–783 (2006).

67. M. Ogasawara, N. Shimomura, J. Takamatsu, S. Yoshitake, K. Ooki, N. Nakayamada, F. Okabe, T. Tojo, and T. Takigawa, "Reduction of long range fogging effect in high accelaration voltage electron beam writing system," *J. Vac. Sci. Technol. B* **17**(6), 2936–2939 (1999).

68. R. Innes, L. H. Veneklasen, A. L. Sagle, S. Babin, and C. Hwa, "Minimization of electron fogging in electron beam lithography," U.S. Patent No. 6,326,635 (2001).

69. N. Shimomura, M. Ogasawara, J. Takamatsu, S. Yoshitake, K. Ooki, N. Nakayamada, F. Okabe, and T. Tojo, "Reduction of fogging effect caused by scattered electrons in an electron beam system," *Proc. SPIE* **3748**, 408–414 (1999).

70. D. K. Atwood, P. Kuo, S. Vaidya, and K. D. Cummings, "Effect of charging on pattern placement accuracy in e-beam lithography," *Proc. SPIE* **1089**, 358–366 (1989).

71. J. Choi, D. S. Nam, B. G. Kim, S.-G. Woo, and H. K. Cho, "Resist charging effect in photomask: its impact on pattern placement error and critical dimension," *J. Vac. Sci. Technol. B* **26**(6), 2345–2350 (2008).

72. J. Choi, S. H. Lee, D. Nam, B. G. Kim, S.-G. Woo, and H. K. Cho, "Image placement error of photomask due to pattern loading effect: analysis and correction technique for sub-45 nm node," *Proc. SPIE* **7028**, 70281X (2008).

73. N. Nakayamada, S. Wake, T. Kamikubo, H. Sunaoshi, and S. Tamamushi, "Modeling of charging effect and its correction by EB mask writer EBM-6000," *Proc. SPIE* **7028**, 20280C (2008).

74. E. M. Engler, J. D. Kuptsis, R. G. Schad, and Y. Tomkiewicz, "Class of e-beam resists based on conducting organic charge transfer salts," U.S. Patent No. 4,338,392 (1982).

75. T. Kamikubo, R. Nishimura, K. Tsuruta, H. Hattori, J. Takamatsu, S. Yoshitake, H. Nozue, H. Sunaoshi, and S. Tamamushi, "Coping with double patterning/exposure lithography by EB mask writer EBM-6000," *Proc. SPIE* **6730**, 673031 (2007).

76. C. Enkrich, G. Antesberger, O. Loeffler, K. Roeth, F. Laske, K. Schmidt, and D. Adam, "Registration measurement capability of the VISTEC LMS IPRO4 with focus on small features," *Proc. SPIE* **7028**, 70282Y (2008).

77. See, for example, NIST Standard SRM 5000, "2D grid photomask."

78. J. Ye, M. Takac, C. M. Berglund, G. Owen, and R. F. Pease, "An exact algorithm for self-calibration of 2D precision metrology stages," *Precis. Eng.* **20**(1), 16–32 (1997).

79. M. R. Raugh, "Absolute 2D sub-micron metrology for electron beam lithography: a theory of calibration with applications," *Precis. Eng.* **7**(1), 3–13 (1985).

80. M. Takac, J. Ye, M. Raugh, R. F. Pease, C. N. Berglund, and G. Owen, "Self-calibration in two dimensions: the experiment," *Proc. SPIE* **2725**, 130–146 (1996).

81. S. Schulze, P. LaCour, and L. Grodd, "OASIS-based data preparation flows: progress report on containing data size explosion," *Proc. SPIE* **5379**, 149–157 (2004).

82. SEMI Standard SEMI P39, "OASIS®—open artwork system interchange standard," www.semi.org.

83. SEMI Standard SEMI P44, "Specification for open artwork system interchange standard (OASISTM) specific to VSB mask writers," www.semi.org.

84. E. Sahouria, S. Schulze, T. Suzuki, and J. Hirumi, "Extending OASIS for the unification of mask data representation," *Proc. SPIE* **5835**, 51–62 (2005).

85. E. Sahouria and S. Schulze, "OASIS-based unification of mask data representation," *Proc. SPIE* **5446**, 431–438 (2004).

86. H. C. Hamaker and P. D. Buck, "Performance of a new high-NA scanned-laser mask lithography system," *Proc. SPIE* **3236**, 42–54 (1997).

87. P. C. Allen, "Laser pattern generation technology below 0.25 μm," *Proc. SPIE* **3334**, 460–468 (1998).

88. B. T. Schafmann, C. H. Geller, and C. E. Franks, "Improved performance of a CORE-2100 through a joint development program," *Proc. SPIE* **1604**, 55–66 (1991).

89. P. C. Allen, M. Bohan, E. R. Christenson, H. Dai, M. Duane, H. C. Hamaker, S. C. Howells, B. Kenan, P. Pirogovsky, M. K. Sadiq, R. Teitzel, and M. White, "ALTA® 4700 system mask patterning performance improvements for X-Architecture and wafer electrical performance interchangeability with 50kV E-beam," *Proc. SPIE* **5835**, 37–50 (2005).

90. C. Spence, C. Tabery, R. Cantrell, L. Dahl, P. Buck, and B. Wilkinson, "A comparison of DUV wafer and reticle lithography—what is the resolution limit?" *Proc. SPIE* **4889**, 177–186 (2002).

91. T. Sandström, P. Askebjer, J. Sallander, R. Zerne, and A. Karawajczyk, "Pattern generation with SLM imaging," *Proc. SPIE* **4562**, 38–44 (2001).

92. T. Sandstrom and N. Eriksson, "Resolution extensions in the Sigma7000 imaging pattern generator," *Proc. SPIE* **4889**, 157–167 (2002).

93. B. M. Rathsack, C. E. Tabery, S. A. Scheer, C. G. Willson, M. Pochkowski, C. Philbin, F. Kalk, C. L. Henderson, and P. D. Buck, "Optical lithography simulation and photoresist optimization for photomask fabrication," *Proc. SPIE* **3678**, 1215–1226 (1999).

94. U. Okoroanyanwu, *Chemistry and Lithography*, SPIE Press, Bellingham, Washington (2010).

95. L. F. Thompson and M. J. Bowden, "A new family of positive electron beam resists—poly(olefin sulfones)," *J. Electrochem. Soc.* **120**, 1722–1726 (1973).

96. R. Dean and C. Sauer, "Characterization and modeling of CD performance with thin PBS," *Proc. SPIE* **3236**, 405–412 (1998).

97. C. Sauer, D. Alexander, and C. Mack, "Electron beam lithography simulation for maskmaking, part II: comparison of the lithographic performance of PBS and EBR900-M1," *Proc. SPIE* **3236**, 413–423 (1997).

98. T. Coleman, D. Alexander, and M. Lu, "Comparison of EBR-900 M1 and ZEP 7000 with plasma-etch processing for MEBES 45000S," *Proc. SPIE* **3236**, 397–404 (1998).

99. F. Letzkus, J. Butschke, C. Koepernik, C. Holfeld, J. Mathuni, L. Aschke, and F. Sobel, "SiO_2 buffer-etch processes with a TaN absorber for EUV mask fabrication," *Proc. SPIE* **5567**, 1407–1416 (2004).

100. G. Han, M. Khan, Y. Fang, and F. Cerrina, "Comprehensive model of electron energy deposition," *J. Vac. Sci. Technol. B* **20**(6), 2666–2671 (2002).

101. M. Tschinkl, C. Bürgel, U. A. Griesinger, B. Jeansannetas, and A. B. E. Vix, "Negative chemically amplified resist (nCAR) for DRAM mask fabrications," *Proc. SPIE* **4889**, 137–146 (2002).

102. W. Jeong, D. Park, E. Park, S. Seo, H. Kwon, J. Kim, S. Jung, and S. Choi, "The comparative evaluation of positive and negative chemically amplified resist characteristics for 90 nm node photomask production," *Proc. SPIE* **5130**, 157–167 (2003).

103. W.-S. Huang, R. Kwong, J. Hartley, W. Moreau, M. Angelopoulos, C. Magg, and M. Lawliss, "A CA resist with high sensitivity and sub 100 nm resolution for advanced mask making," *Proc. SPIE* **4066**, 150–159 (2000).

104. H. Sakarai, T. Abe, M. Itoh, A. Kumagae, H. Anze, and I. Higashikawa, "Resist heating effect on 50 keV EB mask writing," *Proc. SPIE* **3748**, 126–136 (1999).

105. E. S. Park, J. H. Lee, D. I. Park, W. G. Jeong, S. K. Seo, J. M. Kim, S. Choi, and S. Jeong, "Optimum PEC conditions under resist heating effect reduction for 90 nm node mask writing," *Proc. SPIE* **4889**, 792–799 (2002).

106. S. Babin, "Comparison of writing strategies subject to resist heating," *Proc. SPIE* **3546**, 389–397 (1998).

107. "Chromium etchants applications notes," Cyantek Corporation, Fremont, California (1999).

108. B. J. Curtis, H. R. Brunner, and M. Ebnoether, "Plasma processing of thin chromium films for photomasks," *J. Electrochem. Soc.* **130**(11), 2242–2249 (1983).

109. B. Wu, "Photomask plasma etching: a review," *J. Vac. Sci. Technol. B* **24**(1), 1–15 (2006).

110. S. Aoyama, S. Sakamoto, T. Koike, N. Yoshioka, N. Harashima, A. Hayashi, and T. Sasaki, "Advanced Cr dry etching process," *Proc. SPIE* **3748**, 137–146 (1999).

111. J. Butschke, M. Irmscher, H. Sailer, L. Nedelmann, M. Pritschow, H. Loeschner, and E. Platzgummer, "Mask patterning for the 22 nm node using a proton multi-beam projection pattern generator," *Proc. SPIE* **7122**, 712236 (2008).

112. B. Wu and D. Chan, "MoSi etch of phase-shift masks," *J. Microlithog. Microfab. Microsyst.* **2**, 54–60 (2003).

113. C. Constantine, D. J. Johnson, and R. J. Westerman, "Inductively coupled plasma etch of DUV MoSi photomasks: a designed study of etch chemistries and process results," *Proc. SPIE* **3546**, 88–97 (1998).

114. V. Shea, and W. J. Wojcik, "Pellicle cover for projection printing system," U.S. Patent No. 4,131,363 (1978).

115. A. Flamholz, "An analysis of pellicle parameters for step-and-repeat projection," *Proc. SPIE* **470**, 138–146 (1984).

116. P.-Y. Yang, M. Yeung, and H. Gaw, "Printability of pellicle defects in DUV 0.5 μm lithography," *Proc. SPIE* **1604**, 106–117 (1991).

117. J. Ryu, D. Lee, J. Ryu, S. Jeong, S. Kim, and C. Kim, "Simulation analysis of backside defects printability in 193 nm photolithography," *Proc. SPIE* **7122**, 712218 (2008).

118. W. N. Partlo and W. G. Oldham, "Transmission measurements of pellicles for deep-UV lithography," *IEEE Trans. Semicond. Manuf.* **4**(12), 128–133 (1991).

119. R. W. Murphy and R. Boyd, "The effect of pressure differentials on pelliclized photomasks," *Proc. SPIE* **2322**, 187–210 (1994).

120. W. Chen, J. Carroll, G. Storm, R. Ivancich, J. Maloney, O. Maurin, and E. Soudeillet, "Pellicle-induced reticle distortion: an experimental investigation," *Proc. SPIE* **3546**, 167–172 (1998).

121. J. Braat, "Analytical expressions for the wave-front aberration coefficients of a tilted plane-parallel plate," *Appl. Optic.* **36**(32), 8459–8466 (1997).

122. K. Bubke, "Pellicle-induced aberrations and apodization in hyper-NA optical lithography," *Proc. SPIE* **6283**, 628318 (2006).

123. M. K. Yang, R. H. French, and E. W. Tokarsky, "Optical properties of Teflon® AF amorphous fluoropolymers," *J. Micro/Nanolith, MEMS, MOEMS* **7**(3), 033010 (2008).

124. K. Sato, S. Nagai, N. Shinichiro, T. Sato, and M. Itoh, "Influence of pellicle on hyper-NA imaging," *Proc. SPIE* **6924**, 692451 (2008).

125. N. Yamamoto, J. Kye, and H. J. Levinson, "Polarization aberration effect on aerial image using Pauli-Zernike representation," *3rd Int. Sym. Immersion Lithog.* (2006). http://www.sematech.org/meetings/archives/litho/7799/7799abstracts.pdf.

126. B. J. Grenon, C. Peters, K. Battacharyya, and W. Volk, "Formation and detection of sub-pellicle defects by exposure to DUV system illumination," *Proc. SPIE* **3873**, 162–176 (1999).

127. B. J. Grenon, K. Bhattacharyya, W. Vol, and A. Poock, "Reticle surface contaminants and their relationship to sub-pellicle particle formation," *Proc. SPIE* **5256**, 1103–1110 (2003).

128. B. Wu, "Mask haze formation in deep ultraviolet lithography," *Appl. Phys. Lett.* **91**, 163110 (2007).

129. B. Wu, and A. Kumar, "Theoretical study of mask haze formation," *Proc. 24th European Mask Lithog. Conf.* (2008).

130. A. C. Rudack, L. B. Levit, and A. Williams, "Mask damage by electrostatic discharge: a reticle printability evaluation," *Proc. SPIE* **4691**, 1340–1347 (2002).

131. L. Levit, and A. Englisch, "The Canary Reticle—a new diagnostic for reticles and window into the physics of ESD damage to reticles," *EMC 2000: Proc. 17th European Mask Conf.* Munich-Unterhacning, Germany, 51–55 (2000).

132. T. Sebold, "Don't kill Canaries!: introducing a new test-device to assess the electrostatic risk-potential to photo-masks," *Proc. SPIE* **7122**, 71220H (2008).

133. G. Rider, "Estimation of the field-induced damage thresholds in reticles," *Semicond. Manuf.*, 80–94 (February, 2004).

134. G. C. Rider, "Electric field-induced progressive CD degradation in reticles," *Proc. SPIE* **7122**, 71220G (2008).

135. J. Bruley, G. Burr, R. E. Davis, P. Flaitz, W. D. Hinsberg, F. A. Houle, D. C. Miller, M. Pike, J. Rankin, A. Wagner, and A. Watts, "Cr migration on 193 nm photomasks," *Proc. SPIE* **7272**, 727215 (2009).

136. G. Rider, "Protection of reticles against damage from field-induced electrostatic discharge," *Semicond. Manuf.*, 132–142 (September, 2003).

137. A. Tchikoulaeva, A. Holfeld, M. Arend, and E. Foca, "ACLV degradation: root cause analysis and effective monitoring strategy," *Proc. SPIE* **7028**, 702816 (2008).

138. T. Sebald and G. Rider, "High sensitivity electric field monitoring system for control of field-induced CD degradation in reticles (EFM)," *Proc. SPIE* **7379**, 73791P (2009).

139. K. Smith, (2002), "Shedding light on optical comparators—How much better can this type of system get?" *Quality Digest: QCI International*. http://www.qualitydigest.com/may02/html/optcomp.html, retrieved 26 September 2009.

140. J. H. Bruning, M. Feldman, T. S. Kinsel, E. R. Sittig, and R. L. Townsend, "An automated mask inspection system—AMIS," *IEEE Trans. Electron. Dev.* **ED-22**(7), 487–495 (1975).

141. W. Broadbent, I. Yokoyama, P. Yu, K. Seki, R. Nomura, H. Schmalfuss, J. Heumann, and J. Sier, "Field results from a new die-to-database reticle inspection platform," *Proc. SPIE* **6607**, 660714 (2007).

142. A. Dayal, B. Mu, P. Lim, A. Goonesekera, and B. Broadbent, "Results from the KLA-Tencor TeraScanXR reticle inspection tool," *Proc. SPIE* **7122**, 71223G (2008).

143. P. Fiekowsky and D. Selassie, "Defect printability measurement on the KLA-351: correlation to defect sizing using the AVI metrology system," *Proc. SPIE* **3873**, 754–759 (1999).

144. H. Shigemura, T. Amano, Y. Nishiyama, O. Suga, Y. Arisawa, H. Hashimoto, K. Takahara, and K. Usuda, "Evaluation of EUVL mask defect inspection using 199-nm inspection tool with super-resolution method," *Proc. SPIE* **7379**, 73792K1 (2009).

145. A. Sagiv, Y. Shirman, and S. Mangan, "A novel approach to mask defect inspection," *Proc. SPIE* **7122**, 71223E (2008).

146. J. Tison and J. O'Connor, "Status and trends in laser mask repair technologies," *Proc. SPIE* **1809**, 182–186 (1992).

147. R. Haight, D. Hayden, P. Longo, T. Neary, and A. Wagner, "MARS: femtosecond laser mask advanced repair system in manufacturing," *J. Vac. Sci. Technol. B* **17**(6), 3137–3143 (1999).

148. A. Wagner, R. Haight, and P. Longo, "MARS2: an advanced femtosecond laser mask repair tool," *Proc. SPIE* **4889**, 457–468 (2002).

149. J. Melngailis, "Critical review: focused ion beam technology and applications," *J. Vac. Sci. Technol. B* **5**(2), 469–495 (1987).

150. D. Ferranti, A. Graupera, J. Marshman, D. Stewart, and S. Szelag, "Material removal strategies and results for 193 nm lithography using FIB mask repair," *Proc. SPIE* **5256**, 546–555 (2003).

151. Y. Itou, Y. Tanaka, O. Suga, Y. Sugiyama, R. Hagiwara, H. Takahashi, O. Takaoka, T. Kozakai, O. Matsuda, K. Suzuki, M. Okabe, S. Kikuchi, A. Uemoto, A. Yasaka, T. Adachi, and N. Nishida, "Advanced photomask repair technology for 65 nm lithography," *Proc. SPIE* **5992**, 59924Y (2005).

152. M. R. Laurance, "Subtractive defect repair via nanomachining," *Proc. SPIE* **4186**, 670–673 (2001).

153. B. T. LoBianco, R. White, and T. Nawrocki, "Use of nanomachining for 100 nm nanometer mask repair," *Proc. SPIE* **5148**, 249–261 (2003).

154. D. Brinkley, R. Bozak, B. Chiu, C. Ly, V. Tolani, and R. White, "Investigation of nanomachining as a technique for geometry reconstruction," *Proc. SPIE* **4889**, 232–240 (2002).

155. T. Robinson, R. White, R. Bozak, K. Roessler, B. Arruza, D. Hogle, M. Archuletta, and D. Lee, "New tools to enable photomask repair to the 32 nm node," *Proc. SPIE* **7488**, 74880F (2009).

156. T. Liang, E. Frendberg, D. Bald, M. Penn, and A. Stivers, "E-beam, mask repair: fundamental capability and applications," *Proc. SPIE* **5567**, 456–466 (2004).

157. R. L. Kubena, J. W. Ward, F. P. Stratton, R. J. Joyce, and G. M. Atkinson, "A low magnification focused ion beam system with 8 nm spot size," *J. Vac. Sci. Technol. B* **9**(6), 3079–3083 (1991).

158. J. Orloff, M. Utlaut, and L. Swanson, *High Resolution Focused Ion Beams: FIB and its Applications*, Klewer Academic/Plenum Publishers, New York (2003).

159. P. D. Prewett, B. Martin, A. W. Eastwood, and J. G. Watson, "Effects of focused ion beam reticle repair on optical lithography at i-line and deep ultraviolet wavelengths," *J. Vac. Sci. Technol. B* **11**(6), 2427–2431 (1993).

160. K. Hiruta, S. Kubo, H. Morimoto, A. Yasak, R. Hagiwara, T. Adachi, Y. Morikawa, K. Iwase, and N. Hayashi, "Advanced FIB mask repair technology for ArF Lithography," *Proc. SPIE* **4066**, 520–523 (2000).

161. C. Ehrlich, K. Edinger, V. Boegli, and P. Kuschnerus, "Application data of the electron beam based photomask repair tool MeRiT MG," *Proc. SPIE* **5835**, 145–154 (2004).

162. H. Marchman, D. Taylor, S. Hadisutjipto, S. Mackay, R. Cottle, J. Maltabes, and J. Brown, "Mask repair technique assessment and development for the 45 nm lithographic node," *Proc. SPIE* **6382**, 628311 (2006).

163. H. J. Levinson, "Impact of reticle imperfections on integrated circuit processing," *Proc. Third Annual Sym. Bay Area Chrome Users Soc. (BACUS)*, September 14 and 15,1983, Sunnyvale, California, as described in *Semicond. Int.*, 22–23 (December, 1983).

164. J. N. Wiley, "Effect of stepper resolution on the printability of submicron 5×reticle defects," *Proc. SPIE* **1088**, 58–73 (1989).

165. Y. M. Ham, I. B. Hur, Y. S. Kim, D. J. Ahn, Z. Cha, S. S. Choi, H. J. Kimm, and Y. J. Jeon, "Dependence of defects in optical lithography," *Jpn. J. Appl. Phys.* **31**(99), 4137–4142 (1992).

166. F. Gans, M. Jess, and S. Kohlpoth, "Printability and repair techniques for DUV photomasks," *Proc. SPIE* **3236**, 136–141 (1997).

167. K. A. Phan, C. A. Spence, J. Riddick, J. X. Chen, M. J. Lamantia, and H. A. Vialla, "Correlation of reticle defect detectability and repairs to ArF wafer printability for 0.13 μm design rule with binary OPC/SB mask," *Proc. SPIE* **4186**, 183–197 (2001).

168. L. Karklin, "Comprehensive simulation study of the photomask defects printability," *Proc. SPIE* **2621**, 490–494 (1995).

169. P. Yan, J. Langston, J. Neff, and R. Chatterjee, "Mask defect printability and wafer process critical dimension control at 0.25 m design rules," *Jpn. J. Appl. Phys.* **34**(1), 660–6610 (1995).

170. R. A. Budd, D. B. Dove, J. L. Staples, R. M. Martino, R. A. Ferguson, and J. T. Weed, "Development and application of a new tool for lithographic mask evaluation, the stepper equivalent aerial image measurement system, AIMS," *IBM J. Res. Dev.* **41**, 119–130 (1997).

171. A. Zibold, U. Strössner, N. Rosenkranz, A. Ridley, R. Richter, W. Harnisch, and A. Williams, "First results for hyper NA scanner emulation from AIMS^{TM} 45-193i," *Proc. SPIE* **6283**, 628312 (2006).

172. A. M. Zibold, T. Scherübl, A. Menck, R. Brunner, and J. Greif, "Aerial image measurement technique for today's and future 193 nm lithography mask requirements," *Proc. SPIE* **5504**, 12–18 (2004).

173. A. Sagiv, Y. Shirman, and S. Mangan, "A novel approach to mask inspection," *Proc. SPIE* **7122**, 712232 (2008).

Chapter 8
Confronting the Diffraction Limit

As k_1 factors [Eq. (2.8)] fall below 0.8, a number of effects become observable that are not seen when k_1 is larger. Processes with small values of k_1 began to appear in the mid-1990s, and are quite common today, as seen from Table 8.1. Some of the effects seen in patterns generated with low-k_1 processes are discussed in this chapter. Several techniques that are also described—such as off-axis illumination and phase-shifting masks—have been developed to address the shortcomings of optical imaging as feature sizes become smaller than the wavelength of light. Methods to improve image contrast that involve modification of the mask or illumination are referred to collectively as *resolution-enhancement techniques* (RET) and are discussed in this chapter.

8.1 Off-Axis Illumination

As discussed earlier, coherent light that illuminates a grating is diffracted in very specific directions [Eq. (2.1)]. For normally incident light, sufficiently small dimensions result in situations where all beams except the zeroth order are diffracted outside the entrance pupil of the imaging optics (Fig. 8.1). In this case, no pattern is formed, because a single beam is a plane wave, containing no spatial information, as was explained in Chapter 2. For normally incident (on-axis) illumination, the grating is not imaged when the pitch is too small, because only a

Table 8.1 The evolution of k_1 in manufacturing.

Year	Wavelength (nm)	NA	Half pitch (μm)	k_1
1980	436	0.28	1.50	0.96
1983	436	0.35	1.20	0.96
1986	436	0.45	1.00	1.00
1989	365	0.45	0.70	0.86
1992	365	0.54	0.50	0.74
1995	365	0.60	0.35	0.57
1997	248	0.50	0.25	0.50
1999	248	0.63	0.18	0.46
2001	248	0.80	0.13	0.42
2004	193	0.75	0.10	0.39
2007	193	0.93	0.07	0.34

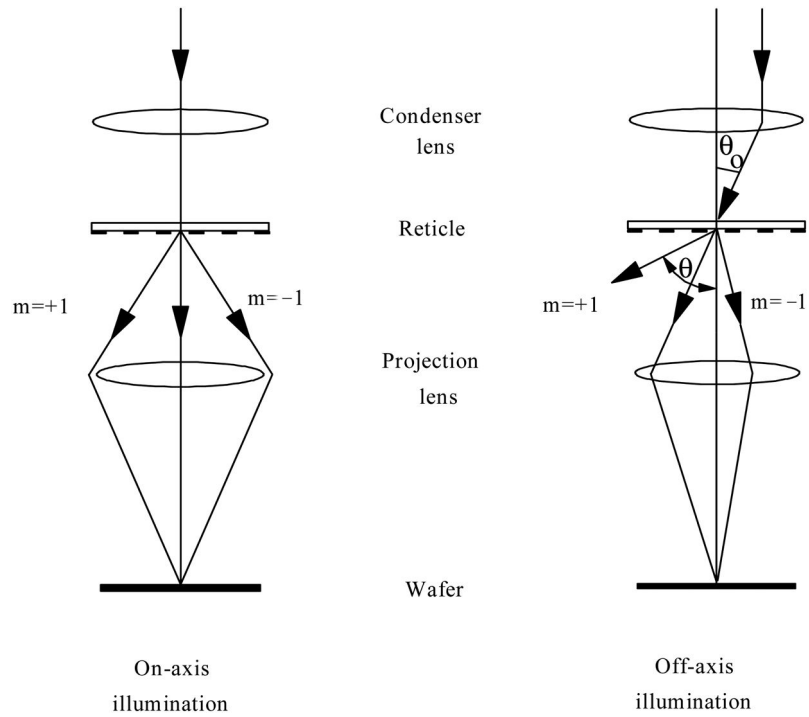


Figure 8.1 On-axis and off-axis illumination. With off-axis illumination, diffraction orders >0 are propagated through the projection lens. The features on the mask are perpendicular to the plane of the illumination.

single beam, the zeroth-order beam, is transmitted through the lens. This illustrates the limitation to resolution imposed by diffraction.

On the other hand, consider the situation in which two plane waves (of unit amplitude) intersect, shown in Fig. 8.2. In this case,

$$\text{Intensity} = \left| e^{i(k_x x + k_z z)} + e^{i(-k_x x + k_z z)} \right|^2 \tag{8.1}$$

$$= 4\cos^2(k_x x)\,. \tag{8.2}$$

With two plane waves intersecting, a pattern with structure can be imaged.

Consider the situation in which the illumination is not normally incident (Fig. 8.1). For such off-axis illumination it is possible for the zeroth-order light and one first-order beam to enter the entrance pupil of the imaging optics. In this situation, there is a pattern imaged onto the wafer. From this simple analysis, one might expect to enhance image contrast by eliminating the illumination with small angles of incidence, since those light rays contribute only to the background light intensity without providing spatial modulation, for very small features.

From Fig. 8.1, one might also expect that off-axis illumination should improve the depth-of-focus, because the angular spread of the light rays in the off-axis

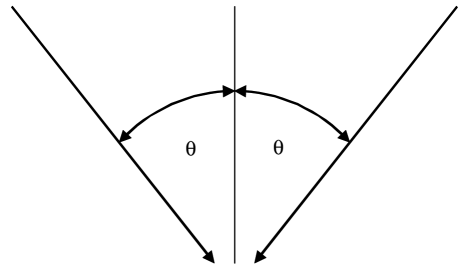


Figure 8.2 Two plane waves intersecting.

situation is less than that for conventional illumination. This is understood by considering the situation depicted in Fig. 8.3. In the limit of geometrical optics, a lens focuses light to a point in the plane of best focus. In another plane, the point broadens to a circle. The diameter of this circle increases with distance z from the plane of best focus. Circles of diameter up to a certain maximum may be tolerated before one says that the image is too blurry, and this establishes the depth-of-focus. Lenses with smaller numerical apertures require a larger distance z' to produce circles of a specified diameter. Hence, images produced with smaller numerical apertures have larger depths-of-focus. As can be seen in Fig. 8.1, a similar effect occurs with off-axis illumination. The angle formed by the intersecting zeroth- and first-order beams is smaller than is obtained in the on-axis case, even if the ±first-order diffraction beams are captured by the lens, enabling pattern image formation. Thus, off-axis illumination has the potential to increase the depth-of-focus, as well as improve resolution.

The depth-of-focus enhancement from off-axis illumination can be analyzed further. Consider the off-axis geometry shown in Fig. 8.4, where only the zeroth-order ray and one first-order diffracted ray enter the lens. These two plane waves can be summed to generate a field amplitude that can be squared to produce the light intensity at the wafer plane. In application, illumination symmetric at $\pm\theta_0$ angles of incidence are used. Summing the light intensities from the two incident rays of illumination, the total light intensity at the wafer plane is given by[1]

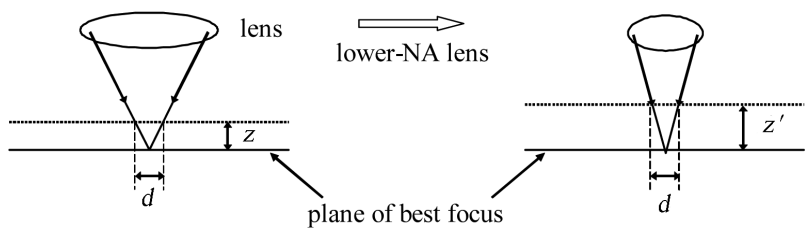


Figure 8.3 Illustration of the depth-of-focus in the context of geometrical optics.

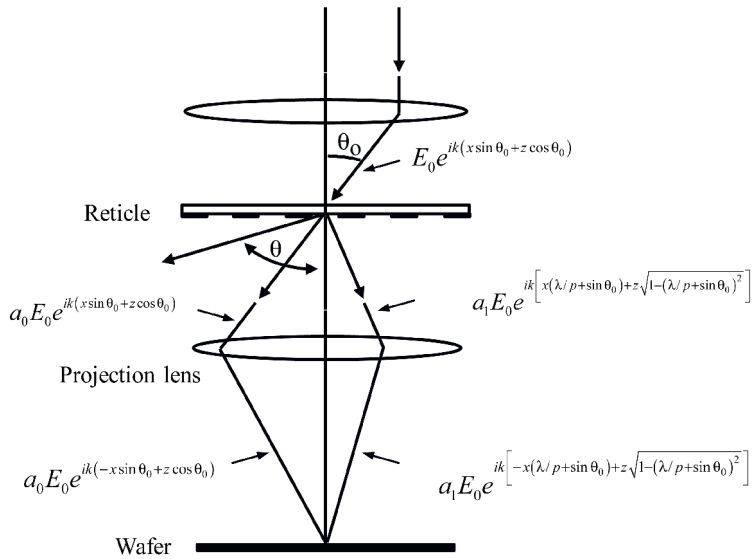


Figure 8.4 Light rays diffracted by a diffraction grating of pitch p. In the figure $k = 2\pi/\lambda$. This example is appropriate for 1:1 optics. Angles and dimensions will be scaled by the lens reduction for reduction optics.

$$I(x,z) = 2a_0^2E_0^2 + 2a_1^2E_0^2 + 4a_1a_0E_0^2\cos\left[\left(\sqrt{1-(\sin\theta_0-\lambda/p)^2}-\cos\theta_0\right)\frac{2\pi z}{\lambda}\right]\cos\left(\frac{2\pi x}{p}\right), \quad (8.3)$$

where a_0 is the amplitude fraction of the light's electrical field in the zeroth-order beam, a_1 is the fraction of the electric field in the first-order beam, E_0 is the amplitude of the incident electric field, p is the pitch of the grating, x is the lateral distance along the wafer, and z is the amount of defocus. This equation is independent of z, i.e., independent of focus, when

$$\cos\theta_0 = \sqrt{1-(\sin\theta_0-\lambda/p)^2}, \quad (8.4)$$

which occurs when

$$\sin\theta_0 = \frac{\lambda}{2p}. \quad (8.5)$$

Equation (8.5) shows an important characteristic of image enhancement when using off-axis illumination: the optimum parameters, such as the angle of incidence for the illumination, are pitch dependent.[2] This indicates that process-window enhancement can be obtained for a particular pitch when using a given angle for the off-axis illumination, but that the enhancement may be less or nonexistent for other pitches.

In practice, infinite depth-of-focus is not obtained, because the illumination consists of more than two rays of light coming from angles $\pm\theta_0$. Instead, the illumination is comprised of light incident over ranges of angles around the optimum. This occurs for several reasons. First, maintaining a reasonable level of light intensity requires that illumination be collected over a finite range of angles. Additionally, a finite range of angles for the illumination results in light that is spread out as it propagates through the lens. Spreading the light out within the projection optics has a number of benefits. One advantage is reduced sensitivity to aberrations in the projection optics. This occurs because phase and amplitude errors of one particular ray of light are averaged with the errors from other rays. The resulting image quality will not be determined by worst-case errors.[3] In addition, with finite collection angles, the light is less concentrated as it propagates through the projection optics, reducing the potential for glass damage (see Section 5.4). Finally, light from a point source is very coherent, which leads to interference issues, as discussed in Chapter 5. For all of these reasons, the illumination is incident over a range of angles. Since this results in less-than-optimal performance, there is pressure on the exposure tool manufacturers to reduce lens aberrations and enhance illuminators to provide high light intensity even when the illumination is directed over a narrow range of angles.

Illumination that contains light incident on the mask in a cone from all angles from zero to a maximum defined by the partial coherence σ is called *conventional illumination*, while illumination from which the normally and near-normally incident light rays are excluded is called *off-axis illumination*. The type of illumination that has been described thus far, where the light is incident on the mask from just two opposing directions, is a particular type of off-axis illumination called *dipole illumination*.

Illumination is typically described by partial coherence σ rather than an angle of illumination. From the defining equation for σ [Eq. (A.11) in Appendix A] and Eq. (8.5), one can see that the optimum illumination angle for dipole illumination is given by

$$\sigma = \left(\frac{1}{NA}\right)\left(\frac{\lambda}{2p}\right). \tag{8.6}$$

Because light is collected over a finite range of angles, off-axis illumination is usually described by a pair of σ values, one representing the outer angle (σ_{outer}) and the other value signifying the inner angle (σ_{inner}). The optimum value given by Eq. (8.6) is usually intermediate between σ_{outer} and σ_{inner}.

The benefits of off-axis illumination for enhancing image contrast have been long known among optical microscopists. Motivated by experience in microscopy, the concept of oblique or off-axis illumination was introduced to microlithography in 1986[4] and it has since been explored extensively.[2–9] It is now commonly used for fabricating integrated circuits.

Features that are parallel to the plane of the illumination (Fig. 8.1) have their imaging degraded with dipole illumination, relative to conventional illumination,

because conventional illumination contains at least some off-axis rays for all geometries, regardless of orientation. For more general geometries, illumination must be configured accordingly. The construction of one such design is illustrated in Fig. 8.5. Because it is difficult to visualize illumination from many angles, it is convenient to depict the illumination in the form of its image in the pupil plane of the projection lens. Figure 8.5 is such a depiction. In such pupil plane representations, light in the center represents on-axis illumination, while light at increasing distances from the center corresponds to illumination at larger angles of incidence on the mask. As can be seen in that figure, there are four regions where good imaging for horizontal features overlaps the good imaging for vertical features. Illumination where light rays propagate only from those overlapping regions is referred to as *quadrupole illumination*. From Fig. 8.5 and Eq. (8.6), it can be seen that the optimum center point for the overlap regions occurs at

$$\sigma = \left(\frac{\sqrt{2}}{2NA}\right)\left(\frac{\lambda}{p}\right). \tag{8.7}$$

Each stepper company has its own name for its particular implementation of quadrupole illumination. Nikon calls their system super-high-resolution illumination control (SHRINC),[8] Canon has named theirs Canon quadrupole effect for stepper technology (CQuest),[10] and ASML's system is called Quasar.[11] Typically, the regions of quadrupole illumination are not square in shape, but are often circular or annular sections.

A common configuration that applies to objects of arbitrary orientation is *annular illumination*, where the light comes down from an annulus, centered on the optical axis, and forming the envelope of a cone. Optimum annular illumination consists of an annulus that approximates all regions of Fig. 8.5, which are good for either vertical or horizontal patterns. While annular illumination is not beneficial for all patterns, nor is it the optimum type of off-axis illumination for other patterns, such as gratings aligned in a single direction, it does provide process-

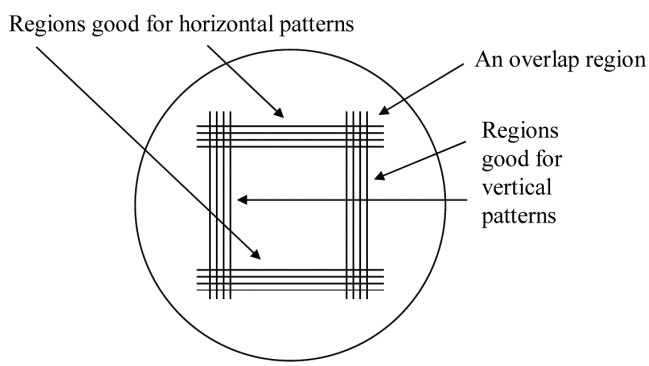


Figure 8.5 Construction of quadrupole illumination. This is the image of illumination in the pupil plane of the projection optics.

window enhancement while still imaging many patterns adequately. Hence, annular illumination has been used extensively.

Annular illumination can be generated by placing an aperture with an annular opening in the path of the illuminator, such as that shown in Fig. 8.6(a). However, this reduces the overall intensity of the light as well, since it blocks much of the available light, reducing the overall exposure-tool productivity. More sophisticated illuminators transmit most of the available light into the annulus of illumination.[12,13] As described in Chapter 5, annular illumination can be generated using axicons in the illuminator optics. Illuminators which provide high off-axis intensity usually can generate light efficiently in portions of an annulus, so quadrupole and dipole illumination often appear closer to what is shown in Fig. 8.7, rather than what is shown in Fig. 8.6.

A full definition of dipole or quadrupole illumination requires specification of the inner sigma, the outer sigma, and the subtended angle of the apertures. This is illustrated in Fig. 8.8. For apertures that are circular, a more restrictive definition is sufficient [Fig. 8.8(b)]. Annular illumination is often described by the ratio $\sigma_{inner}/\sigma_{outer}$, where common ratios are 2:3 and 1:2. Smaller values provide stronger enhancement effects for particular pitches.

The situation described above and leading to Eq. (8.3) is a situation involving *two-beam imaging*. With suitably chosen off-axis illumination, only the zeroth-order and one of the first-order beams enters the entrance pupil of the lens. This leads to Eq. (8.3), in which all dependence of the image in the z direction can be eliminated, i.e., the imaging can be independent of focus. No such conditions exist (where all of the beams are in phase, independent of z) with *three-beam*

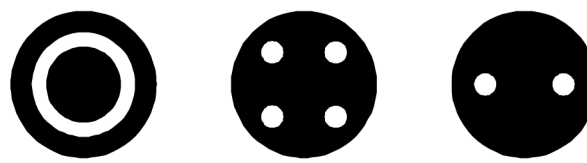


Figure 8.6 Illumination pupil plane images for (a) annular illumination, (b) quadrupole illumination, and (c) dipole illumination.

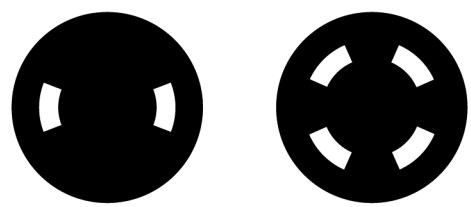


Figure 8.7 (a) Dipole and (b) quadrupole illumination generated from an annular source.

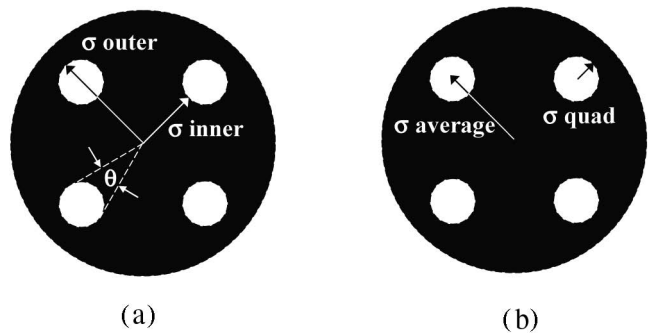


Figure 8.8 Parameters for specifying quadrupole illumination. The parameters shown in (a) can be used to describe the illumination generated from an annular source (Fig. 8.7), while the parameterization shown in (b) is applicable to illumination with circular sections.

imaging, such as occurs with on-axis illumination (Fig. 8.1). This line of reasoning leads to a general conclusion: *resolution-enhancement techniques* (RET) that can produce two-beam imaging provide better performance than methods that generate three-beam imaging. Dipole illumination improves the process window because it leads to "two-beam" imaging, rather than "three-beam imaging." This is a general theme for all techniques to improve resolution and process window: try to create a situation in which two-beam imaging occurs.

Conventional illumination consists of a cone of light, specified by a partial coherence parameter σ. Annular illumination can be viewed as a cone of light with an inner cone removed. The optics used to create any particular illumination usually will not result in a sharp cutoff for the light around the edges of the cone. Instead, the illumination will fall from some large value to zero or near zero over a range of angles. This results in ambiguity over the exact illumination profile for specific values of σ that are given. Consider conventional illumination. Some stepper companies will define σ by the angle at which 90% of the integrated light energy is contained, while others may define it as the angle in which 100% is contained. Neither is correct nor incorrect, but users must be cautious when trying to compare results from two different exposure tools, or between experimental and simulated results. Subtle differences in illumination profiles can lead to very different results.

Results for off-axis illumination from more detailed analyses and measurements found the following:

(1) Enhancement is seen for features smaller than the coherent cutoff limit (k_1 = 0.5). Above this limit there is no difference between conventional and off-axis illumination.
(2) The depth-of-focus improvement for grouped features is much greater than for isolated features.
(3) Grouped/isolated-feature print bias, discussed in more detail in the next section, is affected by off-axis illumination.

Illumination optimized for one pitch may work poorly for other pitches. For patterns with more than one pitch, this is a problem. Illumination sources that include some on-axis illumination, though at lower intensity than for the off-axis light,[14,15] have shown good results for multiple structures. In such cases, the off-axis illumination is used to ensure good patterning of the most critical features, while sufficient additional illumination is added to enable adequate printing of those patterns that are less difficult to print. Dipole illumination produces the biggest process-window enhancement for dense lines oriented in a single direction, while contacts typically image best with illumination that takes into account their intrinsic two-dimensional character. These examples suggest that many different types of illumination might be useful in lithography. Indeed, one of the papers in which off-axis illumination for lithography was first proposed concluded with the speculation that illumination might be custom engineered for individual layers.[5] Dipole, quadrupole, and annular illumination might be considered only as examples of a wide variety of illuminator configurations that could be used to enhance process windows. For example, researchers at Canon have introduced a method, IDEAL, which involves modified illumination and assist features to improve the printing of contacts.[16] More recently, this idea has been generalized to a method that can be used for calculating optimal illumination patterns for a wide range of patterns.[17]

As suggested by Figs. 8.6 and 8.7, illumination can become very complicated. In Chapter 5 it was shown how an axicon can be used to produce annular illumination without losing very much light. More complex illumination can also be produced efficiently using diffractive optical elements (DOE). Locally, a diffractive optical element is a blazed grating, illustrated in Fig. 8.9. Efficiency is achieved by setting the diffraction angle β to the angle at which the light is refracted at the glass-air interface, which occurs when

$$n \sin(\alpha) = \sin(\alpha + \beta), \tag{8.8}$$

where n is the index of refraction of the glass. On a larger scale, diffractive optical elements are more intricate than simple gratings, as needed to produce complex illumination shapes, and the techniques of computer-generated holography are used to design the DOEs.[18,19]

Another method for producing complex illumination is through the use of micromirror arrays, similar in concept to those used in data projectors and flat-panel televisions.[20] While more complex and expensive than a DOE, this approach has the advantage of programmability, which is advantageous for research and development, as well as for manufacturing environments where large numbers of different illumination shapes are required. In the latter case, the higher-capital cost of the programmable illuminator is offset by the expense avoided by not purchasing a large quantity of DOEs.

For patterning a simple diffraction grating, dipole illumination is optimal. For more complex patterns, other types of illumination, such as annular or quadrupole, might be better. It is possible to determine what is the optimal illumination for

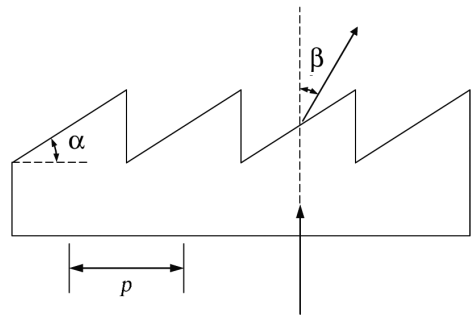


Figure 8.9 A blazed diffraction grating. Beveling leads to high efficiency in light transmitted in the indicated direction. The profille on an actual diffractive optical element will not have the perfect saw-tooth shape shown in this figure, with less than 100% efficiency as a consequence.

a given pattern. For every illumination source point, the quality of the patterning can be assessed. Using criteria set by a lithographer, source points can be retained as part of the illumination or not. While conceptually simple, this is difficult in practice. The set of criteria used to assess the illumination can include linewidth control, image-log slope, line-end shortening (discussed in the next section), and mask-error factor (the topic of Section 8.3). All of these are important, and the engineer needs to weight them as appropriate for the application.

Off-axis illumination provides a means of enhancing optical lithography. Although specific illumination settings improve imaging performance only for a narrow range of pitches, this is still an advantage for many applications, such as memories, where many structures are highly periodic. When there are multiple pitches to be patterned, the problem is more difficult. Some additional complications resulting from patterns with multiple pitches are discussed in the next section.

8.2 Optical Proximity Effects

Light-intensity profiles for features representing two different generations of lithography are shown in Fig. 8.10. The light-intensity profiles of 1.0-μm lines, imaged with 0.35-NA g-line optics ($k_1 = 0.8$), are compared with the images of 0.35-μm lines, patterned with 0.54-NA i-line lenses ($k_1 = 0.5$). For each line size, the light intensities are shown for lines in two different configurations. In the first, the lines are isolated, with no other patterns near the line whose light-intensity profile is plotted; the other configuration is where the line is part of a grating pattern consisting of equal lines and spaces. For the 1.0-μm lines, the light-intensity profiles are nearly the same, whether the line is isolated or part of a grating. Lines that are close to one another are referred to as *grouped* or *dense* lines. The situation is quite different for the 0.35-μm lines, where there is a readily observable difference between the light-intensity profiles of the isolated and grouped lines. Assuming that the linewidth corresponds to the light intensity at the 0.3 level, there is a small difference in size between the grouped and isolated 1.0-μm lines

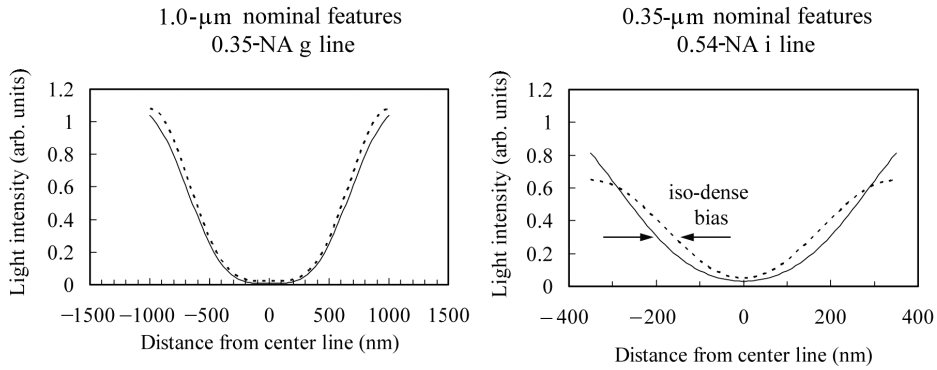


Figure 8.10 Light-intensity profiles for 1.0-μm and 0.35-μm features. Both line sizes in the image are calculated using Prolith for isolated lines and lines that are part of a grating consisting of equal lines and spaces. The solid lines are the light-intensity profiles for isolated lines, while the dashed lines represent lines that are part of gratings that consist of equal lines and spaces.

(0.042 μm, or 4.2%), while the 0.35-μm lines differ in size by 76 nm, or 22%. This difference in line size between isolated and grouped lines is referred to as the *iso-dense bias*. It is an effect of the proximity of features to other geometries and becomes more significant as k_1 becomes smaller. As seen in this example, proximity effects were small for $k_1 = 0.8$ and were not significant for older processes, which had such values for k_1 (see Table 8.1). On the other hand, modern processes are characterized by much smaller values of k_1, and proximity effects are now significant. This proximity effect represents an additional manifestation of diffraction within the context of optical lithography.

Proximity effects can be quite complicated. Just discussed was the iso-dense bias at two extreme pitches, with line-space ratios of 1:1 and 1:∞. Because of the proximity effect, linewidths vary over the intermediate pitches as well, in a nonmonotonic fashion (Fig. 8.11). There are a number of parameters that affect the magnitude of proximity effects. Lens parameters such as wavelength and numerical aperture are important, as are illuminator settings including partial coherence and the use of off-axis illumination. Resist processes can also influence the magnitude of proximity effects. Because models are inexact for resist processes, particularly the post-exposure bake step, for accuracy the magnitudes of proximity effects need to be determined empirically and re-evaluated whenever there are process changes.

The simplest way to compensate for linewidth variations caused by proximity is to have certain geometries on the reticle resized so that all features print at the desired dimensions on the wafers. Lines that print larger because of the proximity effect are made slightly smaller on the reticle. These adjustments of features on the masks to correct for proximity effects are known as *optical proximity corrections*, and are often referred to by their acronym, OPC. While the objective is conceptually simple, the implementation of optical proximity corrections is not trivial. In a typical chip design, features occur in a large number of configurations,

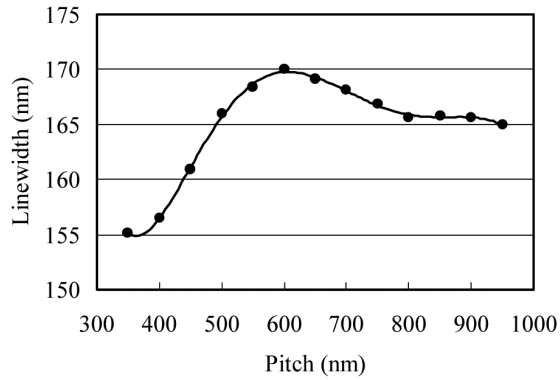


Figure 8.11 Simulated linewidths for 150-nm lines imaged at best focus on varying pitches. The parameters of the simulation (using the commercially available simulator, Solid-C) where $NA = 0.75$, $\lambda = 193$ nm, and $\sigma = 0.6$. An aerial-image threshold of 0.3 was used to determine the linewidth.

not just simple grating structures. Different types of situations that occur are depicted in Fig. 8.12.[21]

Masks generally include geometries of various dimensions. Over a wide range of sizes, the printed features equal approximately the size of the corresponding feature on the mask divided by the lens-reduction factor. However, as feature sizes approach the resolution limit of the lens, nonlinearity is introduced.[22] This is shown in Fig. 8.13. Grating patterns consisting of equal lines and spaces were printed on a Micrascan II (4× lens reduction, 0.5 NA, $\lambda = 250$ nm) over a range of sizes. The resulting spacewidths were measured and plotted. For reticle dimensions ≥1300 nm = 4 × 325 nm, the wafer dimensions transferred 4:1 to the wafer, while for smaller features, the wafer dimensions varied more rapidly as a function of reticle feature size. A feature size of 325 nm corresponds to $k_1 = 0.65$.

When critical geometries are involved, corrections on the mask are necessary to compensate for this nonlinearity if features will be printed over a wide range of sizes. Print bias is a function of feature size, and this can be corrected by resizing

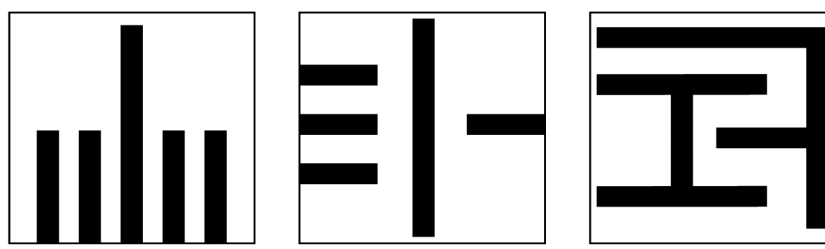


Figure 8.12 Different configurations in which proximity corrections are needed. Configuration (a) is a one-dimensional situation in which linewidth bias is needed. Line-end shortening needs to be addressed in configuration (b), while 2D corner rounding is problematic in configuration (c).

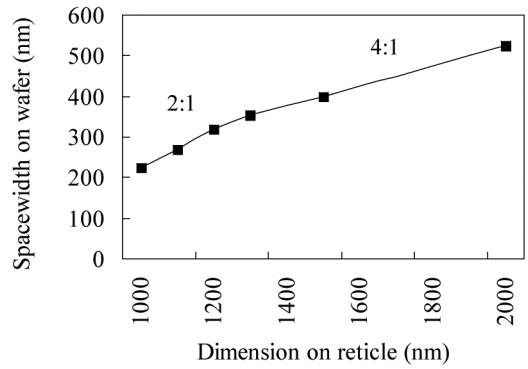


Figure 8.13 Transfer of features from a reticle to the wafer. The features are reduced in size by 4× for large feature size, but the reduction changes for features near the resolution limit of the lens.

features on the mask. Although nonlinearity is not a consequence of proximity, size adjustments on the mask to correct for nonlinear printing are often included under the umbrella term of OPC. It has become common practice to refer to all adjustments made to mask features compensating for the effects of low-k_1 patterning as optical proximity corrections, even though many of these do not directly involve proximity among geometries.

Optical systems act as low-pass filters, removing high-spatial-frequency components of the pattern; in lithography, it is often necessary to compensate for this. The imaging of a diffraction grating under coherent illumination, discussed in Chapter 2, provided an example of spatial filtering by optics. The lens collected diffracted beams representing the low-spatial-frequency parts of the object, while the aperture of the lens cut off beams from high spatial frequencies. Thus, the lens, when forming the image, filters out the high-spatial-frequency components of the object. A similar situation for incoherent imaging was also discussed in Chapter 2, where the intensity of each spatial frequency was transferred from the object to the image but was multiplied by the MTF. Because the MTF is always ≤ 1, decreases monotonically with increasing spatial frequency, and is zero above a cutoff frequency, imaging with incoherent light also results in spatial filtering. As might be expected, spatial filtering occurs when imaging with partially coherent light, although the analysis is more complex.

Another set of corrections that are included under the rubric of OPC are the adjustments made to the mask in order to compensate for the observable spatial filterings when k_1 becomes small. The most familiar of these involves the imaging of contacts, where square patterns on the mask print as circles. Corners, which represent high-spatial-frequency parts of geometries, are effectively filtered by the optics. Small geometries called serifs have long been added to features on contact masks (Fig. 8.14).[23] Serifs, properly sized, are subresolution, and do not print with fidelity. Rather, they modify the image of the contact, producing a pattern that has more discernible corners than would be printed without them. This technique does not have universal applicability. For example, it does not apply in situations

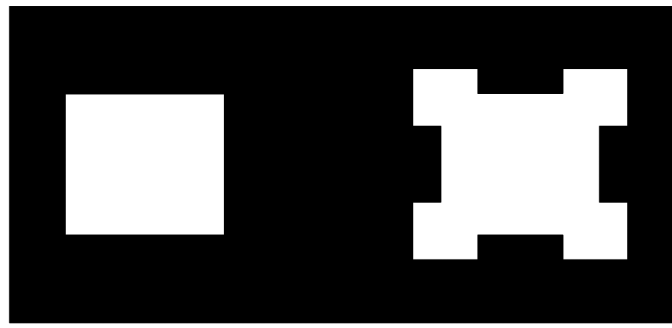


Figure 8.14 Contacts on a reticle: (a) without serifs, and (b) with serifs.

where the spacing between contacts is already a minimum resolvable distance, and additional layout effort is required to create the reticle where serifs have been added to the contacts.

Serifs may be used for structures other than contacts in order to reduce corner rounding. L-shaped features, such as shown in Fig. 8.15, may have serifs added (dark and clear). In this case, a clear serif is often referred to as a *notch*. An example will be shown later that illustrates the need to correct corner rounding in order to maintain gate linewidth control over active regions. As with many techniques that we now apply to semiconductor lithography, serifs were used to reduce corner rounding much earlier in other applications.[24] This simply illustrates the universal nature of physics, since similar physical situations can arise across a diverse range of technological applications.

Subresolution structures, such as serifs, have a significant impact on mask making. When subresolution features were first introduced, they represented a decrease in the minimum feature size on the masks much greater than the usual 0.7× node-to-node reduction. Mask makers were required to improve their resolution capabilities quickly. Fortunately, the linewidth control for subresolution assist features is not as stringent as for critical features that print. The intentional use of subresolution features also complicated the mask-inspection problem. Very high resolution is needed to inspect subresolution features, but then a large number of defects that don't print are also detected, increasing the time for defect review.

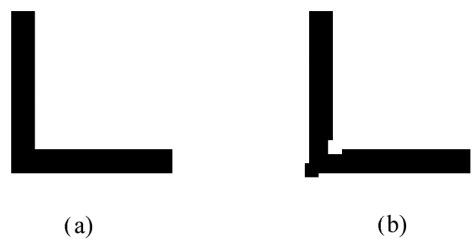


Figure 8.15 An L-shaped structure without serifs. (b) The same L-shaped geometry with serifs to reduce corner rounding.

This illustrates an advantage of mask-defect inspection systems that are based on the aerial image.

Subresolution features are used to solve other optical lithography issues in addition to corner rounding.[25] For example, off-axis illumination is a powerful method that is used to enhance the patterning of periodic structures, but it provides no benefit for isolated lines. By adding subresolution features, such as the scattering bars shown in Fig. 8.16, on the same pitch as dense features that are on the same reticle, the patterning of the isolated lines can benefit from the off-axis illumination.[26] When properly designed, these scattering bars do not print. The subresolution features shown in Fig. 8.16 are sometimes referred to as assist features[27] or subresolution assist features (SRAF). Subresolution features will also reduce iso-dense bias.[28]

Another issue that arises when k_1 becomes small is line shortening,[29] where a rectangle whose width W is printed to size but has a length L that prints too short (Fig. 8.17). Where there is room, rectangles can be biased longer on the reticle to give printed features the desired length. Compensation may also be required for corner rounding[30] for geometries other than contacts. *Hammerheads* (Fig. 8.18) can be added to mask patterns to compensate for both line shortening and corner rounding. In closely packed circuits, such as memories, there is often insufficient room for biasing, and other techniques may be required, such as those techniques discussed below.

For large-k_1 lithography, mask and design patterns could be the same, in order to produce an equivalent result on the wafer. With $k_1 < 0.5$, the mask pattern needs to be changed in order to get the designed patterns on the wafers. These adjustments to the mask are also referred to as optical proximity corrections. However, there are other effects that occur when $k_1 < 0.5$ that limit the degree to which adjustments

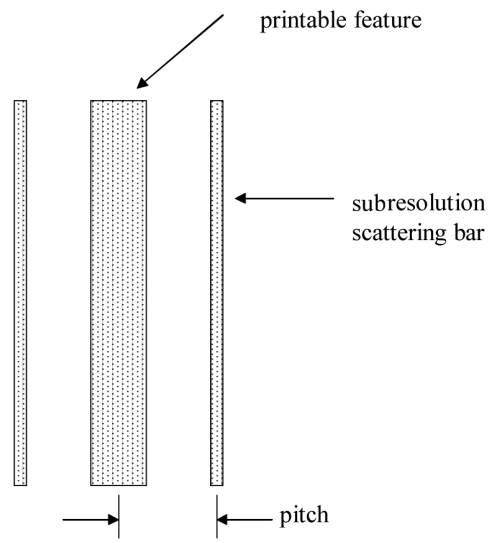


Figure 8.16 Subresolution scattering bars applied to isolated lines.

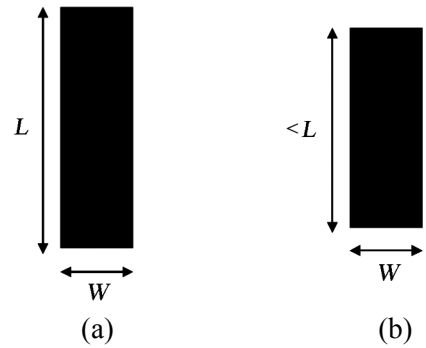


Figure 8.17 Line-end shortening (a) on the mask, and (b) on the wafer. The width of the rectangle is printed to size, but the length is too short.

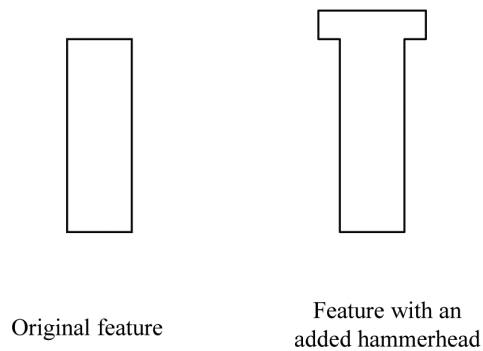


Figure 8.18 Hammerhead used to correct for line shortening and corner rounding.

can be made. One of the most significant of these effects, the mask-error factor, is the topic of the next section.

An example of the application of optical proximity corrections is shown in Fig. 8.19. Without OPC, several problems can be seen. Resist linewidths for dissimilar pitches are different, even though the widths are the same on the mask. When patterning transistors, this leads to devices with different electrical characteristics. Without OPC, the ends of lines do not extend as long as desired, which is the problem illustrated in Fig. 8.17. As narrow lines approach pad areas, the linewidths expand gradually in the resist patterns, not abruptly as in the mask layout. Compensation for this rounding is needed by using notches to ensure good gate linewidth control over the active regions. Finally, the corners of the large pad areas are rounded—a loss of pattern fidelity and desired edge placement. These problems are seen in the resist patterns of Fig. 8.19(b).

All of these problems can be improved by modifying the patterns on the masks with optical proximity corrections, as shown in Fig. 8.19(c). By changing feature sizes on the masks, gate linewidths are adjusted to match transistors on different

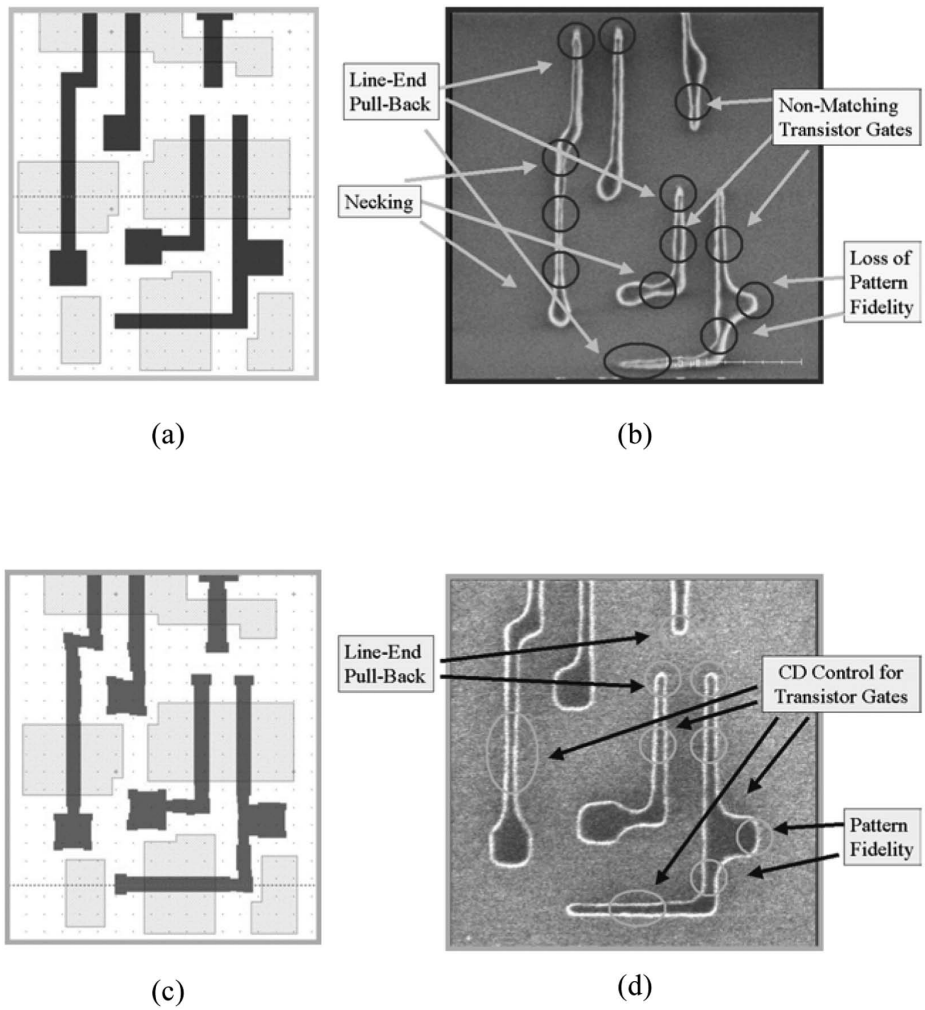


Figure 8.19 Example of optical proximity corrections (OPC).[31] (a) An uncorrected pattern, and (b) the resulting pattern in resist. The corrected mask pattern is found in (c), with the resulting resist pattern in (d). The SEM images are not on the same scale. The transition of features from narrow lines to large pads does not occur abruptly as in the design. Without OPC this results in substantial variation in the length of the gate over the active regions.

pitches, and serifs and hammerheads are added to address the problem of line-end shortening and corner rounding. Notches are introduced to reduce the problem of necking, where linewidths change gradually between narrow lines and large patterns at the ends of the lines.

Implementation of optical proximity corrections is a major task for advanced semiconductor lithography. This is a phenomenon that is associated with low-k_1 lithography, as noted in the introduction to this section. Very little or no OPC engineering was necessary prior to the 250-nm generation, although it required the most engineering resources of any single aspect of lithography development

inside semiconductor companies by the 130-nm generation. Early approaches to OPC, such as those applied to 180-nm technology, were typically *rule-based*. In this type of implementation, iso-dense bias corrections, line extensions, and serifs were applied to features based on a set of rules for each type of feature. As k_1 values became smaller, more sophisticated methods became necessary. While the concepts of optical proximity correction, as presented in this section, are reasonably straightforward, their reliable application to ultra-large-scale integrated circuits is by no means trivial.

Feature biasing and added assist features such as serifs or subresolution features not only affect a particular feature of interest, they also modulate the images of nearby geometries. Thus, there is a need to compute optical proximity corrections based on optical models, and OPC solutions must necessarily be global in nature. This *model-based OPC* implies a formidable computation challenge for highly integrated parts. For example, consider AMD's "Shanghai" quad-core 64-bit Windows™-compatible microprocessor. This chip has ~758 million transistors, and it is necessary to determine the optical proximity corrections for every single transistor. Of these 758 million transistors, a significant number are in cache memory, where the proximity corrections can largely be repeated cell-to-cell, but there are also tens of millions of transistors of random logic that require correction. For such large numbers of features, automation is required for the application of OPC. The software for this automation must be very reliable, since correcting errors on even a small fraction of the features is a major task.

Optical proximity corrections are determined by first calculating the aerial images for all features on the mask. This requires very efficient algorithms. Once the aerial images are calculated, adjustments are made for resist effects such as the diffusion and chemical reactions that take place during post-exposure bake. The physics of this step, while conceptually comprehensible, are very complex, as discussed in Chapter 4. First, bases are typically added to chemically amplified resists, so there are at least two diffusing components: the base and the photoacid. Second, the diffusivities of these moieties depend upon the density of the polymer matrix, and this density is a complex function of the amount of deprotection that has already occurred at the time of interest as well as the thermal history of the resist. This means that the diffusivity is a function of position since the amount of deprotection varies according to the amount of exposure, and it is also a function of bake time. There are multiple chemical reactions that must be considered as well, such as deprotection and acid-base neutralization. The result of this complexity has been difficulty in identifying models of sufficient accuracy, yet simple enough to be readily applied. Because there are no easily implemented models based on first principles for the post-exposure bake step for chemically amplified resists, these adjustments for resist effects are typically based on measurements.[32] Even though these resist effects are nonoptical in nature, adjustments made on the mask to compensate for them are still referred to as part of the OPC, by convention.

As optics are extended to very large numerical apertures, optical simulations must be of increasing sophistication. For numerical apertures greater than 1.0, vector models that take into account topography and polarization-dependent

reflection and refraction at the reticle and wafer are required in order to achieve good accuracy. Such models can be applied to small areas ($\sim$1 μm × 1 μm), but application to full chips is beyond current cost-effective computational capabilities. Thus, approximation methods are needed.

Adding to the OPC challenge is the need to maximize process windows. The optimum illumination condition can be determined for a given mask pattern, but a different illumination condition may prove optimal after applying optical proximity corrections. Thus, adjusted optical proximity corrections must be calculated with the newly determined optimum illumination condition. This is often an iterative process. An illumination that appears to give a good process window is identified, and the optical proximity corrections are calculated. The illumination is then reassessed, and possibly adjusted to provide a larger process window. Then the OPC needs to be recalculated. This process is repeated until solutions converge.

OPC introduces additional complexities in design databases. In a memory array, the memory cells are identically designed. This allows for compression of the data through a hierarchical structure, where memory arrays may be stored compactly as a basic cell along with its repetitive layout. After OPC is applied, the cells on the outside of the arrays do not have the same reticle layout as the inner cells. This perturbs the hierarchy and results in much larger data sets that can be many gigabytes in size.

Optical proximity corrections are implemented by moving edges of features, or parts of edges, on masks. To do this, the edges of the polygons that comprise the mask pattern must be divided into smaller *fragments*, each of which can potentially be moved. An example of this is shown in Fig. 8.20. Even within the framework of model-based OPC, assignment of the fragments is often based upon a set of rules.

How finely designs are fragmented will affect how much computation time is required when using software that moves edges according to the layout fragmentation. Fragmentation that is too fine can lead to an unacceptably long computational time and perhaps even numerical instabilities, while fragmentation, which is too coarse, will not treat corners adequately. However, efficient

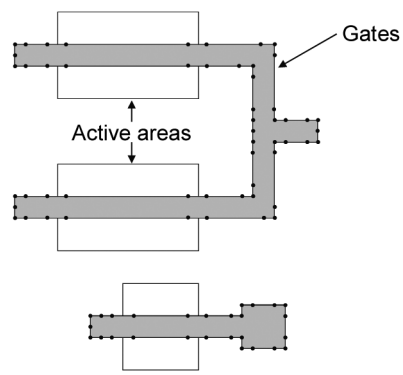


Figure 8.20 An example showing pattern fragmentation. The line between each dot is an edge that can be moved.

computational methods have been developed that enable optical proximity corrections to be calculated on a very fine and uniform grid.[33] Sometimes the improvement in computational time can be accomplished through the use of hardware acceleration.[34,35] Optical proximity corrections calculated on a very fine grid are sometimes referred to as *dense OPC* or *gridded OPC*, to be distinguished from *sparse OPC*. Dense OPC enables calculations without specifying the fragmentation first, but there remains a need for reducing the mask layout to rectangles with established vertices, because a physical mask needs to be made at some point. Fragmenting at a very fine level will increase the number of geometries, which can increase the time and cost for making reticles, often substantially. Hence, there is a practical reason to balance how finely OPC is applied with mask complexity. Adjustments should be made only to the extent that adequate patterning results.

In model-based OPC the errors in the placement of the printed edges [edge-placement error (EPE)] are calculated using a lithographic model, once features have been fragmented. Edges of the features on the mask are moved according to the computed error, and the edge-placement error is recomputed. This process is repeated until the edge-placement errors are less than some preassigned value. This value is determined by the combination of linewidth-control requirements, inherent model accuracies, and computational errors. Solutions that may be optimal at best dose and focus may not be useful, because adequate printing is not obtained over the entire process window. Consequently, optical proximity corrections need to be calculated throughout the range of potential process conditions for processes with very small process windows.[36]

OPC continues to be an area of active research and development. There are a number of companies that provide products and/or services to assist lithographers with OPC. Because of the number of unresolved problems, some of which are discussed above, turnkey solutions are not provided. OPC software should be considered as a powerful engineering tool, but lithography engineers are still required to create working solutions. This is analogous to a word processor, which facilitates writing, but the author's creativity is still required. As chip designs become more complex and linewidth-control requirements tighten, computational methods with increased efficiency are needed, even as it becomes necessary to account for new effects, such as polarization. Resolution-enhancing techniques, such as phase-shifting masks (to be discussed shortly), must be implemented as part of mask generation, of which optical proximity corrections are an integral part. OPC has become a core element of semiconductor lithography.

8.3 The Mask-Error Factor

One of the consequences of the nonlinear printing of features near the diffraction limit is a change in the magnitude by which mask linewidth errors are reduced by lens reduction.[37] What can be inferred from the data in Fig. 8.13 is that variations in linewidths about their nominal size are not reduced by the lens-reduction factor when k_1 gets small. For a set of features of the same designed size, if $\sigma_{reticle}$ is the

variation in these linewidths on the reticle, and σ_{wafer} is the corresponding variation of linewidths on the wafer, then

$$\sigma_{wafer} = \frac{MEF}{N}\sigma_{reticle}, \tag{8.9}$$

where N is the lens-reduction factor and MEF is the *mask-error factor*, defined by the following:

$$MEF = N\frac{\partial CD_{wafer}}{\partial CD_{mask}}, \tag{8.10}$$

where CD_{wafer} is the dimension on the wafer corresponding to a feature of size CD_{mask} on the mask. For large features, $MEF = 1$, but the MEF becomes bigger when k_1 is small. As lithography is pushed to the diffraction limit, the mask-error factor becomes larger. For leading-edge lithographic processes, MEFs as large as four or greater are common. For mask-error factors equal to 4, reticle linewidth variations transfer 1:1 on 4×-reduction exposure tools. The entire benefit of lens reduction, to reduce the impact of reticle variations on wafer linewidth control, is completely absent. Without significant improvements in mask fabrication, mask linewidth variations consume increasing proportions of the wafer linewidth-control budget.

The characteristics of the MEF can be better understood by considering the imaging of a diffraction grating with incoherent light near the diffraction limit. In this case the image is given by the first two terms of the Fourier series expansion of the mask pattern, which is a generalization of Eq. (2.13) to gratings of unequal lines and spaces:[39]

$$I(x) = I_0\left[\frac{CD_{ret}}{P_{ret}} + \frac{2MTF}{\pi}\sin\left(\frac{\pi CD_{ret}}{P_{ret}}\right)\cos\left(\frac{2\pi Nx}{P_{ret}}\right)\right], \tag{8.11}$$

where CD_{ret} and P_{ret} are the sizes of the line and pitch on the reticle, respectively, x measures distance in the wafer plane, and N is the lens reduction. MTF is the modulation transfer function for spatial frequency equal to the pitch of the grating. The grating is symmetric about $x = 0$. When CD_{ret} is close to one-half the pitch, the sine prefactor in Eq. (8.11) is approximately equal to 1.0 and varies little for small variations in CD_{ret} about its nominal value, while the prior term varies linearly with the mask dimension. If a resist edge x_0 corresponds to a threshold exposure dose, then

$$I_{threshold} = I_0\left[\frac{CD_{ret}}{P_{ret}} + \frac{2MTF}{\pi}\sin\left(\frac{\pi CD_{ret}}{P_{ret}}\right)\cos\left(\frac{2\pi Nx_0}{P_{ret}}\right)\right]. \tag{8.12}$$

Differentiating Eq. (8.12) with respect to CD_{ret}, one obtains

$$0 = 1 + 2MTF\cos\left(\frac{\pi CD_{ret}}{P_{ret}}\right)\cos\left(\frac{2\pi Nx_0}{P_{ret}}\right) - 2MTF\sin\left(\frac{\pi CD_{ret}}{P_{ret}}\right)\sin\left(\frac{2\pi Nx_0}{P_{ret}}\right)2N\frac{\partial x_0}{\partial CD_{ret}}. \quad (8.13)$$

Recognizing that the dimension of the feature on the wafer is $2x_0$, then

$$MEF = 2N\frac{\partial x_0}{\partial CD_{ret}}. \quad (8.14)$$

Consequently,

$$MEF = \frac{\frac{1}{MTF} + 2\cos^2\left(\frac{\pi CD_{ret}}{P_{ret}}\right)}{2\sin^2\left(\frac{\pi CD_{ret}}{P_{ret}}\right)}. \quad (8.15)$$

A plot of Eq. (8.15) illustrated in Fig. 8.21 enables us to gain some insight into the mask-error factor. The mask-error factor is a minimum when the linewidth is exactly one half of the pitch. Away from this condition, either the line or the space becomes smaller. It should be noted that the approximations in the above analysis break down when the linewidth departs far from one half of the pitch. In fact, the MEF for an isolated line is typically smaller than for equally sized dense lines. This has a consequence for implementing optical proximity corrections, because isolated lines have lower MEF than similar lines with scattering bars (Fig. 8.16).[37] The optimum process then depends upon photomask quality since the benefits of scattering bars may be offset by linewidth variations from the photomask. Near the diffraction limit, contacts have ~2× the mask-error factor of lines and spaces of similar width. The mask-error factor is another parameter that must be considered when optimizing a process since some combinations of mask layout and illumination can provide very good exposure latitude and depth-of-focus, but may be very sensitive to small changes in mask linewidth. Overall process control

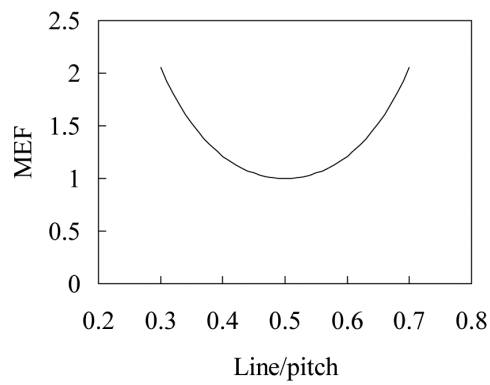


Figure 8.21 Graph of Eq. (8.15) with $MTF = 0.5$. The mask-error factor is minimized when the linewidth is one half of the pitch.[39]

depends on the contributions from the mask and sensitivities to mask linewidth variation depend on the mask-error factor.

The mask-error factor is related to the image log slope. Consider the aerial image for a pattern consisting of lines perpendicular to the x axis. The intensity for this image is a function of x and also is a function of the reticle dimension, CD_{ret}:

$$I = I(x, CD_{ret}). \tag{8.16}$$

Variations in the mask CD from its nominal value $CD_{ret,0}$ will cause a change in the position x_0 of the edge of the resist feature. In the context of the thin-resist model, these changes all occur at constant threshold intensity I_{th}. Accordingly,

$$dI(x, CD_{ret}) = \delta x \left.\frac{\partial I}{\partial x}\right|_{x_0} + \delta CD_{ret} \left.\frac{\partial I}{\partial CD_{ret}}\right|_{CD_{ret,0}} \tag{8.17}$$

$$= dI_{th} \tag{8.18}$$

$$= 0. \tag{8.19}$$

With suitable rearrangement, the use of Eq. (8.10), which defines the mask-error factor, and the fact that the wafer CD is $2x_0$ we obtain[40]

$$MEF = -2N \frac{\partial I}{\partial CD_{ret}} \left(\frac{\partial I}{\partial x}\right)^{-1} \tag{8.20}$$

$$= -\left(\frac{2N}{I_{th}}\right)\left(\frac{\partial I}{\partial CD_{ret}}\right)\left(\frac{1}{ILS}\right). \tag{8.21}$$

Once again, it can be seen that a process is more robust when there is a large image log slope.

As discussed previously, dimensions on the reticle can be resized through design to adjust for nonlinearity in printing. This works to the extent that features are sized perfectly on the reticle. With mask-error factors considerably larger than one, optical proximity corrections are much more difficult to make. Small changes in linewidth on the reticles will cause large changes of linewidth on the wafers when the mask-error factor is large. The mask-error factor affects the precision with which optical proximity corrections can be applied. Consider a 45-nm process that has a 2-nm bias between isolated and dense lines (on the wafer), printed on a 4×-reduction exposure tool. For $MEF = 1$, the linewidths on the photomasks would need to be adjusted by 8 nm to correct for the isolated-dense bias, while 2-nm adjustments would be needed for $MEF = 4$. To achieve the desired level of control, the mask would need to be written on a much finer grid. Depending upon the architecture of the mask-writing tool, this could greatly increase mask-writing time and therefore reticle cost.

The mask-error factor has a significant impact on reticle costs. As long as the $MEF = 1$, it was necessary for improvements in reticle-linewidth control to scale directly to the minimum feature size, in order to hold constant the reticle

contributions to wafer-linewidth variations. It has proven to be a challenge for mask makers to keep pace with decreases in minimum feature sizes that have historically changed at a rate of −11% per year. The challenge is now much greater,[38] since mask-error factors as large as four or greater are now often encountered. For mask-linewidth variations to result in a constant linewidth variation (as a percentage of the critical dimension) on the wafer, multiple improvements in the quality of photomasks are required. To the degree that this is technically feasible, it still comes with a cost. Alternatively, phase-shifting masks can be used, which can reduce the impact of reticle-linewidth variations considerably.[41] Phase-shifting masks are discussed in the next section.

8.4 Phase-Shifting Masks

Another method for addressing the resolution limits imposed by diffraction is phase shifting. While there are now many types of phase-shifting masks, they all employ the same basic concept, which is well illustrated by the original version introduced by Levenson, Viswanathan, and Simpson,[42] and a similar form is used here, in its contemporary embodiment, to introduce the subject. Consider a grating imaged optically using a lens with a particular numerical aperture and partial coherence. Imaging is degraded because light from a clear area on the mask is diffracted into regions that ideally are completely dark (Fig. 8.22). The nominally dark regions have light diffracted into them from the spaces on both the left and right. The idea behind the alternating phase-shifting mask is to modify the reticle so that alternating clear regions cause the light to be *phase shifted* 180 deg (Fig. 8.23).

This is accomplished by recessing the mask substrate by a small amount in the alternating clear regions. With a suitably chosen depth, the light passing through the recessed region will have a 180-deg phase difference from the light that is transmitted through the regions that are not recessed. As a consequence, the light diffracted into the nominally dark area from the clear area to the left of the dark feature destructively interferes with the light diffracted from the right clear area. This improves image contrast relative to non-phase-shifted masks, as shown in Fig. 8.24. Interference is more effective with a high degree of coherence.[43] All phase-shifting masks employ this same basic characteristic, where the destructive interference of light of opposite phases is used to improve image contrast.

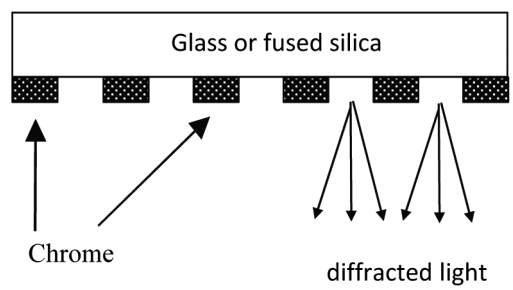


Figure 8.22 The degradation of image contrast because of diffraction.

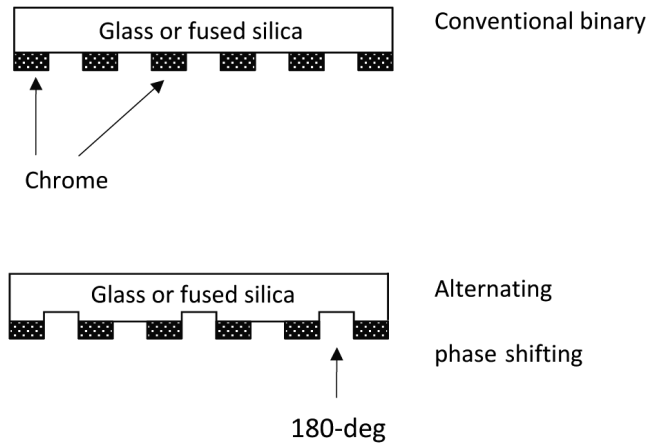


Figure 8.23 Conventional binary chrome-on-glass reticle and alternating, or Levenson-type, phase-shifting mask.

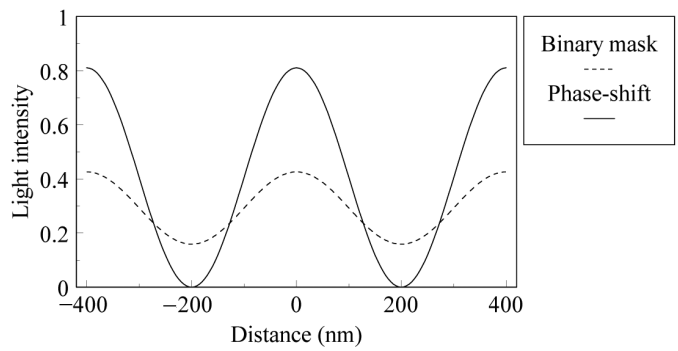


Figure 8.24 Simulated light-intensity distributions of a 400-nm pitch grating with equal lines and spaces, imaged with 0.5-NA optics at a wavelength of 248 nm. For the binary mask image, $\sigma = 0.6$, while $\sigma = 0.3$ for the alternating phase-shifting image.

Light that travels the distance a in air changes phase by

$$\phi_a = \frac{2\pi}{\lambda} a, \tag{8.22}$$

where λ is the wavelength of the light in vacuum. Through the glass material of the substrate the phase changes by an amount

$$\phi_n = \frac{2\pi}{\lambda} na, \tag{8.23}$$

where n is the index of refraction of the mask substrate.

The difference in phase shifting caused by the phase shifter is then

$$\Delta\phi = \frac{2\pi}{\lambda} a\,(n-1)\,. \tag{8.24}$$

To achieve the condition $\Delta\phi = 180$ deg $= \pi$ rad, we have the following relationship for the thickness a of the phase-shifting layer:

$$a = \frac{\lambda}{2\,(n-1)}. \tag{8.25}$$

There are a number of problems with the alternating phase-shift mask. First, additional work is required in order to apply it to anything other than an effectively infinite grating. For example, consider what happens at the edge of a phase shifter, which must exist if the product consists of anything more than completely repetitive structures. This can lead to the situation where there is a transition in the glass from 0 to 180 deg. The light from the clear 0-deg and clear 180-deg adjacent areas will interfere destructively, resulting in the light-intensity profile shown in Fig. 8.25.

Such an optical image prints into photoresist (see Fig. 8.26). Implementation of alternating phase-shifting masks must deal with these edge shifters. Three approaches have been taken. In the first, additional features are introduced between the 0 and 180-deg phase areas to produce phase edges of $\ll 180$ deg, which do not print. For example, transitions can be made 0 deg $\rightarrow$ 60 deg, 60 deg $\rightarrow$ 120 deg, and 120 deg $\rightarrow$ 180 deg. Generating these additional phases on the mask requires additional processing for mask fabrication, and there may not be adequate space to accommodate all of these transitions on masks with very tight pitches. Alternatively, a second mask can be used to expose the phase edges away. Neither approach is completely satisfactory. The first method requires additional

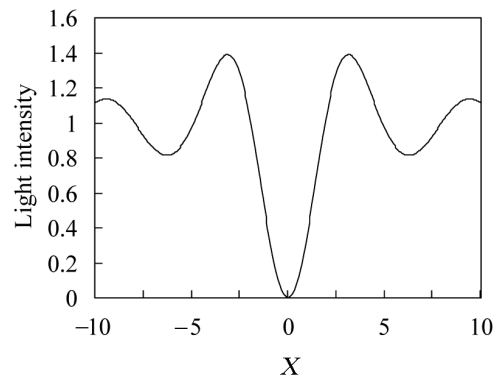


Figure 8.25 Light-intensity distribution at a phase edge for coherent light. $X = 2\pi NAx/\lambda$, where x is the physical distance from the phase edge measured in the same units as λ. The phase edge occurs at $X = 0$. The derivation of the function graphed in this figure is given below.

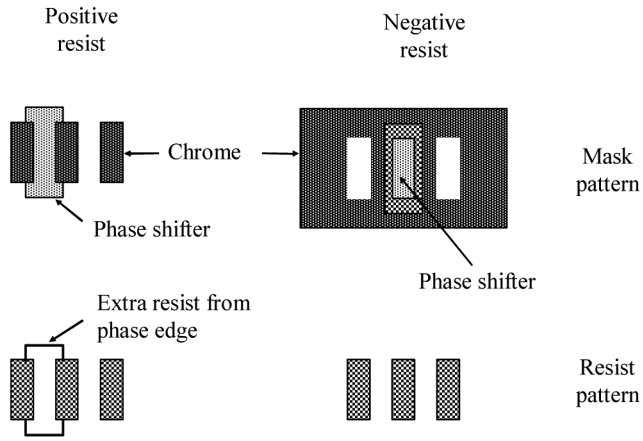


Figure 8.26 On the left, extra resist printed by the phase edge for a positive photoresist. The right side of the figure shows what occurs with negative resist. In this idealized diagram, issues that may exist, such as line shortening and corner rounding, are not indicated.

steps in the reticle fabrication process and is not applicable to very tight pitches, while the second reduces stepper productivity. In all cases, mask design and layout is complex. However, the improved capabilities that result from the alternating phase-shift–type of mask may justify these additional processing steps.

A third approach involves the use of negative photoresist. As one can see from Fig. 8.26, creating islands of resist using negative resist and phase-shifting masks avoids the phase-edge problem. This is one of the reasons that negative resists are again being used for leading-edge lithography.

The printed patterns shown in Figs. 8.25 and 8.26 can be understood as follows.[44] As discussed in Chapter 2, the image of a diffraction-limited lens for coherent illumination can be obtained from the Fourier transform of the mask pattern. In the case of the phase edge, the mask pattern has unit intensity, but half of the pattern has a phase 180 deg (π rad) different from the other half. For this situation, this transmission is given by

$$\begin{aligned} T(x) &= 1 \quad \text{for } x < 1 \\ &= e^{i\pi} \quad \text{for } x > 1. \end{aligned} \tag{8.26}$$

For coherent light, the intensity for this pattern is given by

$$I(x) = \frac{4}{\pi^2}\mathrm{Si}^2\left(\frac{2\pi NAx}{\lambda}\right), \tag{8.27}$$

where the sine integral is defined as

$$\mathrm{Si}(\theta) = \int_0^{\theta} \frac{\sin(z)}{z} dz. \tag{8.28}$$

From Eq. (8.27), one can estimate that the width of the aerial image at a 0.25 intensity is approximately

$$\text{width} = 0.25\frac{\lambda}{NA}. \tag{8.29}$$

The prefactor of 0.25 in Eq. (8.29) should be compared with the prefactors in Eqs. (2.4) and (2.7). With the phase edge, one can obtain over a 50% reduction in the size of a printable feature, compared to non-phase-shifted imaging. *Phase-edge* photomasks can be used to delineate extremely narrow lines,[45] such as gates and interconnects.

Phase-shifting masks address the problem of the mask-error factor.[46] Shown in Fig. 8.27 are simulated isolated linewidths as a function of chrome size on an alternating phase-shifting mask. For small features, the linewidth on the wafer actually becomes insensitive to the chrome linewidth on the mask, as the object begins to resemble a phase edge. In this instance, the mask-error factor can actually be less than 1.0.

One of the problems of a phase-shifting mask with the construction shown in Fig. 8.23 is the potential for the phase shifter to affect light amplitude as well as phase. This difficulty is shown in Fig. 8.28. This may occur as a consequence of etching-induced roughness or three-dimensional optical effects. The resulting line asymmetry can be characterized and corrected by OPC to maintain linewidth uniformity. Linewidth size errors may also occur if the phase shifting is different from 180 deg, which places requirements on the glass-etch process.

Phase edges provide very high resolution, but they have a number of deficiencies, which limit their applicability. Like alternating phase-shifting lithography, applying phase-edge lithography to real patterns leads to unwanted features that must be eliminated by means of a second exposure. It is also difficult

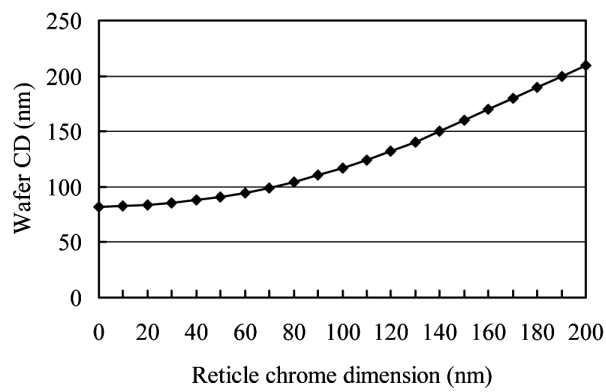


Figure 8.27 The wafer dimension as a function of reticle dimension for an alternating phase-shifting mask. These results were simulated using PROLITH2, assuming a numerical aperture of 0.7, best focus, and an imaging wavelength of 193 nm. The feature on the reticle was an isolated line of chrome. A 30% threshold of the aerial image was used to generate the wafer dimension.

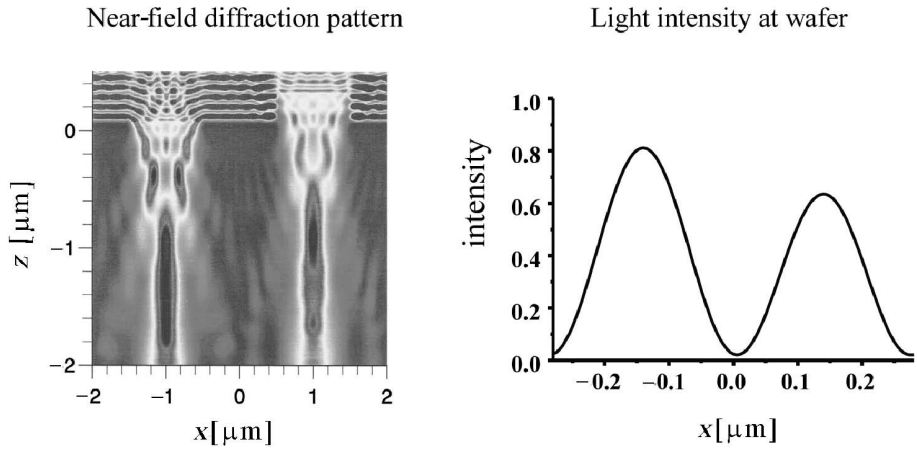


Figure 8.28 Calculated KrF images for an alternating phase-shifted mask for 250-nm lines and spaces, assuming 80-nm-thick chromium. The reduced intensity for the light going through the recessed glass is apparent in the intensities calculated assuming 0.6 NA and $\sigma = 0.5$.[47]

to modulate the linewidth in a controlled way, such as through adjustment of chrome widths on the mask. An approach to phase shifting that addresses these deficiencies of phase-edge lithography is chromeless phase-shifting lithography (CPL), which uses two phase edges to produce a single line in resist, instead of just one phase edge. This type of lithography has also been referred to as chromeless mask (CLM) lithography. It is illustrated in Fig. 8.29, where aerial images are shown for pairs of phase edges. As the phase edges are moved closer together, a single line is formed. The width of this line can be modulated by adjusting the separation between the phase edges. It can be seen that very small features can be produced by this technique.

CPL did not appear very promising from early studies.[44] The degree of enhancement over binary masks was small, and the width printed on the wafer was as much as two times larger than the nominal mask dimension (divided by the lens reduction). This latter deficiency was found to be reduced substantially when larger numerical apertures (NA $\geq$ 0.85) were used.[48] The small resolution enhancement was addressed by the use of off-axis illumination.

With CPL, a line in resist can be formed by the phase sequence of 0/180/0 deg, which avoids the problem of phase assignment and extraneous phase edges. Consequently, with CPL it is possible to expose layers with only one single mask, a considerable advantage over other methods for implementing strong phase shifting. There are also challenges associated with chromeless phase-shifting lithography. Mask patterns formed entirely from glass can have defects that are more difficult to detect and repair than patterns constructed from chrome or other opaque materials. The formation of large features also requires additional work. For example, large features can be formed from arrays of phase edges. Alternatively, chrome can be used on the mask for large features, while pairs of phase edges can be used for narrow lines.[49] This latter approach requires multiple patterning and etch steps

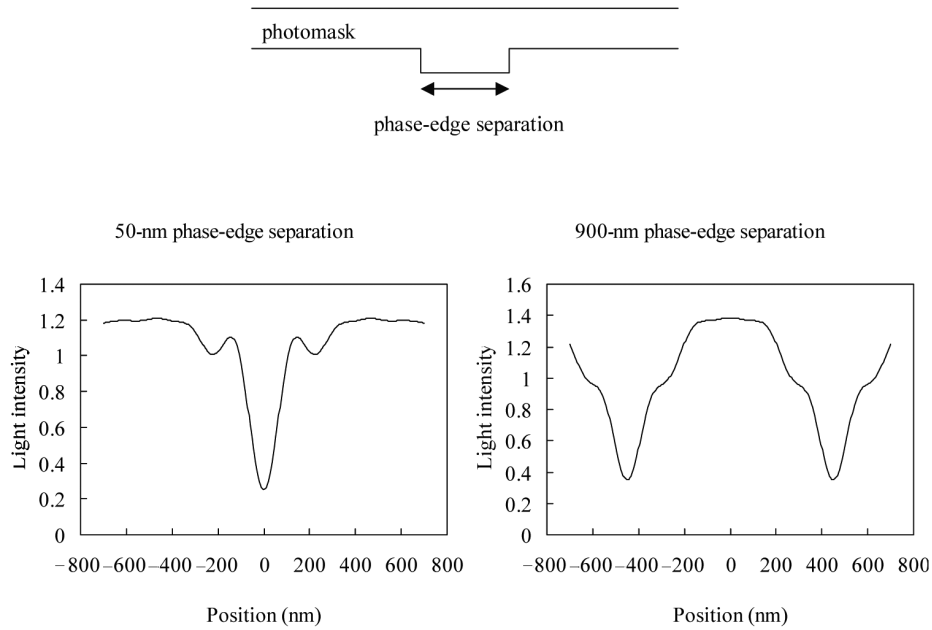


Figure 8.29 CPL example. Shown are calculated aerial images, assuming a 0.75-NA ArF exposure tool with quadrupole illumination ($0.55\,\sigma_{inner}$, $0.85\,\sigma_{outer}$). The dimensions are at the wafer.

when making the mask, but it does simplify design, layout, and OPC. There are also strong optical proximity corrections required with CPL, and these will be different, in general, from the corrections required with other types of masks.[50] An example of considerable line-end shortening observed with the use of CPL is shown in Fig. 8.30.

CPL exhibits considerable variations in linewidths and mask-error factors as a function of pitch,[51,52] which requires substantial optical proximity corrections;

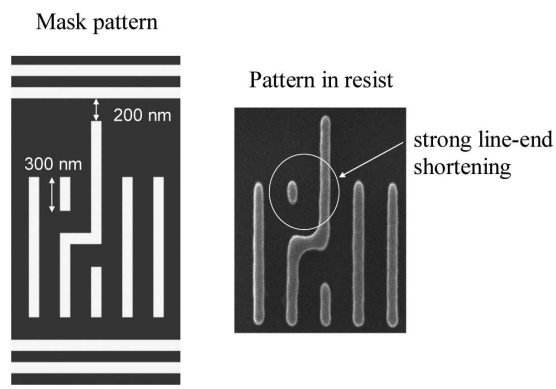


Figure 8.30 Line-end shortening with CPL.[48] A 0.7-NA KrF scanner with quadrupole illumination was used for exposing the wafer.

in this respect, it less advantageous than alternating phase shifting. However, the superior throughput, due to the need for only a single exposure, is a strong advantage of CPL. When contemplating the use of CPL, the existence of patents related to this technology should be taken into consideration.[53]

In addition to chromeless and alternating phase-shifting masks, several other versions of phase-shifting masks have been introduced, and the most common ones are listed in Table 8.2. One of these, the edge shifter, was just discussed, where it was seen as an undesirable complication for applying the alternating phase-shifting mask, and a powerful method for printing isolated lines. The *rim-shift* phase-shifting mask[54] was developed to address the problem of overlay in the mask-fabrication process through the creation of a self-aligned phase-shifting structure. This is illustrated in Fig. 8.31(a). In the fabrication of the rim shifter, the reticle is patterned with resist, and the phase-shifting material is etched anisotropically. Then the chrome is etched so it undercuts the phase-shifting material. Among the issues are rim-size optimization and sensitivities for print biases, slope in the light-intensity profile, and side-lobe intensities,[55] leading back to reticle-fabrication concerns.

Outrigger phase-shifting masks [56] [Fig. 8.31(b)] involve an approach that requires some degree of overlay between the step that patterns the chrome and the step that patterns the phase shifter, but in a mode where the edge of the phase shifter is placed on chrome, as in the case of the alternating phase-shifting mask. These types of masks are considered hybrids between phase-shifting masks and optical proximity corrections of the type discussed in the preceding section.

An extremely attractive type of phase-shifting mask, from the perspective of mask fabrication, is the *attenuated phase-shifting mask*. [57,58] In this type of mask, the nonclear areas are partially transmitting. This can be achieved, for example, with thin layers of chrome. Other materials, such as CrO, CrON, MoSiO, and MoSiON,[59,60] have also been used. Optical superlattices have also been tried.[61] With suitable processing, one can achieve 180-deg phase difference between the

Table 8.2 Types of phase-shifting masks.

Type of phase-shifting mask	Alternate names	Category	Applicable situations
Alternating	Levenson	strong	grouped lines and spaces
Chromeless	Chromeless phase-shifting lithography (CPL), chromeless mask lithography (CLM)	strong	lines and spaces, contacts[51]
Rim-shift	—	weak	contacts, isolated features
Attenuated	leaky chrome, half tone, embedded	weak	contacts, isolated features
Phase-edge	chromeless, unattenuated	strong	narrow lines
Outrigger	subresolution, additional aperture	weak	isolated features
Vortex	—	strong	contacts

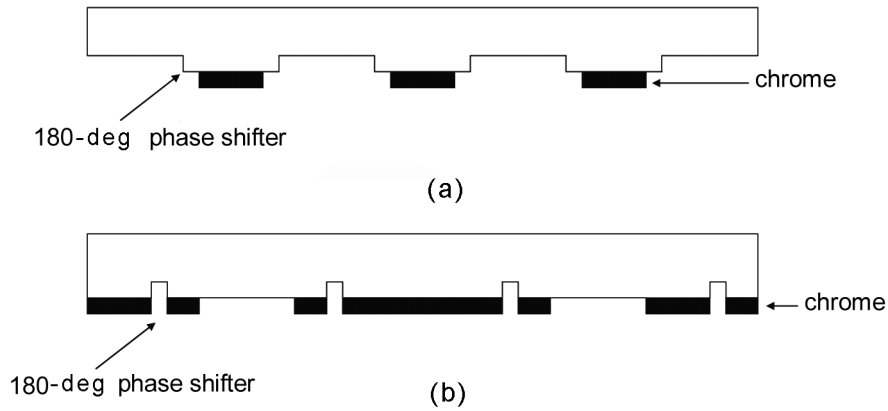


Figure 8.31 (a) A rim-shift phase-shifting mask, and (b) an outrigger phase-shifting mask.

clear and nonclear areas, sometimes with an etch of the quartz to an appropriate depth.

The partial transmission of the nonclear areas is a problem with this type of mask. For example, the threshold exposure dose for significant resist loss must be less than the amount of light that "leaks" through. An optimization of the normalized slope of the aerial image may have unacceptable levels of light in nominally "dark" areas of the mask (Fig. 8.32). This places requirements on the resist and on the transmittance of the partially transmitting areas on the reticle. The attenuated phase-shifting mask is attractive because it can be used in a single exposure step, and is relatively easy to fabricate. It is particularly useful for patterning contacts.

Another phase-shifting method that has been proposed for printing contacts and vias is the *optical vortex*, illustrated in Fig. 8.33.[62] A very dark spot is created at the intersection of all four phases. Using this type of mask and negative resists, it is possible to print small contacts and vias, with the virtue of having a low mask-error

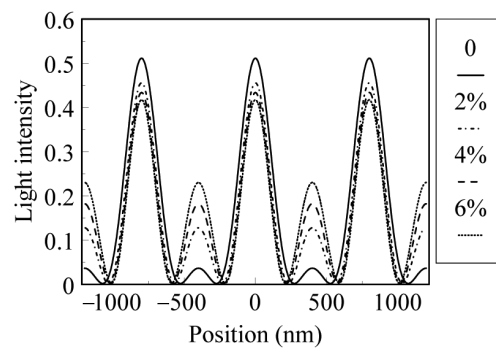


Figure 8.32 Light-intensity distributions from an attenuated phase-shifting mask, calculated with PROLITH2 for various levels of transmission through the "leaky chrome."

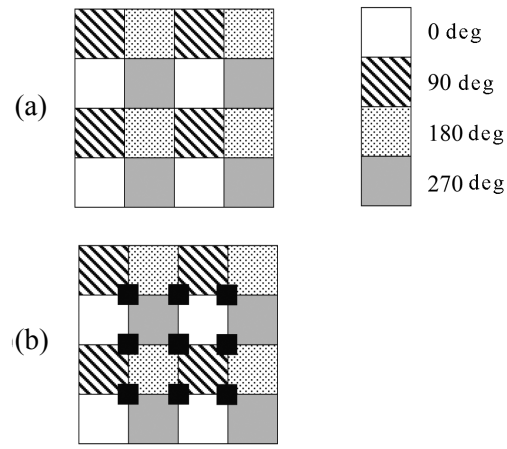


Figure 8.33 Vortex layout for printing an array of contacts using phase shifters of 0, 90, 180, and 270 deg. (a) A maskless implementation, while in (b) an absorber on the mask is added to control the contact diameters.

factor. Requiring four phases, the mask fabrication is more complicated than for a two-phase mask.

As phase-shifting technology becomes more commonplace, a name is needed for the conventional type of mask, which consists only of opaque areas (usually chrome) and quartz or glass. Because conventional masks have only completely clear or completely opaque areas, they are usually referred to as binary intensity masks or chrome-on-glass (COG) masks.

Phase-shifting masks are also classified as *strong* or *weak*, according to their ability to suppress the zeroth-order diffraction component, i.e., the extent to which they produce two-beam imaging. This is seen in Fig. 8.24, where, without a phase-shifting mask, zero intensity cannot be achieved at any point because of a zeroth-order component background. Phase-shifting masks other than the alternating or chromeless are less capable of eliminating the zeroth-order diffraction component. The combination of rim-shifted and attenuated phase-shifting masks and off-axis illumination has been found to be a powerful synergy,[63,64] since each method is only partially capable of reducing the zeroth-order light component by itself.

This can be seen from Eq. (8.3). For a mask pattern comprised of a grating with equal lines and spaces (without phase shift),

$$a_o = \frac{1}{2} \tag{8.30}$$

and

$$a_1 = \frac{1}{\pi}. \tag{8.31}$$

Because $a_0 > a_1$, contrast is less than 1.0 (see Problem 8.1). The use of phase-shifting masks can balance a_0 and a_1, thereby increasing image contrast. This can be understood as follows. With a binary mask the light amplitude varies between zero and one. With an attenuated phase-shifting mask, this amplitude varies between one and some negative value. The coefficient a_0 is proportional to the integral of the light amplitude. This is reduced with attenuated phase-shifting masks, relative to binary masks, because the 0-deg phase-shifted light is reduced in amplitude by the nonzero amount of 180-deg phase-shifted light. Similarly, the coefficient a_1 is proportional to the range of the amplitude, which is greater for attenuated phase-shifting masks than for binary masks. In particular, the combination of off-axis illumination with attenuated phase-shifting masks is common. Chromeless phase-shifting lithography can be thought of as a variation of attenuated phase shifting, with 100% transmission, and off-axis illumination is typically used with this type of mask as well.[48]

From the preceding discussions, it is clear that a particular type of phase-shifting mask or alternative illumination is applicable in specific situations but not in others. For example, alternating phase-shifting masks are useful for dense lines and spaces, but not for isolated contacts. Table 8.2 lists different types of phase-shifting methods and the situations that result in enhanced lithographic performance. Similarly, off-axis illumination was found to improve imaging of dense lines and spaces, but not particularly useful for patterning isolated features. FLEX improves the depth-of-focus for contacts, but not lines and spaces, and was found to reduce ultimate contrast. This appears to be a general feature of many newly developed methods for enhancing optical lithography: the improvement is restricted to particular types of features; there is no "magic bullet."

The fabrication of several types of phase-shifting masks requires additional processing compared to what is necessary for making binary photomasks. In particular, the alternating phase-shifting mask requires at least two patterning steps. This introduces the need for overlay capability in the mask-making process. The introduction of additional types of features on the reticles imposes new demands on inspection capabilities. Such new requirements have significant effects with respect to the capitalization required for making advanced photomasks. There are also phase defects where either glass is etched in areas where it should not have been, or glass may remain where it should have been etched. These glass-on-glass phase defects are often more difficult to detect than defects formed from opaque materials. Detection of phase defects is facilitated by inspection wavelengths close to the stepper exposure wavelengths. This has motivated the development of new generations of mask-defect-inspection tools with wavelengths close to 248 nm and 193 nm.

Attenuated phase-shifting masks are used extensively in manufacturing today. Because they require little additional processing steps compared to binary masks, they can be reasonably priced. Design issues typically revolve around the optical proximity corrections that need to be optimized for the type of mask used.

As illustrated in Fig. 8.26, there are design challenges associated with alternating phase shifting. In addition to the typical opaque or missing-chrome

defects found on binary masks, phase defects on phase-shifting masks present a challenge to inspection, since these are glass defects on a background of glass. The additional steps required for fabrication also result in masks that are more expensive. As a consequence of all of these challenges, alternating phase-shifting masks were not used extensively in volume manufacturing for a long time after the concept was introduced. However, a reported instance of alternating phase-shifting masks used for high-volume manufacturing demonstrated that the technical issues could be overcome,[65] and the use of alternating phase shifting became more widespread.

8.5 Putting It All Together

All of the techniques in this chapter need to be considered when optimizing a process. Lithography engineers must decide whether or not to use phase-shifting masks, what type of illumination to use, and where to place subresolution features. OPC must be applied, and the mask-error factor and other mask-related considerations need to be kept in mind. To make matters more complicated, each of these choices affects the others. Introduction of subresolution features will change mask-error factors and print biases, requiring recalculation of optical proximity corrections. The choice between binary and phase-shifting masks is a fundamental decision, affecting all other factors.

Illumination optimization was discussed in the section on off-axis illumination. The calculation of optical proximity corrections and optimum use of subresolution features requires the assumption of particular illumination. However, after determining the OPC and inserting subresolution features, the illumination may require reoptimization. This may be an iterative process that proceeds until convergence, or at least an adequate solution, is reached.

OPC and subresolution features represent changes to the mask that improve lithographic capability. Just as there is a methodology for optimizing the illumination, given a set of mask patterns, it is natural to ask if a similar methodology can be constructed to optimize the mask. Better yet, can techniques be developed that simultaneously optimize the source and mask, and avoid iteration? Given the potential benefits of such a capability, there have been efforts to develop the means for *source-mask optimization* (SMO).[17,66]

Source optimization is by far simpler than mask optimization or simultaneous source-mask optimization. The general approach for source optimization has been identified and was described in Section 8.1. A modestly sized set of features may be used to determine this optimum (or at least near-optimum) illumination. Once an optimized source is generated, it may be used for generating optical proximity corrections using existing OPC software and infrastructure. There are software tools available from several suppliers for source optimization, and engineers within chip-making companies have developed their own software. Mask optimization is another challenge altogether. By the very nature of mask optimization, software must be capable of handling complete mask designs, which can easily contain

hundreds of millions or billions of figures. Addressing this is a much bigger task than source optimization based on a reduced set of features.

The idea of source-mask optimization is, to some extent, a reformulation of the problem to be solved. With OPC one asks how the design layout can be modified in generating the mask layout so that the printed image more closely matches the desired design layout. The starting point is the design layout, and from that the mask is modified and effectual illumination is sought to produce wafer patterns that closely resemble what is desired. With source-mask optimization an attempt is made to start with the desired layout on the wafer and then solve the inverse problem of determining the mask and source. For this reason it is sometimes referred to as inverse lithography technology. However, this is a subtle distinction since the objective in all cases is to produce a desired pattern on the wafer by appropriate choices of illumination and modifications of the mask patterns.

The results of attempts at solving this inverse problem have led to some interesting results. Because the starting point at which OPC is applied is typically the design layout, the resulting mask usually resembles the design layout, with suitable resizings of features and the addition of serifs and subresolution assist features. Even with these modifications, the original design is usually discernible. Such is not always the case with inverse lithography, as shown in Fig. 8.34. In this case, it is far from obvious from the mask design that the intent is to print contacts in resist. As might be expected, the problem of mask inspection when subresolution assist features are utilized only increases in difficulty when source-mask optimization is used.

One approach to mask optimization has been to break the layout into pixels. If the mask is a binary mask, then the optimization process consists of deciding whether a pixel should be clear or an absorber.[68,69] Others have successfully implemented a pixel-based approach using chromeless phase-shifting masks.[70–72] As might be expected, there are many mask-making challenges associated with a pixelated mask approach.[73]

Another approach to addressing the situation where k_1 is very small is to eliminate features from the design that are fundamentally difficult to print. For example, dipole illumination can enable patterning to values of $k_1 < 0.28$, but only for lines in a single orientation and for very specific pitches. To alter the design to be more lithographically friendly, the design might be changed so that all of

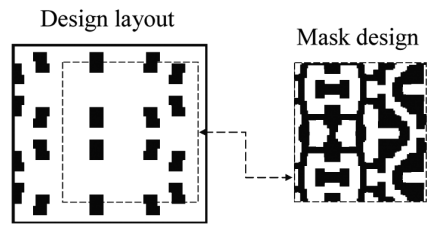


Figure 8.34 Design layout and mask design of a contact layer, resulting from applying inverse lithography techniques.[67] In both cases the black areas represent the design features, namely the contacts or the clear areas on the mask.

the critical lines on a given masking layer are oriented in the same direction.[74] This approach to making designs more printable is often referred to as *design for manufacturability* (DFM). This approach does increase the complexity of developing new processes, since it requires a greater level of interaction between lithographers and designers. However, the benefits are such that all companies making leading-edge integrated circuits use some level of DFM.

The use of resolution-enhancement techniques enables the extension of optical lithography to processes with values of k_1 much smaller than thought possible before the introduction of these methods. For the reader interested in early studies of resolution enhancement, many of the original papers on these techniques have been collected by Schellenberg.[75] Since the initial pioneering work, resolution-enhancement techniques have become a core component of lithographic technology. How far optical lithography can be extended by the use of these techniques will be discussed in Chapter 10.

Problems

8.1 Referring to Eq. (8.3), show that $a_0 = a_1$ is required to achieve an optical contrast of 1.0.

8.2 For dipole illumination, show that the optimum value for $\sigma = 1/4k_1$. If the maximum σ available on a scanner's illuminator is 0.8, what is the minimum k_1 that can be imaged with dipole illumination? For $\sigma = 0.9$?

8.3 The index of refraction of fused silica is 1.51 at a wavelength of 248.4 nm, and $n = 1.56$ at a wavelength of 193.4 nm. For KrF and ArF lithography, show that the fused silica should be recessed 243 nm and 172 nm, respectively, in order to achieve 180-deg phase shifting on an alternating phase-shifting mask.

8.4 If we want to control the phase shift on an alternating phase-shifting mask for ArF lithography to ±2 deg, show that the etch depth in the recessed areas of the mask must be controlled to ±1.9 nm.

8.5 Show that the width of the image of a phase edge is 36 nm for an ArF lens with $NA = 1.35$. (Numerical apertures greater than one are discussed in Chapter 10.)

References

1. W. N. Partlo, P. J. Thompkins, P. G. Dewa, and P. F. Michaloski, "Depth of focus and resolution enhancement for i-line and deep-UV lithography using annular illumination," *Proc. SPIE* **1927**, 137–157 (1993).
2. K. Tounai, S. Hashimoto, S. Shiraki, and K. Kasama, "Optimization of modified illumination for 0.25-μm resist patterning," *Proc. SPIE* **2197**, 31–41 (1994).

3. T. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. Res. Develop.* **41**(1/2), 57–67 (1997).
4. G. O. Reynolds, "A concept for a high resolution, optical lithographic system for producing one-half micron linewidths," *Proc. SPIE* **633**, 228–238 (1986).
5. D. L. Fehrs, H. B. Lovering, and R. T. Scruton, "Illuminator modification of an optical aligner," *KTI Microelectron. Sem.*, 217–230 (1989).
6. C. A. Mack, "Optimum stepper performance through image manipulation," *KTI Microelectron. Sem.*, 209–216 (1989).
7. K. Kamon, T. Miyamoto, M. Yasuhito, H. Nagata, M. Tanaka, and K. Horie, "Photolithography system using annular illumination," *Jpn. J. Appl. Phys. B* **30**(11), 3021–3029 (1991).
8. N. Shiraishi, S. Hirukawa, V. Takeuchi, and N. Magome, "New imaging technique for 64M-DRAM," *Proc. SPIE* **1674**, 741–752 (1992).
9. K. Tounai, H. Tanabe, H. Nozue, and K. Kasama, "Resolution improvement with annular illumination," *Proc. SPIE* **1674**, 753–764 (1992).
10. M. Noguchi, M. Muraki, Y. Iwasaki, and A. Suzuki, "Subhalf micron lithography system with phase-shifting effect," *Proc. SPIE* **1674**, 92–104 (1992).
11. J. Finders, J. van Schoot, P. Vanoppen, M. Dusa, B. Socha, G. Vandenberghe, and K. Ronse, "KrF lithography for 130 nm," *Proc. SPIE* **4000**, 192–205 (2000).
12. T. Ogawa, M. Uematsu, F. Uesawa, M. Kimura, H. Shimizu, and T. Oda, "Sub-quarter micron optical lithography with practical super resolution technique," *Proc. SPIE* **2440**, 772–783 (1995).
13. R. Rogoff, G. Davies, J. Mulkens, J. de Klerk, P. van Oorschot, G. Kalmbach, J. Wangler, and W. Rupp, "Photolithography using the Aerial illuminator in a variable NA wafer stepper," *Proc. SPIE* **2726**, 54–70 (1996).
14. T. Ogawa, M. Uematsu, T. Ishimaru, M. Kimura, and T. Tsumori, "Effective light source optimization with the modified beam for depth-of-focus enhancements," *Proc. SPIE* **2197**, 19–30 (1994).
15. T. Ogawa, M. Uematsu, F. Uesawa, M. Kimura, H. Shimizu, and T. Oda, "Sub-quarter-micrometer optical lithography with practical super-resolution technique," *Proc. SPIE* **2440**, 772–783 (1994).
16. A. Suzuki, K. Saitoh, and M. Yoshii, "Multilevel imaging system realizing $k_1 = 0.3$ lithography," *Proc. SPIE* **3679**, 396–407 (1999).
17. A. E. Rosenbluth, S. Bukovsky, M. Hibbs, K. Lai, A. Molless, R. N. Singh, and A. Wong, "Optimum mask and source patterns to print a given shape," *Proc. SPIE* **4346**, 486–502 (2001).
18. O. K. Ersoy, *Diffraction, Fourier Optics and Imaging*, Wiley Interscience, Hoboken, NJ (2007).

19. J. Leonard, J. Carriere, J. Stack, R. Jones, M. Himel, J. Childers, and K. Welch, "An improved process for manufacturing diffractive optical elements (DOEs) for off-axis illumination systems," *Proc. SPIE* **6924**, 69242O (2008).

20. M. Mulder, A. Engelen, O. Noordman, R. Kazinczi, G. Streuker, B. van Drieenhuizen, S. Hsu, K. Gronlund, M. Degünther, D. Jürgens, J. Eisenmenger, M. Patra, and A. Major, "Performance of a programmable illuminator for generation of freeform sources on high NA immersion systems," *Proc. SPIE* **7520**, 75200Y (2009).

21. F. M. Schellenberg, H. Zhang, and J. Morrow, "SEMATECH J111 project: OPC validation," *Proc. SPIE* **3334**, 892–911 (1998).

22. W. Maurer, K. Satoh, D. Samuels, and T. Fischer, "Pattern transfer at $k_1 = 0.5$: get 0.25-μm lithography ready for manufacturing," *Proc. SPIE* **2726**, 113–124 (1996).

23. A. Starikov, "Use of a single size square serif for variable print bias compensation in microlithography: method, design and practice," *Proc. SPIE* **1088**, 34–46 (1989).

24. *Techniques of Microphotography*, Eastman Kodak Industrial Data Book No. P-52 (1963).

25. J. F. Chen, and J. A. Matthews, "Mask for photolithography," U.S. Patent No. 5,242,770 (1993).

26. G. P. Watson, R. A. Cirelli, A. G. Timko, O. Nalamasu, C. Lockstamphor, S. D. Berger, N. Bassom, and G. Sundaram, "Sub-resolution assist feature and off-axis illumination optimization for 200- and 240-nm contact windows using 248-nm lithography," *Proc. SPIE* **3334**, 131–139 (1998).

27. J. Garofalo, C. J. Biddick, R. L. Kostelak, and S. Vaidya, "Mask assisted off-axis illumination technique for random logic," *J. Vac. Sci. Technol. B* **11**(6), 2651–2658 (1993).

28. J. F. Chen, T. Laidig, K. E. Wampler, and R. Caldwell, "Practical method for full-chip optical proximity correction," *Proc. SPIE* **3051**, 790–803 (1997).

29. P. Chien and M. Chen, "Proximity effects in submicron optical lithography," *Proc. SPIE* **772**, 35–40 (1987).

30. P. Pforr, A. K. Wong, K. Ronse, L. Van den Hove, A. Yen, S. R. Palmer, G. E. Fuller, and O. W. Otto, "Feature biasing versus feature-assisted lithography: a comparison of proximity correction methods for $0.5^*(\lambda/NA)$ lithography," *Proc. SPIE* **2440**, 150–170 (1995).

31. C. Spence, "Mask data preparation issues for the 90 nm node: OPC becomes a critical manufacturing technology," *Future Fab* **16**, 77–79 (2003).

32. B. Tollkühn, A. Erdmann, J. Lammers, C. Nölscher, and A. Semmler, "Do we need complex resist models for predictive simulation of lithographic process performance?" *Proc. SPIE* **5377**, 983–994 (2004).

33. N. Cobb and D. Dudau, "Dense OPC and verification for 45 nm," *Proc. SPIE* **6154**, 615401 (2006).

34. Y. Cao, Y. Lu, L. Chen, and J. Ye, "Optimized hardware and software for fast, full chip simulation," *Proc. SPIE* **5724**, 407–414 (2005).

35. J. Conga, and Y. Zou, "Lithographic aerial image simulation with FPGA-based hardware acceleration," *Proc. 16th Int. ACM/SIGDA Sym. Field Programmable Gate Array*, 67–76 (2008).

36. M. Terry, G. Zhang, G. Lu, S. Chang, T. Aton, R. Soper, M. Mason, S. Best, and B. Dostalik, "Process window and interlayer aware OPC for the 32 nm node," *Proc. SPIE* **6520**, 65200S (2007).

37. A. K. Wong, R. A. Ferguson, and S. M. Mansfield, "The mask error factor in optical lithography," *IEEE Trans. Semicond. Manuf.* **13**(2), 235–242 (2000).

38. A. Vacca, B. Eynon, and S. Yeomans, "Improving wafer yields at low k_1 with advanced photomask defect detection," *Solid State Technol.*, 185–192 (June, 1998).

39. C. A. Mack, "More on the mask error enhancement factor," *Microlithog. World*, 18–20 (Autumn, 1999).

40. C.-K. Chen, T.-S. Gau, J.-J. Shin, R.-G. Liu, S.-S. Yu, A. Yen, and B. J. Lin, "Mask error tensor and causality of mask error enhancement for low-k_1 imaging: theory and experiments," *Proc. SPIE* **4691**, 247–258 (2002).

41. T. Terasawa and N. Hasegawa, "Theoretical calculation of mask error enhancement factor for periodic pattern imaging," *Jpn. J. Appl. Phys* **39**(1), 6786–6791 (2000).

42. 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**(12), 1828–1836 (1982).

43. K. Hashimoto, K. Kawano, S. Inoue, S. Itoh, and M. Nakase, "Effect of coherence factor σ and shifter arrangement for the Levenson-type phase-shifting mask," *Jpn. J. Appl. Phys* **31**, 4150–4154 (1992).

44. C. A. Mack, "Fundamental issues in phase-shifting mask technology," *KTI Microelectron. Conf.*, 23–35 (1991).

45. T. Tanaka, S. Uchino, N. Hasegawa, T. Yamanaka, T. Terasawa, and S. Okazaki, "A novel optical lithography technique using the phase-shifter fringe," *Jpn. J. App. Phys.* **30**, 1131–1136 (1991).

46. M. D. Levenson, "Will phase-shift save lithography?" *Proc. Olin Microlithog. Symp.*, 165–178 (1998).

47. The results shown in this figure were calculated by Dr. Andreas Erdmann and were published in an article by W. Henke, "Simulation for DUV-lithography," *Semicond. Fabtech*, 9th ed., 211–218 (1999).

48. J. F. Chen, J. S. Petersen, R. Socha, T. Laidig, K. E. Wampler, K. Nakagawa, G. Hughes, S. MacDonald, and W. Ng, "Binary halftone chromeless PSM technology for λ/4 optical lithography," *Proc. SPIE* **4246**, 515–533 (2001).

49. J. S. Petersen, W. Conley, B. Roman, L. Litt, K. Lucas, W. Wu, D. Van Den Broeke, J. F. Chen, T. Laidig, K. Wampler, D. J. Gerold, R. Socha, J. van Praagh, and R. Droste, "Development of a sub-100-nm integrated imaging system using chromeless phase-shifting imaging with very high NA KrF exposure and off-axis illumination," *Proc. SPIE* **4691**, 515–529 (2002).

50. D. Van Den Broeke, J. F. Chen, T. Laidig, S. Hsu, K. E. Wampler, R. Socha, and J. S. Petersen, "Complex 2D pattern lithography at λ/4 resolution using chromeless phase lithography (CPL)," *Proc. SPIE* **4691**, 196–214 (2002).

51. A. Kroyan and H.-Y. Liu, "Resolution enhancement technology requirements for 65-nm node," *Proc. SPIE* **5042**, 305–313 (2003).

52. A. Kroyan and H.-Y. Liu, "Phase-shifting methods at 65 nm: a comparison of AAPSM and CLM," *Solid State Tech.*, 59–62 (October, 2003).

53. J. F. Chen, and J. S. Petersen, "System and method of providing optical proximity correction for features using phase-shifted halftone transparent/semi-transparent features," U.S. Patent No. 6,335,130 B1 (2002). The assignee is ASML Masktools Netherlands B.V.

54. A. Nitayama, T. Sato, K. Hashimoto, F. Shigemitsu, and M. Nakase, "New phase-shifting mask with self-aligned phase shifters for a quarter micron photolithography," *IEDM Tech. Dig.*, 57–60 (1989).

55. Z. Cui, P. D. Prewett, and B. Martin, "Optimization of rim phase shift masks for contact holes," *Microelctron. Eng.* **23**, 147–150 (1994).

56. T. Terasawa, N. Hasegawa, T. Kurosaki, and T. Tanaka, "0.3-μm optical lithography using a phase-shifting mask," *Proc. SPIE* **1088**, 25–33 (1989).

57. T. Terasawa, N. Hasegawa, H. Fukuda, and S. Katagiri, "Imaging characteristics of multiphase-shifting and halftone phase-shifting masks," *Jpn. J. App. Phys.* **30**, 2991–2997 (1991).

58. B. J. Lin, "The attenuated phase-shifting mask," *Solid State Technol.*, 43–47 (January, 1992).

59. M. Nakajima, N. Yoshioka, J. Miyazaki, H. Kusunose, K. Hosono, H. Morimoto, W. Wakamiya, M. Murayama, Y. Watakabe, and K. Tsukamoto, "Attenuated phase-shifting mask with a single-layer absorptive shifter of CrO, CrON, MoSiO, and MoSiON film," *Proc. SPIE* **2197**, 111–121 (1994).

60. G. Dao, G. Liu, R. Hainsey, J. Farnsworth, Y. Tokoro, S. Kawada, T. Yamamoto, N. Yoshioka, A. Chiba, and H. Morimoto, "248-nm DUV MoSiON embedded phase-shifting mask for 0.25 micrometer lithography," *BACUS Photomask News* **11**(8), 1–9 (1995).

61. P. F. Carcia, R. H. French, G. Reynolds, G. Hughes, C. C. Torardi, M. H. Reilly, M. Lemson, C. R. Miao, D. J. Jones, L. Wilson, and L. Dieu, "Optical

superlattices as phase-shift masks for microlithography," *Proc. SPIE* **3790**, 23–35 (1999).

62. M. D. Levenson, S. M. Tan, G. Dai, Y. Morikawa, N. Hayashi, and T. Ebihara, "The vortex via process: analysis and mask fabrication for contact CDs < 80 nm," *Proc. SPIE* **5040**, 344–370 (2003).
63. K. Kamon, T. Miyamoto, Y. Muoi, H. Nagata, N. Kotani, and M. Tanaka, "Photolithography system using a combination of modified illumination and phase-shift mask," *Jpn. J. Appl. Phys. B* **31**(12), Pt. 1, 4131–4236 (1992).
64. T. A. Brunner, "Rim phase-shift mask combined with off-axis illumination: a path for $0.5\lambda/NA$ geometries," *Proc. SPIE* **1927**, 54–62 (1993).
65. P. Ackmann, S. Brown, J. Nistler, and C. Spence, "Phase shifting and optical proximity corrections to improve CD control on logic devices in manufacturing for sub-0.35-μm i-line," *Proc. SPIE* **3051**, 146–153 (1997).
66. A. E. Rosenbluth, D. O. Melville, K. Tian, S. Baghheri, J. Tirapu-Azpiroz, K. Lai, A. Waechter, T. Inoue, L. Ladanyi, F. Barahona, K. Scheinberg, M. Sakamoto, H. Muta, E. Gallagher, T. Faure, M. Hibbs, A. Tritchkov, and Y. Granik, "Intensive optimization of masks and sources for 22 nm lithography," *Proc. SPIE* **7274**, 727909 (2009).
67. K. Lai, A. E. Rosenbluth, S. Bagheri, J. Hoffnagle, K. Tian, D. Melville, J. Tirapu-Azpiroz, M. Fakhry, Y. Kim, S. Halle, G. McIntyre, A. Wagner, G. Burr, M. Burkhardt, D. Corliss, E. Gallagher, T. Faure, M. Hibbs, D. Flagello, J. Zimmerman, B. Kneer, F. Rohmund, F. Hartung, C. Hennerkes, M. Maul, R. Kazinczi, A. Engelen, R. Carpaij, R. Groenendijk, J. Hageman, and C. Russ, "Experimental result and simulation analysis for the use of pixelated illumination from source mask optimization for 22 nm logic lithography process," *Proc. SPIE* **7274**, 72740A-1 (2009).
68. B. Kim, Y. Kim, S. Lee, S. Kim, S. Ha, J. Kim, and A. Tritchkov, "Pixel-based SRAF implementation for 32 nm lithography process," *Proc. SPIE* **7122**, 71220T (2008).
69. X. Ma and G. R. Arce, "Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography," *Optic. Exp.* **17**(7), 5783–5793 (2009).
70. Y. Borodovsky, W. Cheng, R. Schenker, and V. Singh, "Pixelated phaser mask as novel lithography RET," *Proc. SPIE* **6924**, 69240E (2008).
71. R. Schenker, S. Bollepalli, B. Hu, K. Toh, V. Singh, K. Yung, W. Cheng, and Y. Borodovsky, "Integration of pixelated phase masks for full-chip random logic layers," *Proc. SPIE* **6924**, 69240I (2008).
72. V. Singh, B. Hu, K. Toh, S. Bollepalli, S. Wagner, and Y. Borodovsky, "Making a trillion pixels dance," *Proc. SPIE* **6924**, 69240S (2008).
73. W. Cheng, J. Farnsworth, W. Kwok, A. Jamieson, N. Wilcox, M. Vernon, K. Yung, Y Liu, J. Kim, E. Frendberg, S. Chegwidden, R. Schenker, and

Y. Borodovsky, "Fabrication of defect-free full-field pixelated phase mask," *Proc. SPIE* **6924**, 69241G (2008).

74. L. Capodieci, P. Gupta, A.B. Kahng, D. Sylvester, and J. Yang, "Toward a methodology for manufacturability-driven design rule exploration," *Proc. 41st Design Automation Conf.*, 311–316 (2004).

75. F. M. Schellenberg, Ed., *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, WA (2004).

Chapter 9
Metrology

Measurement is required for developing lithographic processes and is fundamental to engineering and manufacturing control. The problem of accurate measurement is particularly significant for lithographers, where the extremely small features of state-of-the art processes have pushed measurement capabilities to their limits, and, some might say, beyond. Some of the metrology issues and challenges faced by lithographers are discussed in this chapter.

9.1 Linewidth Measurement

A number of methods have been used to measure the dimensions of resist lines, as well as spaces and holes in resist. Optical methods were used in the earliest days of the semiconductor industry, but were replaced by 1990 by scanning electron microscopes as the primary tools for linewidth measurement. Consequently, these early types of optical methods are not discussed further in this book. However, for linewidth measurement, optical methods have returned in a new form, scatterometry, and this technique is discussed. A number of other methods for measuring linewidths are used in critical, specialized applications, and one of these, electrical linewidth measurement, is also included for discussion.

9.1.1 Linewidth measurement using scanning electron microscopes

The most common tool for measuring linewidths is the scanning electron microscope (SEM). Low-voltage SEMs are capable of measuring resist features in line. In the scanning electron microscope, electron beams are scanned across patterns on wafers. The voltage of the electron beams ranges from a few hundred volts to tens of thousands of volts. For measuring resist features, a typical range is 300–1000 volts. The incident beam is scattered, both elastically and inelastically, producing secondary and backscattered electrons.[1,2] By commonly accepted definition, secondary electrons are those with energy less than 50 eV, while the backscattered electrons are those with energies closer to that of the incident beam. The secondary electrons created by inelastic scattering or the elastically backscattered electrons are detected synchronously with the scan of the incident beam (Fig. 9.1). Because the number and direction of the scattered electrons depend upon the material composition and topography of the features

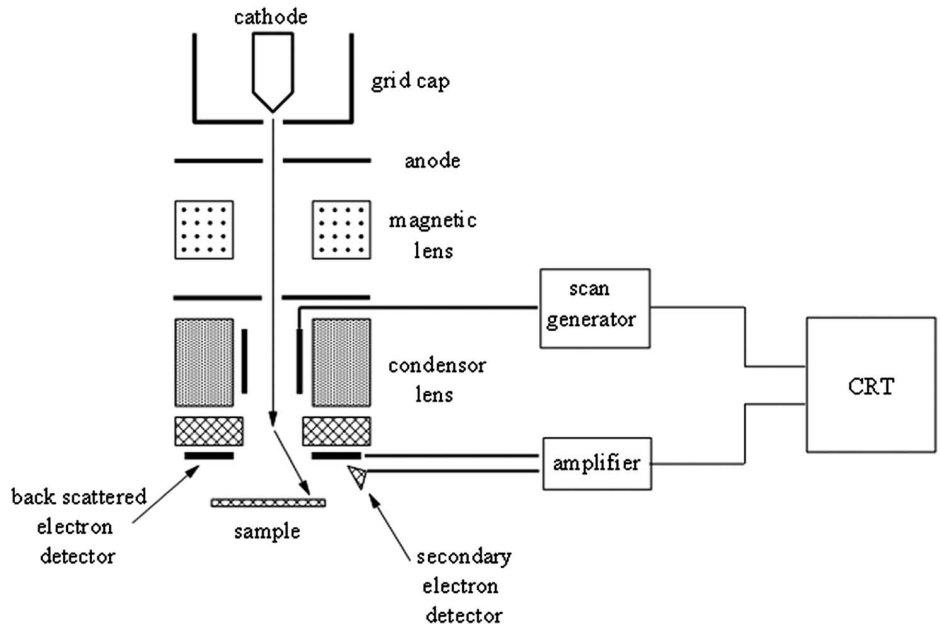


Figure 9.1 Cross-sectional schematic view of a scanning electron microscope.

over which the incident beam is scanned, the intensity of the detected signal varies so that an image can be created.

As the incident beam is scanned across a line, the detected signal varies. Algorithms are required to relate the edges of the line to points on the curve of intensity plotted as a function of position, and these algorithms do not always provide complete accuracy. An idealized signal at the edge of a feature is shown in Fig. 9.2. A mathematical algorithm is required to identify a point on the signal profile to represent the edge of the line. A number of edge-detection methods are used.[3] The simplest algorithm uses a threshold value for identifying this point:

$$\text{threshold level} = (1 - P) \times \text{minimum} + (P) \times \text{maximum}, \tag{9.1}$$

where P is a value between 0 and 1, typically in the neighborhood of 0.5. Other methods for establishing line-edge position include use of maximum slopes, inflection points, linear approximations, minima, or maxima.

Secondary and backscattered electron emissions are proportional to slopes of features, typically varying as $1/\cos(\theta)$ (Fig. 9.3).[4] Steep slopes appear "bright" in scanning electron microscopes. Consequently, the signal depends upon the slope of the feature, something that changes easily with stepper defocus and other process variations for resist features, and variations in etch processes change the slopes of etched features. By measuring linewidths and slopes from cross sections and comparing these results to in-line SEM measurements, a careful study found that measurements of 180–250-nm lines vary by as much as 10 nm per degree change in line-edge slope (Fig. 9.4).[5]

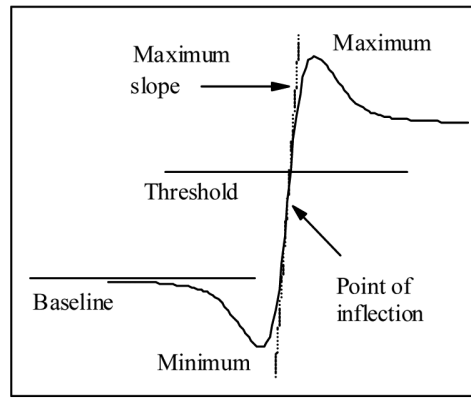


Figure 9.2 The electron signal corresponding to one edge of a feature being measured by a scanning electron microscope. Several attributes of the feature are potentially used to infer the location of the edge of the feature.

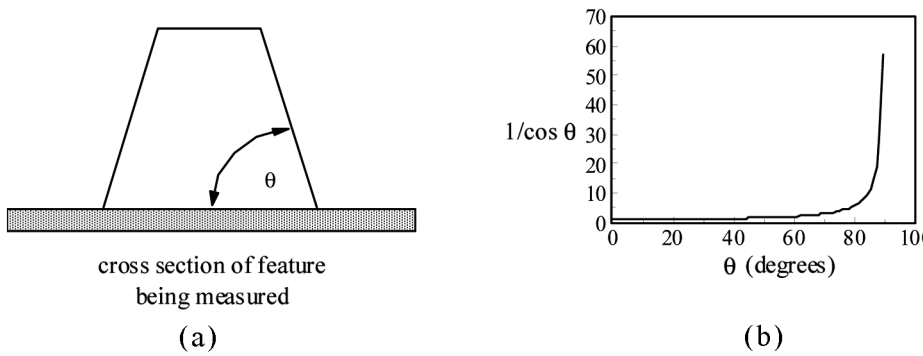


Figure 9.3 (a) Slope of a feature being measured, and (b) the secondary electron production as a function of this slope.

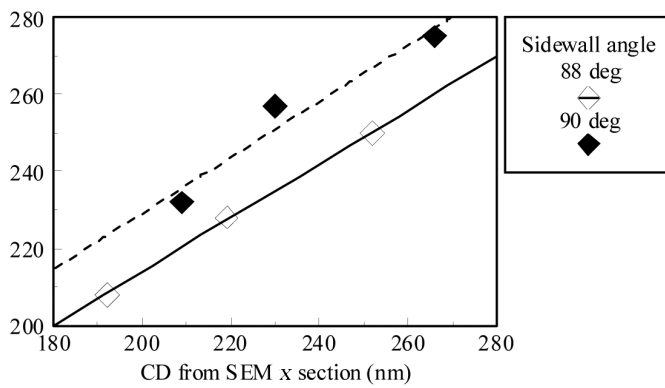


Figure 9.4 Top-down SEM measurements versus measurements from cross sections of the same etched polysilicon features.[5] Measurement error bars are not shown.

Sample charging, another problem associated with SEM linewidth measurements, is a particularly significant problem for measuring photoresist. Consider an electron beam focused on a homogeneous substrate. At low-incident-beam voltages more electrons are introduced into the sample than are emitted in the form of secondary and backscattered electrons. As the beam voltage is increased, this balance changes, and samples acquire a net positive charge during imaging[6] (Fig. 9.5). At even higher voltages, the net charge in the sample reverses yet again. Imaging and measurement is usually most stable at a point where the samples remain electrically neutral. Point E_2 is the more stable of the two energies where sample charge neutrality is maintained, and scanning electron microscopes are typically operated close to this voltage. For conductive materials, this voltage ranges from 2–4 kV, while it is only 0.5–1.0 kV for nonconducting materials such as photoresist.[2] Since imaging is typically superior at higher voltages, SEM metrology for lithography is at a distinct disadvantage.

The preceding description of charging an object on which an electron beam is impinging is appropriate for flat and uniform materials. The charging of features measured in lithography is a complex phenomenon, and the idea of a single voltage at which there is charge neutrality is inadequate in the context of integrated circuit metrology. The incident electrons and emitted secondary and backscattered electrons are not the only sources of electrical charge that need to be considered in order to understand sample charging. Samples in scanning electron microscopes are typically grounded, and there is a current between the surface of the sample and ground. This current is proportional to the voltage difference between the sample surface and ground, and it depends upon the resistivity of the materials that comprise the sample. Highly conductive materials do not charge, because there is

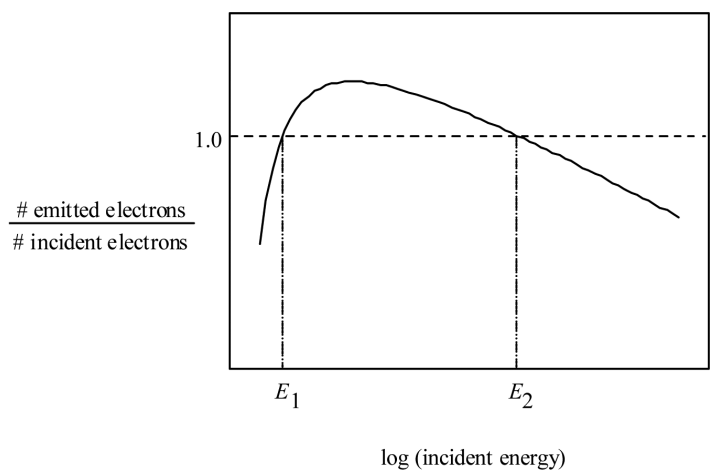


Figure 9.5 At incident energies E_1 and E_2 the sample is electrically neutral while it charges for other energies. The values of E_1 and E_2 are material dependent.

a steady current to ground. This charge balance is written as

$$I_B = (\delta + \eta)\, I_B + \frac{E_{surface}}{R}, \tag{9.2}$$

where I_B is the incident beam current, δ is the secondary electron yield, η is the backscattered electron yield, $E_{surface}$ is the surface potential, and R is the effective resistance to ground. Consequently, the surface potential is given by[7]

$$E_{surface} = I_B R\,(1 - \delta - \eta)\,. \tag{9.3}$$

The surface potential is seen to be proportional to beam current, once the current-to-ground is taken into consideration. Conductivity to ground can be complex in insulating materials, where the excitation of electrons to the conduction band is a function of beam energy and current. Complicating the situation even more is the ability of insulators to trap charge. The time it takes to discharge the trapped charges as a consequence of thermal relaxation is on the order of minutes.

Unfortunately, this simple picture is still incomplete for the types of measurements that occur in lithography applications and typically involve features patterned into photoresist. As noted previously, secondary electron emission depends on the angle between the incident beam and the plane of the surface being probed. Consequently, the electron-beam energy at which there is charge neutrality varies between the top and sides of particular features, because the emission rate of secondary and backscattered electrons from the tops of features differ from the rate of the sides. The photoresist is nearly always developed to the substrate in exposed areas, and the substrate usually has a different value for E_2 than the photoresist.

Moreover, the surface potential may be more important than overall charge neutrality. It is the electrical potential at the surface of the features, relative to the potential at the detector, that determines the flow of secondary electrons from the object being measured to the detector. Overall charge neutrality is important for electrons far from the object, but the distribution of electrons within and near the object will determine the electrical potential at the object's surface. The issue of sample charging remains a topic of current research.

Because photoresist is not electrically conducting it is difficult to dissipate charge once accumulated, and for the reasons discussed above, complete charge neutrality is rarely achieved. Charging also affects the linewidth measurement. A negatively charged line, which occurs for higher voltages, deflects the electron beam and results in a narrower measurement (Fig. 9.6).[8] For this reason, charging is of great significance to linewidth metrology using scanning electron microscopes.

Once secondary electrons are generated, they need to escape from the sample and hit a detector in order to create a signal. SEM measurements of contact holes are difficult for this reason—it is difficult for low-energy secondary electrons to emerge from positively charged holes.[9] For this reason, detection of backscattered electrons is sometimes used for metrology with lithography applications. Because backscattered electron emission peaks at or near the direction of the incident beam,

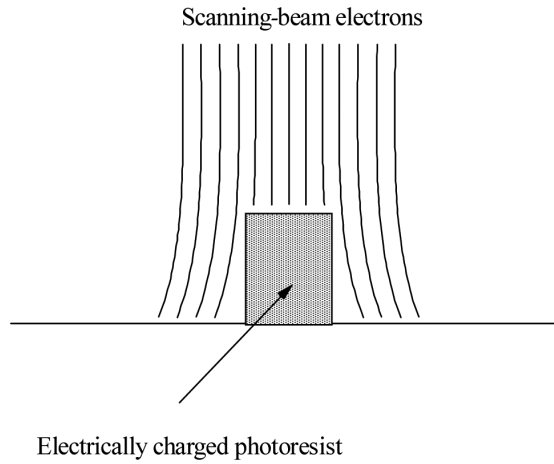


Figure 9.6 Scanning-beam electrons are deflected away from the negatively charged photoresist. As a consequence, the line is measured narrower than it actually is.

detectors are typically placed within the lens.[10] The conversion of a secondary or backscattered electron intensity into a linewidth number is not automatic. It may be argued that measurement precision is the critical requirement, and that absolute accuracy is less important. This is true to a certain extent, but facilities with more than one SEM, or ones that need to transfer processes to or from others, require standards to match SEMs. Consequently, some level of calibration is usually required, and this gauge setting is generally material dependent. Algorithms for calibrations on resist features will differ from those for polysilicon and depend on the thickness and slopes of the resist features. Consequently, there are very few linewidth standards available from organizations such as NIST that are used directly in lithography applications. However, there are standards that can be used for calibrating the magnification of electron microscopes, which has utility for maintaining process control, particularly in facilities with multiple measurement tools.

A current NIST magnification standard reference material, SRM 2800, is fabricated from chromium on quartz, intended primarily for calibrating measurements on photomasks, and has features down to 1 μm.[11,12] There have been efforts to develop standards designed specifically for use in SEMs.[13,14] A standard for calibrating SEM magnification, SRM 8820, is available from NIST. SRM 8820 consists of etched lines of amorphous silicon on an oxide-coated silicon wafer, and it has pitches down to 200 nm.[15] The NIST SRM 2069b is comprised of graphitized rayon fibers and can be used to assess contrast in the SEM.[16] Since SEMs "measure" different linewidths for features made from different materials or thickness, these standards are useful only for ensuring measurements on photomasks or controlling such parameters as the magnification of the SEM. However, the magnification is verified by measuring the pitch of grating patterns, which gives measurements independent of materials and measurement algorithms.[17,18]

Although commercially available SEM-based linewidth-measurement systems may not produce accurate measurements (i.e., the "real" linewidth), they may be precise and repeatable. The difference between accuracy and precision is illustrated in Fig. 9.7. Measurement reproducibility is good to have, but it cannot be the only metric used to judge measurement techniques. Methods of measurement need to be able to distinguish correctly between features of different linewidth.[19]

One problem preventing the creation of good linewidth standards for lithography is carbon deposition during SEM measurement.[20] While the electron beam is scanning the sample, carbon is being deposited. As one might expect, this is a particularly significant problem for photoresist samples. Repeated measurements lead to changes in the measured linewidth, even though the pitch remains unchanged.[21] One approach to dealing with the change in measured linewidth due to carbon deposition on the standard is to predetermine the magnitude of deposition-induced linewidth change per measurement, and subtract that from the monitor values.[22] Statistical process control (SPC) can then be applied to the residuals.

While an SEM reproduces the linewidth on a particular standard artifact, there is some concern that linewidth measurements of other materials, such as photoresist, are not properly calibrated. To this end, atomic force microscopes are sometimes used to measure photoresist linewidths for the purpose of calibration.[23] Atomic force microscopes are purely mechanical and will not have the type of material dependence that scanning electron microscopes have. As with any measurement system, the proper interpretation of the numbers produced by atomic force microscopes requires an understanding of how they measure. For example, because of probe tip size and shape, atomic force microscopes do not provide reliable measurements of structures with very tight pitches. Nevertheless, the atomic force microscope is another useful metrology tool.

Calibration usually requires measurements of multiple linewidths. Within a limited range of dimensions, calibrations can be based on affine relationships. If r is the "real" linewidth, then the measurement linewidth y is given by

$$y = ar + b \tag{9.4}$$

to the first order, where a and b are constants. Measurement of more than one linewidth is required to determine the two calibration parameters, as well as reduce

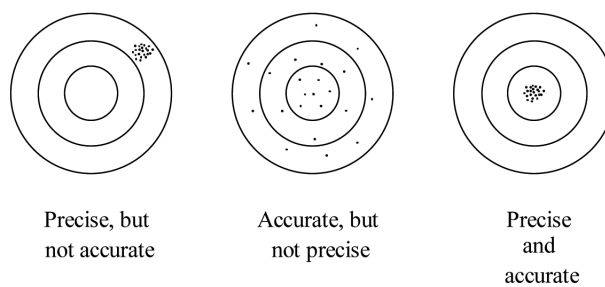


Figure 9.7 Shots at a target, illustrating precision and accuracy.

errors from measurement noise. Calibrations over large ranges of dimensions are typically piecewise linear.

The SEM used for metrology must be properly focused and stigmated in order to obtain good measurements. In the past, this was accomplished by using a sample of very small grains, such as sputtered gold or grainy silicon, and adjustments were made until the image appeared best to the operator. Recent work has shown that the subjective element associated with this approach can be overcome by taking a two-dimensional Fourier transform of the image.[24] Defocus involves a decrease in the magnitudes of the higher-frequency components, and astigmatism shows asymmetries between axes.

The introduction of ArF optical lithographic technology brought along a metrology problem associated with measuring linewidths in resists. Many of the materials comprising ArF resists are modified by electron beams, with the consequence that resist linewidths shrink under exposure to such beams, as occurs in scanning electron microscopes. Measurement precision requires repeatability among measurements in terms of the exposure of the resist to e-beams, and accuracy requires taking the shrinkage into account. The magnitude of the problem is reduced by minimizing the time that the resist is exposed to the electron beam and by operating at low-beam voltages.[25]

Care must be taken to compensate properly for the shrinkage. One method to characterize the shrinkage is to measure and remeasure the same feature repeatedly. When this is done, results similar to the curve in Fig. 9.8 are typically produced. As the resist is exposed initially to the electron beam the changes occur rapidly, but slow down after longer exposures. Correct compensation for resist shrinkage requires that this nonlinear behavior be included properly.

In addition to measuring linewidths, SEMs are used for measuring line-edge roughness (LER) and linewidth roughness (LWR). Although there have been efforts to develop other techniques,[27] currently only SEMs provide the capability for measuring LER and LWR, particularly when there is interest in the power spectral density of the roughness. Even when using a scanning

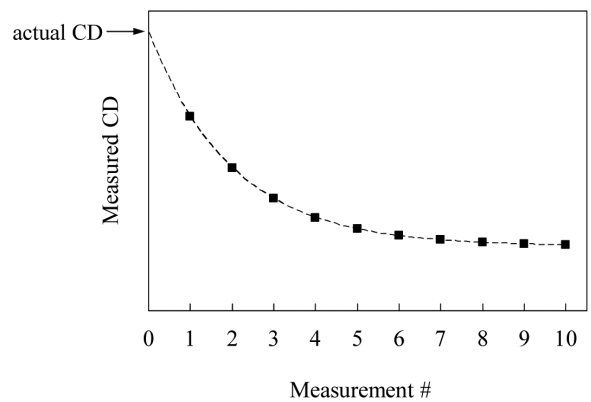


Figure 9.8 Exponential trend in measured critical dimension (CD) for ArF resists after successive measurements in a scanning electron microscope.[26]

electron microscope, there are details that need to be considered in order to get good measurements of line-edge roughness.[28] For example, suppose one wants to measure the power spectral density for a given resist. Prior to making the measurement, it is necessary to determine what domain of spatial frequencies is of interest. Low spatial frequencies are important, because they have big impact on linewidth control, while high-spatial-frequency roughness can affect electrical resistance of metal lines and reliability. Suppose that the data in a scanning electron micrograph are collected at N positions over a length L, where

$$L = \Delta N. \tag{9.5}$$

In this situation the maximum spatial frequency that can be measured is the Nyquist frequency (see Fig. 9.9):

$$f_{max} = \frac{1}{2\Delta}. \tag{9.6}$$

Equation (9.6) is used to determine the value for Δ, which is related to the resolution at which the SEM is operated. Similarly, the minimum measurable spatial frequency is

$$f_{min} \leq \frac{1}{L}. \tag{9.7}$$

In order to measure low spatial frequencies one needs to scan over long distances, which requires a large image field for the SEM. On the other hand, high resolution is usually achieved at high magnification and small image fields. Thus, there is a tradeoff in measurement between f_{max} and f_{min}.

9.1.2 Scatterometry

One configuration for the measurement of linewidth using scatterometry is illustrated in Fig. 9.10. A grating is patterned on the wafer. A beam of light is

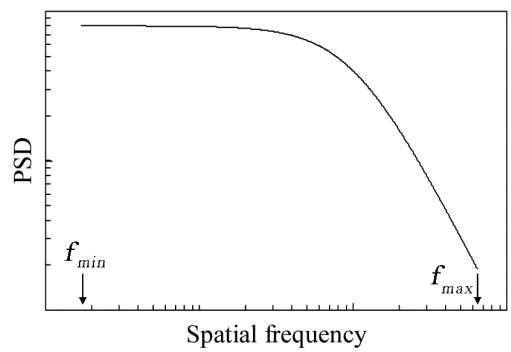


Figure 9.9 Illustration of the minimum and maximum spatial frequencies that can be measured.

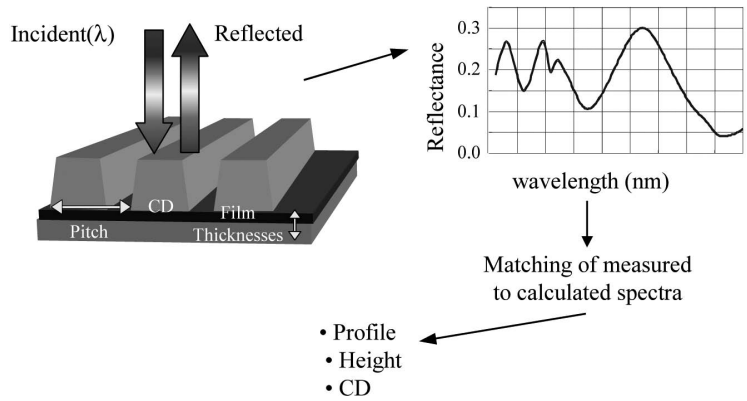


Figure 9.10 Configuration for scatterometry.

then reflected from the grating over a range of wavelengths, and the reflectance is measured as a function of wavelength. The reflectance versus wavelength is also calculated, with various assumed values for the printed pattern's critical dimension (CD) and resist profile, and for the thicknesses of the various films, including the patterned resist. The measured and calculated reflectances are compared for various assumed values of the linewidths, profiles, and film thicknesses, and a good match is considered as identification of these parameters. Another configuration is angle-resolved scatterometry, where the angle of incidence is varied, rather than the wavelength.[29] When wavelength is varied, the method is referred to as spectroscopic ellipsometry. It is also possible to make the measurements using an ellipsometer. In this case, measurements and calculations are performed for $\tan \psi$ and $\cos \Delta$.[30] Each of these techniques has its merits.[31]

Shown in Fig. 9.11 is an example of a ~160-nm resist linewidth measured with scatterometry. The measurement from scatterometry agreed well with the linewidth determined from a cross section measured with a scanning electron microscope. Scatterometry provided more information than possible from a top-down SEM. In addition to the linewidth, resist profile information was obtained. Remarkably, even the foot of the resist near the resist-substrate interface was captured by the ellipsometric measurement. The ability to capture the resist profile indicates the potential of scatterometry to identify out-of-focus exposure conditions.

Scatterometry is a measurement method somewhat complementary to measurement with scanning electron microscopes. Fairly large areas (at least 50 μm × 50 μm) are required to obtain an adequate signal. An array pattern is needed to obtain adequate signal-to-noise, but the pitch may be sufficiently loose for the lines to be effectively isolated from a lithography point of view. With some implementations of scatterometry it is necessary to prepare libraries of calculated curves in advance, so considerable setup is needed. Thus, scatterometry is best suited for use in manufacturing, although it has been used effectively to characterize exposure tools, since simple bare silicon substrates can be used that reduce setup complexity. It is not an appropriate method of measurement when one is attempting to verify OPC of individual transistors. Some attempts have

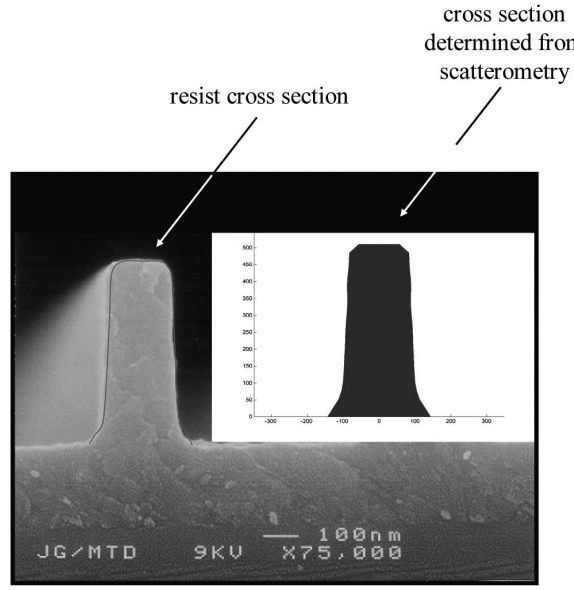


Figure 9.11 Example of a scatterometry linewidth measurement.

been made to apply scatterometry to measuring contacts, but the two-dimensional problem is computationally very challenging.

9.1.3 Electrical linewidth measurement

Linewidths are measured electrically as a supplement to SEM measurements. This method has the advantage of speed; considerable data are collected in a small amount of time once wafers have been patterned and etched. Offline stepper lens setup and characterization often requires that large numbers of linewidths be measured, and delays that are incurred because wafers must be etched are tolerable. For example, the determination of best focus and field tilt requires that linewidths be measured at many points in the field with features of at least two orientations, and at several focus settings. It is useful to measure linewidths in multiple fields in order to average out noise from factors not related to the lens and optical column tilt. It is not practical to measure many thousands of linewidths with a scanning electron microscope, even an automated one.

Electrical linewidth metrology has been used as a practical method for this type of setup.[32] Linewidths of electrically conducting material, such as polysilicon, silicide, or metal, can be measured using structures such as those shown in Fig. 9.12.[33] The electrical linewidth W is related to two measured resistances:

$$W = \frac{R_s}{R_b} L. \tag{9.8}$$

R_s is the sheet resistance and it is determined as follows: a current I is forced between pads 3 and 4, and a voltage V_{25}^{+} is measured between pads 2 and 5. The

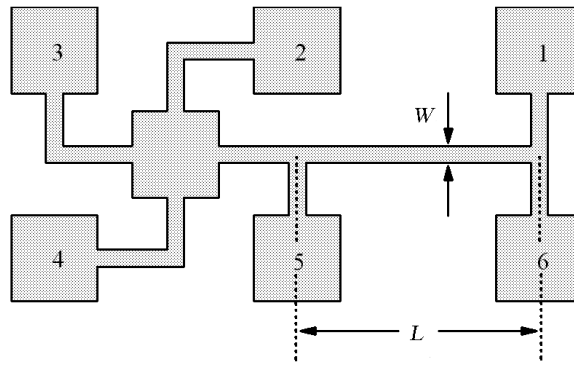


Figure 9.12 The resistor test structure for measuring linewidths electrically.

current is reversed, and a voltage V_{25}^{-} is measured. The current I is then forced alternately between pads 2 and 3, and voltages V_{45}^{+} and V_{45}^{-} are measured. The sheet resistance is given by

$$R_s = \frac{\pi}{\ln 2} \frac{\left|V_{25}^{+}\right| + \left|V_{25}^{-}\right| + \left|V_{45}^{+}\right| + \left|V_{45}^{-}\right|}{4I}. \tag{9.9}$$

R_b is determined by forcing current I_b between pads 1 and 3 and measuring the voltages between pads 5 and 6.

$$R_b = \frac{\left|V_{56}^{+}\right| + \left|V_{56}^{-}\right|}{2I_b}. \tag{9.10}$$

To avoid end effects, typically $W \ll L$.

Electrically measured linewidths, as determined from Eqs. (9.8)–(9.10), tend to disagree in absolute value with measurements from other metrology methods, such as SEMs and atomic force microscopes, and are generally smaller. This is often thought to be the result of nonconducting material on the sides of the pattern following etch. From the perspective of lithography, there is also a relevant etch bias between the resist and final dimensions, and nonuniformities in the etch may introduce across-wafer variations, which need to be taken into account. Despite these complications, electrical linewidth measurements are very useful for characterizing linewidth control across lenses, because the speed of measurement enables a lot of data to be collected.

The structure in Fig. 9.12 is used for measuring the widths of isolated lines. Additional structures are added to provide measurements of linewidths of grouped structures for any pitch that is desired. One can also include subresolution features on the mask, as appropriate. A method has even been developed for measuring contact sizes using electrical resistance measurements[34] by measuring the effective electrical linewidth W of a structure with holes (Fig. 9.13), compared to the electrical linewidth W_{ref} of an identical reference structure without holes. The

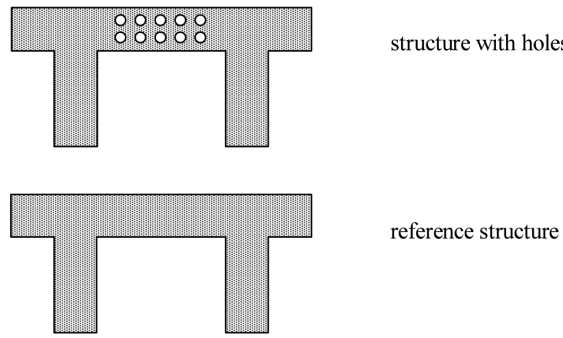


Figure 9.13 Configuration for measuring contact diameter electrically.

diameter d of the holes is given by

$$d = \frac{W_{ref}}{\sqrt{12\pi}} \left[\sqrt{1 + \frac{48}{N}\left(\frac{L}{W} - \frac{L}{W_{ref}}\right)} - 1 \right]^{\frac{1}{2}}, \tag{9.11}$$

where N is the number of contacts in the structure being measured.

For very small dimensions, one must be careful when interpreting the results of electrically measured linewidth data. Equations (9.8)–(9.11) are based on the assumption that the electrical conductivity of the material used is homogeneous. At very small dimensions this is no longer the case. Surface effects and inhomogeneous distributions of dopants become significant as dimensions become smaller than 0.25 μm.[36] For this reason, scatterometry, which has the ability to measure large numbers of linewidths quickly, has supplanted electrical linewidth measurements for exposure tool characterization. Moreover, scatterometry can provide information regarding resist slopes in addition to linewidths, which is useful for measuring defocus.

Electrical metrology can also be used for monitoring SEM stability.[35] Even though electrical linewidth measurements do not provide absolute measurements, a correlation between electrical and SEM measurements provides a method for checking the stability of the SEMs through comparison with electrical measurements.

9.2 Measurement of Overlay

Overlay is usually measured using optical-linewidth-measurement systems. Consider the structure shown in Fig. 9.14. The linewidths x_1 and x_2 can be used to calculate the overlay error in the x direction:

$$\Delta x = \frac{1}{2}(x_2 - x_1). \tag{9.12}$$

The features shown in Fig. 9.14 are large enough that optical-linewidth-measurement systems can be used because of their throughput advantage over

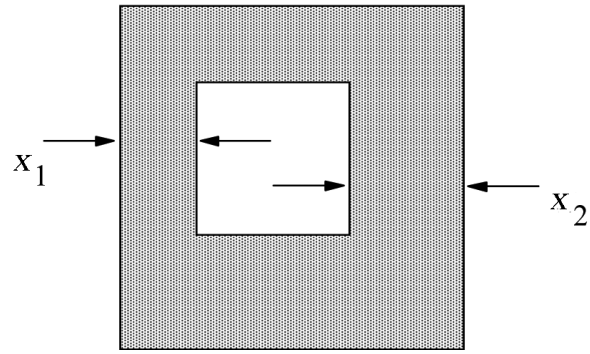


Figure 9.14 A structure for measuring overlay. The inner box is part of the substrate pattern, and the outer box is created in the overlaying photoresist layer.

SEMs. Sufficient data are collected automatically in order to perform the overlay modeling discussed in Chapter 6.

Overlay measurement can falsely introduce apparent overlay errors. The most common of these errors are discussed in this section. This error is called tool-induced shift (TIS), and results from asymmetries in the optics of the overlay measurement system. Because the materials of Layer 1 and Layer 2 in Fig. 9.14 are not the same, the measurements of x_1 and x_2 are not equivalent if there are asymmetries in the measurement equipment. Types of asymmetries that occur are tilted or decentered lenses, nonuniform illumination, lens aberrations, and nonuniform detector response.[37] The presence of TIS is easily verified by rotating wafers from 0 to 180 deg.[38,39] In the absence of TIS,

$$\Delta x_{(0)} \rightarrow -\Delta x_{(180)}, \tag{9.13}$$

where $\Delta x_{(0)}$ is the measurement for the unrotated wafer, and $\Delta x_{(180)}$ is the measurement for the wafer rotated 180 deg. A measure of TIS is therefore

$$TIS = \Delta x_{(0)} + \Delta x_{(180)}, \tag{9.14}$$

which is ideally zero. The TIS value is measured when overlay measurement programs are first established for particular process layers, and automatic compensation is made to reduce the TIS error.[40] Asymmetries in alignment measurement can also be induced when features are too close to each other (x_1 or x_2 is too small), relative to the resolving power of the optics in the measurement tool.[41] The resolution of optical tools using visible light is typically in the range of 0.8–1.0 μm, while systems using ultraviolet light have somewhat better resolution. The features of overlay measurement structures on the wafer must be consistent with the resolution of the optics of the overlay measurement tool.

Wafer processing induces asymmetries in overlay measurement marks. Consider situations where metal is sputtered onto wafers. If the deposition is not collimated, metal builds up preferentially on one side of the overlay measurement marks near

the edges of the wafers (Fig. 6.24). Measurements of overlay are shifted relative to the overlay of the critical features. Errors in overlay measurement caused by wafer processing are referred to as wafer-induced shift (WIS). For the example illustrated in Fig. 6.24, the overlay error is in opposite directions on opposite sides of the wafer and appears to be a wafer scaling error.[42] To the extent that asymmetries in step coverage are repeatable, wafer-induced shifts are corrected by comparing overlay measurements before and after etch.[43]

Chemical-mechanical polishing (CMP) causes problems for correct acquisition of alignment targets (Chapter 6) and also creates difficulties for overlay measurement. Overlay measurement structures need to be optimized to minimize this effect. The bar or frame structures shown in Fig. 9.15 typically lead to less of a problem than the box-in-box structures, though the exact dimensions need to be determined for individual processes. More recently, overlay measurement structures have been developed that involve grating patterns. An example is shown in Fig. 9.16.[44] These new patterns are less sensitive to degradation by CMP. Such grating marks also have the potential for being made small ($\leq 5\ \mu m \times 5\ \mu m$), since gratings provide a high ratio of line edge to area, which can be useful for measuring overlay within product die.[45]

Another motivation for using small marks is the increase in the number of masking steps that has occurred over time. For many years only two metal layers were used to fabricate integrated circuits, while eight or more metal layers are common today. Many implant layers are required to produce the transistors for high-performance microprocessors. As a consequence, even scribe line space has become insufficient for including the targets required to measure overlay for all of the layers. This has motivated the development of overlay measurement marks that can be used to measure overlay among many layers while being space efficient.[46,47]

As discussed in Chapter 6, intrafield registration depends on the illumination conditions and the feature type. Lines and spaces may have different registration at the different illumination settings. This causes a problem for metrology if the critical features are one type, such as contact holes, while other types of features, such as lines, are used for overlay measurement.[48] This type of subtle difference is significant as overlay requirements approach 50 nm or less.

Overlay measurement errors can occur simply because the linewidths of the measurement structure differ from those in the circuit. Because optical methods

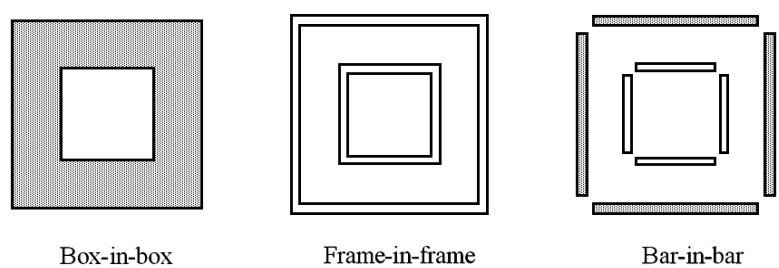


Figure 9.15 Various structures for measuring overlay.

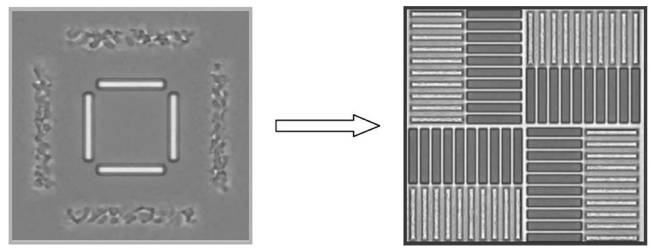


Figure 9.16 On the left is a traditional frame-in-frame measurement structure, damaged by chemical-mechanical polishing, because the width of the frame feature is much larger than the features in the circuit, for which the CMP process is optimized. The new type of overlay measurement structure[44] more closely replicates the features in the circuit.

are used, the feature sizes for these structures have been fairly large, on the order of 1 μm or larger, even when the features in the devices have been submicron. These differences in feature sizes between the overlay measurement structures and features actually in the circuits were of minor consequence prior to the advent of lithography with low k_1. With low-k_1 processes, the placement errors of small features due to lens aberrations will be different from the placement errors of larger features (Fig. 9.17). Without a change in the overlay measurement structures, there will be overlay errors as a consequence.

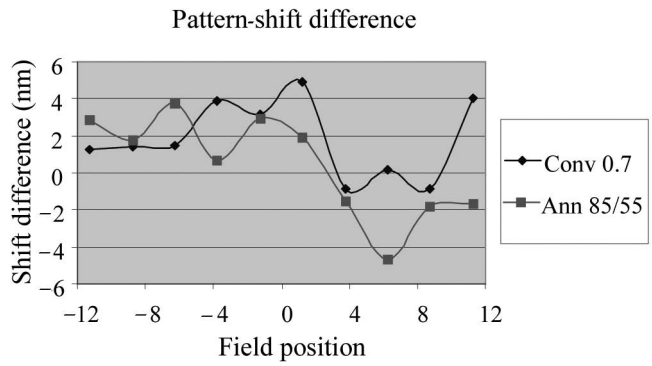


Figure 9.17 Pattern shift of 120-nm line/360-nm pitch features compared to 1-μm features. These were calculated errors based upon measured aberrations of an ArF lens.

Problems

9.1 What are advantages and disadvantages of scanning electron microscopes for measuring critical dimensions, relative to scatterometry tools?

9.2 Why are low-beam voltages (a few hundred eV or less) used when applying scanning electron microscopes for measuring dimensions of features composed of photoresist?

References

1. K. M. Monahan, M. Davidson, Z. Grycz, R. Krieger, B. Sheumaker, and R. Zmrzli, "Low-loss electron imaging and its application to critical dimension metrology," *Proc. SPIE* **2196**, 138–144 (1994).
2. L. Reimer, *Image Formation in Low-voltage Scanning Electron Microscopy*, SPIE Press, Bellingham, WA (1993).
3. R. R. Hershey and M. B. Weller, "Nonlinearity in scanning electron microscope critical dimension measurements introduced by the edge detection algorithm," *Proc. SPIE* **1926**, 287–294 (1993).
4. J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*, 2nd ed., Plenum Press, New York (1984).
5. J. Finders, K. Ronse, L. Van den Hove, V. Van Driessche, and P. Tzviatkov, "Impact of SEM accuracy on the CD-control during gate patterning process of 0.25 μm generations," *Proc. Olin Microlithog. Sem.*, 17–30 (1997).
6. K. M. Monahan, J. P. H. Benschop, and T. A. Harris, "Charging effects in low-voltage SEM metrology," *Proc. SPIE* **1464**, 2–9 (1991).
7. D. C. Joy and C. S. Joy, "Low voltage scanning electron microscopy," *Micron* **27**(3–4), 247–263 (1996).
8. M. Davidson and N. T. Sullivan, "An investigation of the effects of charging in SEM based CD metrology," *Proc. SPIE* **3050**, 226–242 (1997).
9. C. M. Cork, P. Canestrari, P. DeNatale, and M. Vasconi, "Near and sub-half micron geometry SEM metrology requirements for good process control," *Proc. SPIE* **2439**, 106–113 (1995).
10. S. R. Rogers, "New CD-SEM technology for 0.25 μm production," *Proc. SPIE* **2439**, 353–362 (1995).
11. D. A. Swyt, "Certificate of analysis, standard reference material 2800," National Institute of Standards and Technology, (2002).
12. NIST Standard Reference Materials Catalog, National Institute of Standards and Technology, U.S. Department of Commerce. www.nist.gov/srm.
13. M. T. Postek, A. E. Vladar, S. Jones, and W. J. Keery, "Report on the NIST low accelerating voltage SEM magnification standard interlaboratory study," *Proc. SPIE* **1926**, 268–286 (1993).
14. M. T. Postek, "Scanning electron microscope-based metrological electron microscope system and new prototype scanning electron microscope magnification standard," *Scanning Microsc.* **3**(4), 1087–1099 (1989).
15. M. T. Postek, and R. L. Watters, "Report of investigation reference material 8820," National Institute of Standards and Technology (2009).
16. W. P. Reed, "Certificate standard reference material 2069b," National Institute of Standards and Technology (1991).

17. Y. Nakayama and K. Toyoda, "New submicron dimension reference for electron-beam metrology system," *Proc. SPIE* **2196**, 74–84 (1994).

18. B. L. Newell, M. T. Postek, and J. P. van der Ziel, "Performance of the protoype NIST SRM 2090A SEM magnification standard in a low-accelerating voltage SEM," *Proc. SPIE* **2439**, 383–390 (1995).

19. M. Sendlebach and C. Archie, "Scatterometry measurement precision and accuracy below 70 nm," *Proc. SPIE* **5038**, 224–238 (2003).

20. T. W. Reilly, "Metrology algorithms for machine matching in different CD SEM configurations," *Proc. SPIE* **1673**, 48–55 (1992).

21. K. Phan, J. Nistler, and B. Singh, "Metrology issues associated with submicron linewidths," *Proc. SPIE* **1464**, 424–437 (1991).

22. E. E. Chain, M. G. Ridens, and J. P. Annand, "SPC qualification strategy for CD metrology," *Proc. SPIE* **2876**, 218–224 (1996).

23. D. A. Chernoff, "Atomic force microscope (AFM) analysis of photoresist test structures for use in SEM as in-house linewidth standards," *Proc. SPIE* **2439**, 392–400 (1995).

24. M. T. Postek, A. E. Vladar, and M. P. Davidson, "Fourier transform feedback tool for scanning electron microscopes used in semiconductor metrology," *Proc. SPIE* **3050**, 68–79 (1997).

25. B. Bunday, A. Cordes, N. G. Orji, E. Piscani, D. Cochran, J. Byers, J. Allgair, B. J. Rice, Y. Avitan, R. Peltinov, M. Bar-Zivi, and O. Adan, "Characterization of CD-SEM metrology for iArF photoresist materials," *Proc. SPIE* **6922**, 69221A (2008).

26. A. Habermas, D. Hong, M. Ross, and W. Livesay, "193 nm CD shrinkage under SEM: modeling the mechanism," *Proc. SPIE* **4689**, 92–101 (2002).

27. C. Wang, R. L. Jones, E. K. Lin, W. Wu, J. S. Villarrubia, K. Choi, J. S. Clarke, B. J. Rice, M. Leeson, J. Roberts, R. Bristol, and B. Bunday, "Line edge roughness characterization of sub-50 nm structures using CD-SAXS: round-robin benchmark results," *Proc. SPIE* **6518**, 65181O (2007).

28. B. D. Bunday, M. Bishop, D. McCormack, J. S. Villarrubia, A. E. Vladár, T. Vorburger, N. G. Orji, and J. A. Allgair, "Determination of optimal parameters for CD-SEM measurement of line edge roughness," *Proc. SPIE* **5375**, 515–533 (2004).

29. J. Allgair, D. Benoit, R. Hershey, L. C. Litt, I. Abdulhalim, B. Braymer, M. Faeyrman, J. C. Robinson, U. Whitney, Y. Xu, P. Zalicki, and J. Seligson, "Manufacturing considerations for implementation of scatterometry for process monitoring," *Proc. SPIE* **3998**, 125–134 (2000).

30. N. Jakatdar, X. Niu, J. Bao, C. Spanos, S. Yedur, and A. Deleporte, "Phase profilometry for the 193 nm lithography gate stack," *Proc. SPIE* **3998**, 116–124 (2000).

31. P. C. Logoftu and J. R. McNeil, "Measurement precision of optical scatterometry," *Proc. SPIE* **4344**, 447–453 (2001).
32. L. J. Zych, G. Spadini, T. F. Hassan, and B. A. Arden, "Electrical methods for precision stepper column optimization," *Proc. SPIE* **633**, 98–105 (1986).
33. L. W. Linholm, R. A. Allen, and M. W. Cresswell, "Microelectronic test structures for feature placement and electrical linewidth metrology," in *Handbook of Critical Dimension Metrology and Process Control*, K. M. Monahan, Ed., SPIE Press, Bellingham, WA (1993).
34. B. J. Lin, J. A. Underhill, D. Sundling, and B. Peck, "Electrical measurement of submicrometer contact holes," *Proc. SPIE* **921**, 164–169 (1988).
35. J. Kye and H. Levinson, "Electrical linewidth measurement for next-generation lithography," *Proc. SPIE* **4344**, 637–643 (2001).
36. E. E. Chain and M. Griswold, "In-line electrical probe for CD metrology," *Proc. SPIE* **2876**, 135–146 (1996).
37. R. M. Silver, J. Potzick, and R. D. Larrabee, "Overlay measurements and standards," *Proc. SPIE* **3429**, 262–272 (1995).
38. D. J. Coleman, P. J. Larson, A. D. Lopata, W. A. Muth, and A. Starikov, "On the accuracy of overlay measurements: tool and mark asymmetry effects," *Proc. SPIE* **1261**, 139–161 (1990).
39. A. Starikov, D. J. Coleman, P. J. Larson, A. D. Lopata, and W. A. Muth, "Accuracy of overlay measurements: tool and mark asymmetry effects," *Opt. Eng.* **31**, 1298–1310 (1992).
40. M. E. Preil, B. Plambeck, Y. Uziel, H. Zhou, and M. W. Melvin, "Improving the accuracy of overlay measurements through reduction in tool and wafer induced shifts," *Proc. SPIE* **3050**, 123–134 (1997).
41. N. Smith, G. Goelzer, M. Hanna, and P. Troccolo, "Minimizing optical overlay measurement errors," *Proc. SPIE* **1926**, 450–462 (1993).
42. J.-S. Han, H. Kim, J.-L. Nam, M.-S. Han, S.-K. Lim, S. D. Yanowitz, N. P. Smith, and A. M. C. Smout, "Effects of illumination wavelength on the accuracy of optical overlay metrology," *Proc. SPIE* **3051**, 417–425 (1997).
43. Y. Tanaka, M. Kamiya, and N. Suzuki, "New methodology of optimizing optical overlay measurement," *Proc. SPIE* **1926**, 429–439 (1993).
44. J. Allgair and K. Monahan, "Microeconomics of overlay control at the 65 nm technology node," *IEEE Int. Sym. Semicond. Manuf. Conf. Proc.*, 103–106 (2003).
45. S. Girol-Gunia, B. Schulz, N. Smith, and L. Binns, "Using in-chip overlay metrology," *Proc. SPIE* **6922**, 69220N (2008).
46. C. P. Ausschnitt, W. Chu, D. Kolor, J. Morillo, J. L. Mornngstar, W. Muth, C. Thomison, R. J. Yerdon, L. A. Binns, P. Dasari, H. Fink, N. P. Smith, and G. Ananew, "Blossom overlay metrology implementation," *Proc. SPIE* **6518**, 65180G (2007).

47. C. P. Ausschnitt and S. D. Halle, "Combinatorial overlay control for double patterning," *J. Micro/Nanolith. MEMS MOEMS* **8**(1), 011008 (2009).

48. T. Saito, H. Watanabe, and Y. Okuda, "Overlay error of fine patterns by lens aberration using modified illumination," *Proc. SPIE* **3051**, 687–696 (1997).

Chapter 10
Immersion Lithography and the Limits of Optical Lithography

In Chapter 2, where the parameter *numerical aperture* was introduced, it was defined as $n \sin\theta$, where n is the index of refraction of the medium between the lens and the resist on the wafer, and θ is the half angle subtended by the lens. By geometry, and the fact that light travels in a straight line through homogeneous media, the largest value that θ can reach is 90 deg, so that $\sin\theta$ must be less than 1.0. The medium between the lens and the resist is typically air, for which $n \approx 1$. This implies that the numerical aperture cannot exceed 1.0 in air, creating a limit on resolution for a given wavelength.

It has long been known that resolution in optical microscopy can be enhanced by immersing microscope objective lenses in a liquid that has an index of refraction greater than 1.0. This is not something that can be accomplished by using optics that have been designed for imaging in air, but requires lenses that have been designed specifically for operation using an immersion fluid with a well-defined index of refraction. Nevertheless, such optics have been exploited successfully in optical microscopy for over a century.[1]

The same physics can be used to advantage in lithography. Since $\sin\theta$ remains constrained by geometry to being less than 1.0, the index of refraction of the medium between the lens and the wafer represents the theoretical upper limit for the numerical aperture. At a wavelength of 193 nm, ultrapure water has an index of refraction of 1.437.[2] Water is also very transparent at ArF wavelengths, so it is a suitable fluid for ArF immersion lithography, where water fills the space between the bottom of the lens and the wafer. The use of water immersion leads to a significant increase in the numerical aperture over the maximum value possible when imaging in air, and hence greatly extends the resolution capability of ArF lithography. With water as an immersion fluid instead of air, the physical limit for numerical apertures increases from 1.0 to 1.437. In this chapter the basic concepts of immersion lithography will be discussed. The remainder of the chapter will then be concerned with what other opportunities there might be to extend optical lithography.

10.1 Immersion Lithography

While conceptually simple, and despite its long use in microscopy, the application of immersion imaging to lithography is not straightforward. There are a number of issues that are unique to lithography. For example, for the reasons described in Chapter 5, state-of-the-art exposure tools are scanners, so the immersion imaging must be accomplished while the wafer is moving relative to the lens. This is one reason that immersion imaging is more complex in application to lithography than to optical microscopy. The immersion fluid can also exert forces against the lens during scanning, and deformation of the lens—which will induce aberrations—must be avoided. Immersion scanners must be designed to ensure that the immersion fluid remains sufficiently homogeneous during scanning; there can be no significant compression or heating of the fluid, for example.

A configuration in which the immersion fluid is contained locally between the projection optics and the wafer has proven most suitable for use in scanners. This is shown schematically in Fig. 10.1. An alternative design, in which the entire stage is immersed in the fluid, was considered by suppliers of exposure tools, but high-speed scanning proved difficult in that configuration.

In addition to providing the potential for increased resolution, immersion lithography can increase the depth-of-focus at any given resolution, relative to imaging in air. An expression was given in Chapter 2 for the depth-of-focus for small to moderate values of NA and for imaging in air. A more general expression for the depth-of-focus is derived as follows. Consider the situation shown in Fig. 10.2. Light is imaged by a diffraction-limited lens. In the plane of best focus, all rays are in phase. For the spherical wavefront shown in Fig. 10.2, the zeroth- and first-order rays have a phase difference in the z direction of

$$n\Delta z\,(1-\cos\theta)\,\frac{2\pi}{\lambda}, \tag{10.1}$$

where n is the index of refraction of the medium in which image formation is occurring. If the edge of the focus window is said to occur when the phase difference between the zeroth and first-order rays is $\pi/2$, and with

$$NA = n\sin\theta, \tag{10.2}$$

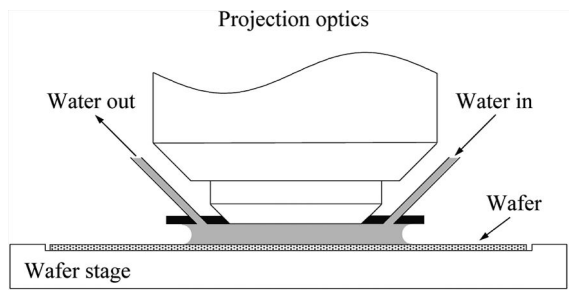


Figure 10.1 Illustration of local containment of the immersion fluid (water in this instance).

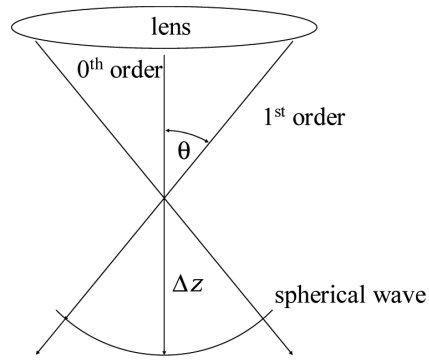


Figure 10.2 Illustration of the depth-of-focus for general numerical aperture.

then the depth-of-focus is given by

$$DOF = 2\Delta z \tag{10.3}$$

$$= \frac{\lambda}{2n\left(1 - \sqrt{1 - \dfrac{NA^2}{n^2}}\right)}. \tag{10.4}$$

For small values of θ, this reduces to

$$\frac{\lambda}{2n\left(1 - \sqrt{1 - \dfrac{NA^2}{n^2}}\right)} \approx \frac{\lambda}{n \sin^2 \theta} \tag{10.5}$$

$$= \frac{n\lambda}{NA^2}. \tag{10.6}$$

As can be seen, the exact expression for depth-of-focus reduces to the expression given in Chapter 2 for small NAs and $n = 1$.

There are several things to be noted from these expressions. First, the exact depth-of-focus is actually less than the depth-of-focus given by the expression that is suitable for low NA. This means that analyses which indicated difficulties with small depths-of-focus for large-NA lithography were actually optimistic! While these differences between Eqs. (10.6) and (10.4) are small for low NAs, the disparity at high NA can be as much as 50%. It can also be seen from Eq. (10.6) that the depth-of-focus is improved by a factor of n or larger. To the extent that optical lithography is limited by the depth-of-focus, immersion can provide improvement.

Silicon wafers will heat during exposure. Modeling has predicted a temperature rise of ~16 K for localized regions of the wafer.[3] Although only some of this heat will be transferred to the immersion fluid, the fluid's index of refraction will be altered. Careful measurements have shown that the index of refraction of water will change by about $-1 \times 10^{-4}/^{\circ}\mathrm{C}$ for ArF light,[2] so only very small increases in the

water's temperature can be tolerated. Fortunately, the heat appears to be confined to a volume of water within 200 nm of the wafer surface[4] during exposure. Suppose that we are considering a lens with a numerical aperture of 1.35. In this case,

$$NA = n \sin \theta \tag{10.7}$$

$$= 1.35. \tag{10.8}$$

Thus, $\theta = 70$ deg assuming $n = 1.437$, the value appropriate for water at a wavelength of 193 nm. The path through which the most oblique angles (see Fig. 10.3) will be incident will have a path length through the volume of heated water of

$$\frac{200 \text{ nm}}{\cos \theta} = 584 \text{ nm}. \tag{10.9}$$

Consequently, a 1 °C temperature rise within this volume will create a phase error of

$$n\,(1\ ^\circ\text{C})\left(-1 \times 10^{-4}/^\circ\text{C}\right) 584 \text{ nm} = -0.0839 \text{ nm}. \tag{10.10}$$

Since phase errors from other aberrations are typically ~1 nm,[5] the additional phase errors induced by water heating are not considered critical during the exposure of a single field. Over the course of exposing an entire wafer, the average water temperature could rise, so the immersion fluid must be circulated, and temperature must be controlled actively.

Heating of the immersion fluid due to optical absorption is not the only temperature concern associated with immersion lithography. Even with good fluid containment, small amounts of immersion fluid—perhaps just monolayers—will remain on the wafer surface. Subsequent evaporative cooling will cause the wafer to contract, affecting overlay (see Problem 10.4). For this reason, considerable engineering resources have been applied to improving overlay on immersion exposure tools.

Sensitivity to defects is considerably greater in application to semiconductor immersion lithography than it would be in microscopy. Not only must there be nearly no defects in lithographic applications, this must be the case over very large areas, in contrast to microscopy, where the fields of view are often only microns on a side, or less. The issue of defects is a critical one for immersion lithography.

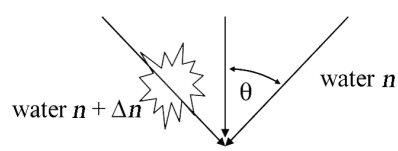


Figure 10.3 Light passing through water and converging on a wafer.

The introduction of immersion lithography has led to an increased focus on wafer-edge defects in particular. This can be understood by looking at Fig. 10.4. The immersion fluid extends well beyond the area being exposed. As a consequence, the fluid can pick up particles from the wafer edge or wafer stage. Because wafers are beveled, particles from the back side of the wafer can even be picked up from fluid in the small gap between the wafer and the chuck. To address scanning at the wafer edge, the wafer is recessed into the chuck so that the tops of the wafer and the stage are at the same level (see Fig. 10.5). Fluid will go into the small gap between the wafer and the chuck and pick up particles. These particles can be deposited on the wafer, reducing yield. Keeping wafers clean on the back side and edges is important for immerson lithography.

The resist is another potential source of defects. Chemicals can be leached from the resist into the immersion fluid. This is a problem particularly if drops of water are left behind as the wafer is scanned. Components essential to functionality can be pulled from the resist, leading to a spot where the resist does not perform properly. After the water has evaporated, material can be left behind, also leading to the formation of defects. An example of what can result from a water droplet is shown in Fig. 10.6. There is greater potential for the water not to be fully contained

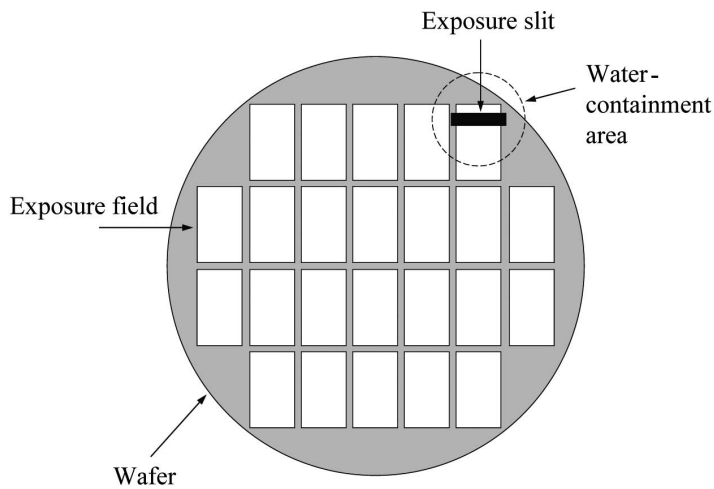


Figure 10.4 Top view of immersion-scanning lithography, illustrating how water will be swept off the side of the wafer and onto the wafer stage.

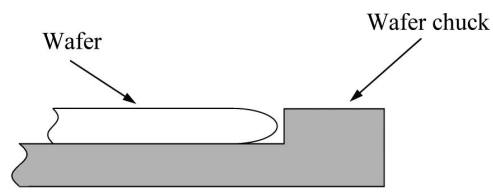


Figure 10.5 Cross-sectional side view of the wafer on the wafer chuck of an immersion scanner.

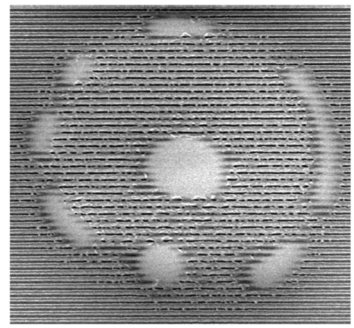


Figure 10.6 Scanning-electron micrograph of a defect resulting from a water drop. The pattern was supposed to consist purely of lines and spaces.

at higher scanning speeds, and this leads to another set of engineering problems for achieving high scanner throughput. The potential for material leaching from the resist also leads to concern over contamination of the bottom lens element.

Two approaches have been taken to address the problem of chemicals leaching from the photoresist. One is to use a topcoat. As discussed in Section 3.5, topcoats have long been used as barriers to base diffusion in chemically amplified resists. For process simplicity, these topcoats are typically water soluble so they can be readily removed prior to resist development, but this property makes them unsuitable for use in water-immersion lithography. Two different types of topcoats have been used in immersion lithography. An early immersion-compatible topcoat required a solvent to be removed, but this was undesirable because of process cost. Later topcoats were soluble in developer but not pure water. The most preferred solution is to have immersion resists that intrinsically leach only very small amounts of chemicals into water.

Bubbles constitute another potential source of defects in immersion lithography because they can scatter light.[6–9] One typically thinks of particles in terms of dense objects, but light can be scattered by objects of low index of refraction, such as bubbles, that are embedded in media of higher optical density.[9] Fortunately, bubble formation and dissipation have been studied extensively, and it has been found that the gases that comprise air are very soluble in water. Once the water has been degassed, i.e., the air dissolved in it has mostly been removed, bubbles have short lifetimes. Bubble lifetime has been predicted theoretically[10] and measured in confirmation of the theory. Results from both are shown in Fig. 10.7. As can be seen from the data, bubbles have very short lifetimes in degassed water; consequently, light scattering due to bubbles is expected to be small. However, even if there are no bubbles, care must be taken to control the amount of air dissolved in the water, since air can affect the index of refraction at a level significant for lithography. A difference in the refractive index at $\lambda = 193$ nm between air-saturated and completely degassed water has been measured to be 6.7×10^{-6}.[11] Light passing through millimeters of water will have differences in phase on the order of nanometers, depending upon whether the water is air saturated or degassed.

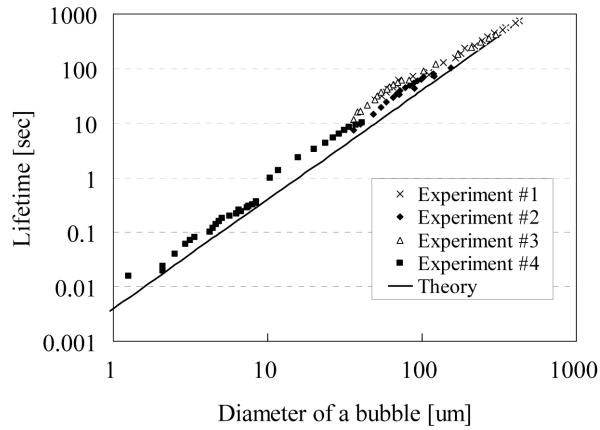


Figure 10.7 Bubble lifetime measurement and theoretical prediction.[12]

It should be noted that water can be used for KrF immersion, at least from an optical point of view. However, the index of refraction of water is only 1.378 at KrF wavelengths, so water is less effective at improving KrF than ArF lithography. Also, it appears that many existing KrF resist platforms do not work well when the resists are immersed in water. Consequently, there has been little effort to develop KrF immersion lithography. At $\lambda = 193$ nm, water immersion provides a significant extension of optical lithography, enabling >40% increase in resolution over imaging in air. How far optical lithography can ultimately be extended is the subject of the remainder of this chapter.

10.2 The Diffraction Limit

In 1979, *Electronics* magazine reported that stepper lithography would be a passing fancy superseded by direct-write electron-beam lithography by the year 1985.[13] It was admitted in a follow-up article, written for that same magazine in 1985, that the demise of optical lithography had been predicted prematurely and that it would take until 1994 for shipments of optical wafer steppers to be of lower volume than those of x-ray step-and-repeat systems.[14] It was expected that optical lithography, once it reached its resolution limit of 0.5 μm, would need to be replaced. Both pronouncements were based upon accepted expert opinion. This most recent edition of this book was written in the year 2010, with optical lithography still going strong. Clearly, accepted expert opinion is not a criterion for validity. Nevertheless, there are some signs that optical lithography is at last reaching its limit, and there are several programs dedicated to developing alternative lithography techniques. It is worth reviewing those earlier arguments that predicted that optical lithography was nearing its end of life and those arguments that are being presented today to justify billions of dollars of investment in new lithography techniques.

The argument that optical lithography has limited resolution is based upon Rayleigh's scaling laws of resolution and depth-of-focus. From Chapter 2,

resolution is given by

$$\text{resolution} = k_1 \frac{\lambda}{NA}, \tag{10.11}$$

where the prefactors of Eqs. (2.4) and (2.7) are replaced by a general factor k_1. Similarly, the expression for depth-of-focus can be written as

$$\text{depth-of-focus} = \pm k_2 \frac{\lambda}{NA^2}. \tag{10.12}$$

It has long been recognized that Rayleigh's and equivalent expressions are inexact predictors of resolution, but do correctly capture the trends associated with wavelengths and numerical apertures. Other factors, such as the resist process, are captured by the coefficients k_1 and k_2. In 1979, the state-of-the-art lens had a resolution of 1.25 μm, a ±0.75-μm depth-of-focus, a numerical aperture of 0.28, and was imaged at the mercury g line (λ = 436 nm). This produced values of 0.80 and 0.13 for k_1 and k_2, respectively. With these values for the coefficients in Eqs. (10.11) and (10.12), the numerical aperture of a g-line lens capable of producing 0.8-μm features would be 0.44, with a ±0.3-μm depth-of-focus. In 1979, long before the use of chemical-mechanical polishing in device fabrication, this was too small a depth-of-focus to provide adequate imaging over the heights of then-typical device topographies and within the ability of steppers to control focus. Thus, it was declared that optical lithography would not be capable of submicron lithography.

There were several mistakes in this argument that were already becoming apparent by the mid-1980s.[15]

1. These early predictions were wrong, because they assumed that the optics found on state-of-the-art steppers in the late 1970s and early 1980s were nearly diffraction limited, and extrapolations were based upon that assumption. The extent to which these assumptions were off the mark has been demonstrated by the application of today's superior lens-designing methods and manufacturing capability. Until recently, these capabilities have been applied only to leading edge, high-numerical-aperture lenses, but with the introduction of very large field systems for mix-and-match applications, actual diffraction limits for smaller numerical apertures can be observed. This enables a direct reassessment of earlier analyses. For example, Nikon produced a 0.3-NA i-line lens for its 4425i stepper. This lens had a resolution of 0.7 μm (k_1 = 0.58) and specified ±2.5-μm depth-of-focus (k_2 = 0.62), while thirty years ago, lenses (g line) with similar numerical apertures had 1.1- to 1.25-μm resolution and ±0.75-μm depths-of-focus. By extrapolating the future of optical lithography from these two different sets of capabilities—old and new—one arrives at significantly different conclusions.
2. It was wrong to assume that the mercury g line would always be used for optical lithography. For a given feature size, depth-of-focus is increased by

using a smaller wavelength. The progression from the 436-nm (mercury g-line) wavelength to the 365-nm (i-line) wavelength, and onto 248-nm (KrF) and 193-nm (ArF) lithography, has enabled feature sizes to shrink while maintaining a useable depth-of-focus.

3. Photoresists have improved, contributing to effective changes to k_1 and k_2. Lithography today routinely operates at values of $k_1 = 0.35$ and smaller. The argument of 1979, revised with a k_1 value of 0.5, predicts that 0.8-μm resolution could be achieved with a numerical aperture of 0.27 with a g-line lens. Even with the previously predicted value for k_2 of 0.13, the resulting depth-of-focus would have been more than adequate.
4. Depth-of-focus requirements have decreased, primarily as a result of the use of chemical-mechanical polishing. Because it is no longer necessary to image through the depth of device topography, good images can be obtained on optical systems with small depths-of-focus. Depths-of-focus of only ±0.1 μm and smaller are now considered adequate.

In addition to the reasons already identified in the 1980s, more recently developed resolution-enhancement techniques, such as phase shifting, off-axis illumination, and the other methods presented in Chapter 8, can further extend optical lithography beyond the limits that were inferred from scaling using the Rayleigh criteria. The field of "wavefront engineering," an umbrella term coined to include all the techniques that seek to modify the effects of diffraction in order to improve resolution and depth-of-focus,[16] addressed many of the challenges at the 100-nm level and below, forestalling the long-forecasted "death of optics."

Regardless of the errors in past predictions, the scaling laws for resolution and depth-of-focus indicate that arbitrarily small features will not be made optically. The issue has never been whether there is a limit to optical lithography; the questions have been of when the end will come, and what the ultimate resolution will be. It is clear that we need to look beyond the basic Rayleigh criteria. Wavefront engineering has benefits for specific feature types, dependent on the optical configuration of stepper and mask. It is important to understand the applicability of these sophisticated enhancement techniques to obtain an estimate of optical lithography's ultimate limits.

It is useful to define more precisely what is meant by "optical lithography." For the purposes of this book, optical lithography is defined as any lithographic technique that

1. Uses photons to induce chemical reactions in a photoresist,
2. Involves a transmission photomask, and
3. Has the potential for image reduction using projection optics.

These definitions help to distinguish optical lithography from other patterning techniques that do not involve photons, or do use photons but are very different in character from the techniques described thus far in this book. The last two requirements separate x-ray and extreme ultraviolet (EUV) lithography from optical lithography, even though these two other types of lithography involve

photons. Several reasons were noted above as to why optical lithography has survived longer than earlier anticipated. It is worth trying to understand whether there are opportunities remaining for improvement along the same paths used to extend optical lithography beyond past expectations, and how much extension will be provided by recently developed techniques.

10.3 Improvements in Optics

As mentioned above, early stepper lenses were far from being diffraction limited, and much of the progress in optical lithography has been due to improvements in the optics. This has included increases in numerical apertures and substantial reductions in the aberrations. Early stepper lenses had high levels of aberrations, by contemporary standards. The severe astigmatism shown in Fig. 2.27 was an example from a Zeiss 10-78-37, 5× lens widely used in the mid-1980s. The wavefront errors from lenses in that era were on the order of 0.1 waves rms.[17] This can be contrasted with the <0.006 wave error typical for contemporary lenses.[5] Given the transition from 436-nm to 193-nm light that has occurred, this is nearly a 40× improvement in absolute wavefront error. It is clear that there has been substantial improvement in lens quality, and this has been manifested as better imaging capability. Unfortunately, further improvements in lens quality do not produce imaging capabilities tending toward infinitely narrow lines, but will bring imaging performance closer to the diffraction limit. While process-window detractors, such as lens aberrations, can be reduced, the limitation imposed by diffraction remains and becomes dominant. The difference between converging toward zero linewidth and the diffraction-limited linewidth is illustrated in Fig. 10.8. The solid curve represents 30% improvement between generations (0.7× scaling), starting with 1.0 μm in the first generation, while the dashed curve is a 30% improvement, *relative to a 40-nm limit* (see Problem 10.3). For early generations, scaling to infinitely small dimensions and the 40-nm limitation were indistinguishable, but the 40-nm limit becomes significant for linewidths less than 0.2 μm. Future reductions in lens aberrations bring capability closer to the asymptote that is the diffraction limit. Another 40× improvement in lens aberrations will not extend optical lithography to the same degree brought about by the last 40× improvement.

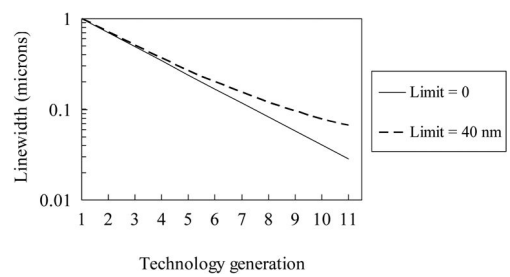


Figure 10.8 Progress in lithography with 0.7× improvement between generations, relative to a limit of zero, or a limit of 40 nm.

Lenses have improved considerably during the past two decades. Unfortunately, the biggest gains have already been made in terms of reductions of aberrations. While further reductions in aberrations will continue to occur, great strides are no longer possible. An extension of optical lithography will have to come from other sources.

10.4 Maximum Numerical Aperture

Among the improvements in lenses over the past three decades has been a substantial increase in numerical apertures. The lens on the first commercially available stepper had a numerical aperture of 0.28. Today, the major stepper suppliers have existing products for lenses with $NA = 1.35$. Increased numerical aperture was a significant means by which optical lithography was extended in the past. How much further numerical apertures can be increased will be discussed in this section. Following that discussion, some other paths by which optical lithography has been extended will be considered, and it will be seen why these approaches to extending lithography are no longer major avenues to higher resolution.

The numerical aperture is given by $n \sin\theta$, where n is the index of refraction of the medium between the lens and wafer, and θ is the half angle subtended by the lens. There are practical considerations regarding how large θ can be made. As the numerical aperture is increased, there are a number of optical engineering challenges that increase in proportion to $\tan\theta$, including the accuracy of optics mounting, the lens diameter, and control requirements for optical coatings (Fig. 10.9). Values for θ and $\tan\theta$ are listed in Table 10.1. From $NA = 1.30$ to 1.35 the increase in difficulty in fabricating the lens is ~30%, while going from $NA = 1.35$ to $NA = 1.40$ represents a further ~60% effort. These increases in NA would require significant investments for <10% improvement in resolution from $NA = 1.30$ to 1.40. Regardless of practicalities, the most that resolution can be increased due to fundamental physical limits is <7% from the current maximum of 1.35, as long as water is the immersion medium. This is significantly less than the 4.8× increase in the past (from $NA = 0.28$ to 1.35).

From Eqs. (2.7) and (10.6), one can see that both resolution and depth-of-focus can be enhanced when the immersion fluid has a large index of refraction. It is therefore natural to consider whether lithographic capability can be improved by using an immersion fluid that has a higher index of refraction than water. When

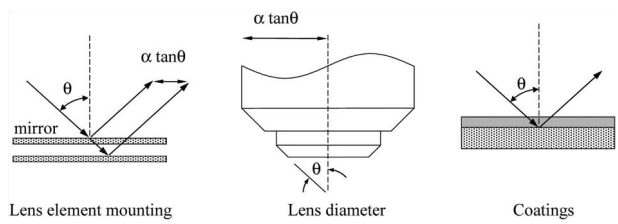


Figure 10.9 Illustration of the factors that increase in difficulty in proportion to $\tan\theta$.

Table 10.1 Values for θ (Fig. 10.9) and $\tan\theta$ for numerical apertures that approach the limit of water-immersion lithography.

Numerical aperture	θ (deg)	$\tan\theta$
1.30	65	2.14
1.35	70	2.75
1.40	77	4.32

considering alternatives there are many technical issues to be addressed. Water is an extremely convenient material for use as an immersion fluid, as a consequence of several important properties:

1. The semiconductor industry has considerable experience in producing ultrapure water inexpensively.
2. Water is nontoxic, which has implications for both worker safety and disposal costs.
3. Water is compatible with resist and materials commonly used to fabricate scanners.
4. The surface tension between water and many materials facilitates containment.[18]
5. Water is stable under exposure to 193-nm light.
6. Water has sufficiently low viscosity to permit fast scanning.

Any fluid that replaces water will need to have these properties, or additional engineering work will be required to address any shortcomings.

Candidate high-index immersion fluids have been developed, with indices of refraction between 1.62 and 1.66,[19] which provides potential for 13–16% higher resolution than water-immersion lithography. However, problems have been encountered with such materials involving fluid decomposition under exposure to ArF light, which can include contamination of the bottom lens element.[20] The problem of decomposition can be mitigated to considerable extent by circulating and repurifying the fluid. Reuse of these fluids is important, since they cost considerably more than water.

As the index of refraction of the immersion fluid is increased, additional physics need to be taken into account. Consider the situation depicted in Fig. 10.10. When the index of refraction of the glass is greater than that of the fluid, the refraction at the glass-fluid interface assists in generating a high numerical aperture, since light is refracted away from normal at the glass-fluid interface. There are practical limits to how large θ_1 can be in a lens, as discussed earlier. Increasing n_{fluid} until it is greater than n_{glass} is a losing proposition, because the increase in NA achieved with a larger fluid refractive index will be offset by a reduction in θ_2. Hence, there is value in increasing n_{fluid} only until it equals n_{glass}. Indices of refraction at 193 nm are given in Table 10.2 for various candidate lens materials. As can be seen from the table, there is little value in increasing the index of refraction of the fluid above

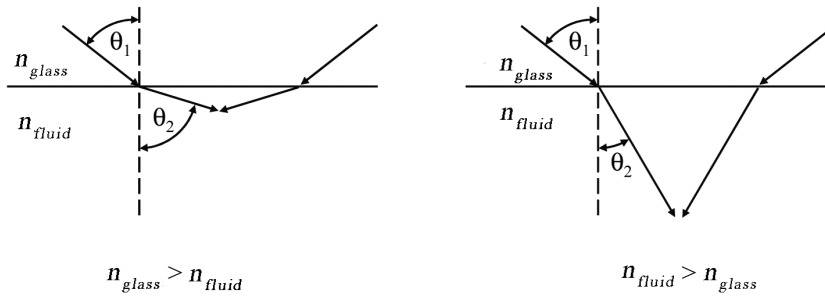


Figure 10.10 Greater numerical apertures are possible when the immersion fluid has an index of refraction less than that of the material of the bottom lens element.

Table 10.2 Materials that can be used for immersion lithography, along with their indices of refraction.

Material	Index of refraction at λ = 193 nm
Water	1.44
Fused silica[21]	1.56
Calcium fluoride[21]	1.50
Aluminum oxide[22]	1.92
Lutetium aluminum garnet (LuAG)[23]	2.14

that of water without also fabricating the bottom lens element with a material other than fused silica and calcium fluoride.

A materials search identified one material, lutetium aluminum garnet ($Lu_3Al_5O_{12}$), that has a high refractive index at $\lambda = 193$ nm and has sufficiently low birefringence that lens makers can consider using it for a lens element. However, progress has been slow in producing crystals with high transparency and good homogeneity. The problem of transparency did not arise from fundamental absorption by the crystal, but rather was a consequence of chemical impurities. With a slow rate of improvement, efforts to produce lens-quality lutetium aluminum garnet crystals have since been reduced. Another reason for slowing down crystal development efforts were challenges in developing an immersion fluid with a high index of refraction which meets all requirements, without which a high-index crystal is not needed.

Assuming that a high-index lens material and a high-index immersion fluid can be identified, and that solutions can be found for all associated technical issues, there is still one more material that needs to be addressed, and that is the photoresist. As noted in Chapter 4, image contrast is reduced at high angles of incidence on the resist for P-polarized light (see Fig. 4.3). The only reason that there is good imaging even at moderate numerical apertures is because of the refraction at the interface between the resist and air or fluid. For this refraction to improve image contrast, it is necessary that the index of refraction of the resist be greater than that of the fluid. For most chemically amplified resists, $n \approx 1.7$, so this condition is satisfied with water as the immersion fluid, as well as most

high-index candidate fluids. However, to achieve better than ~15% improvement in resolution, a resist with an index of refraction >1.7 will be needed.

Given all of the challenges associated with immersion lithography using a fluid with an index of refraction greater than that of water—the fluid itself, a high-index lens material, and high-index resists—it appears quite possible that high-index immersion lithography will not be pursued at 193 nm for use in manufacturing. The potential for higher resolution by going to a wavelength shorter than 193 nm will be considered next.

10.5 The Shortest Wavelength

While the early predictions of the end to optical lithography were incorrect in detail, the fact remains that diffraction imposes limits on the potential of optical lithography. At what feature size does diffraction finally defeat ingenuity and invention? In order to predict the limits of optical lithography, it first needs to be decided what minimum wavelength can be used, since shorter wavelengths will enable better resolution with manageable depth-of-focus, all else remaining equal.

As discussed in Chapter 5, lithography is practiced at wavelengths for which there are intense sources of near-monochromatic light. Below 193 nm there are a few options for light sources, but they will all need further development if they are to qualify as adequate for use in semiconductor manufacturing. F_2 excimer lasers emit 157-nm light. For several years 157-nm lithography was considered a potential solution, but progress was slow. It was not possible to obtain sufficient transparency with fused silica for it to be used as a lens material, which meant that all refractive lens elements needed to be made from crystalline CaF_2. It proved very difficult to produce large crystals of CaF_2 with adequate transparency, homogeneity, and low birefringence. There were also challenges with resists. Fluorine chemistry was required to obtain the necessary resist transparency, but good resist performance remained elusive even with such materials. When it became clear that water-immersion lithography was viable and could extend optical lithography even further than (nonimmersion) 157-nm lithography, F_2 lithography was no longer strongly pursued as a lithographic technology option. (See Problem 10.2 for reasons why the return on investment in F_2 lithography was considered to be limited.)

Below 157 nm, Ar_2^* lasers operating at a wavelength of 126 nm have been demonstrated,[24] but reliability and repetition rates need to be improved.[25,26] Material damage at 126 nm is a particular problem.[27] There is also some possibility of exploiting the hydrogen Lyman-α line at 121.5 nm.[28] However, assuming that a lithography-capable light source can be developed, there are other issues that will more likely prevent the extension of optical lithography to a wavelength of 126 nm or 121.5 nm. At the Ar_2^* and Lyman-α wavelengths, transparent materials are few, and include LiF, CaF_2, and MgF_2.[29] With the exception of CaF_2, these materials are not well developed, and more work is required to demonstrate that they are suitable for use in high-resolution lenses. It should be noted that all of these materials are birefringent, which greatly complicates lens design. Although there

is nothing known currently that cannot be overcome with a focused engineering effort, improvements in CaF_2 came slowly when pursued for F_2 lithography, an indication of the magnitude of the engineering challenge.

Photomasks represent a more formidable problem than light sources, one that is perhaps insurmountable. Optical lithography, as currently practiced, involves transmission photomasks. A departure from the use of transmission masks would be a significant change. Materials for transmission masks must meet several requirements. First, transparency is an obvious prerequisite. This requirement shortens the list of potential substrate materials considerably. Secondly, thermal stability is another essential characteristic, in order to maintain adequate registration during mask making. The problem is particularly acute when electron-beam writers are used for making the masks. As discussed in Chapter 7, substantial energy is deposited into the photomask substrates during e-beam exposure with beam energies between 10 and 50 keV, and this causes the substrate to heat and deform mechanically.[30] Fused silica, the material used currently for photomask substrates, has a very low coefficient of thermal expansion, in contrast to other materials that are transparent at 157 nm and shorter wavelengths, such as CaF_2 and BaF_2. The coefficient of thermal expansion for fused silica is 0.5 ppm/K, compared to the substantially higher value of 19 ppm/K for CaF_2. On a fused silica plate, a 0.1 K temperature change will result in a 5-nm registration error across 100 mm, a magnitude of misregistration that can likely be corrected adequately (assuming 4× reduction optics) if it occurs in a predictable way. For a reticle fabricated from CaF_2, the registration error would be almost 100 nm. Even a 90% effective correction would not be sufficient to meet the overlay requirements of advanced technologies, and temperature control considerably tighter than 0.1 K is not practical. Only fused silica appears to have a sufficiently small coefficient of thermal expansion to be a practical mask material. Because fused silica becomes too absorbing at wavelengths just a few nanometers shorter than 157 nm, 157 nm is the shortest wavelength that could be a candidate to be the last "optical lithography," using the definition of optical lithography set forth earlier.

Of course, this assumes that all elements of a 157-nm lithography process can be developed. There were a number of programs to develop photomasks, photoresists, and exposure tools at 157 nm, and considerable progress was made towards the realization of a working 157-nm optical lithographic technology. However, nearly all work on 157-nm lithography development was suspended, as 193-nm immersion lithography appeared to be more attractive.

With respect to immersion lithography, most attention is currently focused on ArF immersion, because that provides immediate impact, but immersion lithography has been studied for 157-nm lithography.[31,32] For 157-nm lithography, water is far too absorbing for use as an immersion fluid. Identification of a sufficiently transparent fluid at 157 nm has been an area of active research.[33]

Regardless of whether the problems of photomask fabrication and light sources can be overcome, there are no known transparent materials for wavelengths shorter than 110 nm. This therefore represents a true limit to optical lithography, as it has been defined in this chapter. However, in light of the previously discussed

problems with mask thermal expansion, lens materials, and resists, 157 nm will be considered for the remainder of this chapter as the smallest wavelength at which optical lithography can theoretically be practiced in the future, while 193 nm is the shortest wavelength from a practical point of view. The ability to advance optical lithography by moving to a shorter wavelength is very near an end.

10.6 Improved Photoresists

Resists have improved remarkably over the past three decades. This has been one of the reasons why it has been possible to practice low-k_1 lithography. However, there appears to be reduced potential for large gains from future improvements in resists. An example of this is the increase in sidewall angle due to high resist contrast. As shown in Chapter 2, the tangent of the angle of resist sidewalls is directly proportional to the resist contrast. Increases in the resist contrast from low to high values can have a significant impact on the resist profile, but further increases in contrast provide marginal benefit once the resist contrast reaches a moderate value (see Fig. 10.11). Current activities in resist chemistry for the extension of optical lithography are in the development of high-contrast resists with immersion compatibility. However, this simply provides improved capability for water-immersion lithography and does not provide significant enhancement to the lithography beyond that. The major gains from improvements in the resists have already occurred, with efforts now focused on maintaining that position, but in the immersion context. Improved resists are still of value, particularly with respect to reduced line-edge roughness, but significant reductions in k_1, because of improvements in photoresists, are not expected.

In addition to the resist challenges of contrast, there are additional issues that need to be considered as resist features shrink well below 50 nm. One issue results from the blurring that occurs during the post-exposure diffusion step in chemically amplified resists, while the other is the line-edge roughness of patterns after develop. During the post-exposure step, photoacids diffuse and cause deprotection of chemically amplified resists. As long as the distances over which this diffusion occurs is small compared to the linewidth, this is not a fundamental problem.

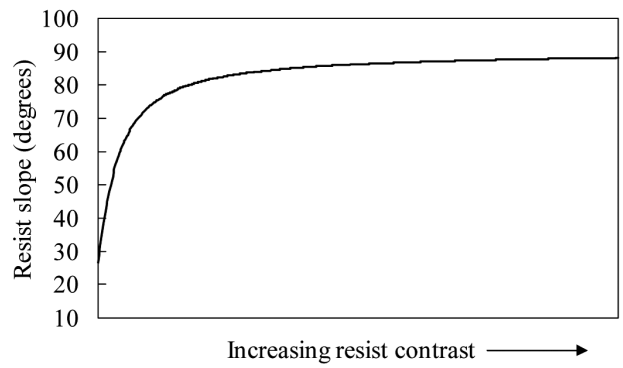


Figure 10.11 Resist sidewall angle as a function of resist contrast.

Recent measurements indicate that the diffusion length for this process is on the order of 10–30 nm. When chemically amplified resists were first used widely, linewidths were ~350 nm. For such features, the blurring from photoacid diffusion was not significant. However, as features have become smaller than 50 nm, the impact of photoacid diffusion becomes quite relevant.

It is possible to reduce the amount of diffusion during the post-exposure baking step, but there are consequences. First, this diffusion ameliorates line-edge roughness (LER) to a certain degree. The larger the diffusion, the more the LER is smoothed. Reducing the diffusion will necessitate additional measures to keep LER at an acceptable level. Also, a significant amount of diffusion enables a large amount of chemical amplification. Thus, reducing diffusion will also reduce resist sensitivity. This is of particular concern for some of the technologies that will be discussed in Chapters 12 and 13.

10.7 Flatter Wafers

For many years device topography presented a challenge to lithographers. This topography led to many of the problems associated with varying resist thickness (discussed in Chapter 4) and variable reflectivity from the substrate. In addition, the depths-of-focus needed to be large enough to provide good imaging throughout the height of the topography. For example, device topography of 0.7 μm (±0.35) was typical. With tool focus control of ±0.4 μm, the optics needed to have at least ±(0.35 + 0.4) μm = ±0.75-μm depth-of-focus in order to provide good imaging. For g-line optics, this implied a maximum NA of approximately 0.4, using the Rayleigh expression to calculate depth-of-focus. This limited resolution to no better than 0.6 μm. Matters were improved by the use of shorter wavelengths, but the large fraction of the depth-of-focus budget from device topography still limited the use of high-numerical-aperture optics.

The advent of chemical-mechanical polishing reduced the problems associated with device topography significantly by eliminating most device topography. The requisite depth-of-focus of the optics was decreased, because it was no longer necessary to image well throughout the height of the device features. This enabled the use of higher numerical aperture optics and allowed the extension of optical lithography further than earlier thought possible, because the depth-of-focus requirement was reduced by the use of chemical-mechanical polishing.

However, wafers can be flattened only once. The benefit of chemical-mechanical polishing has nearly been fully realized, and this process cannot provide additional help for extending optical lithography. As with resist and wavelength, only small gains are possible for extending optical lithography by the means of reducing wafer topography, as used in the past. Optical lithography has been extended beyond earlier expectations, but the means by which this extension occurred provide little further benefit. However, there are new resolution-enhancement techniques, such as those discussed in Chapter 8, that have not yet been fully exploited, enabling a move to lower k_1. The question of how low in k_1 it is possible to go is now considered.

10.8 How Low Can k_1 Go?

An inspection of Fig. 2.13 shows that image contrast becomes zero when the half pitch is $0.25\lambda/NA$ or smaller. This is true even when using powerful resolution-enhancement techniques such as dipole illumination or alternating phase-shifting masks. Thus, there is a limit on k_1 of 0.25 imposed by the laws of physics. It only need be assumed that the optical imaging is linear and that a single exposure step is involved for this to be a true limit imposed by the laws of nature. How close to $k_1 = 0.25$ a manufacturing process can be is a matter of technology.

The value of k_1 used in manufacturing has been declining for years. Since k_1 must always be larger than 0.25, it is interesting to see how $k_1' = k_1 - 0.25$ trends over time. A graph of k_1' is shown in Fig. 10.12. As can be seen, considerable progress has been made in reducing k_1 over the years, and k_1 has descended from a value of ~1.0 in the mid-1980s to ~0.3 today. This decrease in k_1 alone accounts for a reduction in feature sizes of over 3×, which represents three nodes of progress on the International Technology Roadmap for Semiconductors. However, a similar reduction in the future is not possible, since k_1 cannot become smaller than 0.25. The most reduction that is physically possible is

$$\frac{0.25}{0.30} = 0.83, \tag{10.13}$$

and this requires taking resolution to the absolute limit allowed by the laws of physics. A more likely minimum value will be somewhat larger, perhaps $k_1 = 0.28$ or 0.27. Thus, only modest decreases in minimum feature sizes will be possible in the future through reduction in k_1.

As has been described thus far in this chapter, nearly all of the improvements that have been introduced in the past to advance lithography—better resists, lenses with lower aberrations and higher numerical apertures, flatter wafers, and shorter wavelengths—have been exploited nearly completely. How close to the theoretical limit optical lithography can be taken is discussed in the next section.

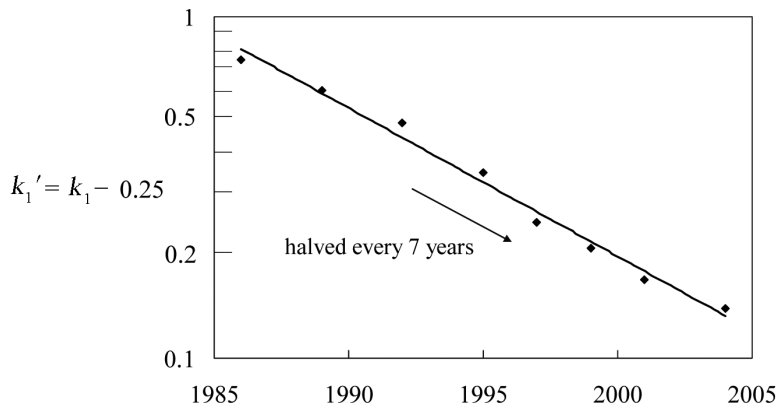


Figure 10.12 The evolution of k_1 over time, using DRAM half pitches to ascertain k_1.

10.9 How Far Can Optical Lithography Be Extended?

Based upon the discussion in this chapter thus far, optical lithography at its ultimate limit consists of the following:

1. A wavelength of 157 nm. While there are likely practical reasons why 157-nm lithography may not be used in manufacturing, $\lambda = 157$ nm does represent the optical lithography wavelength limit, at least in principle.
2. Photoresists sensitive at 157 nm, but with performance comparable to mature KrF resists, in terms of the process k_1.
3. Fully implemented two-beam resolution enhancement techniques, such as dipole.
4. Negligible device topography.
5. An immersion fluid with an index of refraction of 1.65. This assumes that lens materials, such as lutetium aluminum garnet, can be produced with sufficient quality, so the higher-index fluids can be exploited.

These are the assumptions for the remainder of this section. Optical lithography is limited because many parameters cannot be improved indefinitely, but rather to asymptotic limits. The key examples of these are listed in Table 10.3. The mathematical limit of $\sin\theta$ is 1.0. Shorter wavelengths can be used until fused silica transparency becomes limiting. Wafer flatness can be improved until the wafers are flat, and then no further improvement is possible. Optical lithography is limited by the *tyranny of the asymptote*.

Because resolution in optical lithography is limited by depth-of-focus, the DOF that is actually required is at the heart of the problem. The limit of optical lithography is quite different if exposure-tool focus can be controlled within ±25 nm, rather than, say, ±100 nm. Another relevant issue when discussing the ultimate limit of optical lithography is the extent to which lens aberrations can be reduced. A lens with 20-nm field curvature will have less capability than one with only 5-nm field curvature.

Suppose the lens is indeed diffraction limited. Consider:

$$\text{resolution} = k_1 \frac{\lambda}{NA}. \tag{10.14}$$

Table 10.3 Key lithography parameters are their asymptotic limits. θ is defined in Fig. 2.11.

Lithographic parameter	Asymptotic limit
Resolution	Diffraction limit
$\sin\theta$	1.0
Resist performance	Tangent of the resist sidewall angle
Wavelength	157 nm
Wafer flatness	Perfectly flat
Immersion fluid index of refraction	1.65

Today, ArF processes with $k_1 < 0.35$ are common. For these processes, pitches <90 nm are imaged with currently available 1.35-NA lenses. If we assume that similar values of θ can be produced for a 157-nm lens and an immersion fluid with $n = 1.65$ is available, along with glass materials to support such a fluid, it would be possible to produce 27-nm lines and spaces with $k_1 = 0.27$ and a depth-of-focus [Eq. (10.4)] of approximately ±36 nm. Lithography with such depths-of-focus will require lenses with very little field curvature and astigmatism, exposure tools with focus control of 10 nm (or less), negligible device topography, and extremely flat wafers. Nevertheless, such processes are not out of the question. Even if we restrict ourselves to conservative values for NA (1.35), k_1 (0.28), and wavelength (193 nm), a resolution of 40 nm is achievable. All of these estimations indicate that the ultimate limit for optical lithography is somewhere between 27 nm and 40 nm. Beyond this range—the 22-nm node, for example—lithography methods that differ in form from optical lithography as described thus far in this book will be required.

The degree to which optical lithography can be extended to very low values of k_1 depends upon how successfully all of the methods described in Chapter 8—optical proximity corrections, off-axis illumination, and phase-shifting masks—can be implemented. The ability to implement particular techniques depends upon the application. For example, the imaging of highly repetitive structures, such as those found in memories, can be enhanced through the use of off-axis illumination, but the benefit is less for isolated features.

In Chapter 8, several resolution-enhancement techniques were presented: optical proximity corrections, phase shifting, and off-axis illumination. The issues associated with each of these techniques are best understood in light of requirements. The constraints for memories, logic, and application-specific integrated circuits are all different. For memories, the goal is packing a large number of bits into a given area of silicon. Memory is used by the bit, and increasing density provides more bits at lower prices. The memory business is very dependent upon the manufacturing cost per bit.

The requirements for logic are somewhat different. There is certainly an element of importance in manufacturing efficiency that can come from packing density, particularly for microprocessors with significant amounts of on-board cache memory. However, processor speed has a big effect on the value of the part. This is shown for a Windows-compatible microprocessor in Fig. 10.13. Retail prices are seen to be extremely dependent upon processor speed. Clearly, consumers are willing to pay significantly more money for higher-performance microprocessors, which motivates manufacturers of microprocessors to maximize performance. Fast chips require fast transistors, which usually necessitates short gate lengths. The speed at which the processor can operate reliably is usually limited by the slowest transistor on the part. Hence, it is undesirable to have any transistors with long gate lengths. On the other hand, transistors with gates that are *too* short may break down electrically or can result in wasted power consumption due to leakage. Thus, it is undesirable for gates to be either too long or too short. For microprocessors, linewidth *control* is critical.

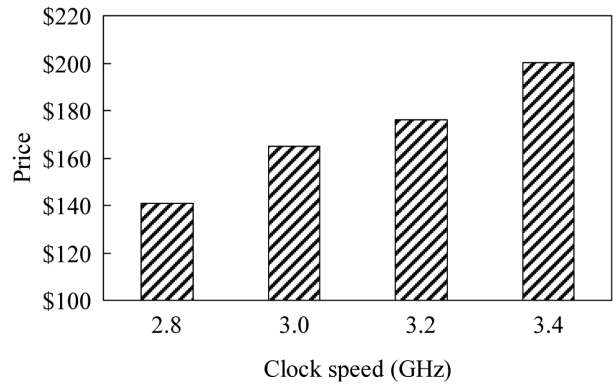


Figure 10.13 AMD Phenom II Deneb processor prices as a function of processor clock speed. The prices were taken on January 11, 2010, from the Web site* of a company that sells computer hardware.

Linewidth-control requirements for logic gates have been incorporated into the International Technology Roadmap for Semiconductors and are summarized in Table 10.4. These requirements are very tight, and even the smallest detractors of linewidth control will need to be addressed. Consider the issue of optical proximity corrections. There are limits to how finely these can be applied, particularly in circumstances where there are large mask-error factors (MEF). The problem can be illustrated by a hypothetical example. Suppose that a 3-nm iso-dense adjustment needs to be made. For 4× reticles, this represents a 12-nm adjustment, if $MEF = 1$ and the beam writer is writing on a 5-nm grid, then there will be a 2-nm residual error following a 10-nm correction on the mask. With $MEF = 1$, this results in a 0.5-nm error on the wafer, which might be considered acceptable. However, with $MEF = 4$, the error on the wafer will be 2 nm, which is quite significant relative to linewidth control requirements in the 45-nm node and beyond.

Table 10.4 Gate critical dimension (CD) control requirements from the year 2009 International Technology Roadmap for Semiconductors.

Node (nm)	Year	Microprocessor gate CD in resist (nm)	Microprocessor gate CD control, lithography contribution only (3σ, nm)
45	2010	41	2.8
32	2013	28	2.1
23	2016	20	1.6
16	2019	14	1.2

Phase shifting does not eliminate the problem of a finite grid size for adjusting iso-dense biases, although it does reduce the mask-error factor significantly. The

*www.Newegg.com

value of this is illustrated in Figs. 10.14 and 10.15, where the linewidth target of 120 nm is simulated over a range of pitches. Without optical proximity corrections, the linewidth variations are large (Fig. 10.14). They are reduced significantly by adjusting the mask CD at several different pitches, as illustrated in Fig. 10.15. However, even with these optical proximity corrections there is some residual linewidth variation, even at best focus. This is a consequence of corrections being made on a grid of only 20 nm on the mask. The mask could be made on a finer grid, but this would increase the cost.

There are also certain "forbidden" pitches[35] where the depth-of-focus is much worse than it is for others. For some technologies, such variations might be tolerable, but not for high-performance microprocessors, where exceptionally

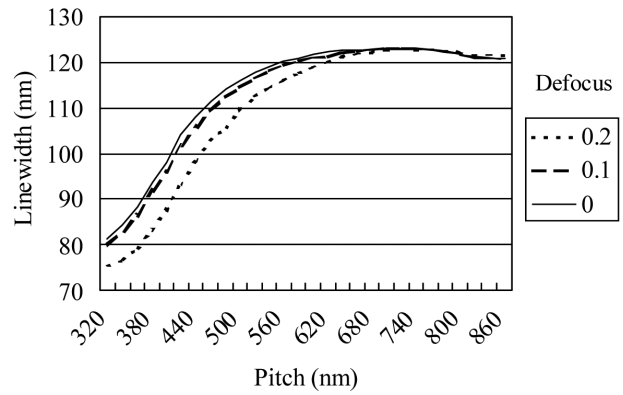


Figure 10.14 Simulated linewidth on the wafer versus pitch for a fixed linewidth (100 nm) on an alternating phase-shifting mask, for several values of defocus, targeting 120 nm on the wafer. The parameters of the calculation assumed $\lambda = 248$ nm, $NA = 0.6$, and $\sigma = 0.5$.[34] The resist was UV5-like. An aberration-free lens was assumed.

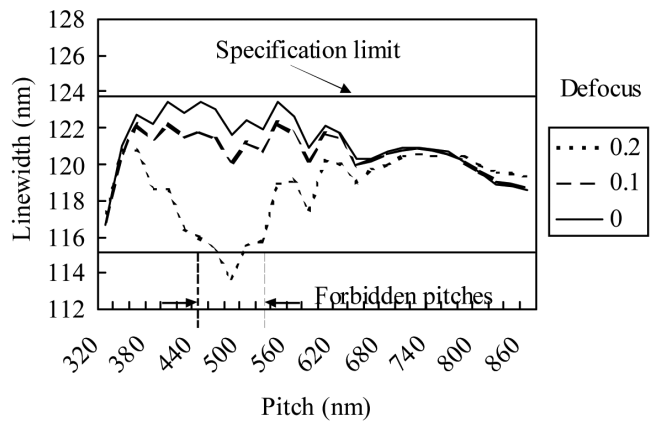


Figure 10.15 Simulated linewidth for the same conditions as for Fig. 10.14, except that the dimensions were proximity corrected on a 5-nm grid (20 nm on the mask for a 4× system). Numerical aperture and partial coherence were selected to optimize the range of the forbidden pitches.

tight gate linewidth control is required. Changing parameters, such as numerical aperture and partial coherence, only increases the variation shown in Fig. 10.15, although the particular range of forbidden pitches may shift. This is not something that can be corrected through optical proximity corrections—the problem is depth-of-focus, not CD variation at best focus, over the range of pitches.

While memories may not have the extraordinary challenges for gate linewidth control that is required for high-performance logic, there are other very significant lithography challenges. Consider the addition of hammerheads to address the problem of line shortening (Fig. 8.17). For patterns on large pitches, there is space to add the hammerheads, but consider the memory-array pattern shown in Fig. 10.16. When features are densely packed, there is little room to add features for optical proximity corrections. Fortunately, densely packed memory arrays are precisely those situations where resolution-enhancement techniques that are applicable to repetitive patterns have the most benefit.

Makers of memories and high-volume microprocessors can match their designs and lithography processes. Memories can be laid out with pitches to enable some degree of optical proximity correction and be set on a pitch that can be well imaged with predetermined illumination conditions. There is a greater challenge for foundries, which are expected to produce products designed at other companies. Enabling the greatest flexibility for designs while also providing leading-edge lithography capability is a challenge, and some restrictions on layout are required to achieve the most dense layouts possible.

Many resolution-enhancement techniques, such as off-axis illumination, have maximum benefit over a narrow range of pitches. Good-quality patterning of exceptionally fine-line features using these pitch-specific techniques is not easily obtained for high-performance logic applications with random logic on layouts that have many pitches, orientations, and configurations. To achieve very tight control of linewidths on tight pitches it is necessary to pattern only a small number of pitches. For this reason, methodologies have been developed to transform designs to layouts that are "lithographically friendly." These approaches are collectively

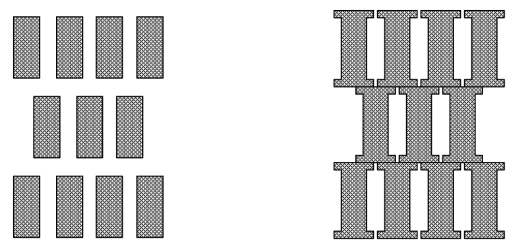


Figure 10.16 When memory arrays are densely packed, there is no room to add features for optical proximity corrections.

referred to as *design for manufacturability* (DFM). Under the umbrella of DFM, there are a number of methodologies for improving yield and reducing variability:

1. *Identification and elimination of design layouts that have very small process windows*. This may be achieved, for example, by determining optical proximity corrections throughout a focus-dose process window and not just at best focus and dose (see Fig. 10.15). This example illustrates that there is no sharp dividing line between what people call OPC and DFM.
2. *Significant restrictions on layouts*. Design rules have long been used to restrict the features that are allowed in designs. For many years these rules were very simple, usually consisting of statements of minimum allowed feature sizes. Over time more rules were added, such as requiring that all contact holes be the same size. More recently, manufacturers of high-performance logic have oriented all transistor gates in a single direction in order to reduce variability. Such major changes to the rules required a change in paradigm, where designers and lithographers worked together to generate the design rules. The greater complexity of design rules requires extensive development work involving both designers and lithographers, and implementation is greatly facilitated by good design tools.
3. *Process-aware design rules, including electrical rules*. Lithographic variations will have an impact on electrical characteristics, often beyond the variations seen in simple line and space patterns. For example, optical proximity effects will modify the shape of a transistor's gate that is embedded within a complex pattern beyond a simple shift in linewidth (Fig. 10.17). In this case the electrical characteristics of the transistor are determined by breaking the gate into segments of variable length.[36]

Design for manufacturability typically requires consideration of more than a single layer, because layout modifications are often involved. An example of this is illustrated in Fig. 10.18. In this case the layout modifications are substantial and involve several device layers.

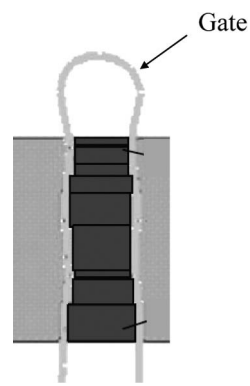


Figure 10.17 The gate shape is the result of simulation. As a consequence of patterns in proximity to this gate, the gate length is not uniform across the active region.

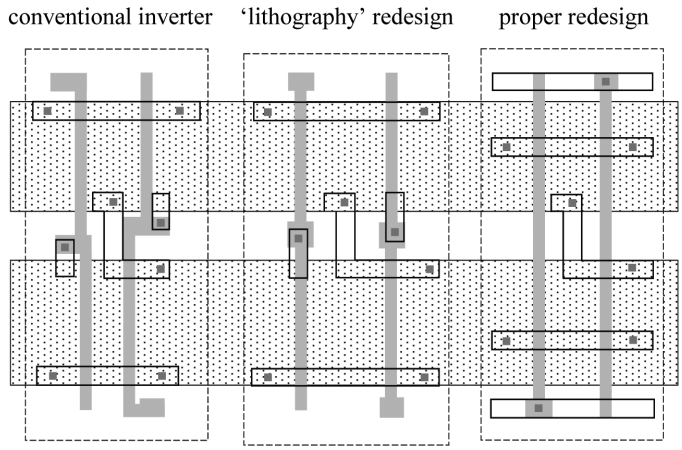


Figure 10.18 Modifying the layout of an inverter to become increasingly lithographically friendly.[37]

There are many challenges that need to be overcome for optical lithography to be extended to the limits of water-immersion lithography, and even greater challenges for extension beyond that. Regardless, 27 nm appears to be a clear limit for conventional optical lithography. A possible optical approach involving a paradigm that has potential for patterning down to the 22-nm node is presented in the next section, and nonoptical technologies will be the subject of this book's final two chapters.

10.10 Double Patterning

The limitations of the resolution of optical lithography are constraints on the minimum pitch that can be printed, not the size of individual features. Consequently, one means of patterning a very dense pitch is to print every-other line on one exposure and then print the lines in between on a succeeding step. Because resist solvent will normally dissolve resist patterns on the wafer, various double-patterning approaches have been adopted to address this issue. One such approach is illustrated in Fig. 10.19. In this case the pattern of the first lithography step is transferred into a hardmask on which the resist of the second patterning operation can be coated, exposed, and developed. This approach clearly represents a considerable increase in process complexity and cost, replacing a single lithography and etch step with two such operations, along with an additional hardmask deposition and removal.

While Fig. 10.19 captures the basic concept of double patterning, there are many details about which lithographers need to be concerned that are not fully captured in the idealized illustration. For example, there is likely to be a difference in average linewidth between the first and second exposures, and this will increase the variance of the linewidth distribution. If the standard deviation of the linewidths from the first patterning (including etch) is σ_1 and the linewidth mean is μ_1, while σ_2 and μ_2 are the standard deviation and mean of the linewidths from the second

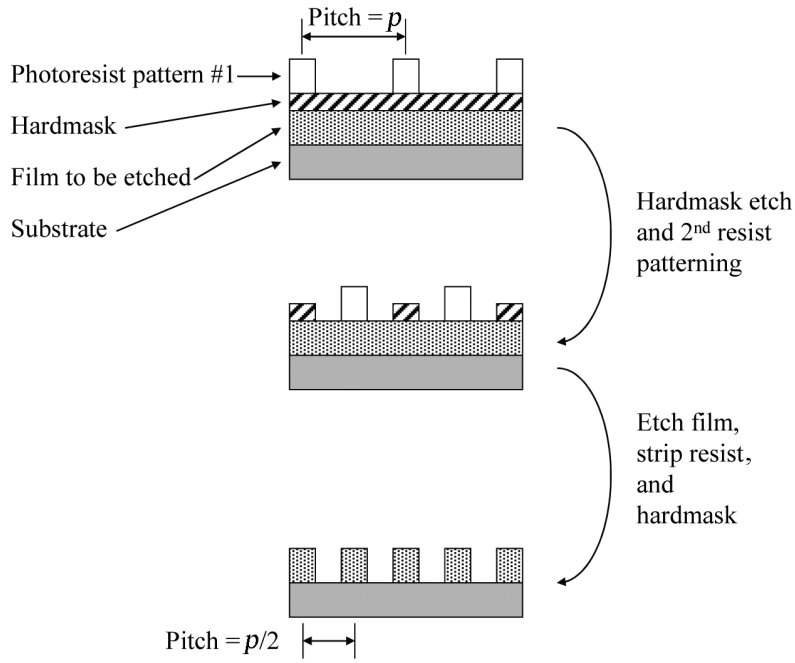


Figure 10.19 Illustration of a double-patterning process that incorporates a hardmask and two etches.

patterning, respectively, then the overall linewidth variation is[38]

$$\sigma_{total} = \sqrt{\frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} + \left[\frac{1}{2}(\mu_1 - \mu_2)\right]^2}. \tag{10.15}$$

If $\sigma_1 = \sigma_2 = \sigma$, this reduces to

$$\sigma_{total} = \sqrt{\sigma^2 + \left[\frac{1}{2}(\mu_1 - \mu_2)\right]^2}. \tag{10.16}$$

It cannot be assumed automatically that $\sigma_1 = \sigma_2$, because the second lithography step takes place on the topography created by the first patterning step. For example, the spin coating of bottom antireflection coatings and resists will be affected by the topography, as described in Chapter 3, so a multilayer resist process might be considered for the second patterning step.

An overlay error between the first and second pattern will affect the dimensional control of the spaces (Fig. 10.20). Because the variance for overlay is usually greater than that for linewidths (from a single patterning), the variance for the spacewidths will be much greater than that for lines for the example shown in Fig. 10.20. If the mean overlay error between the first and second exposure is μ_{OL}

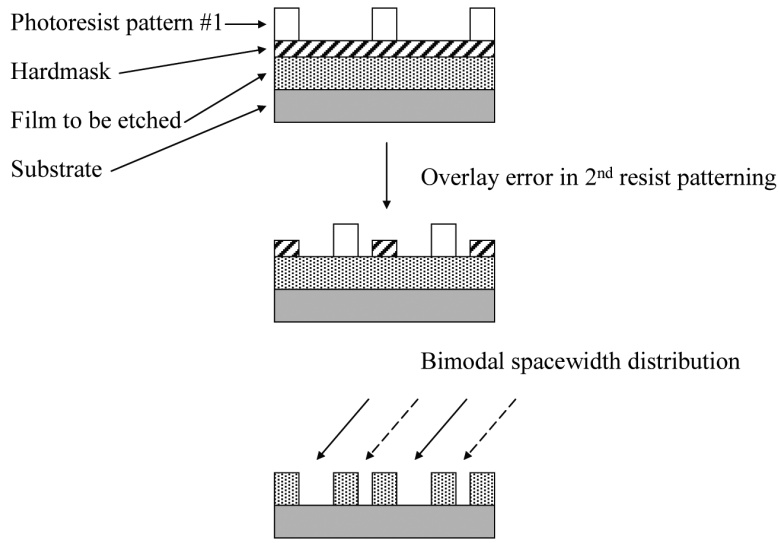


Figure 10.20 Illustration of a double-patterning process with an overlay error.

with a standard deviation of σ_{OL}, the resulting spacewidth variation is given by

$$\sigma_{CD,space} = \sqrt{\frac{\sigma_{CD,line}^2}{2} + \sigma_{OL}^2 + \mu_{OL}^2}. \tag{10.17}$$

For the situation where lines of resist are the features produced by the individual patterning operations, this equation clearly shows the increase in the spacewidth variation caused by overlay.

If the lines are much more critical than the spaces, as is typically true for gate layer, then patterning lines of resist is an appropriate choice for a double-patterning process, since the critical gate length is not affected directly by overlay errors. However, there are circumstances where the spaces might be considered more critical than the lines. For example, in a damascene metal process, the metal lines are defined by the space. If the primary concerns are resistance and perhaps electromigration, then the preceding analysis shows that process control might be severely limited by overlay. A way to avoid this is to print spaces as the primary features, as shown in Fig. 10.21. There are some difficulties with this. For very small features, spaces are more difficult to print using positive resist than lines of resist, even on loose pitches. Negative tone resist processes can provide a solution.[39]

Another way to reduce the impact of overlay on spacewidth control is by the use of spacers to achieve tight pitches.[40] This approach is illustrated in Fig. 10.22. In this approach a pattern on a loose pitch is created, and this is then coated with a conformal thin film. The film is then etched anisotropically in a process very similar to that used for many years to create spacers on gates. As can be seen, the spacer process eliminates the dependence of critical dimensions on overlay.

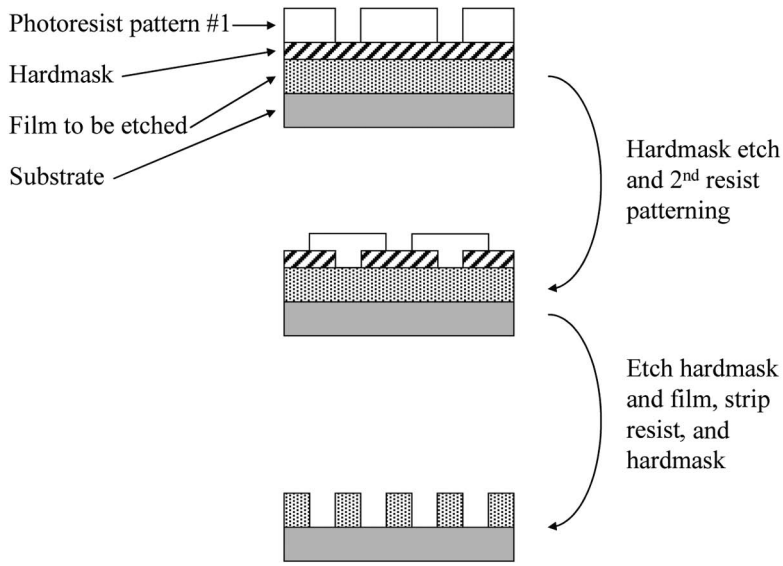


Figure 10.21 Double-patterning process with spaces being patterned directly.

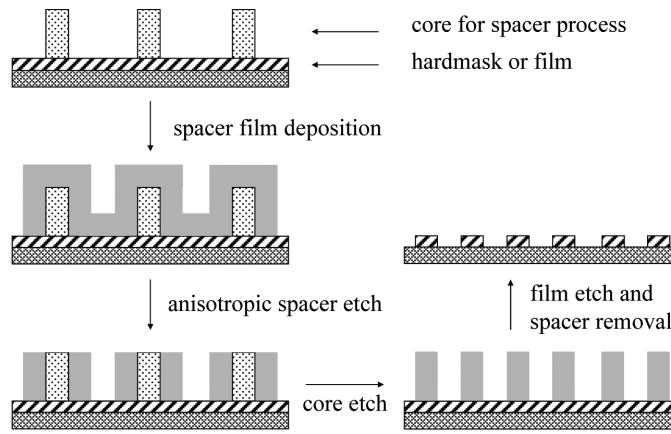


Figure 10.22 Double patterning using a spacer process that halves the pitch dimension.

However, the opposite situation can occur. A mis-sized core pattern used in a spacer process will cause misplacement of the spacer-defined features. Spacer processes will require additional lithography and etch steps to remove features at the ends of lines that may not be desired (Fig. 10.23).

Double patterning has been illustrated in Figs. 10.19–10.23 with line and space patterns. This is appropriate, since the problem that double patterning is attempting to address is that of creating tight pitches, and such processes have already found applications for printing the dense grating structures found in flash memories. However, patterns of real circuits consist of far more than simple lines and spaces, particularly for logic and memories outside of the cores (Fig. 10.24). It may not be possible to decompose layouts for double-patterning processes, and circuit layouts

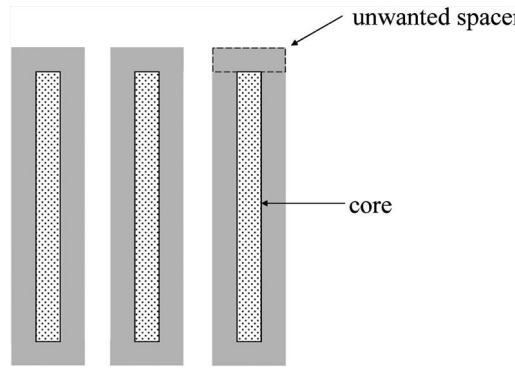


Figure 10.23 Top view of a pattern generated in a spacer process, illustrating an unwanted spacer that needs to be removed with additional lithography and etch steps in order to created a unidirectional grating.

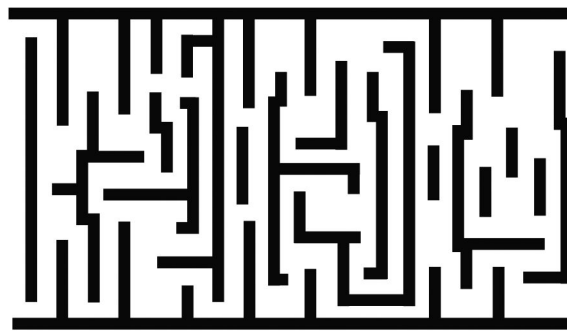


Figure 10.24 Metal layout for random logic, unmodified to be amenable to patterning using spacer double-patterning processes.

will need to be altered if double patterning is going to be used to shrink patterns from those generated using single exposures.

The double-patterning processes described thus far in this section are complex, involving multiple film depositions, lithography steps, and etches. There have been efforts to reduce the number of steps in the process in order to reduce overall cost. One variation of the process illustrated in Fig. 10.19 is to "harden" or "freeze" the resist features from the first patterning step so that another layer of photoresist can be coated directly over the first resist pattern without the resist features already on the wafer being dissolved by the solvent in the resist of the second patterning process. Methods for freezing the resist include vacuum ultraviolet (VUV) flood exposures,[41] chemical hardening,[42] and design of resists that can be hardened with a high-temperature bake without flowing.[43] Similarly, spacer processes can be simplified by depositing spacers at sufficiently low temperatures so that they can be deposited directly onto resist patterns.

Double patterning is a way to overcome the intrinsic physical resolution limits of single exposures and create patterns on wafers with pitches that are smaller than could be patterned in a single lithography step. This approach increases

the complexity and cost of lithography, essentially by doubling the number of processing steps to generate a layer of a circuit, but it does enable a significant extension for optical lithography. Design layouts need to be modified so they can be decomposed into two masking steps. For this reason, there is still motivation to find lithographic technologies with resolution beyond that of ArF immersion-projection lithography. One such method that is used in laboratory settings is described in the next section, and techniques under consideration for use in high-volume manufacturing are the subjects of Chapters 12 and 13.

10.11 Interferometric Lithography

One method that potentially qualifies as an optical-lithography technique is interferometric lithography. This approach relies on the small-pitch grating that results when two coherent plane waves intersect. Suppose two plane waves $Ae^{ik(x\sin\theta - z\cos\theta)}$ and $Ae^{ik(-x\sin\theta - z\cos\theta)}$ intersect (Fig. 10.25), where $k = 2\pi/\lambda$ and λ is the wavelength of the light. The resulting light intensity is given by

$$4A^2\cos^2(kx\sin\theta)\,. \tag{10.18}$$

This light-intensity distribution has a pitch of

$$\frac{\lambda}{2\sin\theta}. \tag{10.19}$$

Continuous-wave DUV light at 213 nm can be generated by 5× frequency multiplying of the 1064-nm light output of a neodymium-YAG laser. For θ = 60 deg, this results in a pitch of 123 nm. Equating the half pitch to the resolution, this technique offers imaging capability below 70 nm.

One potential advantage of interferometric lithography is the possibility of extending it in an immersion configuration[44] using a fluid with a refractive index higher than that of water. In such a system, a prism is used to couple the light (Fig. 10.26). Because the gratings can be created using light of a single polarization and in one direction, the prism can be fabricated of materials with levels of

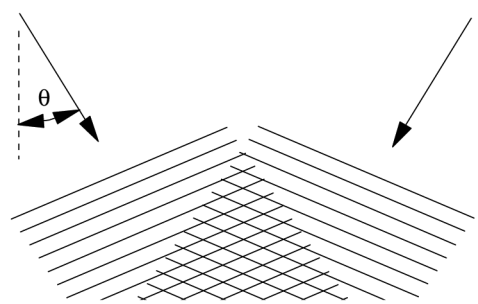


Figure 10.25 In interferometric lithography, a grating pattern is created by the interference of two plane waves.

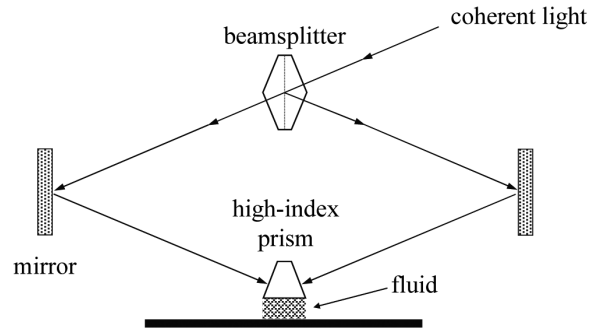


Figure 10.26 Configuration of an immersion interferometric lithography system.

birefringence that are too high for use in lenses. One such material is aluminum oxide, with an index of refraction of 1.92 at $\lambda = 193$ nm.[45]

Interferometric lithography is a technique capable of patterning gratings of very fine pitches, a method of great utility for making high-resolution optical gratings. It is less straightforward to extend this technique to the patterning of random logic patterns. Some approaches have been proposed,[46,47] but there has not yet been a commercial attempt to implement them and verify their practicality. Nevertheless, interferometric lithography has proven to be a useful method for studying fundamental resist issues, such as how resists behave when immersed in fluids during exposure. Interferometric lithography is a useful tool for the laboratory, even if it cannot yet be practically used to pattern integrated circuits.

The extension of optical lithography to its ultimate limits will require that many problems be solved. These include full implementation of the topics presented in Chapter 8, along with further refinement of lenses, exposure tools, and masks. While technical solutions for these problems may be found, their cost becomes a concern. The cost of lithography is the subject of the next chapter.

Problems

10.1 Show that the approximate depth-of-focus given by

$$DOF = \frac{\lambda}{NA^2}$$

is always larger than the exact DOF given by the expression in Eq. (10.4).

10.2 Show that the linewidth reduction possible by changing the wavelength from 193 nm to 157 nm is 21%, assuming equivalence in all other aspects of the lithographic processes. Is this sufficient to advance lithography one node? How does this resolution improvement compare with the benefits of ArF water-immersion lithography?

10.3 Assuming $\lambda = 193$ nm and $NA = 1.35$, show that 40-nm lines and spaces is $k_1 = 0.28$. How much smaller do think is possible from such optics?

10.4 Suppose that a single drop of water is left on a 300-mm-diameter (775-μm-thick) silicon wafer. If the wafer is thermally isolated except for the evaporation of the water, show that the wafer temperature will decrease by ~1 °C due to the evaporation. Note that the latent heat of evaporation for the wafer is 2490 J/g, the specific heat of silicon is 0.7 J/gC, and the density of silicon is 2330 kg/m^3. Do you think that water evaporation could lead to overlay problems with immersion lithography?

References

1. C. A. Winslow, *Elements of Applied Microscopy*, John Wiley & Sons, New York (1905).
2. 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.* **3**(1), 68–72 (2004).
3. G. Nellis and A. Wei, "Preliminary analysis of laser-pulse-induced pressure variation for immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**(1), 84–86 (2004).
4. S. Owa and H. Nagasaka, "Advantage and feasibility of immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**(1), 97–103 (2004).
5. S. Owa, K. Shiraishi, S. Nagaoka, T. Fujiwara, and Y. Ishii, "Full field, ArF immersion projection tool," *Proc.*17th*Ann. SEMI/IEEE Adv. Semicond. Manuf. Conf.*, 63–70 (2006).
6. P. L. Marsten, "Light scattering from bubbles in water," *Proc. Oceans '89*, 1186–1193 (1989).
7. L. Tsang, J. A. Kong, and K. H. Ding, *Scattering of Electromagnetic Waves: Theories and Applications*, Wiley, New York (2000).
8. T. Gau, C. -K. Chen, and B. J. Lin, "Image characterization of bubbles in water for 193-nm immersion lithography–far-field approach," *J. Microlith. Microfab. Microsyst.* **3**(1), 61–67 (2004).
9. D. Ngo and G. Videen, "Light scattering from spheres," *Opt. Eng.* **36**(1), 150–156 (1997).
10. P. S. Epstein and M. S. Plesset, "On the stability of gas bubbles in liquid-gas solutions," *J. Chem. Phys.* **18**(11), 1505–1509 (1950).
11. A. H. Harvey, S. G. Kaplan, and J. H. Burnett, "Effect of dissolved air on the density and refractive index of water," *Int. J. Thermophys.* **26**(5), 1495–1514 (2005).
12. A. Suzuki, "Immersion lithography update," *Sematech* 2nd*Immersion Lithog. Work.* (July, 2003).
13. A. C. Tobey, "Wafer stepper steps up yield and resolution in IC lithography," *Electronics*, 109–112 (August 16, 1979).

14. J. Lyman, "Optical lithography refuses to die," *Electronics*, 36–42 (October 7, 1985).

15. H. J. Levinson and W. H. Arnold, "Focus: the critical parameter for submicron lithography," *J. Vac. Sci. Technol.*, 293–298 (1987).

16. M. D. Levenson, "Wavefront engineering for photolithography," *Physics Today*, 28–36 (July, 1993).

17. Dr. Bernd Geh, Private Communication (1999).

18. A. C. Wei, G. F. Nellis, R. L. Engelstad, C. Chen, M. Switkes, and M. Rothschild, "Microfluidic simulations for immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**(1), 28–34 (2004).

19. M. K. Yang, S. G. Kaplan, R. H. French, and J. H. Burnett, "Index of refraction of high-index lithographic immersion fluids and its variability," *J. Micro/Nanolith, MEMS, MOEMS* **8**(2), 023005 (2009).

20. K. Sakai, T. Iwasaki, S. Mori, A. Yamada, M. Ogusu, K. Yamashita, T. Nishikawara, S. Hara, and Y. Watanabe, "Feasibility study on immersion system using high-index materials," *Jpn. J. Appl. Phys* **47**, 4853–4861 (2008).

21. R. Gupta, J. H. Burnett, U. Griesmann, and M. Walhout, "Absolute refractive indices and thermal coefficients of fused silican and calcium fluoride near 193 nm," *Appl. Optic.* **37**(25), 5964–5968 (1998).

22. B. W. Smith, Y. Fan, M. Slocum, and L. Zavyalova, "25 nm immersion lithography at a 193-nm wavelength," *Proc. SPIE* **5754**, 141–147 (2004).

23. J. H. Burnett, S. G. Kaplan, E. L. Shirley, D. Horowitz, W. Clauss, A. Grenville, and C. V. Peski, "High-index optical materials for 193-nm immersion lithography," *Proc. SPIE* **6154**, 615418 (2006).

24. K. Kurosawa, Y. Takigawa, W. Sasaki, M. Okuda, E. Fujiwara, K. Yoshida, and Y. Kato, "High-power operation of an argon excimer laser with a MgF_2 and SiC cavity," *IEEE J. Sel. Top. Quant. Electron.* **26**(1), 71–76 (1991).

25. S. Neeser, M. Schumann, and H. Langhoff, "Improved gain for the Ar_2^* excimer laser at 126 nm," *Appl. Phys. B* **63**, 103–105 (1996).

26. K. Kurosawa, R. Sonouchi, A. Yokotani, W. Sasaki, M. Kattoh, Y. Takigawa, and K. Nishimura, "Fabrication, characteristics and performance of diamond mirrors for vacuum ultraviolet excimer lasers," *Opt. Eng.* **34**(5), 1405–1409 (1995).

27. M. Katto, M. Okuda, W. Sasaki, K. Kurosawa, and Y. Kato, "Electron beam pumped argon-excimer laser using an unstable resonator," *IEEE J. Sel. Top. Quant. Electron.* **1**(3), 924–930 (1995).

28. Dr. Mordechai Rothschild, Private Communication (2000).

29. *Handbook of Optical Constants of Solids*, E. D. Palik, Ed., Academic Press, Boston (1998).

30. B. Shamoun, R. Engelstad, and D. Trost, "Assessment of thermal loading–induced distortions in optical photomasks due to e-beam multipass patterning," *J. Vac. Sci. Technol. B* **16**(6), 3558–3562 (1998).

31. M. Switkes and M. Rothschild, "Resolution enhancement of 157-nm lithography by liquid immersion," *J. Microlith. Microfab. Microsyst.* **1**, 225–228 (2002).

32. As an historical note, the recognition that ArF water-immersion lithography should be considered seriously became apparent to many lithographers during the question and answer session following a presentation in 2002 on 157-nm immersion lithography during the 3rd International Symposium on 157-nm Lithography in Antwerp.

33. R. R. Kunz, M. Switkes, R. Sinta, J. E. Curtin, R. H. French, R. C. Wheland, C.-P. C. Kao, M. P. Mawn, L. Lin, P. Wetmore, V. Krukonis, and K. Williams, "Transparent fluids for 157-nm immersion lithography," *J. Microlith. Microfab. Microsyst.* **3**(1), 73–86 (2004).

34. H. J. Levinson and J. Kye "Optical lithography: how far can it go?" *Semicon Korea 2001 Technical Symposium* (2001).

35. R. Socha, M. Dusa, L. Capodieci, J. Finders, F. Chen, D. Flagello, and K. Cummings, "Forbidden pitches for 130-nm lithography and below," *Proc. SPIE* **4000**, 1140–1155 (2000).

36. W. J. Poppe, L. Capodieci, J. Wu, and A. Neureuther, "From poly line to transistor: building BSIM models for non-rectangular transistors," *Proc. SPIE* **6156**, 61560P (2006).

37. This figure was provided by Dr. Lars Liebmann.

38. A. J. Hazelton, S. Wakamoto, S. Hirukawa, M. McCallum, N. Magome, J. Ishikawa, C. Lapeyre, I. Guilmeau, S. Barnola, and S. Gaugiran, "Double patterning requirements for optical lithography and prospects for optical extension with double patterning," *Proc. SPIE* **6924**, 69240R (2008).

39. S. Tarutani, H. Tsubaki, and S. Kanna, "Development of materials and processes for double patterning toward 32-nm node 193-nm immersion lithography process," *Proc. SPIE* **6923**, 69230F (2008).

40. C. Bencher, Y. Chen, H. Dai, W. Montgomery, and L. Huli, "22 nm half-pitch patterning by CVD spacer self alignment double patterning (SADP)," *Proc. SPIE* **6924**, 69244E (2008).

41. N. Bekiaris, H. Cervera, J. Dai, R. Kim, A. Acheta, T. Wallow, J. Kye, H. J. Levinson, T. Nowak, and J. Yu, "A lithographic and process assessment of photoresist stabilization for double-patterning using 172 nm photoresist curing," *Proc. SPIE* **6923**, 692321 (2008).

42. M. Hori, T. Nagai, A. Nakamura, T. Abe, G. Wakamatsu, T. Kakizawa, T. Anno, M. Sugiura, S. Kusumoto, Y. Yamaguchi, and T. Shimokawa, "Sub-40nm half-pitch double patterning with resist freezing process," *Proc. SPIE* **6923**, 69230H (2008).

43. K. R. Chen, W. Huang, W. Li, and P. R. Varanasi, "Resist freezing process for double exposure lithography," *Proc. SPIE* **6923**, 69230G (2008).

44. 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), 3306–3309 (1999).

45. B. W. Smith, Y. Fan, M. Slocum, and L. Zavyalova, "25 nm immersion lithography at a 193 nm wavelength," *Proc. SPIE* **5754**, 141–147 (2004).

46. X. Chen and S. R. J. Brueck, "Imaging interferometric lithography–a wavelength division mulitplex approach to extending optical lithography," *J. Vac. Sci. Technol. B* **16**, 3392–3397 (1998).

47. X. Chen and S. R. J. Brueck, "Imaging interferometric lithography for arbitrary systems," *Proc. SPIE* **3331**, 214–224 (1998).

Chapter 11
Lithography Costs

11.1 Cost-of-Ownership

The high price tags of exposure tools have made the cost of lithography a concern since the advent of projection lithography, and lithography costs may ultimately limit patterning capability, more so than technical feasibility. While there will always be a market for electronics where price is secondary to performance, the large personal computer and portable phone markets have proven to be extremely elastic. To meet the demands of the consumer, lithography will need to be cost effective, in addition to providing technical capability. Lithography costs have several components. These include:

1. Capital equipment costs, throughput, and utilization
2. Consumables, such as photochemicals
3. Masks
4. Rework and yield
5. Metrology
6. Maintenance
7. Labor
8. Facilities.

These factors can be considered in various degrees of sophistication. A detailed cost-of-ownership model was generated by SEMATECH, and an enhanced version of this model is commercially available.[1] In this chapter, the basic components of such cost-of-ownership models are introduced and discussed.

Lithography tools are often the most expensive in the wafer fab. Even when they are not, the fact that lithography is required for patterning many layers in IC manufacturing processes, while most other tools are used for only a few steps, means that a large number of lithography tools are needed for each wafer fab, resulting in high total costs for lithography equipment.

Wafer steppers are the most expensive pieces of equipment in the lithography tool set. Their prices have increased by an average of 17% per year since they were introduced in the late 1970s, to the point where the prices for leading-edge step-and-scan systems now exceed $50M (Figs. 11.1 and 11.2).

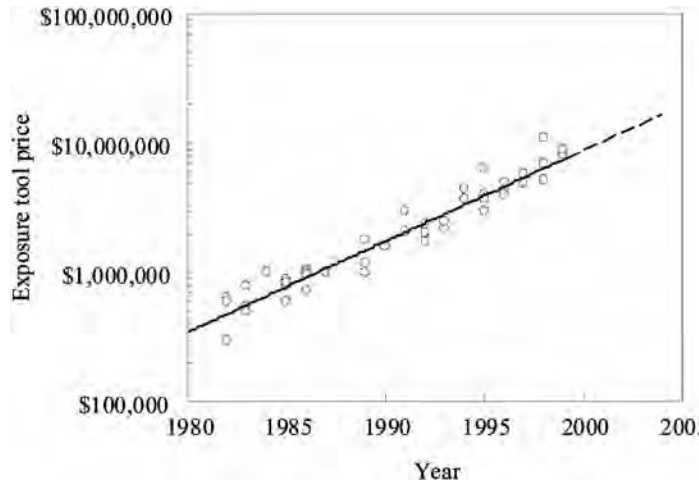


Figure 11.1 Stepper prices over time (in U.S. dollars), collected by SEMATECH.[2] Note the logarithmic scale.

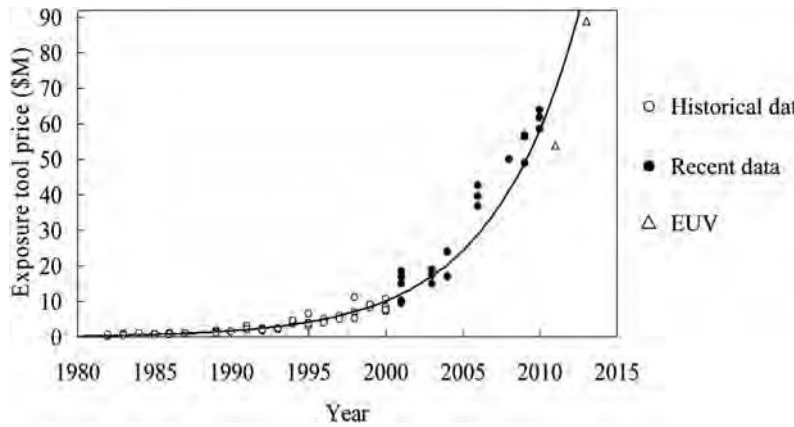


Figure 11.2 Stepper prices over time (in U.S. dollars), with costs from the newest tools added to those included in Fig. 11.1. The data are also plotted on a linear scale. The solid line is an exponential least-squares fit to the SEMATECH data. The prices for the EUV tools (EUV lithography is the subject of the next chapter) are from Wüest, Hazelton, and Hughes.[3]

The cost of lithography can be measured in three ways:

1. The cost per wafer
2. The cost per chip
3. The cost per unit function, such as a bit of memory.

The use for each metric will become more apparent as the components of lithography costs are reviewed.

11.1.1 Capital costs

Assessment of the impact of the increased cost of today's steppers, relative to earlier steppers, must include consideration of the increased throughput and

capability of newer machines. For example, the first commercially available stepper, the GCA DSW4800, cost only $300,000, but it was capable of processing only ten wafers (4 in.) per hour. Modern exposure tools are capable of exposing 140+ wafers (300 mm) per hour. In terms of cost per wafers per hour, the GCA was considerably more cost effective, but one comes to a different conclusion when the amount of silicon area produced per hour is considered (see Table 11.1). The situation over time is plotted in Fig. 11.3. What is interesting about Fig. 11.3 is how relatively constant lithography costs have been, in terms of dollars per silicon area exposed per hour, despite an exponential increase in the prices of individual exposure tools. While there has been some increase in the cost of exposing silicon area with immersion lithography, these increases are significantly less than the exponential growth in exposure-tool prices shown in Figs. 11.1 and 11.2.

There are additional benefits from advanced lithography, because higher resolution enables greater functionality to be packed into every unit area of silicon. Even though exposure-tool costs have increased significantly over time, the cost

Table 11.1 Comparison of capital costs and capability for the GCA DSW4800 and a modern step-and-scan system. The bit throughput assumes that the bit size is proportional to the square of the resolution, which changed from 1.25 μm for the GCA DSW4800 to 40 nm for modern 193-nm immersion systems. This does not take into account any nonlithographic innovations that have enabled reductions in bit size.

	GCA DSW4800	**Modern ArF imersion step-and-scan**
Exposure-tool price	$300K	$50M
Relative prices	1.0	167
Throughput	10 wafers (4 in.) per hour	140 wafers (300 mm) per hour
Relative capital cost/wafer throughput	1.0	11.9
Relative capital cost/silicon area throughput	1.0	1.32
Relative capital cost/bit throughput	1.0	0.001

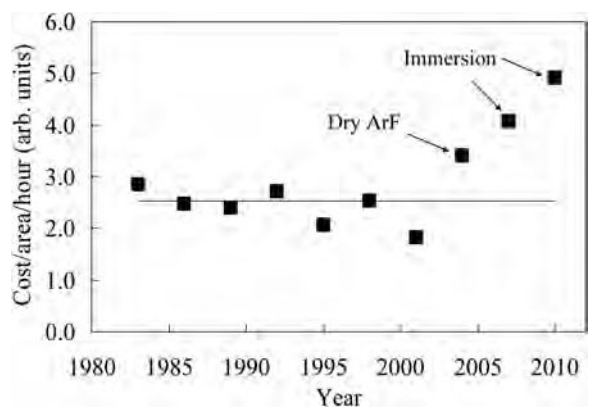


Figure 11.3 Lithography costs per unit area exposed per hour. The solid line is the average for the dry tools.

per bit has plummeted. In terms of the value to the consumer of microelectronics, modern wafer steppers are extremely cost effective. However, for producers of chips, the very high capital costs require enormous investments. This cost has certainly contributed significantly to the transformation of the semiconductor industry from being comprised of a large number of small chip-making companies to only a few large producers and an assortment of companies that contract others to do their manufacturing. Since most cost-of-ownership analyses are oriented toward a time period in which the wafer size is constant, the most common type of cost-of-ownership analysis performed is the cost per wafer, which is the metric discussed in this chapter. Extension to area or bit cost analyses can be obtained by scaling the wafer cost of ownership appropriately.

As noted, the cost of lithography is strongly affected by equipment throughput. The basic capital cost per wafer exposed is

$$\text{cost/wafer} = \frac{C_{ED}}{T_p U}, \tag{11.1}$$

where C_{ED} is the capital depreciation per hour, T_p is the raw throughput of the system (in wafers per hour), and U is the fractional equipment utilization. Raw throughput is the number of wafers per hour that can be aligned and exposed by a stepper operating in the steady state. If lithography equipment is configured to have exposure tools interfaced with resist-processing equipment, then C_{ED} should be the capital depreciation per hour of the total photocluster, and U is the overall photocluster utilization. T_p is the photocluster throughput, which may be determined by the throughput capability of the resist processing equipment, rather than by the exposure tool.

Early efforts to improve lithography productivity centered on stepper throughput. The basic model for the raw throughput of step-and-repeat systems, in wafers per hour, is

$$\text{throughput} = \frac{3600}{t_{OH} + N(t_{exp} + t_{step})}, \tag{11.2}$$

where N is the number of exposure fields per wafer, t_{exp} is the exposure time per field, t_{step} is the amount of time required to step between fields, including the time for vibrations to settle out ("step-and-settle time") and the time required for focusing. Alignment time for a stepper using die-by-die alignment would also be included in t_{step}. The overhead time per wafer t_{OH} is the time required to remove a wafer from the chuck, place a new wafer onto the chuck, and align the wafer. All times in the right-hand side of Eq. (11.2) are measured in seconds, and for this equation it is assumed that only a single reticle is used per wafer.

The GCA4800 had a 10× lens with a 10-mm × 10-mm field size. With such a small field, many exposures [N in Eq. (11.2)] were required per wafer. One of the first steps taken to improve productivity was to decrease the lens reduction to 5×. Field sizes (on the wafer) were increased to diameters > 20 mm, enabling

14-mm × 14-mm and 15-mm × 15-mm fields. Over time, field diameters increased to over 31 mm, capable of supporting 22-mm × 22-mm fields. These bigger fields provided higher throughput and could support large chip sizes. The migration from 10-mm × 10-mm fields to 22-mm × 22-mm fields led to a decrease in N, and greatly increased stepper throughput. It should be noted that the reduction in N does not scale exactly to the ratio of exposure-field areas, because partial fields are often exposed at the edges of wafers, but the transition from 10× to 5× lens reduction reduced N by approximately a factor of 4. With step-and-scan, even larger exposure fields have become available, and tools with exposure fields as large as 26 mm × 33 mm are typical. Nikon and Canon have introduced step-and-repeat systems that match this field size. These tools will be discussed further in Section 11.2.

Decreases in N are beneficial, to the extent that they are not offset by increases in t_{exp} and t_{step}. If the total light energy available to expose a field remained constant, then t_{exp} would need to increase in direct proportion to any increase in field size, because the fixed amount of light would be spread over a larger area, and Nt_{exp} would remain fairly constant, assuming that N roughly scales inversely with field area. Because of edge fields, Nt_{exp} could actually increase without an improvement in total illumination flux, and the only benefit to larger fields would come from the reduction in the Nt_{step} term. Fortunately, another important way in which stepper productivity has improved has been in terms of light flux. If I is the light intensity, and S is the resist sensitivity (exposure dose required to achieve the desired resist dimension on the wafer), then

$$t_{exp} = \frac{S}{I}. \tag{11.3}$$

Early wafer steppers provided <100 mW/cm^2 of light to the wafer plane, while intensities for i-line exposure tools that exceed 4000 mW/cm^2 are common today. These increases in light intensity result in proportional reductions in t_{exp}. Photoresist sensitivity has also improved over time, particularly with the advent of chemically amplified resists. DUV systems, where chemically amplified resists are used, can have exposure intensities <1000 mW/cm^2 while still maintaining practical productivity.[4] Today, fields are often exposed in <0.1 sec. There are practical limits to how short the exposure time can be, since there will be dose-control problems associated with exceedingly short exposure times on step-and-repeat systems, and rapid exposures on step-and-scan systems are limited by the fastest controllable stage-scanning speed. Scanning speeds are also limited by the combination of pulse-to-pulse energy stability and excimer-laser-pulse frequency, as discussed in Chapter 5.

Throughput improvement from large field sizes (smaller N) is somewhat offset by the need to step or scan longer distances. Increases in field sizes require longer stepping distances, so decreases in N are somewhat offset by increases in t_{step}. However, there have been significant improvements in t_{step} as a consequence of improvements in wafer-stage technology. Stages must be capable of high velocity, acceleration, deceleration, and jerk (third time derivative of position).

Prior to exposure, the stage position must be returned to a stable position, and the wafer must also be properly focused and leveled. Advanced mechanical modeling techniques have been used to optimize stage designs, and state-of-the-art control techniques and electronics are used. It should be noted that t_{step} in Eq. (11.2) is an average step-and-settle time, and actual times will differ in the x and y directions, particularly for nonsquare fields. Typical values for the parameters used in Eq. (11.2), for contemporary i-line wafer steppers, are given in Table 11.2.

The overhead time t_{OH} includes wafer transport times and the time for global alignment. There is a clear tradeoff between alignment precision, enhanced by a large number of alignment sites, and throughput. If multiple reticles are used per wafer, then the time to exchange reticles can be included in t_{OH} (see Problem 11.2).

Step-and-scan systems have somewhat modified throughput models. The basic throughput equation, Eq. (11.2), is applicable in modified form. For step-and-scan systems, exposure times are given by

$$t_{exp} = \frac{H_F + H}{v}, \tag{11.4}$$

where H_F is the height of the scanned field, H is the slit height, and v is the scanning speed (Fig. 11.4). All sizes and speeds must consistently refer to either the wafer or reticle. There needs to be a certain amount of scanning startup time, so the stages are moving at a constant controlled speed when exposure actually occurs, and this time can be included in t_{step}. To incorporate resist sensitivity into the model, the exposure dose S needs to calculated. If $I(y)$ is the intensity at position y in the slit, then the exposure dose in a scanner is given by

$$S = \frac{\int_0^H I(y)dy}{v}, \tag{11.5}$$

where H is the slit height. In terms of the average intensity $\bar{I}$,

$$S = \frac{\bar{I}H}{v}. \tag{11.6}$$

Table 11.2 Typical throughput parameters for i-line steppers, for 300-mm wafers.

Parameter	Value
N	76
t_{OH} (single reticle/wafer)	9 sec
I	4000 mW/cm^2
S	200 mJ/cm^2
t_{exp}	0.05 sec
t_{step}	0.07 sec
throughput	200 wafers/hour

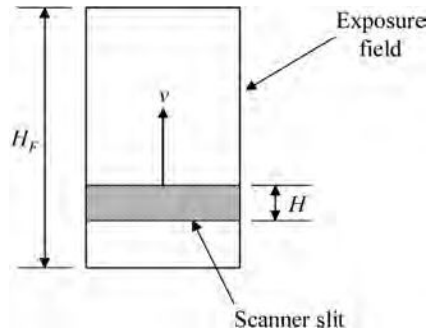


Figure 11.4 The parameters for Eq. (11.4).

The exposure time, in terms of the resist sensitivity, is given by

$$t_{exp} = \left(\frac{S}{\bar{I}}\right)\frac{H_F + H}{H}. \tag{11.7}$$

Comparing this to Eq. (11.3), to achieve the same exposure time for a given exposure dose, the light intensity must be greater for a scanner, compared to a step-and-repeat system, by the factor

$$\frac{H_F + H}{H}. \tag{11.8}$$

For a typical field height $H_F = 25$ mm and slit heights of 5–10 mm, the exposure intensity must be greater by a factor of 3.5–6 in order for the scanner to have the same exposure times as an equivalent step-and-repeat system. However, the light for a scanning exposure tool is being spread over an area that is smaller than on a step-and-repeat system with the same field size by the factor

$$\frac{H}{H_F}. \tag{11.9}$$

Illuminating a smaller area is not difficult with laser-light sources that produce collimated light, but this is a challenge for arc-lamp systems that must gather light emitted in many different directions.

Improvement in throughput with taller fields is not what is predicted by changing N in Eq. (11.2), because longer scans are needed. For example, suppose the number of fields is cut in half by a 2× increase in scanning height:

$$N \to \frac{N}{2}, \tag{11.10}$$

$$H_F \to 2H_F. \tag{11.11}$$

The time to process a wafer for a field height H_F is

$$t_{OH} + N(t_{step} + t_{exp}) = t_{OH} + N\left(t_{step} + \frac{H_F + H}{v}\right), \tag{11.12}$$

$$= t_{OH} + Nt_{step} + N\frac{H_F}{v} + N\frac{H}{v}. \tag{11.13}$$

Upon doubling the scan length, this process time changes to

$$\rightarrow t_{OH} + \frac{N}{2}t'_{step} + N\frac{H_F}{v} + \left(\frac{N}{2}\right)\left(\frac{H}{v}\right). \tag{11.14}$$

The step time is modified as $t_{step} \rightarrow t'_{step}$, indicating slightly longer stepping times with the longer fields. While two terms are cut in half (or nearly half) by the reduction in exposure fields, there is one term, $N(H_F/v)$, that remains proportional to N. Consequently, the benefit of fewer fields is less than proportional to the reduction in the number of fields.

Equipment downtime and setup time detract from the equipment output, and these are included in the utilization factor in Eq. (11.1). A standard set of states has been defined to facilitate the analysis of semiconductor manufacturing equipment productivity (Fig. 11.5).[5] Productivity is enhanced when the amount of time that the equipment is in the productive state is maximized. The most obvious loss of productivity occurs when tools are in a nonfunctional state. This is equipment downtime. Typical downtime for lithography tools is 5–10%. Unscheduled downtime occurs when equipment breaks unexpectedly and requires repair. Scheduled downtime occurs for lamp changes, excimer-laser-window replacements, the cleaning of resist-processing equipment, and similar types of equipment maintenance that are necessary and can be planned. Engineering time is the time the equipment is used for process and equipment engineering. To the extent that process engineering is an objective of the wafer fab, this can be considered useful time. It is not valuable time when process engineering is using the tool to analyze a tool-related problem.

Standby time is often a significant detractor from productivity. It includes the time that no operators are available, the time waiting for the results of production tests, and the time during which there is no product available. When test wafers are needed to set up individual lots, standby time can be high. It should be noted that optimum fab operation will necessarily result in some standby time when there is no product available. This can be understood as follows. Because the times are variable when wafers are being processed outside of lithography, the work available for lithography will also vary. Consider a lithography sector with only a single photocluster. From queuing theory, the number of lots waiting for processing N_q is related to the process time t and average rate R at which lots enter the lithography

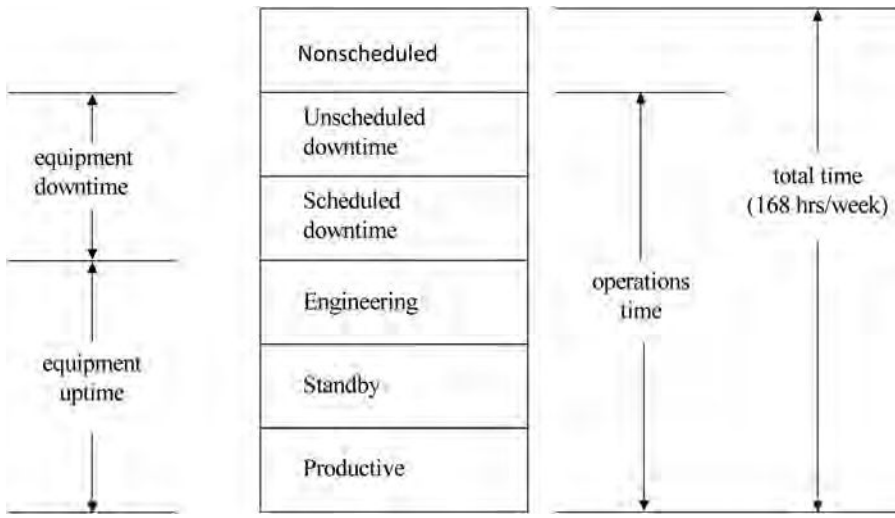


Figure 11.5 Equipment states.

sector by the following equation:[6]

$$N_q = \frac{R^2 \left\langle t^2 \right\rangle}{2\left(1 - \langle t \rangle R\right)}, \tag{11.15}$$

where $\langle \cdots \rangle$ denotes the time average of the quantity in the brackets. From the same theory, the fraction of time F_0 when there are no wafers in process or queue is given by

$$F_0 = 1 - \langle t \rangle R. \tag{11.16}$$

Maintaining 100% equipment utilization is equivalent to having $F_0 = 0$. Comparing Eqs. (11.15) and (11.16), one sees that an attempt to maintain 100% equipment utilization results in the number of lots awaiting processing N_q to tend toward infinity, assuming an infinite source of input. If we assume constant processing time, then

$$\left\langle t^2 \right\rangle = \langle t \rangle^2, \tag{11.17}$$

and Eqs. (11.15) and (11.16) can be solved. The result is plotted in Fig. 11.6. Similar but more complex equations apply for situations with multiple photoclusters, but the basic conclusion remains the same: the queues in front of the lithography operations grow very large unless some time is planned when no wafers are available. With multiple photoclusters, the size of the queue is reduced for a given fraction of time with zero queue. To maximize productivity, this time with no wafers needs to be kept as small as possible, but the optimum value is nonzero. This conclusion is unappealing to managers who want to avoid any

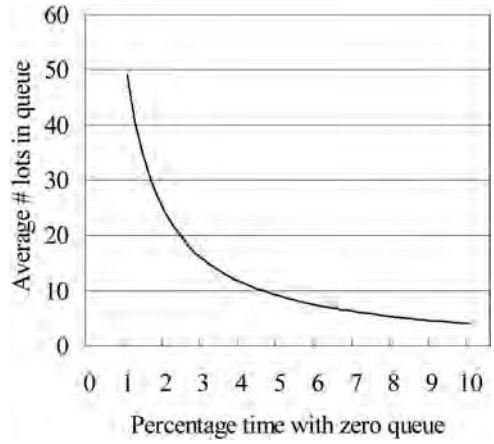


Figure 11.6 The number of lots increases as the percentage of time with no queue goes to zero for a single photocluster.

idle time for their expensive lithography equipment, yet also want to maintain good cycle times, which require small queues. There have been a number of fab managers who have tried to avoid what has been shown mathematically to be inescapable, to their eventual regret and failure.

The utilization factor U in Eq. (11.1) is the sum of the fraction of productive time and engineering times in the equipment-state diagram of Fig. 11.5. Part of the uptime is consumed because of the practical necessity of having a finite standby time to avoid large queues, but unfortunately, this is usually only a small part of the lost productive time. Another contributor is inadequate production organization, and it is clear that some companies are more effective at managing this than others. This is a management issue, not directly one of the science of lithography, and will not be discussed further in this book.

To place lithography capital costs in perspective, consider a photocell consisting of a step-and-scan system costing \$50M, interfaced to resist processing equipment that costs \$9M. If the equipment is depreciated over five years, then the straight depreciation cost is \$1347 per hour. If the raw throughput is 140 wafers per hour and the utilization is 80%, then the capital contribution to the cost per wafer is \$12. All other costs can be considered in comparison to this number. In this analysis, the cost of capital, which fluctuates with interest rates, was not included.

11.1.2 Consumables

Photochemicals are needed on a per-wafer basis. The most expensive chemical is photoresist, with typical costs shown in Table 11.3. Resist coaters use 1–2 cc per wafer (300 mm), so resist costs per wafer, in manufacturing, can run between \$0.10 and \$2.00 per wafer, assuming the lower volume/wafer resist use. Developer costs can add another \$0.20 to \$0.50 per wafer. If antireflection coatings are used, their costs must also be included. If these antireflection coatings are not essential parts of the device integration and are there solely to improve the lithography, their coating

Table 11.3 Typical resist costs.

Type of resist	Typical cost per liter
g or i line	$100–$300
248 nm	$250–$500
193 nm	$300–$2000

or deposition, etching, and stripping costs must also be included in determining lithography cost of ownership.

In addition to photochemicals, there are other consumable components and materials, such as filters, mercury-arc lamps, and excimer-laser gases. Lamp systems have significantly lower cost of ownership than laser systems, since the latter require periodic replacement of expensive laser cavities and laser optical components. The frequencies at which laser components are typically replaced are given in Table 11.4. Replacement frequencies depend upon the usage of the lasers, i.e., how many pulses have been fired. To put the number of pulses in Table 11.4 in perspective, a continuously operating 6-kHz laser fires 189 billion pulses per year. DUV exposure tools will typically be exposing wafers only ~30% of the time, even in volume manufacturing, but this still represents nearly 57 billion pulses annually. The total cost of consumables for a KrF excimer laser system has been estimated to be ~$300,000 per year, compared to only ~$60,000 per year for a lamp-based system,[7] while consumable costs for ArF lasers are higher than for KrF lasers. Lamp-based systems clearly have lower consumables costs.

Disposal costs also contribute to the cost of ownership, since mercury-arc lamps, fluorine filters (for excimer lasers), and used photochemicals all require disposal consistent with appropriate regards for worker safety and the human environment. However, these costs are usually negligible in comparison to other costs, and are often ignored when computing lithography cost of ownership.

11.1.3 Mask costs

Photomasks have become a significant contributor to lithography costs. This was not fully appreciated until one company, AMD, performed a study of internal mask usage, and it was found that an average reticle was used to expose only 1800–2400 wafers.[10,11] From a follow-up survey of SEMATECH member companies it was found that this level of mask usage was not atypical, although there is a wide range of reticle usage that varies from company to company. For makers of application-specific integrated circuits (ASIC), the mask usage can be low; 500 wafers per

Table 11.4 Representative ArF excimer-laser-component replacement frequency.[8,9]

Component	Replacement frequency (# pulses)
Gas	2 billion
Laser cavity	40 billion
Line-narrowing module	60 billion

reticle is considered typical for ASIC companies. For manufacturers of DRAMs or mainstream microprocessors, usage can easily be greater than 5000 wafers per reticle. The contribution of photomasks to the cost of ownership is a function of each company's business.

Binary reticles fabricated on 6-in. substrates cost \$1K–\$200K these days, depending on requirements. Reticles that are not specified to have tight requirements for registration and dimensional control can be made on depreciated equipment, and the costs for such reticles are dominated by substrate costs. Reticles that have tighter specifications for registration and linewidth control will need to be made using the latest mask-making and inspection equipment and consequently will cost more. At a usage rate of 1800 wafers per reticle, the mask contribution to the cost of lithography ranges from \$0.55–\$111.11 per wafer exposure, assuming that mask costs run from \$1K–\$200K. Comparing these costs to the estimates of capital costs given in Section 11.1, it can be seen that the cost of masks can easily exceed the cost of capital equipment. Moreover, photomask costs appear to be increasing faster than equipment costs, making it even more important to consider mask costs when pondering lithography cost of ownership. This is particularly true for manufacturers of application-specific integrated circuits. For a reticle usage rate of only 500 wafers per reticle, these costs skyrocket to between \$2–\$400 per wafer exposure, depending upon mask costs within the range of \$1K–\$200K. Note that these costs refer to single passes through the lithography operation. Because 20 or more masking steps are required to completely fabricate an integrated circuit, these mask costs contribute significantly to total manufacturing costs. With alternating phase-shifting masks or CPL, the mask costs can be double that for binary masks, or more.

These differences in the impact of mask costs on overall lithography costs per wafer indicate a divergence developing in the semiconductor industry. For manufacturers of DRAMs and mainstream microprocessors, mask costs are less significant than for makers of applications-specific integrated circuits. These different classes of manufacturers may pursue different solutions in the future, should mask costs continue to escalate.

The consequences of rising mask prices must be taken in the context of overall costs. It was estimated that the total development cost of a 90-nm technology applications-specific integrated circuit was approximately \$30M.[12] Of this cost, only \$2.4M was for the masks. For 90-nm technology, the expenses for masks, as a percentage of the total development cost, was on par with California's sales tax.

The adoption of new generations of lithographic technologies are worthwhile when transistor counts per chip increase significantly, because the greater expenses associated with new lithographic technologies is justified when it helps to avoid large die sizes. This matters only with integrated devices with large transistor counts. Device design costs will scale with transistor counts, becoming more expensive with increasing numbers of transistors. Although mask costs will be greater for future technologies (22-nm node and beyond), so will design costs. For high-volume products and devices of very great complexity, the increased mask costs can be justified. However, specialty products produced in low quantity and

for which market requirements dictate a low selling price will not be made in the most advanced technologies, because of reticle costs. As discussed in Chapter 13, considerations are also being given to maskless lithographic technologies for products made in low volumes.

11.1.4 Rework

Lithography is unique among semiconductor manufacturing processes in that it can be easily reworked. If patterns in resist are found to be sized or placed incorrectly, the resist can be stripped and the wafer reworked. While reworking is more cost effective than scrapping wafers, at least in the short term, rework still represents a significant cost, since lithography operations need to be performed multiple times to properly pattern the reworked wafers. If f is the fraction of wafers that need to be reworked, lithography costs are increased directly by a factor of $(1 + f)$ relative to costs in the absence of rework.

There are also indirect costs associated with rework. Levels of rework above a few percent indicate that there is poor process control. With high levels of rework it is difficult to organize production, because one cannot be certain when wafers will complete lithography operations. These indirect costs of rework often exceed the direct ones. To ensure that wafers with parameters within specifications are sent to postlithography processing, it is necessary to carefully measure a large fraction of the wafers. Consequently, a lithography operation with poor process control will need to have more metrology equipment. This metrology cost is discussed in more detail in the next section.

11.1.5 Metrology

For low-capability processes, it is necessary to measure wafers (linewidths, overlay, defects) in order to ensure that wafers are processed within specifications. A certain level of measurement is needed even for well-controlled processes, so it is difficult to state precisely which metrology tools have been purchased to meet minimum requirements, and which tools were added to address low process capability. Regardless, metrology costs do need to be factored into the overall cost of ownership, but some of those costs, necessitated by poor process control, are not easily quantified.

Measurement requires sophisticated tools, each costing millions of dollars. These measurement costs need to be included in any estimate of lithography costs. Facilities costs, discussed in Section 11.1.8, are usually low for these tools, but capital costs can be considerable. Since measurements are not made on every patterned wafer, the expression for metrology capital costs is similar to Eq. (11.1), but must be reduced by the fraction of wafers that are actually measured. Suppose that a fraction f_M of all wafers is measured for overlay, linewidths, and defects. Assuming approximately the same equipment utilization U for all metrology tools,

the capital cost per wafer for metrology is

$$\frac{f_M}{U} \sum_i \frac{C_{ED,i}}{T_{P,i}}, \tag{11.18}$$

where the sum is over the different types of metrology tools. If we assume the parameters of Table 11.5 and $f = 0.02$ (2% of the wafers are measured), $U = 0.8$, and straight five-year depreciation, then the metrology cost per wafer is \$1.36, which is greater than 10% of the capital cost from the photocluster. While this is a smaller cost than the photocluster capital cost or the mask cost, it is not insignificant.

11.1.6 Maintenance costs

Maintenance costs were partially captured in the utilization factor of Eq. (11.1). The biggest expense of downtime is the increased capital cost per wafer exposed. Additional maintenance costs include:

1. Parts costs. For certain subsystems, such as excimer lasers, these costs can be large. A big cost of parts is the cost of capital. An inventory of expensive parts is necessary in order to minimize downtime, but these parts have value. Hence, money is invested in spare parts.
2. Service contracts. Lithography tools, particularly wafer steppers and metrology equipment, are extremely complicated, often requiring very specialized expertise to repair them when they are broken. It often proves useful to contract this maintenance to the suppliers of the equipment, who may have maintenance personnel with adequate training.
3. Maintenance performed by one's own maintenance technicians requires that these people be trained and paid. The estimated depreciation for a modern lithography photocluster was previously estimated to be \$1347 per hour. People who can repair broken equipment quickly are essential. Of course, it is even more important that the lithography equipment be intrinsically reliable, to avoid large expenses from equipment downtime.

11.1.7 Labor costs

Labor costs are a relatively small fraction of overall lithography costs. A typical operator can operate two photoclusters. Including employee overhead, this cost is \$20 per hour (or less) per photocluster. At a net throughput of 112 wafers per hour

Table 11.5 Parameters for estimating metrology costs.

Metrology tool type	Capital cost (C_{ED})	Throughput (T_P, wafers/hour)
Linewidth measurement	\$3.5M	10
Overlay measurement	\$2M	80
Defect detection	\$8M	4

per photocluster (140 wafers per hour, 80% utilization), labor represents $0.18 per wafer. This is ~1.5% of capital equipment costs. Direct labor costs are easily offset by small differences in tool utilization, so a highly skilled workforce is a competitive advantage. Improving productivity by only one or two percent can completely compensate for direct labor costs.

11.1.8 Facilities costs

Facilities costs include the cost of cleanroom space, electricity, compressed air or nitrogen, deionized wafer, chilled water, and exhaust. Modern step-and-scan systems designed to expose 300-mm wafers have sizable footprints, as do the tools for resist processing that are interfaced to the exposure systems. Floor space for a complete photocluster can approach 60 m^2, including space for equipment access. A large modern wafer fab with 30,000 m^2 of cleanroom space can cost $0.5B in building costs alone. The cost of floor space of the photocluster is then

$$\$0.5\text{B} \times \frac{60}{30{,}000} = \$1\text{M}. \tag{11.19}$$

For an 80% utilized photocluster with raw throughput capability of 140 wafers per hour, the cost per wafer of this floor space is $0.20, assuming five years straight depreciation for the building. This is small compared to equipment capital costs and mask costs, but not insignificant.

11.2 Mix-and-Match Strategies

One way to mitigate the high cost of lithography is to use leading-edge (and expensive) machines only for the most critical layers, while using cheaper tools for the less-critical layers. This has been a common practice for many years. Exposure systems intended for use on layers with relatively low resolution requirements are usually optimized for throughput and price, rather than resolution and overlay. Representative noncritical-layer exposure tools are listed in Table 11.6. These systems achieve high throughput by having large field sizes, providing the potential to expose wafers in a smaller number of steps. In addition, they are all step-and-repeat systems. Although these are not leading-edge exposure tools, they still have excellent imaging capabilities. Lenses with such large fields and good imaging capability are the result of significant advances in lens technology in recent years, such as aspheric lens elements.

Another way to reduce capital cost is to use steppers for several generations of technology. As high-resolution steppers get older, they will no longer be capable of imaging the smallest geometries. Continuing to use these steppers, but not on the most critical layers, is an extremely cost-effective approach, since these steppers have already been purchased and are often fully depreciated.

There are many applications where the highest resolution steppers cannot be used for some masking layers. The numerical apertures required to image critical features <100 nm have limited depths-of-focus at larger feature sizes and are not

Table 11.6 Noncritical layer steppers.

	Nikon NSR-SF155	Nikon NSR-SF200	Canon FPA-5550iZ
Reduction	4:1	4:1	4:1
Maximum NA	0.62	0.63	0.57
Wavelength (nm)	365	248	365
Resolution	0.28 μm	0.15 μm	0.35 μm
Field size	26 × 33 mm	26 × 33 mm	26 × 33 mm

optimum for printing low-resolution or noncritical layers. Exposure tools capable of patterning such small features are usually DUV systems, and there are few DUV resists of sufficiently high viscosity for coating thick layers. Lower-numerical-aperture i-line steppers are useful for printing many implant masks or pad masks, where the photoresist can be thick. In these instances, high-resolution steppers may actually perform poorly because of their limited depths-of-focus. In general, lower-NA lenses are cheaper to design and build than higher-NA lenses, particularly when linewidth uniformity and distortion requirements are somewhat relaxed. These different stepper requirements, together with the potential for high throughput, can be exploited to reduce the overall cost of the lithography. It is common today to use different types of steppers to address the dissimilar requirements of different masking layers, while minimizing overall lithography costs. This practice is referred to as *mix-and-match* lithography.

When the field area expands by a factor of four, the light intensity will drop by about the same amount without changes in illuminator design and improved light sources. For example, the Nikon 4425i, a tool that has a 44-mm × 44-mm field size, was reported to have one third to one half the light intensity in the wafer plane as the corresponding 22-mm × 22-mm-field i-line stepper. Nevertheless, total stepping time was reduced, due to the decreased number of steps and reduced settling times, even though the stepping distance between fields was twice as long. Since overlay requirements are also relaxed for noncritical-layer exposure tools, settling time can be shortened, contributing further to high tool throughput.

While often worth the effort, implementing mix-and-match lithography is nontrivial. It may involve the use of steppers from more than one company, or steppers that use dissimilar operator interfaces. This requires additional training for operators, engineers, and maintenance technicians, and a separate spare parts inventory will need to be maintained. Making full use of very large fields provided by steppers may require nonconcentric fields, if one type of exposure tool has a smaller field than the other (Fig. 11.7). While the overlay requirements for the layers patterned by noncritical layer steppers are not the tightest, for a given technology, design rules for implant layers usually assume overlay capability within a factor of at least two of the most critical requirements. Achieving good overlay is complicated by the nonconcentric geometry shown in Fig. 11.7, particularly when one of the lenses in use has a nonrandom distortion component.[13] Even with concentric fields, there are overlay issues that occur only in mix-and-

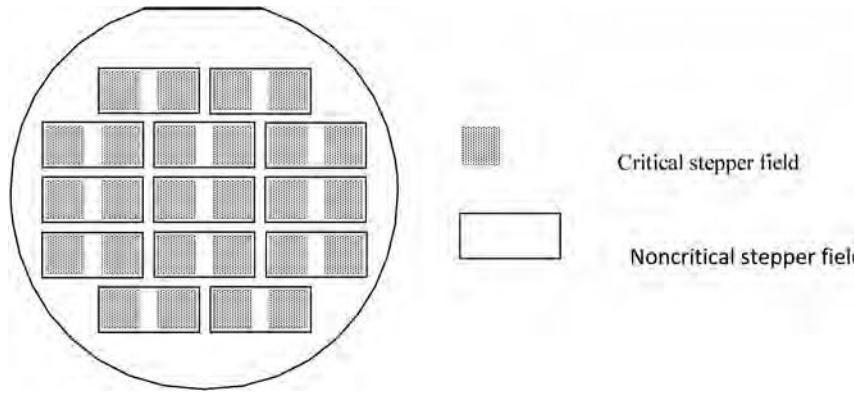


Figure 11.7 Nonconcentric fields when the exposure tools used for patterning critical and noncritical layers have different field sizes.

match situations. This is discussed in Chapter 6. In spite of these problems, mix-and-match lithography is usually worth the effort, since differences in capital costs between critical and noncritical layer machines can exceed $30M. Such savings in capital costs can pay for a great deal of training, spare parts, and engineering time to address the technical issues.

For a number of years, high-resolution step-and-repeat systems did have smaller fields than the tools used for noncritical layers, often leading to the type of situation illustrated in Fig. 11.7 when mix-and-match lithography was attempted. Today, modern high-performance systems are step-and-scan tools with large fields, which simplifies mix-and-match lithography. However, there may be a return in the future to smaller field sizes for the highest resolution tools in order to contain lens costs. If so, there will also be a return to the situation depicted in Fig. 11.7.

The cost of optical lithography equipment has escalated over the past two decades. Fortunately, as this technology has matured, there have been commensurate increases in productivity to offset the higher tool costs. As the diffraction limit is reached, it will be necessary for lithography to move to nonoptical technologies. Alternatives to optical lithography will need to be introduced with very high levels of productivity in order to have the same costs as their optical predecessors. Some of the possible successors to optical lithography are discussed in the next two chapters.

Problems

11.1 The total cost of consumables for an excimer-laser system has been estimated to be ~$300,000 per year and ~$60,000 per year for a lamp-based system. For an exposure system with a raw throughput capability of 140 wafers per hour and utilization of 80%, what is the contribution of light-source consumables to the per-wafer cost of ownership for excimer-laser light sources? And for lamp sources?

11.2 Derive an extension of Eq. (11.2) for a stepper appropriate for exposing two reticles per wafer. Assume that the wafer remains on the exposure chuck while the reticles are exchanged.

References

1. W. Trybula and D. Dance, "Cost of mask fabrication," *Proc. SPIE* **3048**, 211–215 (1997).
2. M. Mason, Sematech Next Generation Lithography Workshop (1998).
3. A. Wüest, A. J. Hazelton, and G. Hughes, "Estimation of cost comparison of lithography technologies at the 22-nm half-pitch node," *Proc. SPIE* **7271**, 72710Y (2009).
4. N. Deguchi and S. Uzawa, "150-nm generation lithography equipment," *Proc. SPIE* **3679**, 464–472 (1999).
5. SEMI E10–96 Standard, "Guideline for definition and measurement of equipment reliability, availability, and maintainability (RAM)," Semiconductor Equipment and Materials International, San Jose, CA.
6. A. O. Allen, *Probability, Statistics, and Queuing Theory: With Computer Science Applications*, Academic Press, Harcourt Brace Jovanovich, Boston (1990).
7. R. DeJule, "More productivity at subcritical layers," *Semicond. Int.* **50**, (April,1998).
8. H. Tsushima, M. Yoshino, T. Ohta, T. Kumazaki, H. Watanabe, S. Matsumoto, H. Nakarai, H. Umeda, Y. Kawasuji, T. Suzuki, S. Tanaka, A. Kurosu, T. Matsunaga, J. Fujimoto, and H. Mizoguchi, "Reliability report of high power injection lock laser light source for double exposure and double patterning ArF immersion lithography," *Proc. SPIE* **7274**, 72743L (2009).
9. K. O'Brien, W. J. Dunstan, R. Jacques, and D. Brown, "Lithography line productivity impact using Cymer GLX^{TM} technology," *Proc. SPIE* **7274**, 72743N (2009).
10. K. Early and W. H. Arnold, "Cost of ownership for soft x-ray lithography," OSA Topical Meeting on EUV Lithography, Monterey, CA (May, 1993).
11. K. Early and W. H. Arnold, "Cost of ownership for 1× proximity x-ray lithography," *Proc. SPIE* **2087**, 340–349 (1993).
12. D. Mansur, "Components of IC design cost relative to the market opportunity," *Adv. Reticle Symp.*, San Jose (2003).
13. M. E. Preil, T. Manchester, and A. Minviell, "Minimization of total overlay errors when matching nonconcentric exposure fields," *Proc. SPIE* **2197**, 753–769 (1994).

Chapter 12
Extreme Ultraviolet Lithography

Since the resolution capability of lithography can be extended by using short-wavelength light, at least in principle, a number of concepts involving light with wavelengths much shorter than 193 nm have been proposed. Considerable effort has been applied to the development of one of these approaches, referred to as extreme ultraviolet (EUV) lithography. In this chapter, the basic concepts underlying EUV technology are discussed.

12.1 Background and Multilayer Reflectors

As the wavelength of light decreases significantly below 193 nm, all known materials become too absorbing to be used for fabricating effectual refractive optical elements (see Fig. 12.1). Moreover, at such short wavelengths the reflectivity of all homogenous materials becomes very small, at least at the near-normal angles of incidence relevant to high-resolution imaging optics (see Fig. 12.2). However, in the 1980s layered coatings were developed to provide practical reflectivities at wavelengths <15 nm.[1,2] This development led to proposals for lenses with all-reflecting optics that could be used for projection lithography.[3,4]

Reflection occurs at interfaces between materials that have different indices of refraction. The larger the difference between the refractive indices of the materials, the greater the reflectivity. At wavelengths <50 nm, all materials have indices of refraction ≈1. Thus, it is difficult to create high reflectance from a single interface, except at grazing angles of incidence. At EUV wavelengths, it has proven possible to make mirrors with moderate reflectivity at near-normal angles of incidence, in the range of 60–70%, by the use of coatings comprised of multiple layers. Multilayer reflectors are made by depositing alternating layers of high-Z (Z is atomic number) and low-Z materials, giving a small but effective difference between refractive indices at each interface. The net effect of small reflectivity at each interface can lead to moderately high reflectivity overall when the stack has a sufficient number of layers (Fig. 12.3), provided the layer stacking satisfies, at least approximately, the Bragg condition:[2]

$$d = \frac{m\lambda}{2\cos\theta\sqrt{1 - \dfrac{2\bar{\delta} - \bar{\delta}^2}{\cos^2\theta}}}, \quad m = 1, 2, \ldots, \tag{12.1}$$

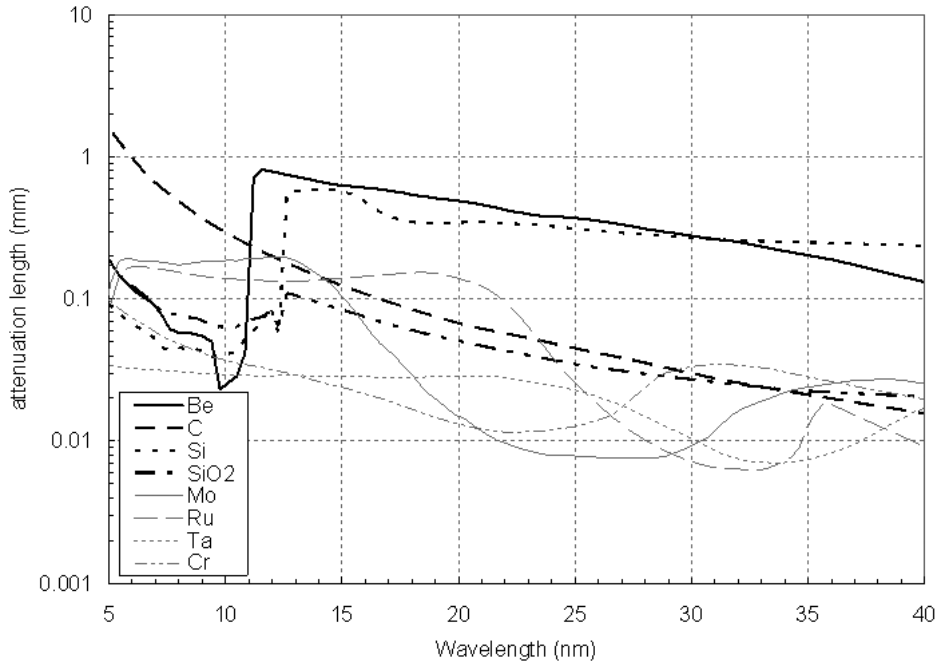


Figure 12.1 Attenuation length for selected materials at short wavelengths.[5,6] The attenuation length is the distance at which the intensity of light propagating through a material falls to $1/e$ of its initial value.

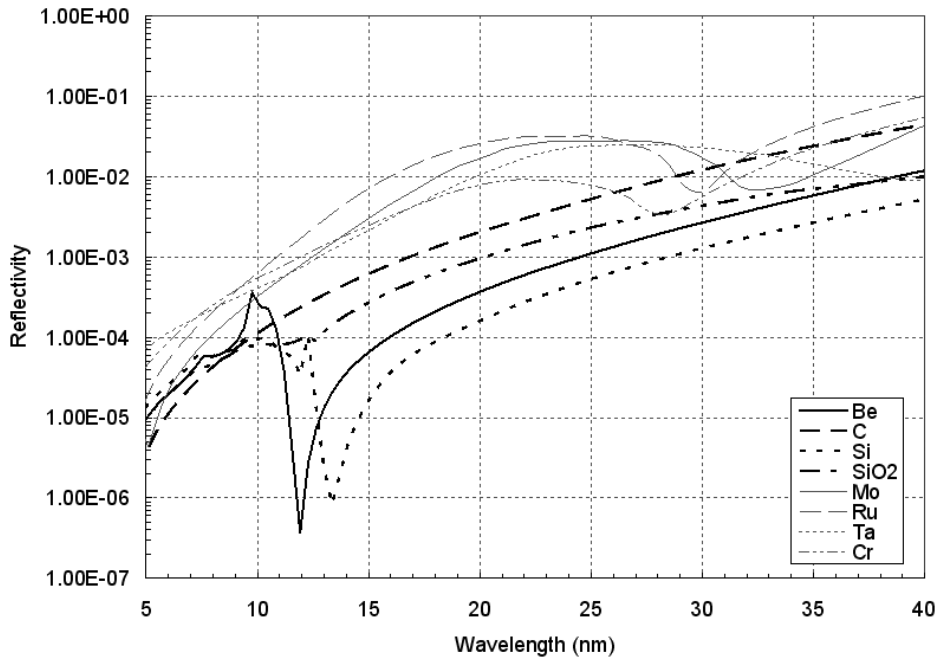


Figure 12.2 Reflectivity of selected materials at short wavelengths for normal incidence.[5,6]

where

$$\bar{\delta} = \frac{\delta_A d_A + \delta_B d_B}{d_A + d_B}. \tag{12.2}$$

The index of refraction of each layer is given by $n = 1 - \delta + i\beta$, while d_A, d_B, and θ are defined in Fig. 12.3, and $d = d_A + d_B$.

Although all materials are highly absorbing at EUV wavelengths, some are less absorbing than others. The materials chosen to comprise the multilayer stack must be relatively weak absorbers of EUV light, since the light must be able to penetrate to the lower layers of the film stack in order achieve high total reflectivity. Several stacks have been identified as possible reflectors in the EUV, and these are listed in Table 12.1. Because of beryllium's toxicity, there is a reluctance to use that metal for EUV lithography, and nearly all effort in EUV lithography today is focused on Mo/Si reflectors. For Mo/Si multilayers designed for near-normal reflectance, the molybdenum layers are ~3 nm thick, while the silicon layers are ~4 nm thick.[7] Because reflectance is angle dependent [Eq. (12.1)], graded depositions are needed on optics that have light incident at varying angles over the mirror surfaces in order to maintain constant reflectance at the operating wavelength.

Reflectance versus wavelength for a Mo/Si multilayer stack is shown in Fig. 12.4. As can be seen in this figure, reflectance peaks around a given

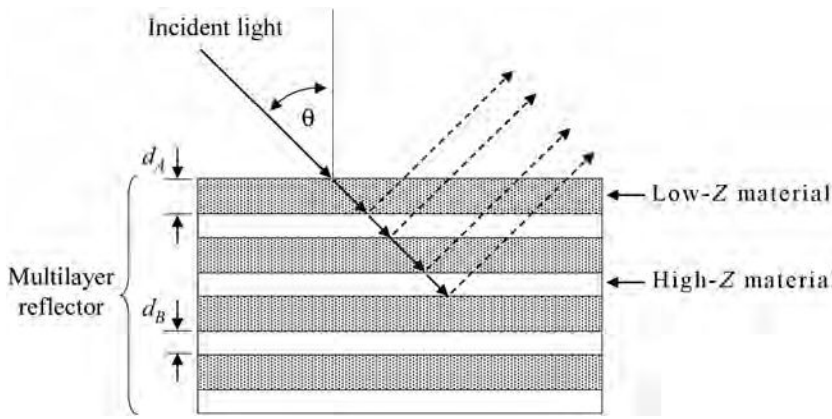


Figure 12.3 Illustration of the way in which high total reflectivity is achieved from reflections from mulitiple interfaces.

Table 12.1 EUV multilayer reflectors.[8–11]

Multilayer film stack	Peak wavelength of near-normal incidence	Reflectance achieved
Mo/Si	13.4	68.2%
$Mo/B_4C/Si$	13.5	70.0%
Mo_2C/Si	13.0	61.8%
Mo/Be	11.3	70.2%
MoRu/Be	11.3	69.3%

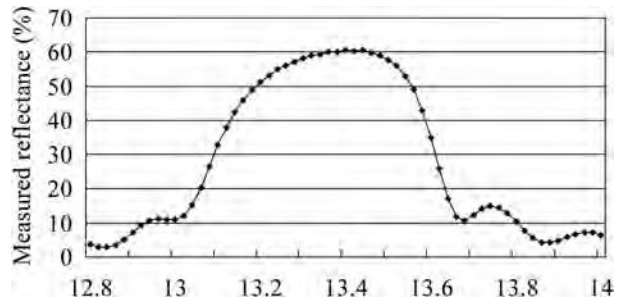


Figure 12.4 Reflectance versus wavelength for a multilayer reflector. The measurements were made at the Advanced Light Source (synchrotron) at Lawrence Berkeley Laboratory on a substrate that was made into an EUV mask at Advanced Micro Devices.

wavelength, in this case ~13.4 nm. From Eq. (12.1), it is apparent that only a very small change in the thickness of layer films will cause a shift in peak wavelength (see Problem 12.3). Even if every mirror in the EUV system has very high peak reflectance, the overall system transmission can be low if these peak reflectances do not occur at nearly the same wavelength. A similar statement is true regarding the need to match the reflectance properites of the masks to those of the projection-optics mirrors.[12] When specifying the "operating" wavelength, it is important to pay attention to detail. As can be seen in Fig. 12.4, the curve of reflectance versus wavelength is asymmetric. It is standard to specify the operating wavelength as the center of the full-width half-maximum range (Fig. 12.5).[13] Because of the asymmetry, this median wavelength typically differs from the peak wavelength.

As discussed in Chapter 5, optical lithography has been practiced where there are light sources that satisfy certain key requirements, particularly narrow bandwidth and high intensity. For each optical lithography technology the optics and masks have been engineered around the wavelength where such a source exists. For EUV lithography there are few options for masks and optics. EUV lithography must necessarily be practiced at wavelengths where there are multilayer reflectors with high reflectance. For EUV lithography it is the light source that needs to be engineered around the wavelength chosen on the basis of multilayer reflector capability, rather than the other way around. Because EUV lithography is practiced at wavelengths where high-intensity light sources were not established first, productivity due to low light intensity at the wafer is a concern with EUV lithographic technology, and considerable attention must be paid to maximizing multilayer reflectance and source-output power. More will be said in Section 12.4 about EUV light sources.

For EUV lithography, it appears that the semiconductor industry has settled on an operating wavelength of 13.5 nm. This choice was dictated in part by the selection of MoSi multilayers, which limited the range for the operating wavelength. The operating wavelength was further pinned down by matching this wavelength to the peak outputs of candidate light sources at the time the wavelength was selected.[14]

High mirror reflectance is important for EUV exposure-tool productivity. Consider a system with six mirrors in the projection optics, two near-normal reflection mirrors in the illuminator, and a reflection mask. The EUV light reflects nine times in such a system. A system in which each multilayer has 70% reflectance results in nearly 30% more light at the wafer plane than in a system where each multilayer reflector has 68% reflectance. Thus, even a small difference in reflectance is important for system productivity. Measured reflectances from multilayer stacks have typically been a few percent lower than the theoretical maximum. This is due in large measure to the nonideal density and homogeneity of the individual layers, as well as interface roughness and diffusion between layers.[2] Efforts to improve multilayer reflectance are focused on these factors.

One of the problems associated with multilayer films is their instability at elevated temperatures due to material intermixing, which reduces the reflectivity of the multilayer film stack. This can be mitigated to some extent by engineering the interfaces. In Mo/Si multilayers, the molybdenum and silicon films begin to intermix at ~150 °C, while Mo_2C/Si films remain stable at temperatures approaching 600 °C, with only a slightly lower peak reflectance (66.8%) than high-quality Mo/Si multilayers (68.6%).[15] With a B_4C layer inserted between the molybdenum and silicon films a reflectance of 70% was achieved at a wavelength of 13.5 nm.[16,17] Even if the peak reflectance is maintained during heating, there may be a shift of the wavelength at which peak reflectance occurs, as was observed in Mo/C/Si multilayers.[18] Films that are expected to experience high temperatures can be annealed prior to use, thereby shifting the wavelength at which peak reflectance occurs to the desired value. Instability at elevated temperatures limits the processes used to deposit and etch the absorbers on EUV masks. Without interface-engineered multilayers, only low-temperature processes can be used for fabricating EUV masks, including cleaning operations.

The invention of multilayer reflectors made EUV lithography conceptually possible. In the remainder of this chapter, the key elements of EUV lithographic

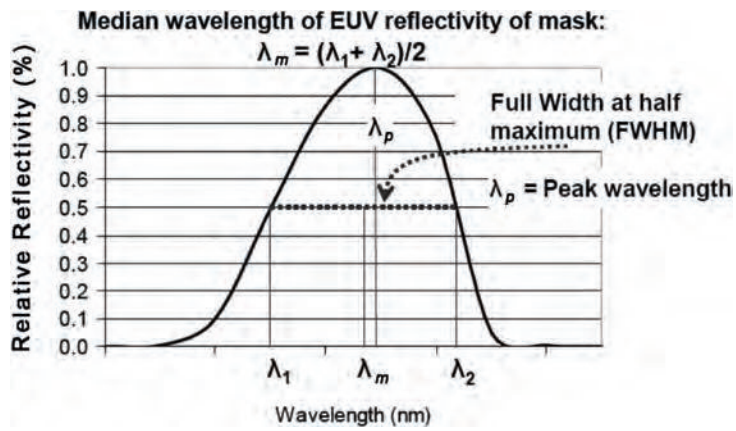


Figure 12.5 Curve showing the definition of median wavelength (λ_m) and how it can differ from the peak wavelength (λ_p).[13]

technology—masks, exposure tools, and resists—will be described in more detail, beginning with a general overview of EUV lithography systems.

12.2 EUV Lithography System Overview

Conceptually, an EUV exposure tool is very similar to the optical scanners described in Chapter 5. Key elements are: a source of light, collector and illuminator optics, projection optics, precision wafer and reticle stages, an alignment system, and mechanical handling systems for wafers and reticles. The basic configuration for EUV lithography is shown in Fig. 12.6. Light from an EUV source is collected by condenser mirrors. The condenser may also contain a spectral filter to remove unwanted long-wavelength light and heat. The light beam from the condenser optics illuminates a reflective mask that has areas that are either of high or low reflectance, according to the pattern to be printed, much as a conventional mask for optical lithography has areas of high and low transmission. The resulting optical pattern is imaged by a lens comprised of several mirrors. Low-NA ($\leq$0.1) EUV lenses usually have two or four mirrors, while higher-NA systems will have six mirrors or more. The image is focused onto a wafer. Because EUV systems are all-reflective, the light incident on the reticle must necessarily be off axis, that is, at a nonnormal angle of incidence. For current six-mirror systems with $NA \leq 0.35$, the angle of incidence is ~6 deg, while the angle is expected to be greater for larger-numerical-aperture imaging. There are implications of this nontelecentric imaging that will be discussed later.

While EUV exposure tools are similar in concept to optical scanners, for manufacturers and users of exposure tools, EUV presents numerous challenges. A central requirement is that exposures at EUV wavelengths must take place in a

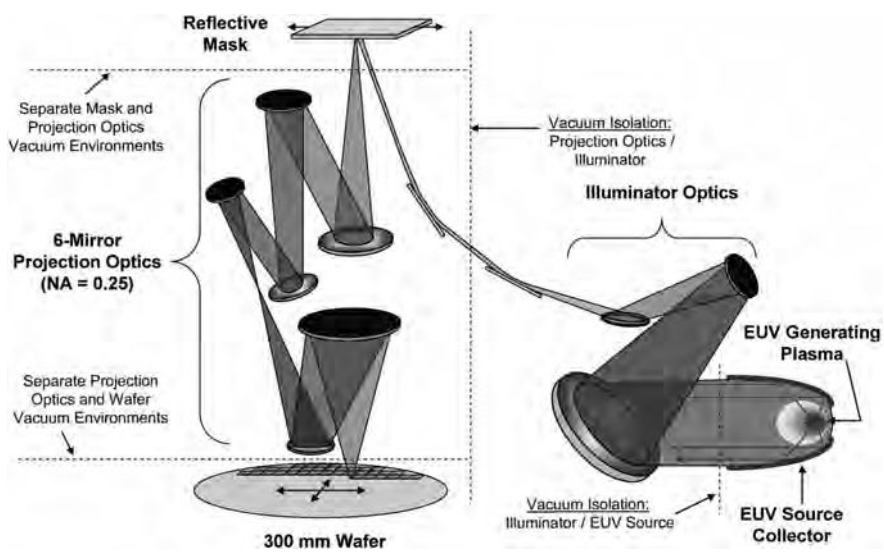


Figure 12.6 Schematic of an EUV exposure system.[19] Actual lenses may contain a different number of mirrors.

vacuum, since EUV light will not propagate in air. (The transmission of 13.5-nm light through 0.1 mm of air at atmospheric pressure is only ~7%.) Moreover, the vacuum must be very good, because photon-induced carbon deposition on mirror and mask surfaces or surface oxidation can result from the presence of very low levels of hydrocarbons in the system.[20,21] For example, if every mirror in a six-mirror lens system becomes coated with 1-nm-thick carbon layers, the transmission of the lens will be degraded by over 7% (see Problem 12.4). This imposes the requirement that all components within the vacuum chamber be constructed of ultrahigh vacuum-compatible materials. It also means that capability must be designed into exposure tools to prevent optics from becoming contaminated by the inevitable outgassing from resists. One idea for preventing the optics from becoming contaminated is to use the flow of an inert gas at a low partial pressure, between the optics and wafer (Fig. 12.7).[22,23] Contaminants will be carried along by the flow of inert gas, greatly reducing optics contamination. Some small amount of EUV light will be absorbed by the gas, estimated to be <3% for an optimized flow, because good protection is possible even with a low partial pressure of argon.

The requirement for good vacuum affects other parts of EUV scanners. While it is possible to construct air bearings for the stages that can be used in a vacuum,[24–26] magnetic bearings have obvious advantages for EUV applications.[27–29] Vacuum load locks are also required for moving wafers and reticles in and out of scanners efficiently, increasing system mechanical complexity, with possible consequences for reduced reliability and increased cost.

Another interesting consequence of exposing wafers in a vacuum is the need to change the method for chucking wafers and masks. In optical scanners, vacuum clamping is used nearly universally to hold wafers and reticles onto their chucks. The force for vacuum clamping is actually provided by atmospheric pressure, so some other approach for holding wafers and masks is needed in EUV

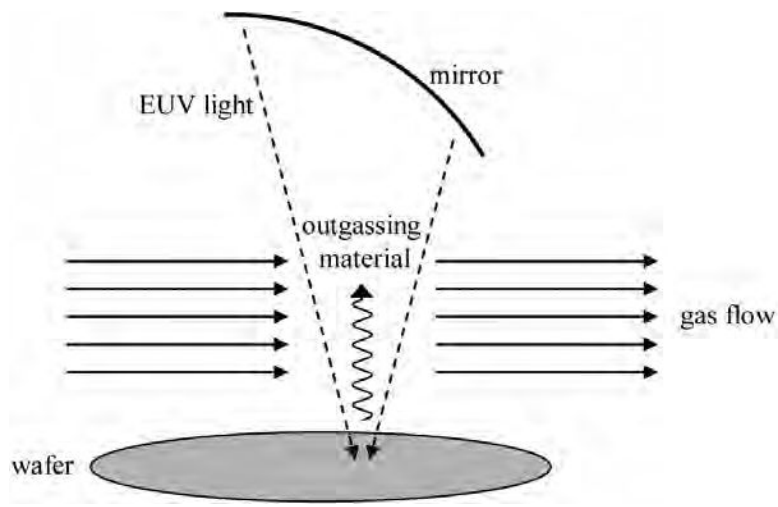


Figure 12.7 Illustration of a gas curtain. Outgassing material is swept along with the gas flow, thereby reducing contamination of the optics.

exposure tools. Electrostatic chucks have been used extensively in other types of semiconductor equipment where the processes occur in a vacuum, and this appears to be the approach taken for EUV systems as well. Electrostatic chucks are considered further in the next section.

As discussed in Chapter 5, very small levels of optical absorption can cause lenses to heat, leading to exposure-dependent variations in focus and overlay. The multilayer coatings on the mirrors that comprise EUV lenses are ~70% reflective, at best, and most of the light energy that is not reflected is absorbed. This can lead to substantial heating of lens elements, and very good active compensation is needed to maintain good focus and overlay control. Masks will also absorb an appreciable amount of light energy, which could lead to expansion and contraction of masks with attendent overlay errors.[30] For this reason, EUV masks are fabricated from ultralow expansion materials, with coefficients of thermal expansion measured in parts-per-billion.[13]

In the excimer lasers used for lithographic applications, the DUV light is produced by the plasmas of very corrosive gases. Such plasmas do not harm any optics because the plasmas are contained, with the light being emitted from a sealed chamber through transparent windows, typically comprised of CaF_2. Such windows do degrade over time and need to be replaced periodically, but they are inexpensive relative to optical components.

The situation is completely different for EUV. Because there are no sufficiently transparent materials at EUV wavelengths, there can be no windows to segregate gases and plasmas from the rest of the optical system. The same line-of-sight paths by which EUV photons travel from the point at which they are generated to the mirrors of the collection optics can also be pathways for contamination and debris. Since such contamination can potentially reduce the reflectivity of such mirrors substantially, the illumination systems of exposure tools must be designed with appropriate mitigation schemes.

Contamination can affect EUV light sensors that depend on the absorption of light to produce some type of electrical signal. This is because the photoelectrical properties of all solids are very sensitive to surface conditions. As described in Chapter 5, steppers use light sensors in real time to control exposure dose. Contamination of EUV light sensors will reduce the accuracy to which exposure doses can be controlled in EUV exposure tools.

Recently, manufacturers of lithographic equipment, ASML and Nikon, have built 0.25-NA full-field exposure tools.[31,32] Using such tools, confidence that EUV lithography will be viable for use in high-volume manufacturing was increased by the fabrication of electrically functional SRAMs and other complex circuits,[33,34] because all parts of EUV technology needed to work in order to produce these devices. The remainder of this chapter will cover particular key elements of EUV lithographic technology in more detail.

12.3 EUV Masks

As discussed previously, there are no high-transmission materials at EUV wavelengths, due to the limitation imposed by atomic absorption. For EUV lithography, not only must all optical elements in the lens be reflecting, so must be the mask. A typical configuration for an EUV mask is shown in Fig. 12.8. Masks are fabricated by first depositing multilayers on a flat substrate. Absorbing materials are then deposited and patterned on the multilayer, completing the fabrication of the EUV mask. A buffer layer is often deposited on the multilayer prior to deposition of the absorber to facilitate etching and repair without damaging the multilayer. Since all materials are absorbing at EUV wavelengths, many materials can be used for the mask absorber, including standard chromium.[35] Early results were obtained with TiN (Fig. 12.8), while materials such as TaN have been used more recently because they have higher absorption than TiN, thereby enabling the use of thinner absorbers.[36,37] As will be discussed, there are advantages to having thin absorbers. In addition to absorption, criteria for choosing an EUV mask absorber include deposition, inspectability, etch, and ease of cleaning.

The nontelecentricity of EUV exposure systems has certain implications for EUV masks. Even though the angle of incidence of the illumination on the mask is small (~6 deg), it does result in differences in the imaging of features that are parallel to the plane of incidence relative to those that are perpendicular.[38,39] This is because features perpendicular to the plane of incidence partially shadow the incident light (Fig. 12.9). For sub-50-nm features, horizontal–vertical print biases of 3–6 nm were seen in simulations, the magnitude depending upon the absorber composition and thickness.[40] Image placement shifts of 4–6 nm were also calculated. Clearly, some form of OPC is required for EUV lithography, in spite of the intrinsic resolution of the technology. It is possible to minimize the consequences of off-axis illumination by using the thinnest absorbers possible.

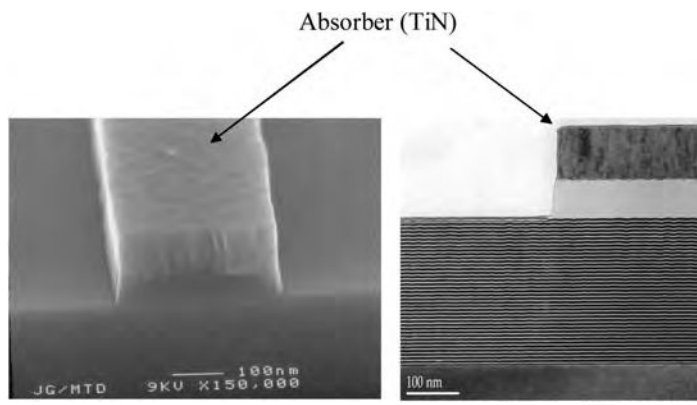


Figure 12.8 Micrographs of an early EUV mask fabricated at Advanced Micro Devices. The TiN absorber sits on a SiO_2 buffer layer that has been coated directly onto a MoSi multilayer. The left micrograph is a tilt SEM, while the right picture is a TEM, showing the multilayer reflector.

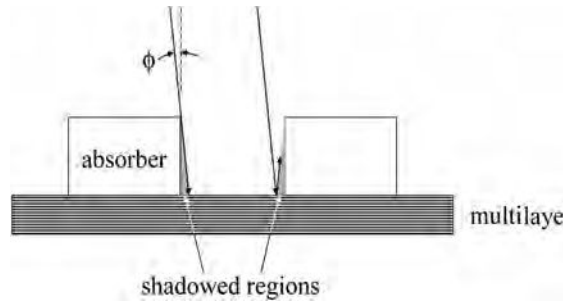


Figure 12.9 Lines that run perpendicular to the plane of incidence of the nonnormally incident light are shadowed. For $NA \leq 0.35$, $\phi = 6$ deg and will be greater for larger NA.

While all materials absorb EUV light, some are more absorbing than others, and this influences the choice of materials for EUV masks. TaN[41] and TaBN[42] are examples of materials with good absorption, along with other desirable characteristics, such as ease of cleaning.

Because EUV masks are ~30% absorbing even in the reflecting areas, these masks will heat up during exposure. Such heating and subsequent cooling will lead to overlay errors due to thermal expansion and contraction of the mask, so EUV masks should be fabricated on substrates with extremely low coefficients of thermal expansion (<30 ppb/K). Glass materials with such low coefficients of thermal expansion are available and have been characterized for use as EUV mask substrates.[43,44] Examples of such glass materials are ULE® from Corning, Inc. and Zerodur® from Schott. These and other EUV mask requirements are detailed in SEMI Standards P37 (specification for extreme ultraviolet lithography mask substrates), P38 (specification for absorbing film stacks and multilayers on extreme ultraviolet lithography mask blanks), and P40 (specification for mounting requirements and alignment reference locations for extreme ultraviolet lithography masks).

The use of reflective masks also imposes requirements on mask flatness. As can be seen in Fig. 12.10, mask nonflatness will result in image-placement errors. For consistency with the overlay requirements of the 32-nm node, EUV masks must be flat to within 36 nm (peak to valley) (see Problem 12.6).[45,46] Flatness of this magnitude must be maintained while the masks are in use, which means that EUV mask chucks must also be very flat, and great care must be taken to avoid particles on the back side of the masks that can lead to front-side mask nonflatness after chucking (see Fig. 12.11).[47]

The use of electrostatic chucks in EUV exposure systems imposes requirements on the backsides of EUV masks. In a basic bipolar Coulomb electrostatic chuck, electrodes within the chuck are charged. When the backside of a substrate is electrically conducting, charges can move freely and lead to electrostatic attraction between the chuck and the substrate (Fig. 12.12). Thus, there is a requirement in SEMI Standard P38 that the backsides of EUV masks be electrically conducting.

There is very limited capability for repairing a multilayer-film stack.[49] Consequently, blanks need to be made essentially defect free. This is a significant

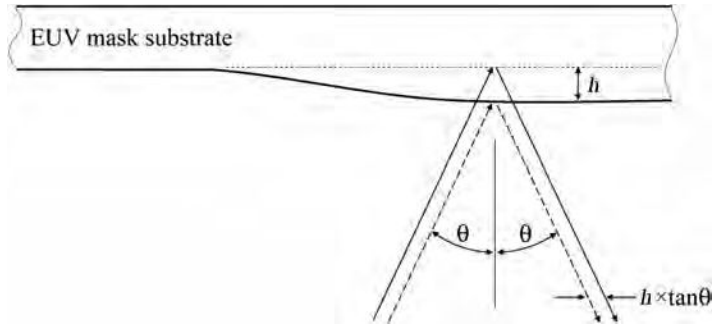


Figure 12.10 Mask nonflatness leads to image-placement errors. The lateral displacement $h \times \tan\theta$ is reduced on the wafer by the reduction of the projection optics.

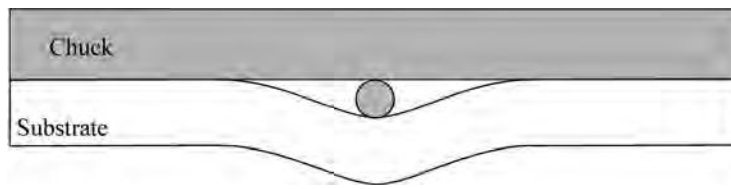


Figure 12.11 Mask deformation due to a particle between the mask substrate and the chuck.

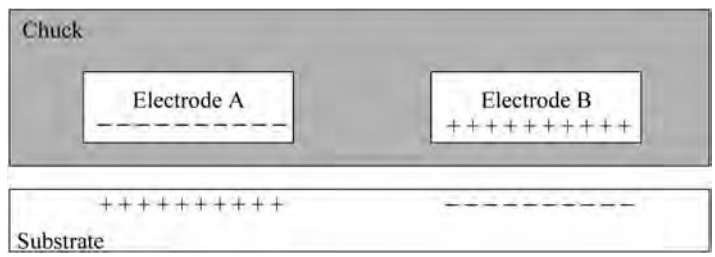


Figure 12.12 Basic configuration of a bipolar Coulomb electrostatic chuck.[48]

challenge for EUV lithography. With wafer feature sizes of 32 nm, and 4× reduction lenses, mask defects need to be much less than 4×32 nm = 128 nm. There is also the potential for phase defects illustrated in Fig. 12.13. Steps can result from particles or scratches on the substrate surface on which the multilayer film is deposited, and these steps can form phase defects. The extent to which such substrate defects propagate through the multilayer is dependent upon the deposition method, and there are initial data indicating that substrate defects can be smoothed over with ion-beam deposition techniques.[50] Phase defects can be created by extremely small particles, which may be smaller than can be detected with optical (visible or DUV) light. Indeed, defects have been detected using actinic inspection tools that could not be found with conventional visible-light defect inspection tools.[51] Consequently, it may be necessary to inspect masks with tools that use EUV light to provide sensitivity to phase defects. This will further

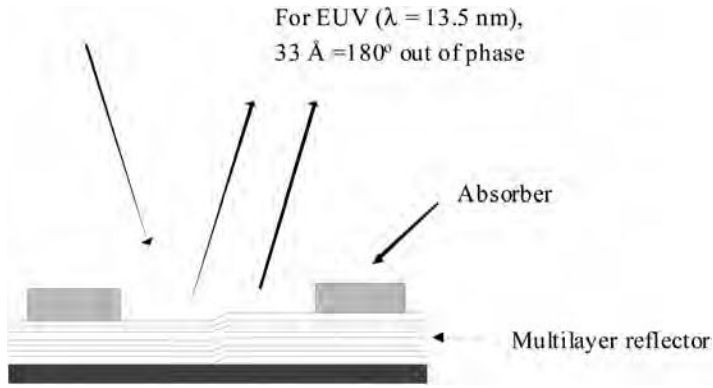


Figure 12.13 Illustration of an EUV mask defect.

add to the cost and complexity of implementing EUV lithography.[51,52] The mask is one of the biggest challenges for EUV lithography, even with the advantage of reduction lenses.

The absence of highly transparent materials at EUV wavelengths implies that conventional pellicles cannot be used. The transparency requirement for EUV lithography is high because light needs to pass twice through a pellicle in the reflection geometries currently implemented. This lack of a pellicle is another problem that needs to be addressed in the deployment of EUV lithography.

12.4 Sources and Illuminators

Light sources for EUV lithography also represent an area of active research and development. Synchrotrons are commonly used as sources of EUV light,[53,54] particularly for R&D and metrology. While such sources have proven to be very useful for laboratory applications, it appears that synchrotrons cannot produce a sufficient amount of light energy within a narrow bandwidth for high-volume manufacturing while still being cost effective.[55] Consequently, more compact sources have been developed, categorized as either discharge-produced-plasma (DPP) sources or laser-produced-plasma (LPP) sources. Laser-produced-plasma sources are conceptually quite simple: very-high-intensity pulsed-laser light is focused onto a material, which creates a plasma containing very highly charged ions (Fig. 12.14). When electrons recombine with the ions, high-energy photons are emitted.[56]

Xenon was the target material ("fuel") most commonly used in the light sources of early EUV systems, because it produces reasonably strong EUV emissions and, being chemically inert in its electrically neutral state, does not directly generate debris that contaminates and degrades the EUV optics. However, EUV light is produced primarily by the highly charged Xe^{+10} ionic state, so EUV light sources need to generate substantial numbers of very highly charged ions. When such ions hit a solid surface considerable sputtering can occur.[57] EUV exposure tools that use Xe-based light sources still need a scheme to mitigate damage done by these highly

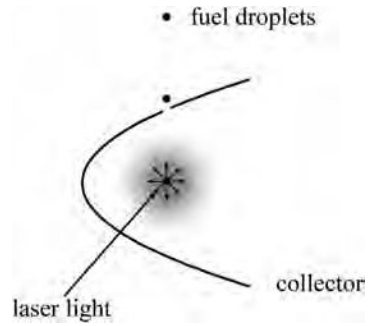


Figure 12.14 Schematic of a laser-produced plasma (LPP) EUV light source. The collector optics, which may be more complex than illustrated in this figure, focus the light onto the intermediate focus of the condenser optics.

charged xenon ions to the condenser optics. Given a need to address condenser optics damage even with the use of a noble gas, fuel materials that provide higher EUV light output than Xe have come into use. Lithium was considered,[58,59] but the dominant fuel in use today is tin, primarily due to its potential for high conversion efficiency to EUV light at the wavelengths where Mo/Si multilayers provide good reflectivity.[60]

LPP EUV sources were among the first to be considered, and such a source, using Nd-YAG lasers, was used on the first full-field EUV exposure tool.[61] Early cost estimates for LPP sources indicated that they would be too expensive for practical use, and discharge sources were subsequently pursued.[62–65] A conceptual drawing of one such discharge source is shown in Fig. 12.15, and a picture of the source in operation in shown in Fig. 12.16 (see Color Plates). Discharge sources involve plasmas generated between electrodes, which can be a significant engineering constraint that limits the ability to cool the electrodes and collect EUV light.

As DPP sources came into use on exposure tools,[32] lithographers became very conscious of the low intensities that were being achieved, and LPP sources were reconsidered. Sources have since been developed using CO_2 lasers that appear to have lower cost of ownership than had been projected earlier for LPP sources that

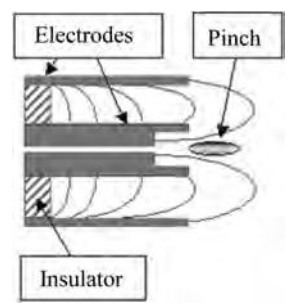


Figure 12.15 Illustration of an EUV discharge source. (Reproduced courtesy of Cymer, Inc.)

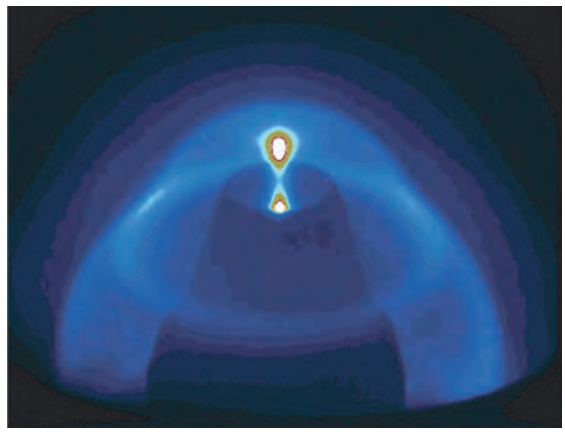


Figure 12.16 An EUV discharge source in operation. (Reproduced courtesy of Cymer, Inc.) (See Color Plates.)

employed Nd-YAG lasers. A picture of an LPP source that uses CO_2 lasers is shown in Fig. 12.17.

There is a thermal challenge for EUV light sources. In addition to EUV light, EUV sources generate even more light at longer wavelengths, including infrared radiation, i.e., heat. The intrinsically compact arrangement of components that produce the plasmas in discharge sources makes it difficult to remove heat. One clever idea is to fabricate the electrodes from the same liquid tin used to fuel the plasma.[66] This liquid tin can continously be resupplied, addressing the problem of electrode erosion in the harsh plasma environment. However, even with such ingenuity, the output of discharge sources has remained too low for use in high-volume manufacturing. The ultimate power output of discharge sources may be inherently limited, in part by the difficulty of dissipating the heat that is generated.

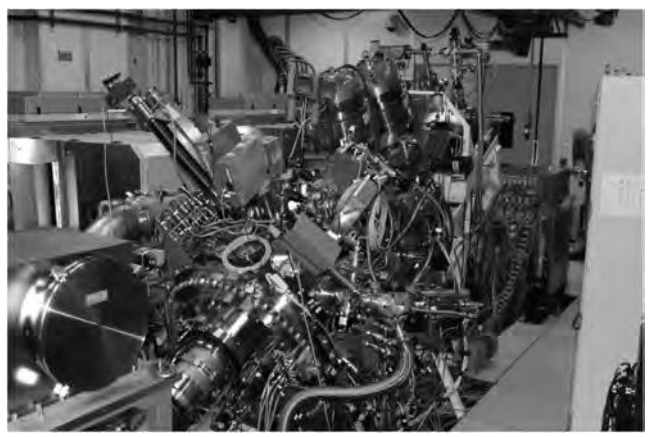


Figure 12.17 An LPP source shipped by Cymer, Inc., to a manufacturer of exposure systems for integration into a full-field exposure tool. (Figure courtesy of David Brandt of Cymer, Inc.)

Source efficiency is another important consideration. Over 200 W of collected in-band EUV power will be required to achieve cost-effective throughput. For LPP xenon-based EUV light sources, efficiencies less than 1% are typically reported for converting laser light to EUV light.[67] With laser conversion efficiencies of 10% (electricity to light), 200 kW of electrical power will be required to sustain light sources producing 200 W of collected in-band power. Such electrical consumption will be very expensive, motivating investigations of more efficient light sources, such as those where tin is used.

The type of optics used to collect EUV light depends upon the source type. As seen in Fig. 12.14, the collector mirror for a LPP source involves near-normal angles of incidence and reflection. On the other hand, grazing-mirror collection optics are needed with DPP sources (Figs. 12.15 and 12.16).

Unlike lasers, which produce very intense light within a narrow bandwidth, DPP and LPP EUV sources generate considerable amounts of light at wavelengths outside of the narrow band that is of interest for EUV lithography. It can be seen in Fig. 12.18 that Mo/Si multilayers reflect considerable amounts of light at deep ultraviolet wavelengths, and such light often is capable of exposing EUV resists. As mentioned in Chapter 5, when mercury-arc lamps are used to produce light for i-line exposure systems, it is necessary to filter out nonactinic light. This is readily accomplished using mature optical-filter technology that has been developed over centuries. Because mercury-arc lamps produce a considerable amount of i-line emission, exposure-tool throughput is not compromised greatly even when such filters reduce the amount of light at the desired wavelength. On the other hand, EUV lithography is a maturing technology, and EUV photons are precious. Filtering out unwanted light is a double challenge for EUV illumination systems; new technology needs to be used, and filters must reduce the amount

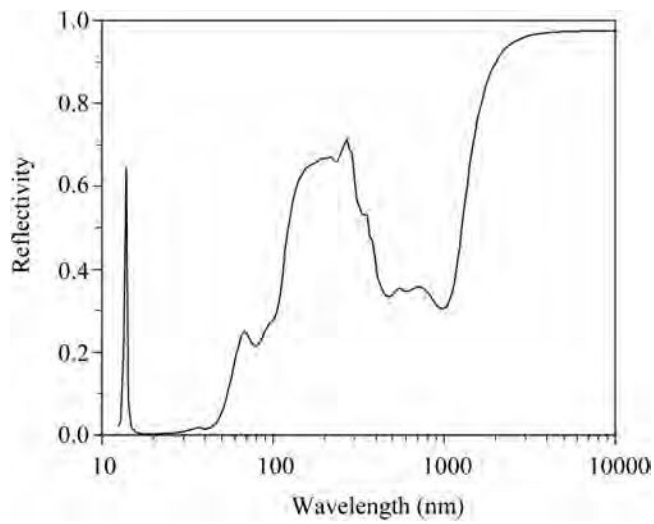


Figure 12.18 Calculated reflectivity of normally incident light from a 40-pair Mo/Si multilayer film.[68]

of actinic light by the smallest amount possible. As discussed in Chapter 5, light from excimer lasers needs to be narrowed for lithographic applications, due to chromatic aberrations in refractive lenses or, to a lesser extent, in catadioptric lenses. The motivation for filtering EUV light is quite different. EUV lenses are all reflective, so the problem of chromatic aberrations is reduced considerably. Moreover, reflections from manifold multilayer mirrors narrow the bandwidth of the actinic light (see Problem 12.8). The primary reasons for filtering the light from EUV sources is to remove long-wavelength light—DUV light that can degrade imaging and infrared light that can cause lens heating.

As mentioned previously, one must avoid contaminating or eroding the optical elements that are used to collect the EUV light generated by the source. Active steps must be taken, since such elements are within line of sight of the plasma, which produces the EUV photons. All of these DPP and LPP sources that use materials other than noble gases currently generate debris that coat the optics and reduce mirror reflectivity.[69] Various methods for protecting the collection optics have been proposed, an example of which is shown in Fig. 12.19.[70] In this system, low pressure (a few torr) of inert gas slows down particles from the plasma, as well as provides directional changes so the particles collide with the foils and get trapped. Argon and krypton are relatively weak absorbers of light at a wavelength of 13.5 nm.

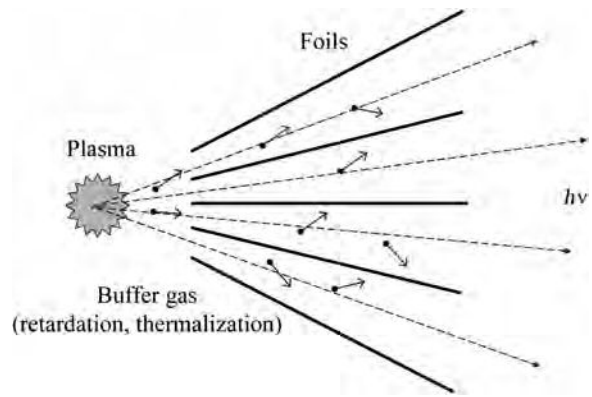


Figure 12.19 Example of a system for trapping debris from the plasma used to produce EUV light.[78]

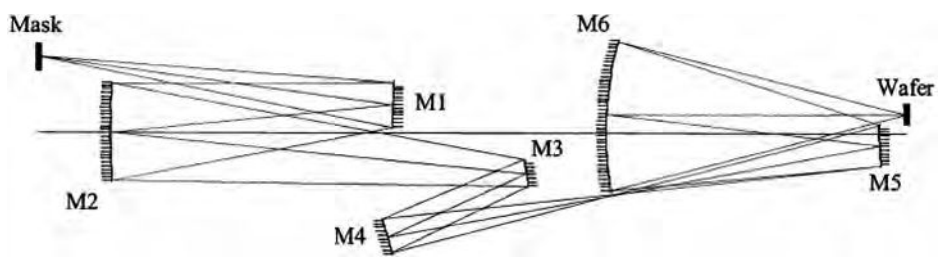


Figure 12.20 A 4× EUV lens design with $NA = 0.25$.[83]

12.5 EUV Optics

Lenses for EUV lithography are fully reflecting, i.e., all of the optical elements are mirrors. These mirrors are typically fabricated by polishing glass substrates to the desired shape (Fig. 12.20). The polished substrates are then coated with multilayer films that provide reflectivity at EUV wavelengths. Because these multilayer coatings only reflect ~70% of the light, a considerable amount of radiation is absorbed by the mirrors in the lenses. To minimize thermal deformation of mirror elements during exposure, the mirrors are typically fabricated from glass materials with low coefficients of thermal expansion such as ULE® from Corning, Inc. and Zerodur® from Schott.

Fabrication of lenses for EUV lithography is extremely challenging. To realize benefits from the shorter wavelength, the wavefront error must remain constant as a fraction of the wavelength. Thus, the transition from 193-nm lithography to EUV lithography requires an improvement in *absolute* wavefront error of (193/13.5)× or over 14×. The net result is a requirement for mirror-figure errors (the amount by which surfaces depart from their design values) of <0.25-nm rms in order to achieve high-resolution EUV optics. As if this were not challenging enough, EUV optics must also be smooth at spatial frequencies that are not a concern for optical lithography. Mid-spatial-frequency roughness (Table 12.2) leads to scattered light, which reduces image contrast, while high-spatial-frequency roughness reduces mirror reflectivity.[71–74] Fortunately, EUV imaging systems can be built with far fewer lens elements than are required for DUV lenses. A typical DUV-stepper lens has 20 elements or more, while 0.25-NA EUV lenses are comprised of only six mirrors. Thus, considerably more effort can be put into the fabrication of each EUV lens element, relative to DUV lens elements, without increasing overall lens cost beyond consideration.

Even with great care in polishing and with only a few elements in the lenses, flare in EUV systems is usually much higher than encountered in conventional optical lithography. An initial set of optics for a full-field exposure tool had approximately 40% long-range flare, and improved optics still had 17% flare.[77,78] The next generation of full-field exposure tools had 7–10% flare.[32,79] This is in contrast to state-of-the-art KrF and ArF lenses, where long-range flare <1% is routine. While EUV lithography is expected to be somewhat more tolerant of flare than optical lithography, because of the larger k_1 (and therefore larger image log

Table 12.2 Requirements and recent results for EUV mirrors.

	Spatial frequency range	**Requirements[75] (nm, rms)**	**Current results[76] (nm, rms)**
Figure	∞–1 mm^{-1}	0.08	~0.04
Mid-spatial-frequency roughness	1 mm^{-1}–1 μm^{-1}	0.14	~0.08
High-spatial-frequency roughness	1 μm^{-1}–50 μm^{-1}	0.10	~0.07

slope), modeling studies show significant reductions in process windows for EUV lithography with levels of flare of 10% or larger (see Problem 12.6).[80]

The reason for the greater level of flare in EUV lenses is the short wavelength involved and the use of mirrors. Consider a surface which produces an rms level of phase variation rms_{phase} in light reflected from the surface. The total integrated scattered (TIS) light from such a surface is given by[81]

$$TIS = 4\pi^2 \left(\frac{rms_{phase}}{\lambda} \right)^2 . \tag{12.3}$$

For a given level of surface roughness, a mirror will produce greater phase error than a refracting surface, as the light passes twice past surface roughness from a reflector but only once with a refracting surface. Because of the very short EUV wavelength, compared to DUV light, the amount of scatter is much greater for EUV optical systems for a given level of surface roughness. [Note the denominator of Eq. (12.3).]

Further complicating the polishing of mirror surfaces is the need to use aspheric mirrors in order to minimize the number of mirrors in the lens. With less than 70% reflectance per mirror, each pair of mirrors added reduces light intensity at the wafer by over half. Fabrication of high-quality aspheric mirrors is relatively new technology, although it has matured considerably.

To obtain the smoothest optical surfaces possible, particular care must be taken in the polishing process. In addition, the multilayer deposition process can be tuned to have the effect of smoothing roughness in the underlying substrate.[82] Because of the very stringent requirements for mirror-surface figure, the multilayer films must be low stress so that they do not deform the carefully polished glass substrates.

Increasing the numerical aperture of all-reflective optics involves a number of challenges. For example, consider the lens design shown in Fig. 12.20. Increasing the numerical aperture necessitates an increase in maximum angles for rays of light, but an increase in the size of certain lens elements will cause light rays to get blocked, a problem known as self-vignetting. To solve this problem, larger off-axis angles are required, and this makes it more difficult to maintain mechanical stability. Also, it becomes more difficult to compensate for aberrations with larger angles of incidence and reflection on the mirrors, necessitating the addition of mirrors to maintain low levels of aberrations. As a general rule, the lenses for EUV lithography systems have an even number of mirrors. This is necessary to separate the reticle and wafer stages physically to opposite sides of the lens. This is a practical consideration because of the large size of the stages, particularly for systems that involve multiple wafer stages. Consequently, the minimum increment in the number of mirrors in a lens is two. Since mirror reflectance is <70%, adding a single pair of additional mirrors will reduce lens transmission of light by over half. In order to maximize exposure-tool throughput, additional mirrors should be added only when absolutely necessary.

One interesting consequence of the use of all-reflective optics is the shape of the slit in scanning-exposure tools. In KrF and ArF reduction scanners with

dioptric lenses, the slit is rectangular. In these situations the imaging is done through the center of the lens, where aberrations are usually smallest. The slit remains rectangular in shape when optical scanners have catadioptric lenses, but the slit height is necessarily reduced in height due to the central obscuration found in catadioptric lenses (Fig. 12.21). As mentioned in Section 5.9, aberrations can often be minimized in a slit that is curved. By imaging in such a curved slit, fewer optical elements are required to correct for aberrations. Since minimizing the number of reflections to maximize throughput is an important consideration for EUV lithography, EUV scanners typically have curved slits.[84] Shown in Fig. 12.22 is a mask that was fabricated so it could be exposed in a static as well as a scanning mode. Consequently, sections of the mask were arranged to follow the slit, and the curvature is evident in Fig. 12.22.

As described in Section 12.3, off-axis illumination causes different horizontal and vertical print biases. The ring-field configuration increases the complexity of this problem, because the direction of the light incident on the mask changes across the slit. While the angle of incidence remains constant, it acquires an azimuthal component, as illustrated in Fig. 12.23. Consequently, the light may not be strictly perpendicular or parallel to horizontal or vertical features on the mask. To compensate for differences in the printing of horizontal and vertical features, OPC needs to vary across the slit to achieve full accuracy.

12.6 EUV Resists

The demands for EUV resists differ from the requirements of photoresists used at longer wavelengths in a few ways. Because EUV lithography is intended to be used to produce features and pitches beyond the capability of ArF lithography, EUV resists must be capable of very high resolution and have low line-edge roughness. There is also an enhanced need for sensitivity because of low output of EUV light sources. For DUV resists, sensitivity has been obtained through chemical amplification. The greater extent to which photoacids can diffuse and produce deprotection, the fewer photons required to expose the resist. However, substantial diffusion will effectively blur the image, reducing resolution. To maintain an image

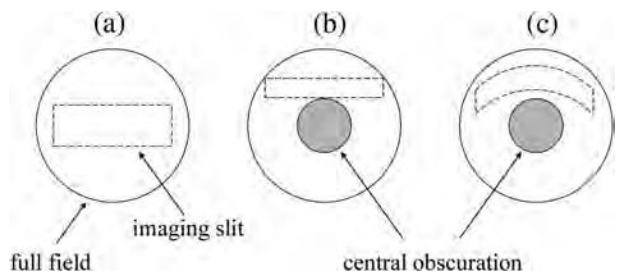


Figure 12.21 Examples of slits for different types of lenses: (a) refractive lens, such as is typically found in nonimmersion optical scanners; (b) catadioptric lens, such as is found in immersion optical scanners or very high-NA nonimmersion scanners; and (c) ring field, as is usually used in EUV exposure systems.

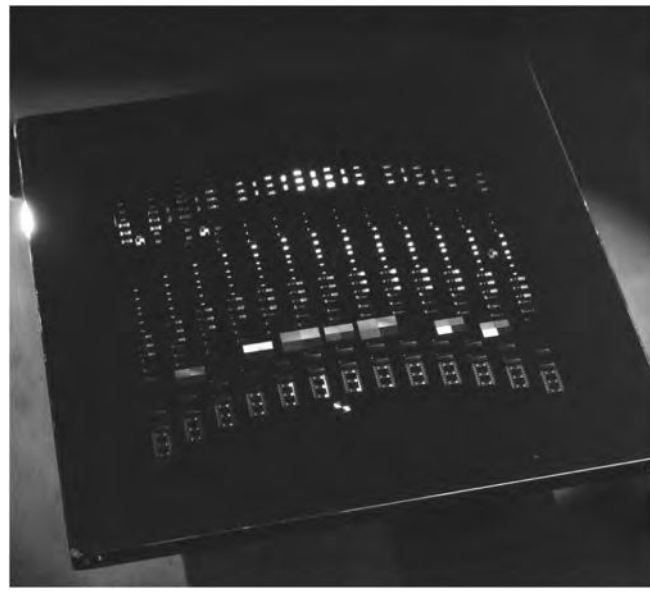


Figure 12.22 An EUV mask with features arranged to follow a curved slit to enable static as well as scanning exposures.

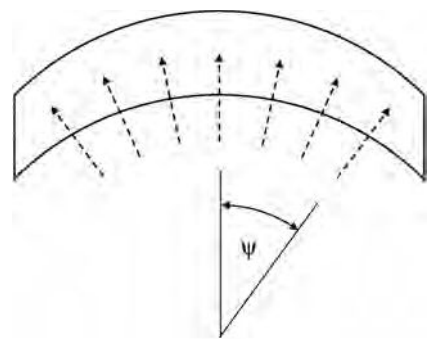


Figure 12.23 The illumination is incident on the mask with an azimuthal angle ψ across the slit.

after post-exposure bake, the diffusion length should be less than ~20% of the pitch. Using this estimate as a guide, for 22-nm half-pitch technology, the diffusion length will need to be less than 9 nm and must be even smaller for later nodes.

The problem of shot noise discussed in Chapter 3 is particularly acute for EUV lithography,[85,86] since there is an order of magnitude fewer photons per mJ of EUV light than for DUV light. Moreover, because EUV lithography is intended for use at smaller features than those created using ArF lithography, the requirements for LER will be more stringent for EUV lithography. The problem of shot noise in EUV lithography is illustrated in Fig. 12.24, which is a graph of LER versus exposure dose for a number of tested EUV resists, along with a curve derived from a simple model of shot noise that predicts the minimum LER that can be achieved

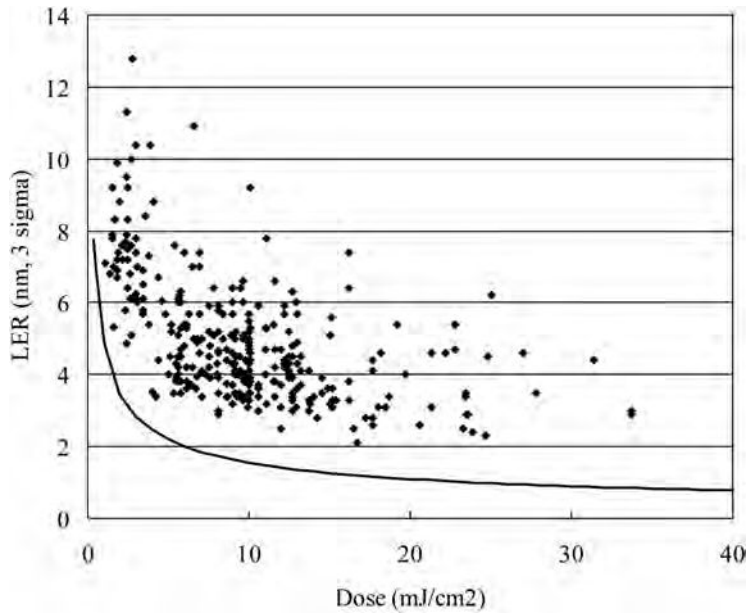


Figure 12.24 Line-edge roughness versus exposure dose for various EUV resists. The diamonds are measured results for individual resists,[87] and corrected for mask absorber and multilayer-roughness contributions to LER,[88] while the solid curve was produced by a model of LER resulting purely from shot noise. (Figure courtesy of B. LaFontaine.)

for a given exposure dose. From the data and model it appears that adequate LER cannot be obtained with extremely sensitive resists.

The model used to produce the solid curve in Fig. 12.24 relating LER and resist sensitivity is derived as follows. Suppose a uniform light beam is normally incident on the surface of a resist film. The average number N of EUV photons that crosses a surface section of size $d \times d$ into a resist film below the surface is given by

$$N = 0.68Ed^2, \tag{12.4}$$

where E is the exposure dose measured in $\mathrm{mJ/cm^2}$ and d is measured in nanometers (see Problem 12.7). Because of quantum fluctuations, the number of photons will vary around the average, leading to variations in dose. The numbers of photons vary according to Poisson statistics,[89] expressed as

$$\frac{\sigma_N}{N} = \frac{1}{\sqrt{N}}, \tag{12.5}$$

where σ_N is the standard deviation of the number of photons in a beam with an average number of photons N. This is a fundamental quantum phenomenon and is usually not considered for lithographic processes, because the value of N is sufficiently large at the doses typically encountered in lithography that quantum fluctuations are negligible. However, the situation in EUV lithography

is substantially different. First, because the energy per EUV photon is much higher than that for photons with wavelengths of 248 or 193 nm, there are far fewer EUV photons for the same doses as measured in mJ/cm^2. Second, EUV light sources are much weaker than the excimer lasers used for DUV lithography, so doses for EUV lithography tend to be lower than those in longer-wavelength lithography, even when measured in mJ/cm^2.

The deviation in the placement of a line-edge Δx due to the variation in dose ΔE is given by the expression (see Section 2.3):

$$\Delta x = \left(\frac{1}{ILS}\right)\left(\frac{\Delta E}{E}\right), \tag{12.6}$$

where ILS is the image log slope. Since photon absorption is also a statistical event, there is another source of variation that must be taken into account. If we assume that half of the incident photons are absorbed by the resist as a binomial process of probability 0.5, the total absorbed dose variation becomes

$$\frac{\Delta E}{E} = \frac{\sqrt{1^2 + 0.5^2}}{\sqrt{N}} \tag{12.7}$$

$$= \frac{1.12}{\sqrt{N}}. \tag{12.8}$$

From Eqs. (12.4), (12.6), and (12.8),

$$\Delta x = \left(\frac{1}{ILS}\right)\left(\frac{1.36}{d\sqrt{E}}\right). \tag{12.9}$$

Setting $LER(3\sigma) = 3\Delta x$ leads to the following relationship between line-edge roughness and dose E:

$$LER = \left(\frac{1}{ILS}\right)\left(\frac{4.08}{d\sqrt{E}}\right). \tag{12.10}$$

Both the image log slope and d will scale with the critical dimensions being patterned, i.e., LER can be expected to be greater for smaller features, at least without better optics. The length d establishes the minimum distance over which LER is evaluated, and this distance over which we are concerned will get smaller as critical dimensions shrink. Estimating that the normalized image log slope (NILS) is given by

$$NILS = CD \times ILS = 2.5, \tag{12.11}$$

and

$$d = \frac{CD}{3}. \tag{12.12}$$

Putting together Eqs. (12.10)–(12.12) gives,

$$LER = \frac{4.9}{\sqrt{E}}. \tag{12.13}$$

In this expression the LER is measured in nanometers, and the dose E is in units of mJ/cm^2. From this simple model we see that shot-noise-limited LER will vary with the inverse of the square root of the dose because of fundamental photon statistics. According to the International Technology Roadmap for Semiconductors, LER will need to be less than 2 nm for half pitches below 40 nm. From Fig. 12.24 it can be inferred that EUV light sources over time will need to support resists that require doses of 10 mJ/cm^2 and higher as dimensions continue to shrink. To some extent, Eq. (12.13) is optimistic. As EUV lithography is pushed to lower values of k_1, the normalized image log slope will decrease and the LER will increase.

None of the measured values for LER shown in Fig. 12.24 are less than 2 nm, even for the highest doses. This suggests that there are sources of LER other than shot noise, and these will also need to be identified and addressed to achieve the needed low levels of LER.[90] One such source that has been identified is mask roughness. Resist materials are likely also contributing to LER. Regardless, solving the problem of the shot-noise contribution to LER will require higher-power sources.

Another consequence of the high absorptivity of all materials at EUV wavelengths is the need for ultrathin resists (<100 nm). This is also necessary because of resist-collapse issues that accompany large resist aspect ratios. Fortunately, it has been shown that processes suitable for manufacturing are possible with ultrathin resist.[91] Moreover, resists very similar to conventional DUV materials are used for EUV lithography. Development of EUV resists is facilitated by the fact that most absorption at EUV wavelengths is dominated by atomic absorption, and the particulars of the molecular bonding have only a very small effect on optical absorption. Consequently, EUV resists can be developed without the types of restrictions on allowed chemical bonds encountered in the development of KrF and ArF resists.

Another key difference between DUV and EUV resists is the mechanism of the radiation-induced chemistry. For DUV chemically amplified resists, photoacids are generated as a direct consequence of photon absorption, either by the photoacid generator or polymers, which can absorb the photons and transfer energy to neighboring photoacid generators for ensuing decomposition,[92] in contrast to EUV resists, where the initial step is the generation of photoelectrons. It is these

photoelectrons that appear to be responsible for subsequent radiation-induced chemical reactions.[93–96] This difference in radation chemistry is significant.

A photoelectron will propagate from its point of origin and subsequently scatter, which can result in the generation of additional energetic electrons. Many of these secondary electrons also have the potential to induce chemical reactions. As a consequence, the quantum yield for EUV resists can actually exceed 1.0,[97] where

$$\text{quantum yield} = \frac{\text{number of photoacids generated}}{\text{number of photons absorbed}}. \tag{12.14}$$

The photoelectrons will travel some distance before being scattered. As a result, there is blur of the original optical image in the resist, even prior to post-exposure bake. Although the mean free path in organic materials for electrons with energy ~100 eV is <1 nm,[98] there can be multiple scattering events, leading to ranges for photoacid generation potentially much larger than 1 nm from the point of initial photon absorption.[99] Determining actual photoelectron ranges in EUV resists is still an area of active research,[100,101] and the answer will have implications for the ultimate resolution capability of EUV lithography. Although existing data are incomplete, an upper boundary for photoelectron-limited resolution can be made from exposure made using EUV interferometric lithography. (Inteferometric lithography is discussed in the next chapter.) While differing optically from conventional projection lithography, interferometric lithography is subject to the same resolution limits caused by photoelectrons. Using calixerene-type negative resists, which are not subject to photoacid blur, patterns with half pitches below 15 nm have been obtained.[102]

Many polymers tend to cross-link when exposed to radiation. As a consequence, resists can simultaneously exhibit positive resist behavior (usually at low to moderate doses) and negative resist behavior (at higher doses).[103] This is another resist issue associated with EUV lithography of which practitioners need to be mindful.

EUV lithography is a promising technology, but one where development is still in progress to enable high-volume manufacturing of integrated circuits. Only a survey of EUV technology has been provided here. For further reading on the basic engineering science of EUV radiation, the interested reader is referred to the excellent book by Dr. David Atwood.[104] In addition, there are a number of survey papers, as well as entire books dedicated to the subject of EUV lithography.[105,106]

Problems

12.1 Show that the Rayleigh resolution of an EUV ($\lambda = 13.5$ nm) lithography system with an $NA = 0.25$ is 33 nm, and that the Rayleigh depth-of-focus is 216 nm. Show that the resolution and depth-of-focus are 20.5 nm and 84 nm, respectively, for an EUV lens with 0.4 NA.

12.2 Show that Eq. (12.1) reduces to the conventional Bragg condition

$$d = \frac{m\lambda}{2\cos\theta},$$

when the indices of refraction of the materials in a multilayer film stack $\rightarrow 1.0$.

12.3 Using the above formula for the conventional Bragg condition, show that each pair in a multilayer reflector should be 6.8 nm thick to produce peak reflectivity at a wavelength of 13.5 nm when the incident light is 6 deg from normal. (Assume that $m = 1$.) Show that a change in the film-pair thickness of 1 Å changes the peak wavelength by 0.2 nm.

12.4 The attenuation length of carbon for light of wavelength 13.5 nm is 155 nm. Show that the transmission of light through a six-mirror lens where each mirror is coated with 1 nm of carbon is 92.6% of an identical lens for which the mirrors have no carbon coating. (Remember that the light will travel twice through the carbon coating on each mirror.)

12.5 The ITRS contains a requirement of 36-nm peak-to-valley mask flatness for the 32-nm node. Show that this level of nonflatness contributes an overlay error of ~1 nm for systems which illuminate the mask at a mean angle of 6 deg and have projection optics with a reduction ratio of 4:1.

12.6 In Eq. (2.32) the change in linewidth ΔL when the dose $E(x)$ is changed proportionally $E(x) \rightarrow (1 + \Delta)E(x)$ is given by

$$\Delta L = 2\Delta\,(ILS)^{-1},$$

where *ILS* is the image log slope. Show that the same relationship holds for an added dose $E(x) \rightarrow E(x) + \Delta$, independent of position.

12.7 Derive Eq. (12.4), keeping in mind that the energy of a single EUV photon is 1.47×10^{-14} mJ.[107]

12.8 Suppose the reflected light produced by a multilayer mirror as a function of wavelength (Fig. 12.4) is approximated by a Gaussian

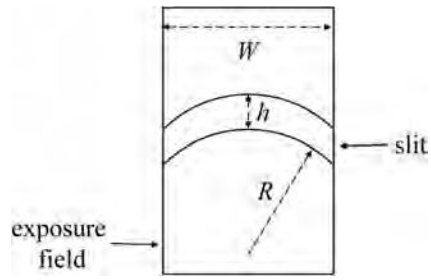
$$R(\lambda) = R_0 e^{\frac{-(\lambda-\lambda_0)^2}{a(\Delta\lambda)^2}},$$

where $a = 1/(4\ln 2)$ and $\Delta\lambda$ is the full-width at half-maximum of the light. Show that the full-width at half-maximum is $\Delta\lambda/\sqrt{N}$ after N reflections.

12.9 Show that a scanner with a curved slit must scan a distance

$$d = R\left(1 - \sqrt{1 - \frac{W^2}{4R^2}}\right)$$

longer than a scanner with a rectangular slit of the same height h, where R is the curved slit's radius of curvature and W is the width of the exposure field.



References

1. T. W. Barbee, S. Mrowka, and M. C. Hettrick, "Molybdenum-silicon multilayers for the extreme ultraviolet," *Appl. Optic.* **24**, 883 (1985).
2. S. Yulin, "Multilayer interference coatings for EUVL," in *Extreme Ultraviolet Lithography*, B. Wu and A. Kumar, Eds., McGraw Hill, New York (2009).
3. A. M. Hawryluk and L. G. Seppala, "Soft x-ray projection lithography using an x-ray reduction camera," *J. Vac. Sci. Technol. B* **6**, 2161–2166 (1988).
4. A concise history of EUV lithography written by two pioneers of the technology can be found in H. Kinoshita and O. Wood, "EUV lithography: an historical perspective," in *EUV Lithography*, V. Baksi, Ed., SPIE Press, Bellingham, Washington (2009).
5. B. La Fontaine, "EUV optics," in *Extreme Ultraviolet Lithography*, B. Wu and A. Kumar, Eds., McGraw Hill, New York (2009).
6. B. L. Henke, E. M. Gullikson, and J. C. Davis, "X-ray interactions: photoabsorption, scattering, transmission, and reflection at E = 50–30000 eV, Z = 1–92," *Atomic Data and Nuclear Data Tables* **54**(2), 181–342 (July 1993). See also http://henke.lbl.gov/optical_constants/.
7. E. Spiller, S. L. Baker, P. B. Mirkarimi, V. Sperry, E. M. Gullikson, and D. G. Stearns, "High-performance Mo-Si multilayer coatings for extreme-ultraviolet lithography by ion-beam deposition," *Appl. Optic.* **42**(19), 4049–4058 (2003).
8. A. A. Krasnoperova, R. Rippstein, A. Flamholz, E. Kratchmer, S. Wind, C. Brooks, and M. Lercel, "Imaging capabilities of proximity x-ray lithography at 70 nm ground rules," *Proc. SPIE* **3676**, 24–39 (1999).
9. J. A. Folta, S. Bajt, T. W. Barbee, R. F. Grabner, P. B. Mirkarimi, T. Nguyen, M. A. Schmidt, E. Spiller, C. C. Walton, M. Wedowski, and C. Montcalm, "Advances in multilayer reflective coatings for extreme-ultraviolet lithography," *Proc. SPIE* **3676**, 702–709 (1999).
10. S. Bajt, "Molybdenum-ruthenium/beryllium multilayer coatings," *J. Vac. Sci. Technol. A* **18**(2), 557–559 (2000).

11. S. Bajt, J. Alameda, T. Barbee Jr., W. M. Clift, J. A. Folta, B. Kaufmann, and E. Spiller, "Improved reflectance and stability of Mo/Si multilayers," *Proc. SPIE* **4506**, 65–75 (2001).
12. S. D. Hector, E. M. Gullikson, P. Mirkarimi, E. Spiller, P. Kearney, and J. Folta, "Multilayer coating requirements for extreme ultraviolet lithography masks," *Proc. SPIE* **4562**, 863–881 (2001).
13. SEMI Standard P37, "Specification for extreme ultraviolet lithography substrates and blanks," Semiconductor Equipment and Materials International, 3081 Zanker Road, San Jose, California (2002).
14. M. Richardson, "EUV sources," from Chapter 3 in *Extreme Ultraviolet Lithography*, B. Wu and A. Kumar, Eds., McGraw-Hill, New York (2009).
15. S. Yulin, T. Kuhlmann, T. Feigl, and N. Kaiser, "Damage-resistant and low-stress EUV multilayer mirrors," *Proc. SPIE* **4343**, 607–614 (2001).
16. S. Bajt, J. Almeda, T. Barbee Jr., W. M. Clift, J. A. Folta, B. Kaufmann, and E. Spiller, "Improved reflectance and stability of Mo-Si multilayers," *Opt. Eng.* **41**(8), 797–1804 (2002).
17. S. Braun, H. Mai, M. Moss, R. Scholz, and A. Leson, "Mo/Si multilayers with different barrier layers for applications and extreme ultraviolet mirrors," *Jpn. J. Appl. Phys.* **41**, 4074–4081 (2002).
18. S. Yulin, N. Benoit, T. Feigl, and N. Kaiser, "Interface-engineered EUV multilayer mirrors," *Microelectron. Eng.* **83**, 692–694 (2006).
19. S. Wurm, and K. Kemp, "SEMATECH pushes extreme UV lithography forward," SPIE Newsroom (April, 2006) [doi:10.1117/2.1200602.0079].
20. K. Boller, R. P. Haelbich, H. Hogrefe, W. Jerk, and C. Kunz, "Investigation of carbon contamination of mirror surfaces exposed to synchrotron radiation," *Nucl. Instrum. Method.* **208**, 273–279 (1983).
21. H. Meiling, V. Banine, H. Harned, B. Blum, P. Kürz, and H. Meijer, "Development of the ASML EUV alpha demo tool," *Proc. SPIE* **5751**, 90–102 (2005).
22. M. P. Kanouff and A. K. Ray-Chaudhuri, "Wafer chamber having a gas curtain for extreme-UV lithography," U.S. Patent No. 6,198,792B1 (2001).
23. M. P. Kanouff and A. K. Ray-Chaudhuri, "A gas curtain for mitigating hydrocarbon contamination of EUV lithographic optical components," *Proc. SPIE* **3676**, 735–742 (1999).
24. K. Suzuki, T. Fujiwara, K. Hada, N. Hirayanagi, S. Kawata, K. Morita, K. Okamoto, T. Okino, S. Shimizu, and T. Yahiro, "Nikon EB stepper: its system concept and countermeasures for critical issues," *Proc. SPIE* **3997**, 214–224 (2000).
25. T. Novak, D. Watson, and Y. Yoda, "Nikon electron projection lithography system: mechanical and metrology issues," *Am. Soc. Precis. Eng. Proc.* (2000). http://www.aspe.net/publications/Annual_2000/Annual_00.html.

26. T. H. Bisschops, et al., "Gas bearings for use with vacuum chambers and their application in lithographic projection apparatuses," U.S. Patent No. 6603130 (2003).

27. A. T. A. Peijnenburg, J. P. M. Vermeulen, and J. van Eijk, "Magnetic levitation systems compared to conventional bearing systems," *Microelectron. Eng.* **83**(4–9), 1372–1375 (2006).

28. P. T. Konkola, "Magnetic bearing stages for electron beam lithography," MS Thesis, Massachussetts Institute of Technology (1998). http://hdl.handle.net/1721.1/9315.

29. M. Williams, P. Faill, P. M. Bischoff, S. P. Tracy, and B. Arling, "Six degrees of freedom Mag-Lev stage development," *Proc. SPIE* **3051**, 856–867 (1997).

30. A. Abdo, B. La Fontaine, and R. Engelstad, "Predicting the thermomechanical distortion of extreme ultraviolet lithography reticles for preproduction and production exposure tools," *J. Vac. Sci. Technol. B* **21**(6), 3037–3040 (2003).

31. H. Meiling, E. Boon, N. Buzing, K. Cummings, O. Frijns, J. Galloway, M. Goethals, N. Harned, B. Hultermans, R. de Jonge, B. Kessels, P. Kürz, S. Lok, M. Lowisch, J. Mallman, B. Pierson, K. Ronse, J. Ryan, E. Smitt-Weaver, M. Tittnich, C. Wagner, A. van Dijk, and J. Zimmerman, "Performance of the full field EUV systems," *Proc. SPIE* **6921**, 69210L (2008).

32. T. Miura, K. Murakami, K. Suzuki, Y. Kohama, K. Morita, K. Hada, Y. Ohkubo, and H. Kwai, "Nikon EUVL development progress update," *Proc. SPIE* **6921**, 69210M (2008).

33. B. La Fontaine, Y. Deng, R. Kim, H. J. Levinson, S. McGowan, U. Okoroanyanwu, R. Seltmann, C. Tabery, A. Tchikoulaeva, T. Wallow, O. Wood, J. Arnold, D. Canaperi, M. Colburn, K. Kimmel, C. Koay, E. McLellan, D. Medeiros, S. P. Rao, K. Petrillo, Y. Yin, H. Mizuno, S. Bouten, M. Crouse, A. van Dijk, Y. van Dommelen, J. Galloway, S. Han, B. Kessels, B. Lee, S. Lok, B. Niekrewicz, B. Pierson, R. Routh, E. Schmitt-Weaver, K. Cummings, and J. Word, "The use of EUV lithography to produce demonstration devices," *Proc. SPIE* **6921**, 69210P (2008).

34. G. F. Lorusso, J. Hermans, A. M. Goethals, B. Baudemprez, F. Van Roey, A. M. Myers, I. Kim, B. S. Kim, R. Jonckheer, A. Niroomand, S. Lok, A. Van Dijk, J. F. de Marneffe, S. Demuynck, D. Goossens, and K. Ronse, "Imaging performance of the EUV alpha demo tool at IMEC," *Proc. SPIE* **6921**, 69210O (2008).

35. A. R. Pawloski, B. La Fontaine, H. J. Levinson, S. Hirscher, S. Schwarzl, K. Lowack, F.-M. Kamm, M. Bender, W.-D. Domke, C. Holfeld, U. Dersch, P. Naulleu, F. Letzkus, and J. Butschke, "Comparative study of mask architectures for EUV lithography," *Proc. SPIE* **5567**, 762–773 (2004).

36. M. Takahashi, T. Ogawa, E. Hoshino, H. Hoko, B. T. Lee, A. Chiba, H. Yamanashi, and S. Okazaki, "Tantalum nitride films for the absorber material of reflective-type EUVL mask," *Proc. SPIE* **4343**, 760–770 (2001).

37. P. Y. Yan, G. Zhang, A. Ma, and T. Liang, "TaN EUVL mask fabrication and characterization," *Proc. SPIE* **4343**, 409–414 (2001).

38. K. Otaki, "Asymmetric properties of the aerial image in extreme ultraviolet lithography," *Jpn. J. Appl. Phys.* Pt. 1 **39**(12B), 6819–6826 (2000).

39. S. B. Bollepalli, M. Khan, and F. Cerrina, "Image formation in extreme ultraviolet lithography and numerical aperture effects," *J. Vac. Sci. Technol. B* **17**(6), 2992–2997 (1999).

40. Y. Deng, B. La Fontaine, H. J. Levinson, and A. R. Neureuther, "Rigorous EM simulation of the influence of the structure of mask patterns on EUVL imaging," *Proc. SPIE* **5037**, 302–313 (2003).

41. P. Y. Yan, G. Zhang, P. Kofron, J. Powers, M. Tran, T. Liang, A. Stivers, and F. C. Lo, "EUV mask absorber characterization and selection," *Proc. SPIE* **4066**, 116–123 (2000).

42. T. Shoki, T. Kinoshita, N. Sakaya, M. Hosoya, R. Ohkubo, Y. Usui, H. Kobayashi, and O. Nagarekawa, "Damage-free extreme ultraviolet mask with TaBN absorber," *J. Vac. Sci. Technol. B* **21**(6), 3021–3026 (2003).

43. R. E. Hrdina, B. Z. Hanson, P. M. Fenn, and R. Sabia, "Characterization and characteristics of an ULE glass tailored for the EUVL needs," *Proc. SPIE* **4688**, 454–461 (2002).

44. I. Mitra, M. J. Davis, J. Alkemper, R. Müller, L. Aschke, E. Mörsen, S. Ritter, H. Hack, and W. Pannhorst, "Thermal expansion behavior of proposed EUVL substrate materials," *Proc. SPIE* **4688**, 462–468 (2002).

45. S. D. Hector, "EUVL masks: requirements and potential solutions," *Proc. SPIE* **4688**, 134–149 (2002).

46. International Technology Roadmap for Semiconductors (2009).

47. V. Ramaswamy, R. L. Engelstad, K. T. Turner, A. R. Mikkelson, and S. Veeraraghavan, "Distortion of chucked extreme ultraviolet reticles from entrapped particles," *J. Vac. Sci. Technol. B* **24**(6), 2829–2833 (2006).

48. Electrostatic chucks may be of Coulomb or Johnsen-Rahbek type, as described further in M. Sogard, A. Mikkelson, M. Nataraju, K. Turner and R. Engelstad, "Analysis of Coulomb and Johnsen-Rahbek electrostatic chuck performance for EUV lithography," *J. Vac. Sci. Technol. B* **25** (6), 2155–2161 (2007), and M. R. Sogard, A. R. Mikkelson, V. Ramaswamy, and R. L. Engelstad, "Analysis of Coulomb and Johnsen-Rahbek elecrostatic chuck performance in the presence of particles for EUV lithography," *Proc. SPIE* **7271**, 72710H (2009).

49. S. Hau-Riege, A. Barty, P. Mirkarimi, S. Baker, M. A. Coy, M. Mita, V. E. Robertson, T. Liang, and A. Stivers, "Repair of phase defects in extreme-ultraviolet lithography mask blanks," *J. Appl. Phys.*, 6812–6821 (2004).

50. C. C. Walton, P. A. Kearney, P. B. Mirkarimi, J. M. Bowers, C. Cerjan, A. L. Warrick, K. Wilhelmsen, E. Fought, C. Moore, C. Larson, S. Baker, S. C. Burkhart, and S. D. Hector, "Extreme ultraviolet lithography–reflective mask technology," *Proc. SPIE* **3997**, 496–507 (2000).

51. S. Jeong, L. Johnson, Y. Lin, S. Rekawa, P. Yan, P. Kearney, E. Tejnil, J. Underwood, and J. Bokor, "Actinic EUVL mask blank defect inspection system," *Proc. SPIE* **3676**, 298–308 (1999).

52. Y. Tezuka, M. Ito, T. Terasawa, and T. Tomie, "Actinic detection of multilayer defects on EUV mask blanks using LPP light source and dark-field imaging," *Proc. SPIE* **5374**, 271–280 (2004).

53. G. Dattoli, A. Doria, G. Piero Gallerano, L. Giannessi, K. Hesch, H. O. Moser, P. L. Ottaviani, E. Pellegrin, R. Rossmanith, R. Steininger, V. Saile, and J. Wüst, "Extreme ultraviolet (EUV) sources for lithography based on synchrotron radiation," *Nucl. Instr. Method. Phys. Res. Sec. A* **474**(3), 259–272 (2001).

54. D. C. Ockwell, N. C. E. Crosland, and V. C. Kempson, "Synchrotron light as a source for extreme ultraviolet lithography," *J. Vac. Sci. Technol. B* **17**(6), 3043–3046 (1999).

55. V. Y. Banine, J. P. H. Benschop, and H. G. C. Werij, "Comparison of extreme ultraviolet sources for lithography applications," *Microelectron. Eng.* **53**(1–4), 681–684 (June, 2000).

56. R. Colombant and G. Tonon, "X-ray emission from laser plasmas," *J. Appl. Phys.* **44**, 3524–3537 (1973).

57. P. A. Grunow, L. E. Klebanoff, S. Graham Jr., S. J. Haney, and W. M. Clift, "Rates and mechanisms of optics contamination in the EUV engineering test stand," *Proc. SPIE* **5037**, 418–428 (2003).

58. 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 on pinch plasmas: a comparison of different concepts," *Proc. SPIE* **3997**, 162–168 (2000).

59. W. Partlo, I. Fomenkov, R. Olivier, and D. Birx, "Development of an EUV (13.5 nm) light source employing a dense plasma focus in lithium vapor," *Proc. SPIE* **3997**, 136–156 (2000).

60. T. Tomie, T. Aota, Y. Ueno, G. Nimi, H. Yashiro, J. Lin, I. Matsushima, K. Komiyama, D. Lee, K. Nishigori, and H. Yokota, "Use of tin as a plasma source material for high conversion efficiency," *Proc. SPIE* **5037**, 147–155 (2003).

61. D. A. Tichenor, A. K. Ray-Chaudhuri, W. C. Replogle, R. H. Stulen, G. D. Kubiak, P. D. Rockett, L. E. Klebanoff, K. L. Jefferson, A. H. Leung, J. B. Wronsky, L. C. Hale, H. N. Chapman, J. S. Taylor, J. A. Folta, C. Montcalm, R. Soufli, E. Spiller, K. Blaedel, G. E. Sommergren, D. W. Sweeney, P. Naulleau, K. A. Goldberg, E. M. Gullikson, J. Bokor, P. J. Batson, D. T. Attwood, K. H. Jackson, S. D. Hector, C. W. Gwynn, and P. Yan, "System integration and performance of the EUV engineering test stand," *Proc. SPIE* **4343**, 19–37 (2001).

62. L. Rymell, M. Berglund, B. A. M. Hansson, and H. M. Hertz, "X-ray and EUV laser-plasma sources based on cryogenic liquid-jet target," *Proc. SPIE* **3676**, 421–424 (1999).

63. M. Klosner, H. Bender, W. Silfvast, and J. Rocca, "Intense plasma discharge source at 13.5 nm for extreme ultraviolet lithography," *Optic. Lett.* **32**, 34–36 (1997).

64. N. R. Fornaciari, J. J. Chang, D. R. Folk, S. Gianoulakis, J. E. M. Goldsmith, G. D. Kubiak, B. C. Long, D. J. O'Connell, G. Shimkaveg, W. T. Silfvast, and K. D. Stewart, "Development of an electric capillary discharge source," *Proc. SPIE* **3997**, 120–125 (2000).

65. 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 on pinch plasmas: a comparison of different concepts," *Proc. SPIE* **3997**, 162–168 (2000).

66. M. Corthout, R. Apetz, J. Brudermann, M. Damen, G. Derra, O. Franken, J. Jonkers, J. Klein, F. Küpper, A. Mader, W. Neff, H. Scheuermann, G. Schriever, M. Schürmann, G. Siemons, R. Snijkers, D. Vaudrevange, E. Wagenaars, P. van de Wel, M. Yoshioka, P. Zink, and O. Zitzen, "Sn DPP source-collector modules: status of alpha sources, beta developments, and the scalability to HVM," *Proc. SPIE* **6921**, 69210V (2008).

67. U. Stamm, I. Ahmad, I. Balogh, H. Bimer, D. Bolshukhin, J. Brudermann, S. Enke, F. Flohrer, K. Gäbel, S. Götze, G. Hergenhan, J. Kleinschmidt, D. Klöpfel, V. Korobotchko, J. Ringling, G. Schriever, C. Tran, and C. Ziener, "High-power EUV lithography sources based on gas discharges and laser-produced plasmas," *Proc. SPIE* **5037**, 119–129 (2003).

68. S. A. George, "Out-of-band exposure characterization with the SEMATECH Berkeley 0.3-NA microfield exposure tool," Lawrence Berkeley National Laboratory (2009). http://escholarship.org/uc/item/3g458337.

69. K. Takenoshita, C.-S. Koay, S. Teerawattanasook, and M. Richardson, "Debris studies for the tin-based droplet laser-plasma EUV Source," *Proc. SPIE* **5374**, 954–963 (2004).

70. L. A. Shmaenok, C. C. de Bruijn, H. Fledderus, R. Stuik, A. A. Schmidt, D. M. Simanovskii, A. A. Sorokin, T. A. Andreeva, and F. Bijkerk,

"Demonstration of a foil trap technique to eliminate laser plasma atomic debris and small particulates," *Proc. SPIE* **3331**, 90–94 (1998).

71. B. La Fontaine, T. P. Daly, H. N. Chapman, D. P. Gaines, D. G. Stearns, D. W. Sweeney, and D. R. Kania, "Measuring the effect of scatter on the performance of a lithography system," in *OSA Trends in Optics and Photonics*, vol. 4, *Extreme Ultraviolet Lithography*, Glenn D. Kubiak and Don R. Kania, Eds., OSA, 203–205 (1996).
72. D. G. Stearns, D. P. Gaines, B. La Fontaine, G. E. Sommargren, D. W. Sweeney, D. R. Kania, and N. M. Ceglio, "Nonspecular scattering in EUV lithography: determining specifications for surface finish," in *OSA Trends in Optics and Photonics*, vol. 4, *Extreme Ultraviolet Lithography*, Glenn D. Kubiak and Don R. Kania, Eds., OSA, 167–168 (1996).
73. D. P. Gaines, T. P. Daly, D. G. Stearns, B. La Fontaine, D. W. Sweeney, H. C. Chapman, and D. Fuchs, "Image degradation from surface scatter in EUV optics," in *OSA Trends in Optics and Photonics* vol. 4, *Extreme Ultraviolet Lithography*, Glenn D. Kubiak and Don R. Kania, Eds., OSA, 199–202 (1996).
74. P. Nalleau, K. A. Goldberg, E. M. Gullikson, and J. Bokor, "Interferometric at-wavelength flare characterization of extreme ultraviolet optical systems," *J. Vac. Sci. Technol. B* **17**(6), 2987–2991 (1999).
75. B. Wu and A. Kumar, "Extreme ultraviolet lithography: a review," *J. Vac. Sci. Technol. B* **25**(6), 1743–1761 (2007).
76. K. Murakami, T. Oshino, H. Kondo, H. Chiba, H. Komatsuda, K. Nomura, and H. Iwata, "Development status of projection optics and illumination optics for EUV1," *Proc. SPIE* **6921**, 69210Q (2008).
77. D. A. Tichenor, W. C. Replogle, S. H. Lee, W. P. Ballard, A. H. Leung, G. D. Kubiak, L. E. Klebenoff, S. Graham, J. E. M. Goldsmith, K. L. Jefferson, J. B. Wronsky, T. G. Smith, T. A. Johnson, H. Shields, L. C. Hale, H. N. Chapman, J. S. Taylor, D. W. Sweeney, J. A. Folta, G. E. Sommargren, K. A. Goldberg, P. Naulleau, D. T. Attwood, and E. M. Gullikson, "Performance upgrades in the EUV engineering test stand," *Proc. SPIE* **4688**, 72–86 (2002).
78. S. H. Lee, P. Naulleau, C. Krautschik, M. Chandbok, H. N. Chapman, D. J. O'Connell, and M. Goldstein, "Lithographic flare measurements of EUV full-field projection optics," *Proc. SPIE* **5037**, 103–111 (2003).
79. K. Murakami, T. Oshino, H. Kondo, M. Shiraishi, H. Chiba, H. Komatsuda, K. Nomura, and J. Nishikawa, "Development progress of optics for EUVL at Nikon," *Proc. SPIE* **7271**, 72711Z (2009).
80. M. Lowisch, U. Dinger, U. Mickam, and T. Heil, "EUV imaging—an aerial image study," *Proc. SPIE* **5374**, 53–63 (2004).
81. E. M. Gullikson, S. L. Baker, J. E. Bjorkholm, J. Bokor, K. A. Goldberg, J. E. Goldsmith, C. Montcalm, P. P. Naulleau, E. A. Spiller, D. G. Stearns,

J. S. Taylor, and J. H. Underwood, "EUV scattering and flare of 10X projection cameras," *Proc. SPIE* **3676**, 717–723 (1999).

82. E. Spiller, S. L. Baker, P. B. Mirkarimi, V. Sperry, E. M. Gullikson, and D. G. Stearns, "High-performance Mo-Si multilayer coatings for extreme-ultraviolet lithography by ion-beam deposition," *Appl. Optic.* **42**(19), 4049–4058 (2003).
83. R. M. Hudyma, "An overview of optical systems for 30 nm resolution lithography at EUV wavelengths," *Proc. SPIE* **4832**, 137–148 (2002).
84. D. M. Williamson, "Evolution of ring field systems in microlithography," *Proc. SPIE* **3482**, 369–376 (1998).
85. G. M. Gallatin, F. A. Houle, and J. L. Cobb, "Statistical limitations of printing 50 and 80 nm contact holes by EUV lithography," *J. Vac. Sci. Technol. B* **21**(6), 3172–3176 (2003).
86. R. L. Brainard, P. Trefonas, J. H. Lammers, C. A. Cutler, J. F. Mackevich, A. Trefonas, and S. W. Robertson, "Shot noise, LER and quantum efficiency of EUV photoresists," *Proc. SPIE* **5374**, 74–85 (2004).
87. The resist data were provided by Dr. Tom Wallow and Dr. Patrick Naulleau.
88. Dr. Patrick Naulleau has estimated that 1.43 nm of LER results from the mask absorber LER, and this has been factored out of the data by assuming that the measured LER is the root sum of squares of the mask LER and the LER due to all other factors. It is the LER due to nonmask sources that is plotted. See P. P. Naulleau, D. Niakoula, and G. Zhang "System-level line-edge roughness limits in extreme ultraviolet lithography," *J. Vac. Sci. Technol. B* **26**(4), 1289–1293 (2008).
89. F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York (1965).
90. C. N. Anderson, "Don't always blame the photons: relationships between deprotection blur, LER, and shot noise in EUV photoresists," Lawrence Berkeley National Laboratory (2009). Retrieved from: http://www.escholarship.org/uc/item/5fj2s0jn.
91. K. B. Nguyen, C. Lyons, J. Schefske, C. Pike, K. Phan, P. King, H. Levinson, S. Bell, and U. Okoroanyanwu, "Characterization of the manufacturability of ultrathin resist," *J. Vac. Sci. Technol. B* **17**(6), 3039–3042 (1999).
92. J. F. Cameron, N. Chan, K. Moore, and G. Pohlers, "Comparison of acid generating efficiencies in 248 and 193 nm photoresists," *Proc. SPIE* **4345**, 106–118 (2001).
93. R. L. Brainard, C. Henderson, J. Cobb, V. Rao, J. F. Mackevich, U. Okoroanyanwu, S. Gunn, J. Chambers, and S. Connolly, "Comparison of the lithographic properties of positive resists upon exposure to deep- and extreme-ultraviolet radiation," *J. Vac. Sci. Technol. B* **17**(6), 3384–3389 (1999).

94. R. L. Brainard, G. G. Barclay, E. H. Anderson, and L. E. Ocola, "Resists for next-generation lithography," *Microelectron. Eng.* **61–62**, 707–715 (2002).

95. T. Kozawa, Y. Yoshida, M. Uesaka, and S. Tagawa, "Study of radiation induced reactions in chemically amplified resists for electron beam and x-ray lithography," *Jpn. J. Appl. Phys.* **31**, L1574–L1576 (1992).

96. R. Brainard, E. Hassanein, J. Li, P. Pathak, B. Thiel, F. Cerrina, R. Moore, M. Rodriguez, B. Yakshinskiy, E. Loginova, T. Madey, R. Matyi, M. Malloy, A. Rudack, P. Naulleau, A. Wüest, and K. Dean, "Photons, electrons, and acid yields in EUV photoresists: a progress report," *Proc. SPIE* **6932**, 693215 (2008).

97. R. L. Brainard, P. Trefonas, J. H. Lammers, C. A. Cutler, J. F. Mackevich, A. Trefonas, and S. W. Robertson, "Shot noise, LER and quantum efficiency of EUV photoresists," *Proc. SPIE* **5374**, 74–85 (2004).

98. S. Tanuma, C. J. Powell, and D. R. Penn, "Calculations of electron inelastic mean free paths V: data for 14 organic compounds over the 50-2000 eV range," *Surf. Interface Anal.* **21**, 165–176 (1994).

99. T. Kozawa, S. Tagawa, H. B. Cao, H. Deng, and M. J. Leeson, "Acid distribution in chemically amplified extreme ultraviolet resist," *J. Vac. Sci. Technol. B* **25**(6), 2481–2485 (2007).

100. T. Kozawa, A. Saeki, and S. Tagawa, "Modeling and simulation of chemically amplified electron beam, x-ray, and EUV resist processes," *J. Vac. Sci. Technol. B* **22**(6), 3489–3492 (2004).

101. T. Kozawa, and S. Tagawa, "Point spread function for the calculation of acid distribution in chemically amplified resists used for electron-beam lithography," *Jpn. J. Appl. Phys.* **46**, L1200–1202 (2007).

102. H. H. Solak, Y. Ekinci, P. Kaser, and S. Park, "Photon-beam lithography reaches 12.5 nm half-pitch resolution," *J. Vac. Sci. Technol. B* **25**(1), 91–95 (2007).

103. K. E. Gonsalves, M. Thiyagarajan, J. H. Choi, P. Zimmerman, F. Cerrina, P. Nealey, V. Golovkina, J. Wallace, and N. Batina, "High performance resist for EUV lithography," *Microelectron. Eng.* **77**, 27–35 (2005).

104. D. Attwood, *Soft X-rays and Extreme Ultraviolet Radiation, Principles and Applications*, Oxford University Press, New York (2000).

105. *Extreme Ultraviolet Lithography*, B. Wu and A. Kumar, Eds., McGraw Hill, New York (2009).

106. *EUV Lithography*, V. Bakshi, Ed., SPIE Press, Bellingham, Washington (2009).

107. A convenient energy converter can be found at http://heasarc.gsfc.nasa.gov/cgi-bin/Tools/energyconv/energyConv.pl.

Chapter 13
Alternative Lithography Techniques

In addition to EUV lithography, a number of other alternatives to optical lithography have been or are being pursued. These alternative techniques have sometimes been called *next-generation* lithographies, and are frequently referred to by the acronym NGL. A number of alternatives to optical lithography have been conceived, but none has yet been developed to the point that it is ready for implementation in manufacturing. Several of these alternative lithographic techniques—proximity x ray, electron-beam direct write, electron projection, ion-projection lithography, nanoimprint lithography, and directed self-assembly—are discussed in this chapter. Each of these approaches has technical challenges that must be overcome before they will be usable in semiconductor manufacturing. In this chapter, the basic concepts underlying several of these technologies are discussed, and the challenges that need to be addressed are highlighted.

13.1 Proximity X-ray Lithography

Optical lithography is limited by diffraction, which is most significant when objects are comparable in size to the wavelength of light. This fact of physics has driven decreases in the wavelength of light used in optical lithography. Similarly, the lithographic use of wavelengths in the x-ray portion of the electromagnetic spectrum was motivated by the idea that diffraction effects could be effectively neutralized by using photons with extremely short wavelengths. However, at x-ray wavelengths there are no known materials for making image-forming lenses or mirrors. Consequently, x-ray lithography involves the use of proximity printing, where the mask is brought to within a few microns of the wafer and the x rays are passed directly through the mask and onto the wafer (Fig. 13.1). This is in contrast to optical lithography, which has the potential for projection of the image by a lens.

Since there are no materials that are highly transparent, x-ray masks are made on very thin membranes (thickness <2 μm) comprised of low-atomic-number materials, on which the circuit patterns are placed in the form of high-atomic-number material (Fig. 13.1). A large percentage of the x rays pass through the low-atomic-number material, but the x rays are absorbed or scattered by the high-atomic-number materials, thus generating pattern contrast. Silicon carbide

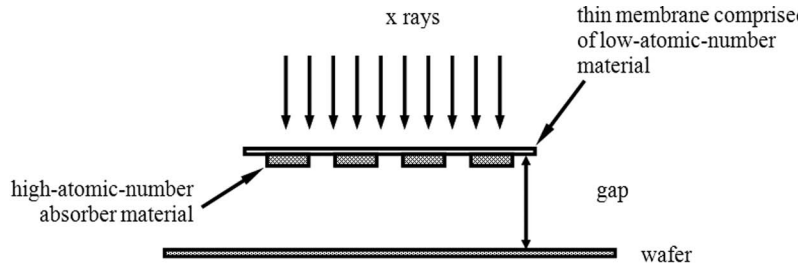


Figure 13.1 X-ray proximity lithography with a collimated light source.

is a typical membrane material, and silicon nitride films were used early in the development of x-ray lithography.

Silicon carbide is a good material to use for the membrane because it has a high Young's modulus (450 GPa), a characteristic that minimizes mechanical distortion, and it is not damaged by long exposure to x rays.[1] Gold[2] and tungsten[3] have been used as absorbers, but the best success has been found with compounds of Ta, such as TaN,[4] $TaSiN_x$,[5] and Ta_4B,[6] because they are compatible with various etch and cleaning processes.[7]

The use of thin membranes for masks introduces a set of challenges. Such films deform because of stresses, and there has been extensive work to understand and control them.[8,9] Mask deformation is particularly problematic for x-ray lithography, because there is no reduction of the image between the mask and wafer. This 1:1 pattern transfer necessitates very tight tolerances for the masks, relative to 4:1 or 5:1 reduction printing. On the other hand, with x rays there are no lens distortions or feature size–dependent pattern-placement errors, so a greater part of the overlay budget can be allocated to the mask in x-ray lithography. However, thin-film masks are susceptible to vibrations when stepped or scanned, and this needs to be addressed in any x-ray exposure system.[10] Diamond was pursued as an x-ray-mask-membrane material because its extremely high Young's modulus (900 GPa) reduces mask mechanical distortion.[11]

X-ray masks are typically made from silicon wafers. The membrane-mask area is fabricated in the center of the wafer. The mask fabrication process is outlined in Fig. 13.2. Because the membrane area is fragile, and silicon wafers are too thin to provide stability, frames or rings have been adopted for x-ray masks. The fragile silicon wafer is bonded to the ring to provide mechanical strength. An example of such a frame, the ARPA-NIST standard, [12] is shown in Fig. 13.3. The frame is made of Pyrex to facilitate bonding between the frame and the silicon on which the mask is fabricated.

Until the recent interest in EUV lithography, there has been greater investment in x-ray lithography than any other potential successor to optical technology. There have been programs at several universities, such as MIT and the University of Wisconsin, and at companies such as IBM, AT&T, and NTT. X-ray technology has also received support from the United States and Japanese governments. X-ray step-and-repeat and step-and-scan systems were made available commercially.[13,14]

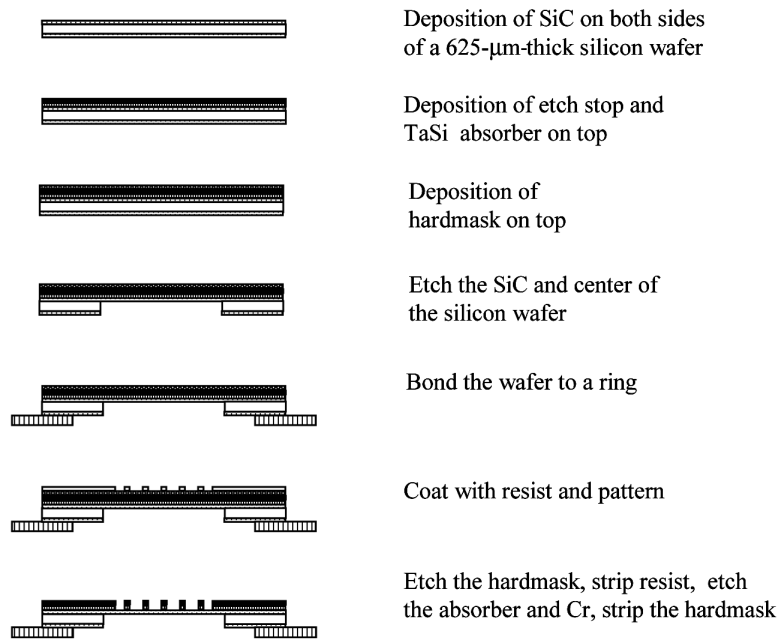


Figure 13.2 The x-ray mask fabrication process.[12]

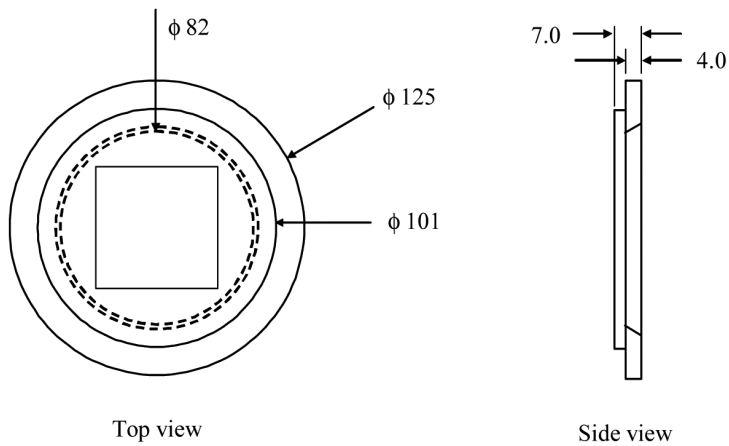


Figure 13.3 The ARPA/NIST x-ray-mask standard mounting fixture.[12] All dimensions are given in mm.

For many years, x-ray sources were a problem for lithographic applications. Point sources of x rays, such as those used in medicine or dentistry, have long been available. However, such sources are far from ideal for use in lithography. This is illustrated in Fig. 13.4, which shows that there are pattern-placement effects that do not occur with a collimated source. Since $L \gg r$, the displacement d is given to good approximation:

$$d = r\frac{g}{L}. \tag{13.1}$$

What this shows is that the pattern shift is a function of position on the mask and depends upon the gap. For a point source, the x-ray masks need to be fabricated with adjustment for this magnification, and a stringent requirement for gap control is imposed in order to maintain good overlay.

Because of this deficiency of point x-ray sources, synchrotron radiation was often adopted for use in x-ray lithography. Synchrotron radiation is produced by moving electrons at speeds approaching the speed of light (relativistic) along a circular arc. When this is done, synchrotron radiation is emitted in a narrow range of angles in the forward direction, along a tangent to the arc and in the same plane. By the use of extremely powerful magnets, relativistic electrons are bent in an arc, causing them to emit synchrotron radiation down a beamline connected to an x-ray exposure tool (Fig. 13.5). This directionality, intrinsic to synchrotron radiation, is useful for x-ray lithography because mirrors and lenses cannot be used to gather

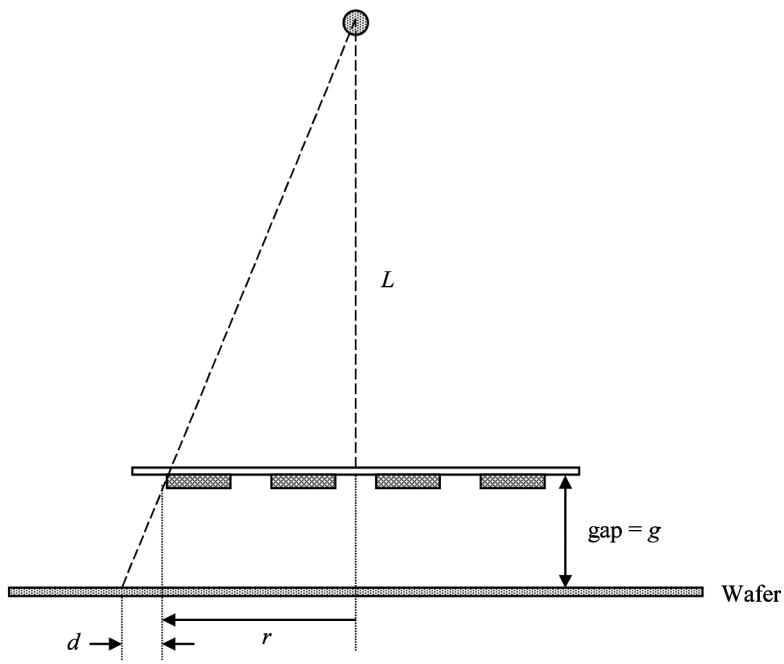


Figure 13.4 Geometry with a point source.

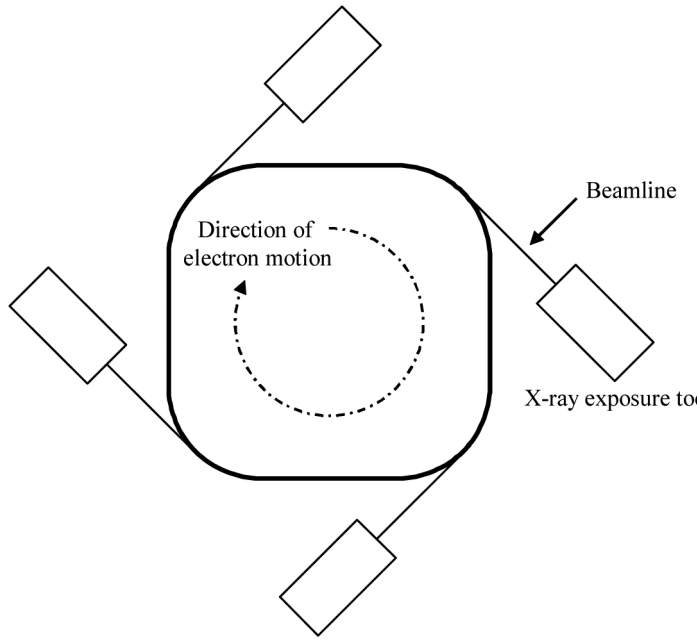


Figure 13.5 Electron storage ring in a lithography facility.

photons efficiently at x-ray wavelengths and direct them toward the exposure tool. Intensity is proportional to the current of electrons. The combination of collimation and intensity make synchrotrons good sources of light for x-ray lithography.

Synchrotron light sources are more complex and expensive than arc lamps or excimer lasers, but are not unreasonable sources of x rays for use in lithography. Because they have no moving parts, synchrotrons are very reliable.[15] A single storage ring can supply x-rays for several steppers, spreading the light-source cost across multiple exposure systems.

Because x-ray light sources also produce photons at long wavelengths, a filter is needed. Beryllium is transparent for wavelengths shorter than 15 Å. Beryllium windows also enable an ultrahigh vacuum to be maintained on the source side of the window, while allowing some level of atmosphere in the exposure tools. This transmission cutoff for beryllium sets a limit for the longest wavelengths used for x-ray lithography. Another limit on the range of wavelengths for x-ray lithography is set by the membrane materials used for the x-ray mask. The silicon-K absorption edge occurs at a wavelength of 6.7 Å, setting this as a lower bound on the usable wavelengths when the mask membrane is silicon carbide. There has been recent work on diamond membranes, which would enable the use of shorter wavelengths for x-ray lithography, but this technology is yet unproven. Consequently, x-ray lithography involves the use of photons with wavelengths between 6.7 Å and 15 Å.

A compact x-ray source has been developed that produces collimated x rays, overcoming a number of problems associated with earlier point sources.[16] The x rays are produced initially by hitting a copper foil with high-power light from a

Nd-YAG laser (1064-nm wavelength). The resulting short-wavelength spectrum is shown in Fig. 13.6. The light is then collected by arrays of fine glass polycapillary fibers, an example of which is shown in Fig. 13.7. These enable much of the light to be collected, and the capillaries also collimate the x rays. This light source has the collimation advantages of synchrotrons but is still a compact source.

Even with wavelengths between 6.7–15 Å, diffraction is significant for feature sizes in the range of interest, <100 nm. Gaps between the mask and wafer typical of x-ray lithography produce configurations in which Fresnel diffraction occurs. For optical lithography, the minimum printable linewidth is related to the numerical aperture of the lens, the wavelength λ of the light, and the resist process, through

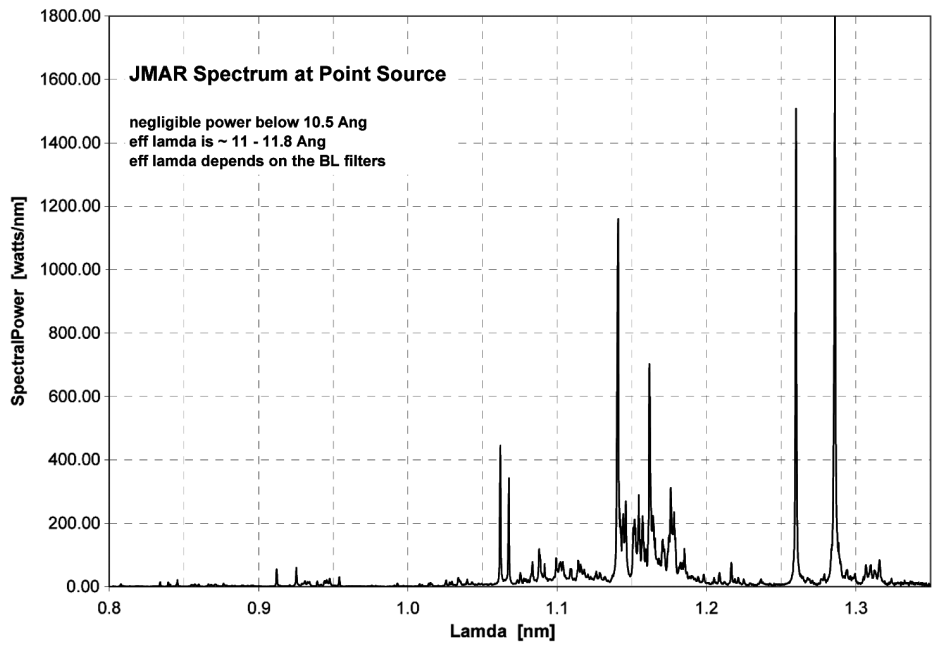


Figure 13.6 Spectrum from compact x-ray light source. The x rays are produced by hitting a copper foil with high-intensity light from a Nd:YAG laser.[17]

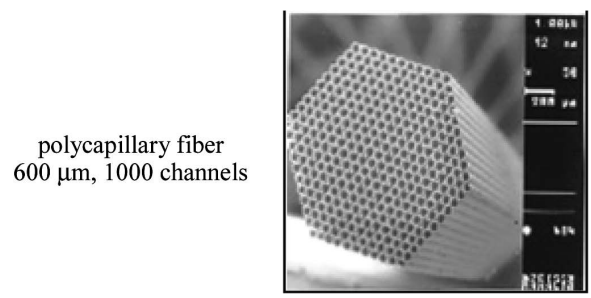


Figure 13.7 Example of a polycapillary fiber used to collimate x rays.[18]

the parameter k_1:

$$\text{minimum linewidth} = k_1 \frac{\lambda}{\text{NA}}. \tag{13.2}$$

For x-ray lithography there is a similar expression:[19]

$$\text{minimum linewidth} = \sqrt{\frac{\lambda g}{\alpha}}, \tag{13.3}$$

where g is the gap between the x-ray mask and the wafer and α is a parameter that captures the contributions from the resist process. Typically, α ranges between 0.5 and 1.5. In order to achieve resolution <100 nm, gaps less than 10 μm are required. Maintaining such tight gaps is not beyond current capability, but it is a nontrivial task. In optical lithography, there is a focus-control requirement in order to maintain good linewidth control. There is a similar requirement for gap control in x-ray lithography.[20] With a 15-μm nominal gap, linewidth variations of 3 nm per micron of gap change were measured for 150-nm features.[21] This problem of linewidth variation caused by changes in the gap increases as the targeted feature size gets smaller.

Many DUV resists, such as UVIIHS,[22] work quite well as x-ray masks. Doses range from 100–300 mJ/cm^2 with this particular chemically amplified resist. This simplified the x-ray lithography development effort, since much work could be done using commercially available resist materials.

Of all the challenges to x-ray lithography, the greatest involves the mask. Because there is no potential for image reduction, defects, linewidth variation, and misregistration are transferred from the mask to the wafer unmitigated by the reduction found in optical steppers. There is also no potential for a pellicle that will keep particulate defects out of the depth-of-field. On the other hand, because there are no optics, there are no lens contributions to linewidth variation and misregistration, so 1× x-ray masks need to have linewidth variations and misregistration only about one third that of 4× reticles. However, since leading-edge mask-making capability is needed just to meet the requirements of 4× reticles, this tightening of requirements by a factor of three has made 1× x-ray masks virtually impossible to make. As a consequence, most x-ray programs have been scaled back considerably or terminated altogether.

13.2 Electron-Beam Direct-Write Lithography

Electron-beam methods, such as those discussed in Chapter 7 for patterning photomasks, are used to create patterns on wafers directly. Electrons, being particles with mass, have extremely short wavelengths at the energies typically used for electron-beam direct-write lithography (5–100 keV), and require features of atomic dimensions to be diffracted.[23] Electron lithography therefore has potential for very high resolution, producing features 10 nm and smaller.[24] Electron

beams also image with extremely large depths-of-focus, providing relief from one the most challenging problems of optical lithography.

Conceptually, electron-beam direct-write lithography is particularly appealing for manufacturing products made in extremely low volume. In Chapter 11, the impact of mask prices on cost of ownership was discussed. With complete mask sets for products often exceeding $2M and increasing with time, masks contribute significantly to the cost of products produced in low volume (<1000 wafers/reticle set). For these types of products, electron-beam direct-write lithography looks attractive because there are no mask costs. However, direct-write lithography is currently not very efficient for high-volume manufacturing. The high efficiency of optical lithography, discussed in Chapter 1, is obtained because many features on the mask are transferred to the wafer in parallel. On the other hand, direct-write lithography fundamentally involves serial processes, making it slow. This low speed is tolerable for mask making, where only a few chips are patterned on reticles, but writing time does become problematic when trying to pattern many chips per wafer and on more than just a few wafers.

A brief analysis serves to illuminate the magnitude of the problem. Consider a 22-nm technology patterned with a raster scanning tool. If the pattern is divided into pixels and we want to be able to adjust the position of the edge of every feature, then 11×11-nm pixels are needed, of which there are 8.26×10^{11} such pixels in one square centimeter. Writing at a rate of one-billion pixels per second, it would take nearly a week to completely pattern a single 300-mm-diameter wafer.

Increasing direct-write throughput significantly also represents a challenge in data-transfer technology. In order to achieve 1-nm edge placement for 22-nm technology, a 5-bit per-pixel data representation is required to achieve the required level of gray scaling.[25] The total design information density therefore exceeds a terabit (10^{12}) per cm^2. Comparable data volumes would be required if vector scanning was used, assuming similar flexibility in pattern placement. To pattern one 300-mm wafer per minute, it is necessary to transfer information at a rate of approximately

$$\frac{\pi \times (150\ \text{mm})^2}{(11\ \text{nm})^2} \times \frac{5\ \text{bits}}{60\ \text{sec}} = 48.7\ \text{Tb/sec}. \tag{13.4}$$

Such data rates require well-engineered data-storage systems from which information can be rapidly retrieved. Lossless data compression is also needed, since even the fastest electronics and optical-transmission systems cannot handle the data rate estimated in Eq. (13.4). Data compression techniques have been specifically developed for application to direct-write lithography.[26] To satisfy data-rate-transfer requirements, data decompression will need to be performed using specialized high-speed electronics rather than software. As can be seen from these estimates, writing complex circuits at advanced nodes on large-diameter wafers using a single electron beam is a complex and slow process.

In addition to writing directly on the wafer with a single beam, other approaches have been considered to improve throughput, including:

1. Multiple-beam direct write
2. Cell projection
3. Projection with use of a scattering mask.

Each of these types of electron-beam lithography involves some level of parallelism in the data transfer, and each will be considered in this chapter, following a more detailed discussion of single-beam systems.

Before proceeding to discuss each of these forms of electron-beam lithography at greater length, there are a few general properties that need to be considered when patterning wafers using electron beams. First, it should be recalled that electrons are scattered by solids (Fig. 7.11), leading to broadening of the beam as the electrons pass through resist. Such forward scattering limits the resolution capability of electron-beam lithography. Suppose that such broadening is quantified by a standard deviation σ_f of a Gaussian that approximates the broadening of the beam. Monte Carlo simulations of this broadening, in terms of electron energy E (for $E < 50$ keV) and resist thickness t (for $t > 0.25$ μm), have been fit to the following expression:[27]

$$\sigma_f = \left(\frac{9.64t}{E}\right)^{1.75}, \tag{13.5}$$

where σ_f and t are both measured in units of microns and E is in units of keV. Due to forward scattering alone, high resolution requires thin resist and high-energy electrons (see Problem 13.4).

Such issues of resolution degradation and proximity effects were discussed in the context of mask fabrication in Chapter 7. There are, however, differences when writing directly onto wafers. First, because masks are typically used for reduction lithography, the feature sizes of interest on the mask are usually bigger than those on the wafer, even when subresolution features are included. This eases the requirements for mask making, relative to direct writing on wafers.

Another problem associated with directly writing on wafers is energy deposition. The scattering of electrons in the resist is reduced when using high-energy electrons, something desirable for high-resolution electron-beam lithography. However, most high-energy electrons pass through the resist film and deposit their energy in the underlying substrate. For mask making, this is not particularly problematic, for several reasons.[28] First, electrons more energetic than 50 keV are rarely used for mask making because the resolution requirements for mask making are less stringent than for writing directly on wafers, a consequence of reduction lithography. For some of the concepts for direct writing on wafers, electrons with energy greater than 50 keV may be required (Section 13.2.4), and more energy is thus deposited into wafer substrates than typically occurs in mask making. More importantly, reticles are usually made from fused silica, which has a low coefficient of thermal expansion (~0.5 ppm/°C), nearly an order of magnitude smaller than for silicon. Typical photomasks also have much greater thermal capacity than wafers. The combination of a lower coefficient of thermal expansion

and a greater thermal capacity offsets fused silica's lower thermal conductivity relative to silicon. Even with electron energies as low as 5 keV, wafer thermal distortion is large enough to affect overlay at a level of significance for 22-nm technology and beyond,[29] unless great care is taken for mitigation.

Another important consideration relevant to the problem of thermal distortion is throughput. Write times of several hours are considered acceptable for photomasks, since far fewer masks need to be fabricated than wafers. For patterning wafers, such throughputs are acceptable for a few applications in research or prototyping, but a serious application of electron lithography to manufacturing requires throughputs of several wafers per hour. This implies that the rate of energy deposition for wafer patterning should be more than two orders of magnitude greater than for mask making. The heat generated by such a greater rate of energy deposition cannot be conducted away readily, leading to greater temperature changes. For all of these reasons, silicon wafers will mechanically distort more than photomasks as a consequence of energy deposition, and the effect on registration is more. For patterning wafers with electron beams, overlay is a significant issue.

Resist sensitivity is a factor that directly influences throughput. As noted in the previous chapter, lithographic processes involving very sensitive resists will also be subject to high line-edge roughness and dimensional variation. For example, suppose we are patterning 22×22 nm contacts, and wish to achieve <5% 3σ dose variation as a consequence of shot noise. This necessitates resists with sensitivity of 120 $\mu C/cm^2$ or greater (see Problem 13.5). With each new generation, linear dimensions scale as ~0.7×, and areas scale as ~0.5×. This means that every new generation of e-beam technology will require ~2× increase in dose (as measured in $\mu C/cm^2$). However, the beam will also be ~2× smaller in area in order to achieve the needed resolution, so this higher dose can be achieved by focusing a fixed current into a smaller spot size. Thus, in order to maintain throughput node to node, currents and scanning rates of the electron beam must necessarily increase ~2×. It is for this reason that enthusiasm for direct-write electron-beam lithography has waxed and waned over time. While electron-beam systems have at times met the throughput requirements for manufacturing, the ability to meet such needs for subsequent nodes has involved engineering challenges not always overcome in time to meet throughput and lithographic requirements.

Highly energetic electrons deposit a net negative charge to the substrate (see Fig. 9.5). If the substrate builds up a negative charge, the electron beam is eventually deflected electrostatically. In this situation, the electrical conductivity of silicon is an advantage, and pattern-placement errors due to such hypothesized deflections have been smaller than measurement capability (20 nm) at the time the experiments were performed.[30] However, for the very tight overlay requirements of the 22-nm node and beyond, electrically conducting layers under the resist may be required. On the other hand, electron-beam lithography systems are operated in vacuum, so the interferometers used for controlling wafer-stage position are not subject to air-induced noise. This enables the use of somewhat simpler wafer stages than for very advanced optical-exposure tools.

13.2.1 Single-beam direct-write systems

Direct-write electron-beam writers very similar to the mask-making systems described in Chapter 7 have been developed and built for fast-turn applications and research and development. These systems are typically vector-scan and/or shaped-beam systems when throughput is the priority, while they are Gaussian-beam systems when the highest resolution is desired. One of the first commercially available systems for writing directly on wafers was the AEBLE 150 from Perkin-Elmer's Electron Beam Technology Division.[31,32] Since then, several electron-beam direct-write systems have been produced. Some examples of direct-write systems currently commercially available are listed in Table 13.1.

In addition to shaped-beam systems, there are Gaussian-beam tools that provide higher resolution at the expense of throughput. Examples of such systems are Vistec's EBPG5000plus, JEOL 9300FS, and the Raith150-TWO, which have been used to create features ≤10 nm. Such beam writers are useful where the number of features that must be patterned is very small.

In addition to commercially produced electron-beam systems, there was a long-term electron-beam direct-write program at IBM that produced a succession of tools, EL-1 through EL-4,[33,34] based upon shaped beams.[35] This technology evolved into the EL-5, which was used to make 1× x-ray masks.[36]

Another application of direct-write e-beam lithography is in technology research, where wafer throughput does not need to be very high. For example, a Leica SB 320-50 SW shaped beam tool (the Leica e-beam operations are now part of Vistec) was used to fabricate SRAM cells with 65-nm design rules at a time when 90-nm technology had not yet reached manufacturing.[37] E-beam patterning was used on the critical active, gate, contact, and metal layers. Throughput was very low, 18–24 h for a single 200-mm wafer, but this was nevertheless useful for early development of advanced technologies, where few wafers were needed. A negative resist, Sumitomo NEB-33, was used for clear-field active and gate layers, while a positive resist, Fujifilm FEP-171, was used for contact and metal layers. The use of a negative resist for clear-field layers minimized writing time.

As discussed previously, in order to expose resist with sufficient doses to avoid shot noise–induced line-edge roughness, the highest possible beam currents are advantageous. The electron density can therefore be high within the electron-optical column, and also where the electron beam is ultimately focused on the wafer. When an electron beam is confined to a small volume, individual electrons experience the electric fields of other electrons, resulting in beam broadening.

Table 13.1 Commercially available direct-write electron-beam systems.

Supplier	System type	Model
Vistec	Shaped beam	SB3050
Vistec	Gaussian beam	EBPG5000plus
JEOL	Gaussian beam	9300FS
Raith GmbH	Gaussian beam	Raith150-TWO

This leads to a fundamental tradeoff between resolution and large-current electron beams capable of supporting high throughput.[38] While the resulting average field can be compensated, there is stochastic scattering—the random electron-electron interactions that will tend to broaden the beam.[39] Beam blur is a consequence of angular divergences within the electron beam from its nominal direction within the electron column. For an electron beam of diameter $2r_b$ traversing a column of length L, the beam-angular divergence has been estimated as[40]

$$\Delta\alpha = \frac{k}{4\pi\varepsilon_0}\left(\frac{m}{4}\right)^{1/3}\frac{I^{2/3}L}{V^{4/3}r_b^{4/3}}, \tag{13.6}$$

where m is the mass of the electron, I is the beam current, V is the beam voltage, ε_0 is the permittivity of free space, and k is a constant ~1.5–2. The amount of beam blur depends upon the electron optical design, but Eq. (13.6) shows that the blur increases with current, regardless of the design. Single columns produce beams with blur that is too great to support the resolution requirements of the 22-nm node and beyond when the beam current reaches 10 μA. Hence, there is a fundamental tradeoff between resolution and throughput. Attempts to neutralize the beam with ions do not eliminate the stochastic beam blurring, which results from random interactions among particles.

The time t required to expose an area A covered with a resist of sensitivity S, using a beam of current I, is given by

$$t = \frac{AS}{I}. \tag{13.7}$$

This shows the direct relationship between beam current and throughput. Limitations on shot noise–induced LER prevent the resist sensitivity S from being reduced significantly to improve throughput.

Another factor limiting throughput is the ability to scan the beam controllably at high speeds. Components for scanning the electron beam and blanking it on and off involve electrostatic or electromagnetic elements, and the intrinsic resistances, capacitances, and inductances of such components ultimately limit the speed at which deflections can occur.[41]

Single-beam electron lithography is proven technology and provides useful capability when patterning small areas is adequate. However, there are several obstacles to achieving the throughputs required for producing integrated circuits cost effectively with a single beam, including the need for high exposures to avoid shot noise–induced LER, beam blur, and limitations on electronics speed. In order to achieve writing times of less than one hour per wafer, some degree of parallel imaging is required.

13.2.2 Multiple-electron-beam direct-write systems

To circumvent the limitations of a single electron beam, approaches have been proposed that involve multiple-electron optical columns[42] or multiple beams.[43]

Multi-electron-beam lithography is a challenging technology to implement, requiring extremely good reliability and innovations to produce adequate calibrations among all of the beams and columns. Nevertheless, the high costs of masks and the need for more flexible manufacturing have led to efforts to develop muliple-beam systems.

Several approaches have been proposed by Vistec,[44] IMS[45,46] and MAPPER.[47] All of these approaches involve micromachining to produce multiple electron beams in reasonably sized systems. However, there are also some differences among the approaches taken by each company, and these are described in this section.

The IMS and Vistec systems both involve multiple 50-keV electron beams projected through single electron columns. This method is illustrated in Fig. 13.8. An aperture array is used to generate separated electron beams. To avoid excessive heating of the aperture array plate, the energy of the electrons impinging on the plate (5 keV for the IMS system) is usually significantly less than the final 50 keV of the electrons that are focused onto the wafers. The use of multiple beams provides increased separation of the electrons as they traverse the electron-optical column, relative to single-beam systems, except at the crossover points. This reduces—but does not eliminate—the stochastic beam blur. Relative to a single-electron-beam system, this allows for a substantial increase of total current before stochastic beam blur becomes limiting for the resolutions of the 22-nm node and below.

The MAPPER system involves the use of both separate beams and electron lenses to avoid the limitation of stochastic blurring of high current densities within a single column. The multiple beams are generated by use of a set of apertures to form multiple beams from a single uniform and collimated electron beam (Fig. 13.9). The apertures are arranged as shown schematically in Fig. 13.10. A deflector array is used to blank beams on and off and scan the beams, thereby

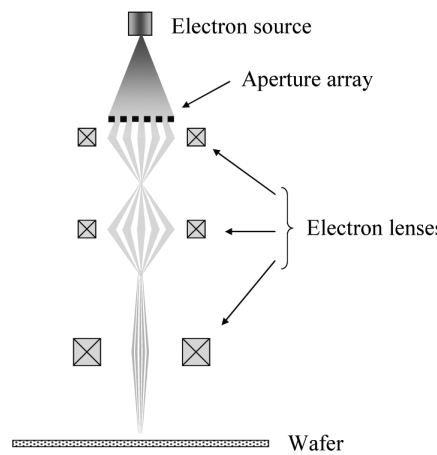


Figure 13.8 Schematic of a multi-electron-beam system using a single-column electron lens.

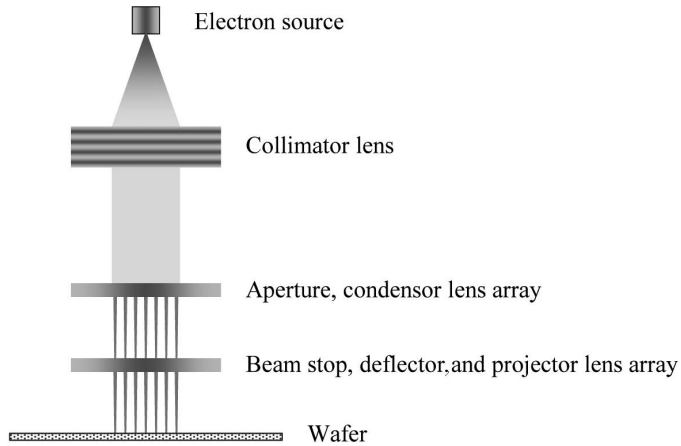


Figure 13.9 Schematic of the MAPPER multi-electron-beam lithography system.

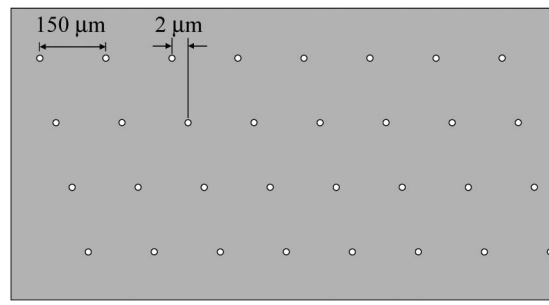


Figure 13.10 Schematic of the MAPPER aperture that generates separate electron beams.

providing patterning capability. The apertures in the rows are offset by 2 μm, and full-wafer coverage is accomplished by scanning each beam across 2 μm. To minimize wafer heating, 5-keV electrons are used in the MAPPER system, in contrast to the tens of kilovolt beams typically used for electron-beam lithography. However, from Eq. (13.5), such low energies lead to imaging blurring due to forward scattering. This can be limiting for 22-nm lithography and beyond, unless very thin resist layers are used (see Problem 13.4).

In order to achieve throughputs of 10 wafers per hour or more, MAPPER systems with tens of thousands of beams have been proposed. This large number of beams also enables redundancy to compensate for a few beams that might not be functioning properly. With many beams, the total beam current necessary for throughput of $\geq$10 wafers per hour can be achieved with low current per beam, thereby avoiding unacceptable resolution lost due to stochastic beam blurring.

The multi-electron-beam systems from IMS, Vistec, and MAPPER are based on concepts that circumvent some of the inherent current limitations of single-beam-electron lithography systems. However, increased currents will heat the wafers, and the resulting thermal expansion will cause overlay errors without compensation.

The temperature increase ΔT of an area A of a silicon wafer exposed for time t with an e-beam is given by:

$$\Delta T = \frac{Pt}{hA\rho c}, \tag{13.8}$$

where h is the wafer thickness, ρ and c are the density and specific heat of silicon, respectively, and P is the net flow of power into the section of the wafer under consideration during the time of exposure. There is a flow of energy into the wafer by the electron beams, while energy is conducted away in the form of heat, and P is the difference between the rates of the two.

A detailed analysis of thermal wafer distortion is complex, but two extreme cases can be considered, between which the overlay problem due to wafer heating by the electron beams is bounded. The following analysis is adapted from Ref. 29. If we can imagine that the volume hA of silicon is thermally isolated, then P equals the power input from the electron beam. With a throughput target of 10 wafers per hour, and a 120 $\mu C/cm^2$ resist sensitivity, from Eq. (13.7) a beam current of at least 0.24 mA is required. Equation (13.8) can be used to calculate the temperature rise resulting from heating by the electron beam, with examples given in Table 13.2.

The case opposite of thermal isolation is that where heat is conducted away readily. An infinitely thick wafer serves as a model for this case, since there are no thermal boundaries in that instance. The heated area can be considered to be a wafer of effective thickness h_{eff} equal to the thermal diffusion length, which is given by[29]

$$h_{eff} = \sqrt{\frac{\pi kt}{\rho c}}, \tag{13.9}$$

where k is the thermal conductivity of silicon. The exposure time t is calculated using Eq. (13.7) and equals 4.16 sec when the resist sensitivity is 120 $\mu C/cm^2$, the

Table 13.2 Wafer heating and resulting uncompensated registration errors for two cases of electron-beam direct write.

Parameter	5-kV beam voltage	50-kV beam voltage
Total power on wafer at 50% pattern density (W)	0.6	6.0
Temperature rise (°C) assuming no heat conduction	2.4	24
Maximum position error (nm) at the edge of a ±16-mm-long field assuming no heat conduction	±96	±960
Temperature rise (°C) assuming an infinitely thick wafer (ideal heat conduction)	0.058	0.58
Maximum position error (nm) at the edge of a 32-mm-long field (ideal heat conduction)	±2.3	23

field size is 26 × 32 mm, and the total beam current is 0.24 mA. This leads to an effective thickness of 32 mm. Even in this ideal case, overlay errors greater than 2 nm can result from electron beam–induced heating, even with beam energies as low as 5 keV. This is a significant fraction of the overlay requirements for the 22-nm node and beyond. The problem is substantially more challenging with high-energy beams. In principle, these registration errors can be corrected to some degree by detailed modeling of the wafer heating, and deflecting the electron beams to compensate.

13.2.3 Cell-projection lithography

Another way to circumvent the thoughput problem that is a consequence of writing pixel-by-pixel is to use some type of mask. Electron optics very similar to those used for single-beam direct systems can be used by keeping the field size small (a few microns on a side) for systems that follow this approach. One such system with a 5-μm × 5-μm field size that was commercially available, the HL-800D from Hitachi,[48,49] was used to make protoype high-bit-count DRAMs[50] before adequate optical lithography capability became available. With a resolution of 100 nm, such systems with a 5-μm × 5-μm field size on the wafer represented an increase in the "parallelism" of electron-beam lithography by a factor of 2500 over single-beam direct-write systems. More recently, Advantest has been producing systems with cell-projection capability.[51]

For cell-projection e-beam lithography, the second aperture (Fig. 7.1) is a mask of the pattern to be printed. These are typically stencil masks, where holes are made in a membrane to allow electrons to be transmitted in some areas while blocked in others. With electron optics, reduction imaging is possible, avoiding the 1× problem of masks encountered in x-ray lithography. Cell projection e-beam systems usually also have conventional single-beam lithography capability for printing the nonrepetitive parts of the pattern.

Such an approach is clearly helpful for patterning memories, where much of the circuit area consists of repeated memory cells. As design rules have moved to smaller geometries, it is now possible to fit logic cells in a field that is only a few microns on a side, so that logic circuits can be patterned using cell-projection lithography, at least through the first few metal layers. Interconnects among logic cells will naturally remain random, and less advantage can be applied to metal layers that connect the logic cells. Nevertheless, cell-projection lithography has potential for improving the throughput of electron-beam lithography, at least to the point where the throughput becomes limited by the wafer heating discussed in the prior section. There are opportunities for improving throughput by choosing design rules that are optimized for cell-projection lithography.[52]

13.2.4 Scattering-mask electron-projection lithography

Optics have been designed for extending electron-projection lithography beyond small field sizes (5 × 5 μm).[53–56] However, there are two significant problems with using large-area stencil masks to exploit such optics.[57] First, stencil masks

are very fragile,[58] and it has proven difficult to fabricate such masks with large areas. Another potential limitation of projection-electron lithography is mask heating. To achieve adequate throughput, high currents are needed. For masks that create opaque regions by simply blocking and absorbing the electrons, the masks become hot. To overcome this problem, a very clever scheme has been proposed,[59] where nearly all electrons are allowed to pass through the mask. Instead of absorbing electrons in those portions of the mask that are supposed to correspond to unexposed areas on the wafer, the electrons are scattered by high-atomic-number materials, such as tungsten, tantalum, or other materials used for x-ray masks. At a focal plane within the electron optics is a physical aperture, through which unscattered electrons pass. However, only a very small fraction of the scattered electrons pass through the aperture (Fig. 13.11). Those portions of the mask corresponding to regions of the design that are supposed to be exposed on the wafer must allow electrons to penetrate with little scattering. As an alternative to stencil masks, the scattering materials are placed on a thin membrane of low-atomic-number material, such as silicon nitride or diamond-like carbon.[60] Fabricating masks on a continuous membrane, rather than by use of stencils, also increases mask manufacturability and durability. This combination of a scattering mask, in conjunction with a focal plane aperture in the electron optics, has been given the name SCALPEL, the acronym for SCattering with Angular Limitation-Projection Electron-beam Lithography. The SCALPEL approach has demonstrated patterning capability. Shown in Fig. 13.12 are 80-nm contacts created using SCALPEL.[61]

Membrane and stencil masks are inherently thin and fragile. Most obviously, there is the potential for damage to such masks. Another problem associated with large-field membrane masks is distortion, a problem that was discussed in the section on x-ray lithography. An innovative solution is to break the design into 1×12-mm sections on the masks, separated by tall struts. SCALPEL masks are usually made from silicon wafers, so the strut heights equal the thickness of a standard silicon wafer (0.75 mm for SEMI-standard 200-mm-diameter wafers).

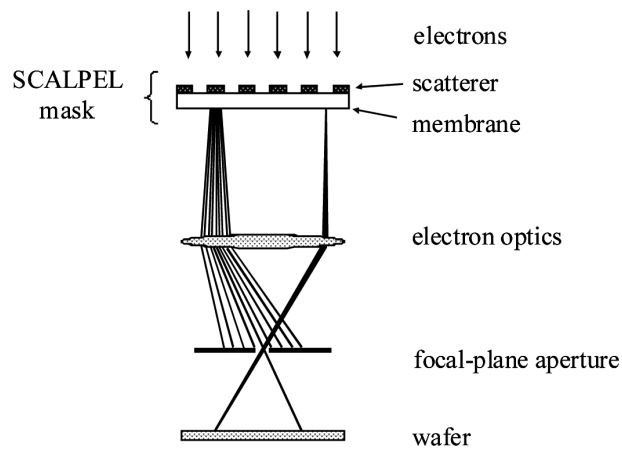


Figure 13.11 Schematic of an electron-scattering projection system.

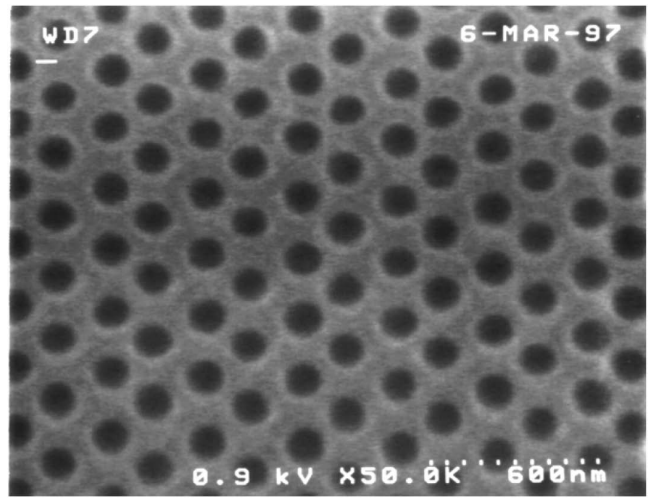


Figure 13.12 80-nm contacts produced by SCALPEL.

With 4× lens reduction, the SCALPEL field size on the wafer is 0.25 mm × 3.0 mm, so the thin membrane area between the struts on the mask is 1.0 mm × 12.0 mm. To create a complete chip pattern on the wafer, the individual 0.25-mm × 3.0-mm fields that comprise the total field need to be stitched together.

This stitching is not a trivial challenge for the SCALPEL technology, since the left half of features, written in one field, need to line up with the right half, with a tolerance that is a small fraction of the linewidth. What complicates this stitching is the heating of the mask and wafer that occurs during exposure. While very little energy is deposited into the SCALPEL mask, the membranes, by being so thin, have very little thermal mass, and they heat appreciably during exposures even though little total energy is deposited. Simulations show that the temperature increase, which occurs nonuniformly across the mask, exceeds 7 °C, causing deformations of ~20 nm.[62] Similarly, most of the energy of the 100-keV electrons used in the SCALPEL technology is deposited into the silicon wafer, causing temperature rises calculated to be several °C.[63] This problem of wafer heating has already been discussed in this chapter. In principle, adjustments can be performed with the electron-beam optics, rather than through mechanical motion, enabling very rapid corrections. Compensation for thermal distortions needs to be accomplished through software, which is known to be the least-reliable component of semiconductor manufacturing equipment.

As with the other next-generation lithography technologies discussed thus far, such as x-ray and EUV lithography, SCALPEL masks cannot be protected by pellicles from particulate contamination. It is apparent that mask-defect mitigation will be a general problem for postoptical lithography.

Throughput is another challenge for SCALPEL. Because of the very high currents involved in a large-field approach to e-beam lithography, the problem of stochastic scattering is particularly severe. The amount of blur that can be tolerated decreases as feature sizes shrink. As a consequence, the beam current must be

reduced in order to image smaller features with a given electron-optical system. This leads to a familiar connection between throughput and resolution: throughput decreases in order to print smaller features. It is possible to design electron optics that minimize electron-electron interactions, but experience in this area is limited, and the extent to which electron-optical design can address the issue of beam blur remains to be seen.

In addition to the SCALPEL program originating from Bell Labs, there was another program, PREVAIL, at IBM.[53] PREVAIL included the essential elements of SCALPEL, a scattering mask and focal plane aperture, but incorporated IBM's electron-optical technology. PREVAIL stands for PRojection Exposure with Variable Axis Immersion Lenses. Nikon engaged with IBM to produce exposure tools based on the PREVAIL concepts,[64] but this program was eventually suspended.

13.3 Ion-Projection Lithography

Ions can also be projected through stencil masks.[65] Because ions scatter very little in solids, they can potentially result in very high resolution. However, because they are much more massive than electrons (the lightest ion, H^+, is approximately 2000× more massive than an electron), ions cannot be deflected at the same speed as electrons. This is a consequence of basic physics:

$$\vec{F} = m\vec{a}. \tag{13.10}$$

A much larger force $\vec{F}$ is required to accelerate an ion than to accelerate an electron. Consequently, the deflection schemes that have been proposed for SCALPEL cannot be easily adopted for ion-beam exposures. For this reason, large-field (12.5 mm×12.5 mm) ion-projection-lithography (IPL) systems have been pursued.

A schematic of an IPL system is shown in Fig. 13.13.[66] Ions with a small spread in energy are produced by a recently developed ion source.[67,68] Electrostatic lenses are then used to produce a uniform beam of ions that covers the area of the mask. Beam energies are typically ≥ 250 keV. Another set of electrostatic lenses then reduces the size of the overall pattern 4× and focuses the ions onto the wafer surface.

Ion-projection lithography requires stencil masks;[69] there is no potential for a membrane-mask option, unlike with electron-projection lithography. One of the problems with stencil masks is the *donut* problem. Consider the mask shape shown in Fig. 13.14. This cannot be made with a single stencil mask, as the center portion will be unsupported and will fall out. Creation of the geometry shown in Fig. 13.14 requires the use of at least two masks. While not fundamentally limiting, the need for double exposures will reduce throughput on some layers.

Because ion masks must block ions in order to generate masked patterns, there is considerable energy deposited into the masks. With ion beams, there is also the potential for sputtering of the masks. Research has shown that deposition of a

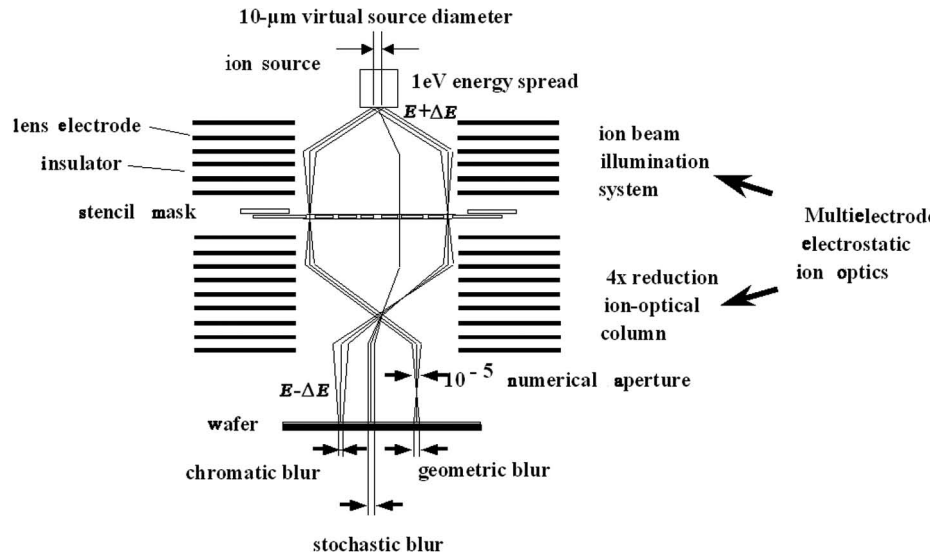


Figure 13.13 Schematic of an IPL system.[66]

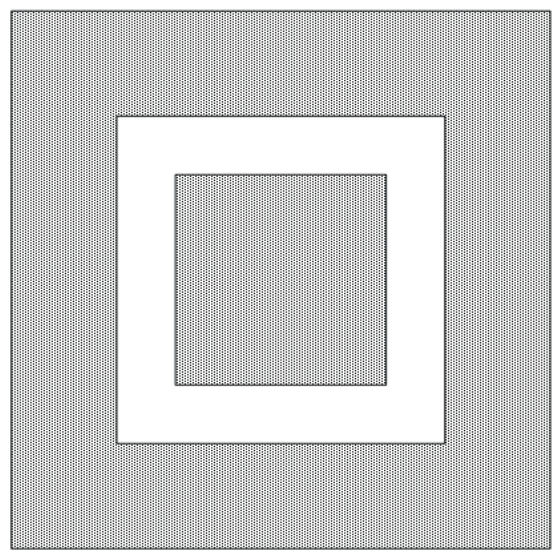


Figure 13.14 Geometry that illustrates the *donut* problem. The dark areas represent the opaque part of the mask, while the white area is open.

carbon film provides high emissivity and resistance against ion damage to enable radiative cooling of the ion-projection masks.[70]

13.4 Imprint Lithography

A nonprojection lithographic technology, imprint lithography, has been introduced.[71] Imprint lithography is essentially a micromolding process, shown schematically in Fig. 13.15. First, a template of the pattern is created. This consists of recessed areas on a piece of a hard substance that will eventually correspond to portions of the wafer where material will remain after etch. In one form of this technology, step-and-flash imprint lithography (SFIL),[72] the template is formed from glass. It is formed from nickel or silicon in another form that was called nanoimprint lithography (NIL).[73] Over time, nanoimprint has become the term applied to all imprint lithography directed at the formation of features with sizes significantly smaller than 1 μm.

To create patterns on the wafer, a small puddle of liquid is dispensed on the wafer, and the template is pressed against the liquid to "imprint" the pattern. After imprinting, the pattern is then fixed on the wafer. With SFIL, the resist material is solidified by exposure with ultraviolet light, while heat and pressure are used with the NIL version of imprint lithography. The template is then released, and the substrate can then be etched.

With this technology, extremely high-resolution lithography has been achieved. Not only have 20-nm features been produced, but small fabrication errors in the template of just a few nanometers in size have been replicated (Fig. 13.16). This technology clearly has high-resolution capability.

Imprint lithography has some definite advantages and disadvantages.[75] Because there is no high-resolution lens required, imprint patterning tools are considerably cheaper than high-performance step-and-scan exposure tools. Offsetting this advantage is significantly less throughput than typical of optical step-and-scan systems, a consequence of the mechanical nature of imprint lithography. Also, due

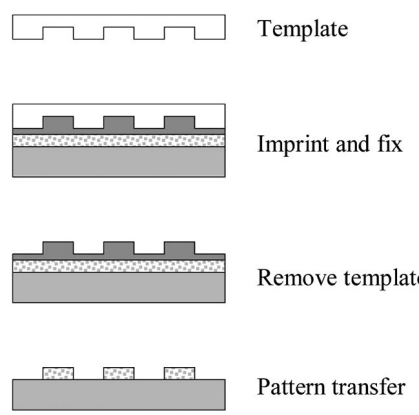


Figure 13.15 Schematic of imprint lithography.

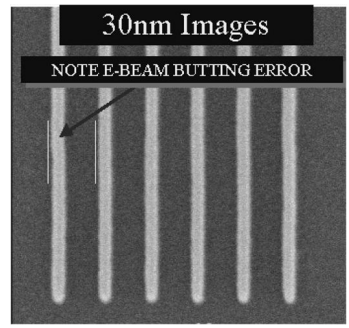


Figure 13.16 30-nm features produced by imprint lithography.[74]

to the 1× nature of the template, difficulties similar to those experienced with x-ray masks might be expected. However, there are some aspects of template fabrication for imprint lithography that make them easier to produce than x-ray masks. In particular, it was necessary to fabricate x-ray masks on thin membranes, which made registration control very difficult. In contast, imprint templates are formed on rigid glass substrates. Even on glass substrates, the 1× nature of the template represents a formidible challenge.

Imprint lithography is a contact-patterning method. Projection optical lithography was developed as a replacement for optical-contact printing because defect levels were too high with contact printing to support high levels of integration. Nevertheless, imprint lithography is used in applications which are defect tolerant, have loose or no requirements for overlay, and low levels of integration. One example of the application of imprint lithography is the generation of patterned media for magnetic storage.[76,77]

13.5 Directed Self-Assembly

Researchers have attempted to create patterns that might be used to fabricate integrated circuits by leveraging the inherent tendency of specific molecules to assemble into regular patterns. Self-assembly is based upon the use of polymer chains comprised of linked immiscible blocks (Fig. 13.17). Because of the mutual immiscibility of the different blocks, polymer chains deposited on a flat surface will be driven thermodynamically to a single orientation, creating a regular pattern. By using blocks that provide relative etch selectivity, this pattern can be transferred to the substrate. For example, one block may be silicon containing, while the other is purely hydrocarbon. An oxidizing plasma etch will remove the hydrocarbons while leaving silicon-containing material. A common combination involving two hydrocarbon materials is polystyrene and polymethylmethacrylate.[78]

To be useful for fabricating semiconductor devices, it is necessary to register the arrays produced by self-assembly with patterns existing on the wafer. For example, metal lines need to be on top of contacts. One way to do this is to coat the wafer with a material that attracts one of the block types. This material can then be patterned with loose-pitch optical (or other type) lithography. The resulting

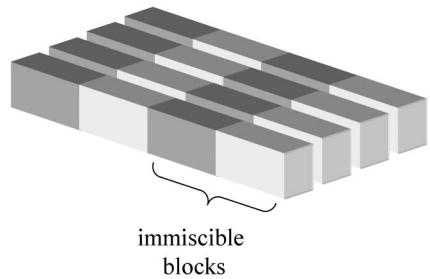


Figure 13.17 Illustration of polymer chains comprised of immiscible blocks.

self-assembled pattern will resemble that in Fig. 13.18, where pitch splitting is achieved. The use of an underlayer film to guide the self-assembly[79] is referred to as *chemical epitaxy*. It is also possible to use physical features, as illustrated in Fig. 13.19. Without anchoring the pattern, self-assembling materials will form random patterns, such as seen in the left side of Fig. 13.20. Properly directed, self-assembled films can form useful patterns.

It is also possible to create arrays of dots or holes by using blocks of different sizes, as illustrated in Fig. 13.21. The polymers will be oriented perpendicular to the substrate. Depending on properties of the specific block polymers, square or hexagonal arrays can be formed. The resulting arrays are useful for creating patterns for magnetic storage, for example.[81]

13.6 The Ultimate Future of Lithography

The lithographic technologies described in this chapter are currently the focus of research and development to see if they can be made useable for the large-scale fabrication of semiconductor and other devices requiring the formation of specific and very small patterns. The extraordinary accomplishments by lithographers over the past several decades has made it very challenging to match the productivity and capability of optical lithography. Because of the inherent limits to optical

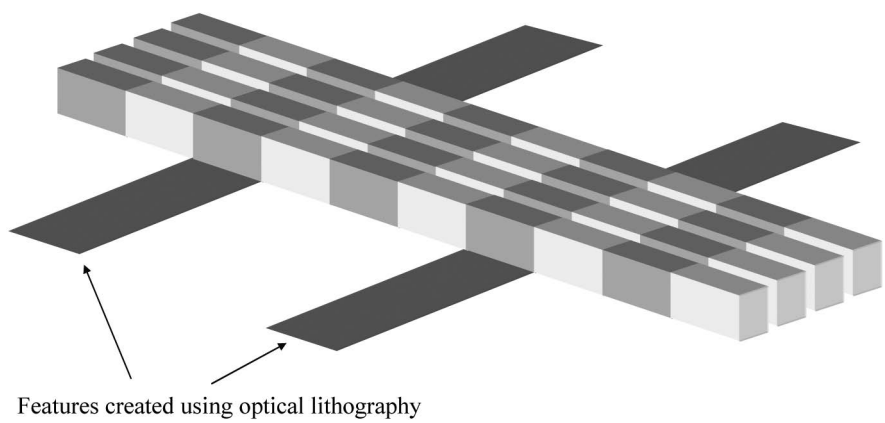


Figure 13.18 Pitch splitting by the use of directed self-assembly and chemical epitaxy.

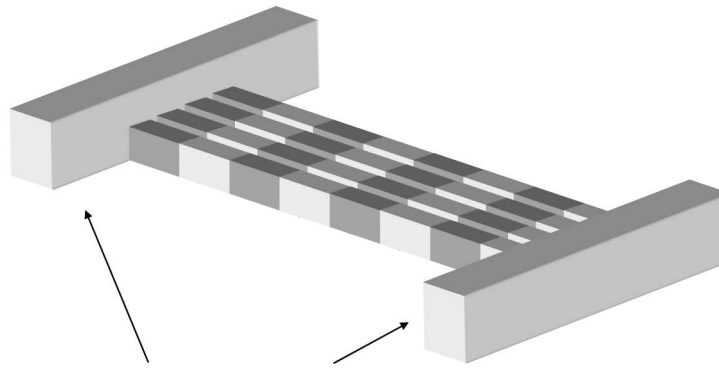


Figure 13.19 Pitch splitting by the use of directed self-assembly and graphoepitaxy.

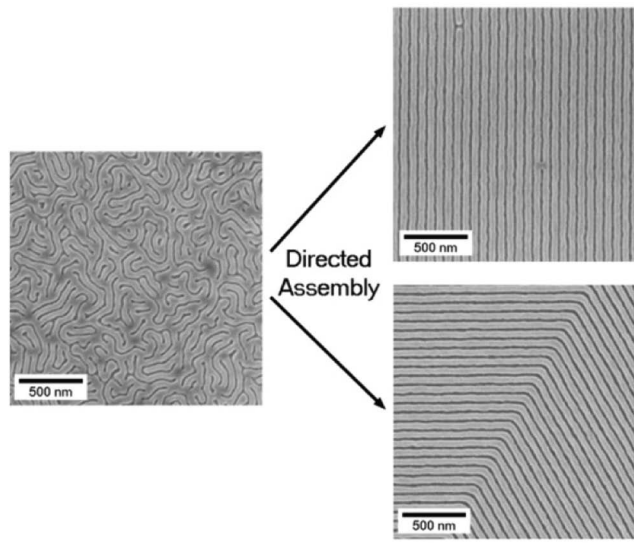


Figure 13.20 Materials with capability for self-assembly will form random patterns (left), while achored features (right) form patterns useful for circuit fabrication.[80]

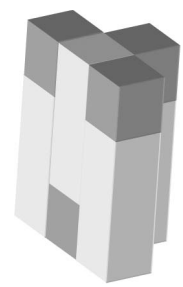


Figure 13.21 Creation of contact or hole arrays.

lithography and its high cost at its most capable, there are motivations to try alternatives. This is particularly true for making products in low volumes, where mask costs become prohibitive. Whether any of the techniques described in this chapter ultimately are used in high-volume manufacturing is something that we can look forward to seeing in the future.

Problems

13.1 Consider an x-ray lithography system with a gap of 20 μm and a distance of 2 m between the point source and mask. If the distance between the center and edge of the exposure field is 10 mm, what is the pattern shift for geometries at the edges of exposure fields, compared to the center due to the noncollimated light source [Eq. (13.1)]?

13.2 For x-ray lithography, $\lambda \approx 1$ nm. What gap between the mask and wafer is required for a resolution of 50 nm, assuming $\alpha = 1$? Is control of such a gap practical?

13.3 Show that the deBroglie wavelength of a 5-keV electron is 1.7×10^{-2} nm and is 5.0×10^{-3} nm for a 50-keV electron. Comment on the resolution limit of electron lithography due to the wavelength of electrons.

13.4 The full-width half-maximum (FWHM) of a Gaussian distribution with a standard deviation of σ is 2.36σ. Suppose that we want to pattern 22-nm features using 5-keV electron beams, and assume that 22 nm ≈ FWHM of the beam. Using Eq. (13.5), show that the resist can be no thicker than 36 nm to achieve this resolution.

13.5 For electron-beam lithography, show that a dose of at least 120 $\mu C/cm^2$ is required to maintain <5% 3σ dose variation in 22 × 22-nm contacts.

13.6 Assuming that all other technical difficulties can be overcome, show that the time required to expose one tenth of the area of a 300-mm wafer covered with a resist with 120 $\mu C/cm^2$ sensitivity and using a beam current of 10 μA is 14 min.

References

1. P. A. Seese, K. D. Cummings, D. J. Resnik, A. W. Yanfo, W. A. Johnson, G. M. Wells, and J. P. Wallace, "Accelerated radiation damage testing of x-ray mask membrane materials," *Proc. SPIE* **1924**, 457–466 (1993).
2. G. E. Georgiou, C. A. Jankowski, and T. A. Palumbo, "DC electroplating of sub-micron gold patterns on X-ray masks," *Proc. SPIE* **471**, 96–102 (1984).
3. K. D. Cummings, D. J. Resnick, J. Frackoviak, R. R. Kola, L. E. Trimble, B. Grant, S. Silverman, L. Haas, and B. Jennings, "Study of electron beam patterning of resist on tungsten x-ray masks," *J. Vac. Sci. Technol. B* **11**(6), 2872–2875 (1993).

4. J. T. Sheu, A. Chu, J. H. Ding, and S. Su, "Characteristics of sputtered TaX absorbers for x-ray masks," *Proc. SPIE* **3676**, 42–45 (1999).
5. J. M. Rocque, D. M. Puisto, D. J. Resnick, K. D. Cummings, W. Chuc, and P. A. Seese, "SNR200 chemically amplified resist optimization," *Proc. SPIE* **3048**, 90–99 (1997).
6. C. J. Brooks, K. C. Racette, M. J. Lercel, L. A. Powers, and D. E. Benoit, "Advanced refractory-metal and process technology for the fabrication of x-ray masks," *Proc. SPIE* **3676**, 14–23 (1999).
7. J. Leavey and P. J. S. Mangat, "Mask and wafer inspection and cleaning for proximity x-ray lithography," *Proc. SPIE* **3331**, 179–188 (1998).
8. E. P. Cotte, R. L. Engelstad, E. G. Lovell, and C. J. Brooks, "Predicting mechanical distortions in x-ray masks," *Proc. SPIE* **3676**, 429–440 (1999).
9. M. F. Laudon, D. L. Laird, R. L. Engelstad, and R. Cerrina, "Mechanical response of x-ray masks," *Jpn. J. Appl. Phys.* **32**, 5928–5932 (1993).
10. M. P. Schlax, R. L. Engelstad, and E. G. Lovell, "Dynamic characterization of step-induced vibrations of x-ray mask membranes," *Proc. SPIE* **3331**, 629–637 (1998).
11. C. J. Brooks, L. A. Powers, R. E. Acosta, D. Molly, F. Faili, and J. A. Herb, "Characteristics of chemical vapor deposition diamond films for x-ray mask substrates," *J. Vac. Sci. Technol. B* **17**(6), 3144–3148 (1999).
12. D. Laird, M. Laudon, and R. Engelstad, "Practical considerations in x-ray mask mounting methodology," *J. Vac. Sci. Technol. B* **11**(6), 2953–2957 (1993).
13. K. Uda, N. Mizusawa, Y. Tanaka, Y. Watanabe, H. Ina, and S. Uzawa, "X-ray stepper development for volume production at Canon," *Proc. SPIE* **3331**, 689–697 (1998).
14. Q. Leonard, J. Wallace, O. Vladimirsky, K. Simon, and F. Cerrina, "X-ray mask replication using a Suss stepper at CXrL," *Proc. SPIE* **3048**, 299–303 (1997).
15. J. P. Silverman, "Proximity x-ray lithography: a white paper for the 1998 SEMATECH next generation lithography workshop," SEMATECH (1998).
16. C. J. Gaeta, H. Rieger, I. C. E. Turcu, R. A. Forber, K. L. Cassidy, S. M. Campeau, M. J. Powers, J. R. Maldonado, J. H. Morris, R. M. Foster, H. I. Smith, and M. H. Lim, "High power laser-plasma x-ray source for lithography," *Proc. SPIE* **4688**, 818–833 (2002).
17. This figure was provided by John Ricardi of JMAR Technologies, Carlsbad, CA. (www.jmar.com).
18. Such polycapillary fibers are produced by X-ray Optical Systems (XOS), Inc. of East Greenbush, New York. (www.xos.com). The figure was provided by John Ricardi of JMAR Technologies.

19. H. I. Smith and F. Cerrina, "X-ray lithography for ULSI manufacturing," *Microlithography World*, 10–15 (Winter, 1997).
20. B. J. Lin, "A new perspective on proximity printing: from ultraviolet to x-ray," *J. Vac. Sci. Technol. B* **8**, 1539–1546 (1990).
21. A. A. Krasnoperova, R. Rippstein, D. Puisto, and J. Rocque, "ACLV control in x-ray lithography," *Proc. SPIE* **3331**, 165–178 (1998).
22. A. A. Krasnoperova, R. Rippstein, A. Flamholz, E. Kratchmer, S. Wind, C. Brooks, and M. Lercel, "Imaging capabilities of proximity x-ray lithography at 70 nm ground rules," *Proc. SPIE* **3676**, 24–39 (1999).
23. E. E. Anderson, *Modern Physics and Quantum Mechanics*, W.B. Saunders, Philadelphia (1971).
24. H. G. Craighead, "10-nm resolution electron-beam lithography," *J. Appl. Phys.* **55**(12), 4430–4435 (1984).
25. V. Dai and A. Zakhor, "Lossless layout compression for maskless lithography systems," *Proc. SPIE* **3997**, 467–477 (2000).
26. H.-I. Liu, V. Dai, A. Zakhor, and B. Nikolić, "Reduced complexity compression algorithms for direct-write lithography systems," *J. Micro/Nanolith. MEMS MOEMS* **6**(1), 013007 (2007).
27. A. N. Broers, "Resolution, overlay and field size for lithography systems," *IEEE Trans. Electr. Dev.* **ED-28**(11), 1268–1278 (1981).
28. B. Shamoun, R. Engelstad, and D. Trost, "Assessment of thermal loading-induced distortions in optical photomasks due to e-beam multi-pass patterning," *J. Vac. Sci. Technol. B* **16**(6), 3558–3562 (1998).
29. S. W. H. K. Steenbrink, B. J. Kampherbeek, M. J. Wieland, J. H. Chen, S. M. Chang, M. Pas, J. Kretz, C. Hohle, D. van Steenwinckel, S. Manakli, J. Le-Denmat, and L. Pain, "High throughput maskless lithography: low voltage versus high voltage," *Proc. SPIE* **6921**, 69211T (2008).
30. M. Bai, and R. F. Pease, "Resist charging in electron beam lithography," *SRC Techcon* (September, 2000).
31. A. M. Carroll and J. L. Freyer, "Measuring the performance of the AEBLE 150 direct-write e-beam lithography equipment," *Proc. SPIE* **537**, 25–33 (1985).
32. Perkin-Elmer spun off their electron-beam lithography division to Etec, Inc., which was later acquired by Applied Materials. Applied Materials is currently not producing new electron-beam lithography systems.
33. H. C. Pfeiffer, "Advanced e-beam systems for manufacturing," *Proc. SPIE* **1671**, 100–110 (1992).
34. H. C. Pfeiffer, D. E. Davis, W. A. Enichen, M. S. Gordon, T. R. Groves, J. G. Hartley, R. J. Quickle, J. D. Rockrohr, W. Stickel, and E. V. Weber, "EL-4, a new generation electron-beam lithography system," *J. Vac. Sci. Technol. B* **11**(6), 2332–2341 (1993).

35. H. C. Pfeiffer, "Variable spot shaping for electron beam lithography," *J. Vac. Sci. Technol.* **15**(3), 887–890 (1978).

36. M. A. Sturans, J. G. Hartley, H. C. Pfeiffer, R. S. Dhaliwal, T. R. Groves, J. W. Pavick, R. J. Quickle, C. S. Clement, G. J. Dick, W. A. Enichen, M. S. Gordon, R. A. Kendall, C. A. Kostek, D. J. Pinckney, C. F. Robinson, J. D. Rockrohr, J. M. Safran, J. J. Senesi, and E. V. Tressler, "EL5: one tool for advanced x-ray and chrome on glass mask making," *J. Vac. Sci. Technol. B* **16**(6), 3164–3167 (1998).

37. Y. Pain, M. Charpin, Y. Laplanche, D. Herisson, J. Todeschini, R. Palla, A. Beverina, H. Leininger, S. Tourniol, M. Broekaart, E. Luce, F. Judong, K. Brosselin, Y. Le Friec, F. Leverd, S. Del Medico, V. De Jonghe, D. Henry, M. Woo, and F. Arnaud, "Advanced patterning studies using shaped e-beam lithography for 65 nm CMOS pre-production," *Proc. SPIE* **537**, 560–571 (2003).

38. G. H. Jansen, *Interactions in Particle Beams*, Academic Press, Boston (1990).

39. L. R. Harriott, S. D. Berger, J. A. Liddle, G. P. Watson, and M. M. Mkrtchyan, "Space charge effects in projection charged particle lithography systems," *J. Vac. Sci. Technol. B* **13**(6), 2404–2408 (1995).

40. M. M. Mkrtchyan, J. A. Liddle, S. D. Berger, L. R. Harriott, A. M. Schwartz, and J. M. Gibson, "An analytical model of stochastic interaction effects in projection systems using a nearest-neighbor approach," *J. Vac. Sci. Technol. B* **12**(6), 3508–3512 (1994).

41. M. Gesley and P. Condran, "Electron beam blanker optics," *J. Vac. Sci. Technol. B* **8**(6), 1666–1672 (1990).

42. T. H. P. Chang, "Electron beam microcolumns for lithography and related applications," *J. Vac. Sci. Technol. B* **14**(6), 3774–3781 (1996).

43. S. Arai, "Fast electron-beam lithography system with 1024 beams individually controlled by a blanking aperture," *Jpn. J. Appl. Phys.* **32**, 6012–6017 (1995).

44. M. Slodowski, H.-J. Doering, T. Elster, and I. A. Stolberg, "Coulomb blur advantage of a multi shaped beam lithography approach," *Proc. SPIE* **7271**, 72710Q (2009).

45. C. Klein, E. Platzgummer, H. Loeschner, G. Gross, P. Dolezel, M. Tmej, V. Kolarik, W. Klingler, F. Letzkus, J. Butschke, M. Irmscher, M. Witt, and W. Pilz, "Projection mask-less lithography (PML2): proof-of-concept setup and first experimental results," *Proc. SPIE* **6921**, 69211O (2008).

46. C. Klein, E. Platzgummer, J. Klikovits, W. Piller, H. Loeschner, T. Bejdak, P. Dolezal, V. Kolarik, W. Klingler, F. Letzkus, J. Butschke, M. Irmscher, M. Witt, W. Pilz, P. Jaschinsky, F. Thrum, C. Hohle, J. Kretz, J. T. Nogatch, and A. Zepka, "PML2: the maskless multibeam solution for the 22 nm node and beyond," *Proc. SPIE* **7271**, 72710N (2009).

47. M. J. Wieland, G. de Boer, G. F. ten Berge, R. Jager, T. van de Peut, J. J. M. Peijster, E. Slot, S. W. H. K. Steenbrink, T. F. Teepen, A. H. V. van Veen, and B. J. Kampherbeek, "MAPPER: high throughput maskless lithography," *Proc. SPIE* **7271**, 72710O (2009).

48. Y. Nakayama, S. Okazaki, N. Saitou, and H. Wakabayashi, "Electron-beam cell projection lithography: a new high-throughput electron-beam direct-writing technology using a specially tailored Si aperture," *J. Vac. Sci. Technol. B* **8**(6), 1836–1840 (1990).

49. Y. Nakayama, Y. Sohda, N. Saitou, and H. Itoh, "Highly accurate calibration method of electron-beam cell projection lithography," *Proc. SPIE* **1924**, 183–192 (1993).

50. H. Yasuda, K. Sakamoto, A. Yamada, and K. Kawashima, "Electron beam block exposure," *Jpn. J. Appl. Phys.* **30**(11B), 3098–3102 (1991).

51. S. Manakli, H. Komami, M. Takizawa, T. Mitsuhashi, and L. Pain, "Cell projection use in mask-less lithography for 45 nm and 32 nm logic nodes," *Proc. SPIE* **7271**, 72710K (2009).

52. T. Maruyama, Y. Machida, S. Sugatani, H. Tsuchikawa, H. Hoshino, M. Ito, H. Tago, L. L. Chau, S. Lee, and H. Komami, "Design for electron beam: a novel approach to electron beam direct writing throughput enhancement for volume production," *J. Vac. Sci. Technol. B* **27**(6), 2532–2536 (2009).

53. H. C. Pfeiffer and W. Stickel, "PREVAIL—an e-beam stepper with variable axis immersion lenses," *Microelectron. Eng.* **27**, 143–146 (1995).

54. H. C. Pfeiffer, R. S. Dhaliwal, S. D. Golladay, S. K. Doran, M. S. Gordon, T. R. Groves, R. A. Kendall, J. E. Lieberman, P. F. Petric, D. J. Pinckney, R. J. Quickle, C. F. Robinson, J. D. Rockrohr, J. J. Senesi, W. Stickel, E. V. Tressler, A. Tanimoto, T. Yamaguchi, K. Okamoto, K. Suzuki, T. Okino, S. Kawata, K. Morita, S. C. Suziki, H. Shimizu, S. Kojima, G. Varnell, W. T. Novak, D. P. Stumbo, and M. Sogard, "Projection reduction exposure with variable axis immersion lenses: next generation lithography," *J. Vac. Sci. Technol. B* **17**(6), 2840–2846 (1999).

55. W. Stickel and G. O. Langner, "PREVAIL: theory of the proof of concept column electron optics," *J. Vac. Sci. Technol. B* **17**(6), 2847–2850 (1999).

56. M. S. Gordon, J. E. Lieberman, P. F. Petric, C. F. Robinson, and W. Stickel, "PREVAIL: operation of the electron optics proof-of-concept system," *J. Vac. Sci. Technol. B* **17**(6), 2851–2855 (1999).

57. O. R. Wood II, W. J. Trybula, M. J. Lercel, C. W. Thiel, M. J. Lawliss, K. Edinger, A. Stanishevsky, S. Shimizu, and S. Kawata, "Benchmarking stencil reticles for electron projection lithography," *J. Vac. Sci. Technol. B* **21**(6), 3072–3077 (2003).

58. K. Uchikawa, S. Takahashi, N. Kagtakura, T. Oshino, S. Kawata, and T. Yamaguchi, "Pattern displacement measurements for Si stencil reticles," *J. Vac. Sci. Technol. B* **17**(6), 2868–2872 (1999).

59. S. D. Berger and J. M. Gibson, "New approach to projection-electron lithography with demonstrated 0.1 micron linewidth," *Appl. Phys. Lett.* **57**, 153–155 (1990).

60. I. Amemiya, H. Yamashita, S. Nakatsuka, M. Tsukahara, and O. Nagarekawa, "Fabrication of a continuous diamondlike carbon membrane mask for electron projection lithography," *J. Vac. Sci. Technol. B* **21**(6), 3032–3036 (2003).

61. W. Waskiewicz, C. J. Biddle, M. I. Blakey, K. J. Brady, R. M. Camarda, W. F. Connelly, A. H. Crorken, J. P. Custy, R. Demarco, R. C. Farrow, J. A. Felker, L. A. Fetter, R. Freeman, L. R. Harriott, L. C. Hopkins, H. A. Huggins, R. J. Kasica, C. S. Knurek, J. S. Kraus, J. A. Liddle, M. Mkrtchyan, A. E. Novembre, M. L. Peabody, L. Ruberg, H. H. Wade, G. P. Watson, K. S. Wender, D. L. Windt, R. G. Tarascon-Auriol, S. D. Berger, and S. W. Bowler, "SCALPEL proof-of-concept system: preliminary lithography results," *Proc. SPIE* **3048**, 255–263 (1997).

62. C. J. Martin, W. H. Semke, G. A. Dicks, R. L. Engelstad, E. G. Lovell, J. A. Liddle, and A. E. Novembre, "Mechanical and thermal modeling of the SCALPEL mask," *J. Vac. Sci. Technol. B* **17**(6), 2878–2882 (1999).

63. B. Kim, R. L. Engelstad, E. G. Lovell, S. T. Stanton, J. A. Liddle, and G. M. Gallatin, "Finite element analysis of SCALPEL wafer heating," *J. Vac. Sci. Technol. B* **17**(6), 2883–2887 (1999).

64. K. Okamoto, K. Suzuki, H. C. Pfeiffer, and M. Sogard, "High throughput e-beam stepper lithography," *Solid State Technol.* **43**(5), 118–122 (May, 2000).

65. G. Stengel, H. Loschner, W. Maurer, and P. Wolf, "Current status of ion-projection lithography," *Proc. SPIE* **537**, 138–145 (1985).

66. This figure was provided by Dr. Hans Löschner of IMS–Ionen Mikrofabrikations Systeme GmbH in Vienna.

67. Y. Lee, R. A. Gough, K. N. Leung, J. Vujic, M. D. Williams, and N. Zahir, "Plasma source for ion and electron beam lithography," *J. Vac. Sci. Technol. B* **16**(6), 3367–3369 (1998).

68. K. Leung, "Plasma sources for electron and ion beams," *J. Vac. Sci. Technol. B* **17**(6), 2776–2778 (1999).

69. A. Ehrmann, A. Elsner, R. Liebe, T. Struck, J. Butschke, F. Letzkus, M. Irmscher, R. Springer, E. Haygeneder, and H. Löschner, "Stencil mask key parameter measurement and control," *Proc. SPIE* **3997**, 373–384 (2000).

70. P. Hudek, P. Hrkut, M. Drzik, I. Kostic, M. Belov, J. Torres, J. Wasson, J. C. Wolfe, A. Degen, I. W. Rangelow, J. Voigt, J. Butschke, F. Letzkus, R. Springer, A. Ehrmann, R. Kaesmaier, K. Kragler, J. Mathuni, E. Haugeneder, and H. Löschner, "Directly sputtered stress-compensated carbon protective layer silicon stencil masks," *J. Vac. Sci. Technol. B* **17**(6), 3127–3131 (1999).

71. D. J. Resnick, W. J. Dauksher, D. Mancini, K. J. Nordquist, T. C. Bailey, S. Johnson, N. Stacey, J. G. Ekerdt, C. G. WIllson, S. V. Sreenivasan, and N. Schumaker, "Imprint lithography: lab curiosity or the real NGL?" *Proc. SPIE* **5037**, 12–23 (2003).

72. M. Colburn, S. Johnson, M. Stewart, S. Damle, T. Bailey, B. Choi, M. Wedlake, T. Michaelson, S. V. Sreenivasan, J. Ekerdt, and C. G. Willson, "Step and flash imprint lithography: a new approach to high-resolution printing," *Proc. SPIE* **3676**, 379–389 (1999).

73. S. Y. Chou, P. R. Krauss, and P. J. Renstrom, "Nanoimprint lithography," *J. Vac. Sci. Technol. B* **14**(6), 4129–4133 (1996).

74. D. J. Resnick, W. J. Dauksher, D. Mancini, K. J. Nordquist, E. Ainely, K. Gehoski, J. H. Baker, T. C. Bailey, B. J. Choi, S. Johnson, S. V. Sreenivsasan, J. G. Ekerdt, and C. G. Willson, "High resolution templates for step and flash imprint lithography," *Proc. SPIE* **4688**, 205–213 (2002).

75. D. J. Resnick, W. J. Dauksher, D. Mancini, K. J. Nordquist, T. C. Bailey, S. Johnson, N. Stacey, J. G. Ekerdt, C. G. Willson, S. V. Sreenivsasan, and N. Schumaker, "Imprint lithography for integrated circuit fabrication," *J. Vac. Sci. Technol. B* **21**(6), 2624–2631 (2003).

76. L. J. Guo, "Nanoimprint lithography: methods and materials requirements," *Adv. Mater.* **19**, 495–512 (2007).

77. B. D. Terris, "Fabrication challenges for patterned recording media," *J. Magn. Magnetic Mater.* **321**(6), 512–517 (2009).

78. E. W. Edwards, M. P. Stoykovich, P. F. Nealey, and H. H. Solak, "Binary blends of diblock copolymers as an effective route to multiple length scales in perfect directed self-assembly of diblock copolymer thin films," *J. Vac. Sci. Technol. B* **24**(1), 340–344 (2006).

79. W. Edwards, M. P. Stoykovich, M. Müller, H. H. Solak, J. J. de Pablo, and P. F. Nealey, "Mechanism and kinetics of ordering in diblock copolymer thin films on chemically nanopatterned substrates," *J. Polymer. Sci. Polymer. Phys.* **43**(23), 3444 (2005).

80. This figure was provided by Prof. Paul Nealey of the University of Wisconsin.

81. R. Ruiz, H. Kang, F. A. Detcheverry, E. Dobisz, D. S. Kercher, T. R. Albrecht, J. J. de Pablo, and P. F. Nealey, "Density multiplication and improved lithography by directed block copolymer assembly," *Science* **321**, 936–939 (2008).

Appendix A
Coherence

Coherence refers to the degree that light waves are correlated. Coherence has been the subject of entire books on optics, or at least lengthy chapters. A few essential elements are summarized in this Appendix. There are two types of coherence—temporal coherence and spatial coherence, both of which are relevant to lithography.

Temporal coherence refers to the correlation of light in time.[1] There are several situations in lithography where temporal coherence is relevant. Consider the situation depicted in Fig. A.1. A beam of light is incident on a resist-coated silicon surface. Consider, for the moment, very transparent resist. The instantaneous amplitude of the incident wave at $\vec{x}$ is given by $A(t)$. The amplitude of the reflected light is given by the amplitude of the light after the time it has propagated to the interface and reflected:

$$\rho A\left(t + \frac{2hn}{c}\right), \tag{A.1}$$

where ρ is the reflectivity of the resist-silicon interface, c is the speed of light, and n is the index of refraction of the resist.

The time-averaged intensity at point $\vec{x}$ is given by:

$$I(\vec{x}) = \left\langle \left| A(t) + \rho \times A\left(t + \frac{2h}{c}\right) \right|^2 \right\rangle, \tag{A.2}$$

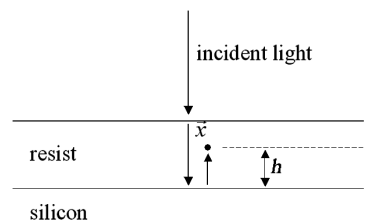


Figure A.1 The geometry of light in a resist film.

where $\langle \cdots \rangle$ indicates time averaging and all quantities are evaluated at the point $\vec{x}$.

$$I(x) = \left\langle |A(t)|^2 \right\rangle + \left|\rho\right|^2 \left\langle \left| A\left(t + \frac{2hn}{c}\right)\right|^2 \right\rangle + \left\langle A(t)\rho * A * \left(t + \frac{2hn}{c}\right)\right\rangle + \left\langle A * (t)\rho A\left(t + \frac{2hn}{c}\right)\right\rangle \tag{A.3}$$

$$= \left(1 + \left|\rho\right|^2\right) I_0 + 2Re\left[\rho\Gamma\left(\frac{2hn}{c}\right)\right], \tag{A.4}$$

where

$$I_0 = \left\langle |A(t)|^2 \right\rangle \tag{A.5}$$

and

$$\Gamma(\tau) = \langle A(t+\tau)A * (t)\rangle . \tag{A.6}$$

Note that $I_0 = \Gamma(0)$. From the Schwartz inequality, it follows that

$$|\Gamma(\tau)| \leq I_0. \tag{A.7}$$

That is, the coherence can only degrade in time. The light intensity in the resist will depend upon the quantity Γ, that is, upon the coherence properties of the light.

For completely incoherent light, $\Gamma = 0$, in which case the total intensity is the sum of the intensity of the incident and reflected light. For completely incoherent light, light intensity is independent of the relative phases of the incident and reflected waves. An example of completely coherent light is a plane wave:

$$A = A_0 e^{i(kx+\omega t)}. \tag{A.8}$$

For a plane wave,

$$|\Gamma(\tau)| = 1 \tag{A.9}$$

for all values of τ. In this case, the light intensity at any point in the resist film depends significantly on the relative phases between incident and reflected waves. It has been shown that propagating light maintains a high degree of coherence over a time, referred to as the coherence time, which is inversely proportional to the bandwidth of the light. For light with a Lorentzian distribution, such as that produced by a high-pressure mercury-arc lamp, the coherence time is given by[1]

$$\tau_c = \frac{0.318}{\Delta\nu}, \tag{A.10}$$

where $\Delta\nu$ is the full-width half-maximum of the frequency spectral distribution. It can be shown that the light used in lithography has high temporal coherence (see Problem A.1).

Another view of coherence—spatial coherence—relates to the interference of light originating from different points in space. Consider the situation shown in Fig. A.2. Light from a source passes through different parts of a mask before being imaged on the wafer. Just as in Eq. (A.4), there will be three terms that give the light intensity, the intensity from the individual beams, and an interference term.

Excimer lasers are suitable sources of light for lithography because they have low spatial coherence, a consequence of their intrinsic multimode character.[2,3] Light with high spatial coherence will tend to produce the phenomenon of speckle,[1] where there is light interference over short-distance surface roughness and particles, resulting in light nonuniformity.[4,5] Such nonuniformity would significantly reduce lithography process control and is therefore undesirable. Fortunately, excimer lasers have low intrinsic spatial coherence, as lasers go, nearly approximating that of arc lamps.[3]

Spatial coherence depends upon the details of the illumination optics, as well as the light source. The spatial coherence of illumination in lithography can be modulated by the geometry of the illumination and projection optics. Consider the optical configuration shown in Fig. A.3. The (spatial) coherence of the light is characterized by the ratio

$$\sigma = \frac{\sin\theta_i}{\sin\theta_0}, \tag{A.11}$$

usually referred to as the *partial coherence*.

The degree of partial coherence is a measure of how much of the entrance pupil of the optical system is filled. For perfectly coherent light, $\sigma = 0$, while $\sigma = \infty$ for completely incoherent light. The light used in wafer steppers is partially coherent, with typical value of σ between 0.3 and 0.8. The simplest way to reduce σ is to insert an aperture in the illuminator to reduce the θ_c. However, this has the effect of reducing light intensity as well. Illuminators have been designed that redirect the light into smaller cones in order to achieve low values of σ, thereby achieving

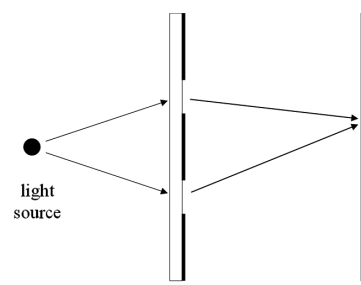


Figure A.2 Situation in which spatial coherence plays a role.

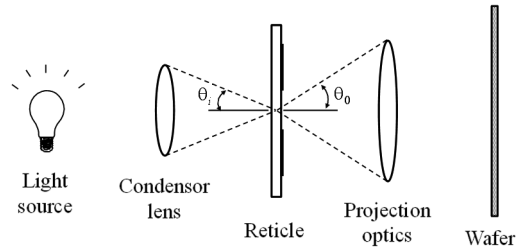


Figure A.3 Illumination geometry.

the desired degree of coherence without sacrificing light intensity, and therefore throughput and productivity.

Problems

A.1 From the values for spectral width given in Chapter 5, and using Eq. (A.10), does the light remain highly coherent throughout the depths of resist films used in semiconductor lithography (0.2–2.0 μm thick)? For resist films used to fabricate thin-film heads for magnetic recording (10–20 μm thick)? For the resist films used in micromachining (20–200 μm thick)?

References

1. J. W. Goodman, *Statistical Optics*, John Wiley and Sons, New York (1985).
2. J. Hecht, *The Laser Guidebook*, 2nd ed., McGraw Hill, New York (1999).
3. K. Jain, *Excimer Laser Lithography*, SPIE Press, Bellingham, Washington (1990).
4. Y. Ozaki, K. Takamoto, and A. Yoshikawa, "Effect of temporal and spatial coherence of light source on patterning characteristics in KrF excimer laser lithography," *Proc. SPIE* **922**, 444–448 (1988).
5. T. E. Jewell, J. H. Bennewitz, G. C. Escher, and V. Pol, "Effect of laser characteristics on the performance of a deep UV projection system," *Proc. SPIE* **774**, 124–132 (1987).

Index

Harry J. Levinson is a Senior Fellow and manager of GLOBALFOUNDRIES's Strategic Lithography Technology Department, which is responsible for advanced lithographic processes and equipment. He started his career in bipolar memory development at AMD, then spent some time at Sierra Semiconductor and IBM, before returning to AMD—now GLOBALFOUNDRIES—in 1994. During the course of his career, Dr. Levinson has applied lithography to many different technologies, including bipolar memories, 64Mb and 256Mb DRAM development, the manufacturing of applications-specific integrated circuits, thin-film heads for magnetic recording, flash memories, and advanced logic. He was one of the first users of 5× steppers in Silicon Valley and was an early participant in 248-nm and 193-nm lithography. He also served for several years as the chairman of the USA Lithography Technology Working Group that participates in the generation of the lithography chapter of the International Technology Roadmap for Semiconductors. He has published numerous articles on lithographic science, on topics ranging from thin-film optical effects and metrics for imaging, to overlay and process control, and he is the author of two books, *Lithography Process Control* and *Principles of Lithography*. He holds over 40 U.S. patents. He is an SPIE Fellow and formerly chaired the SPIE Publications Committee. He has a BS in engineering from Cornell University and a PhD in Physics from the University of Pennsylvania.

COLOR PLATES

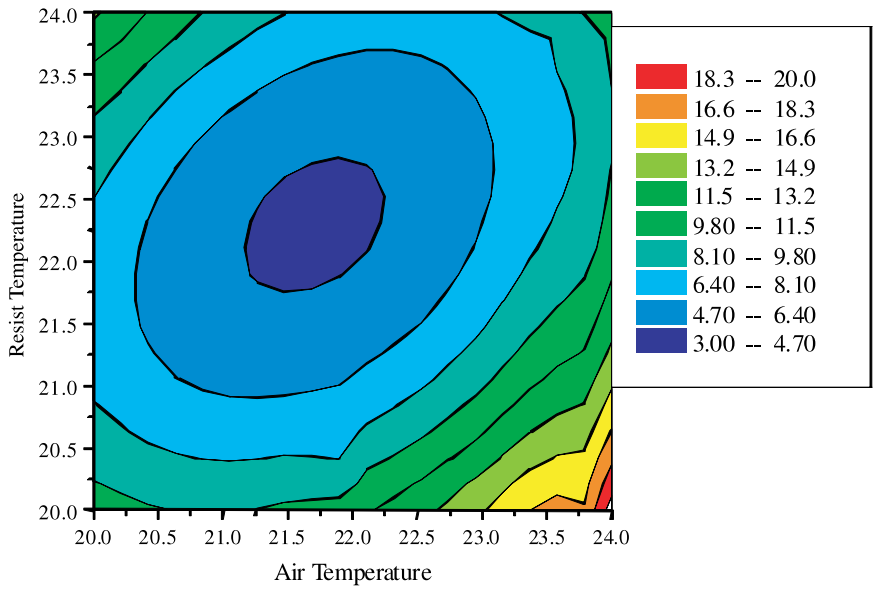


Figure 3.12 Resist uniformity contours as a function of air and resist temperatures for Shin-Etsu 430S resist coated on an SVG ProCell. The data are single-standard deviations, in units of Angstroms.[21] The initial wafer temperature was 22 °C. The most uniform coatings are produced with resist and air temperatures slightly different from the initial wafer temperature (see p. 60).

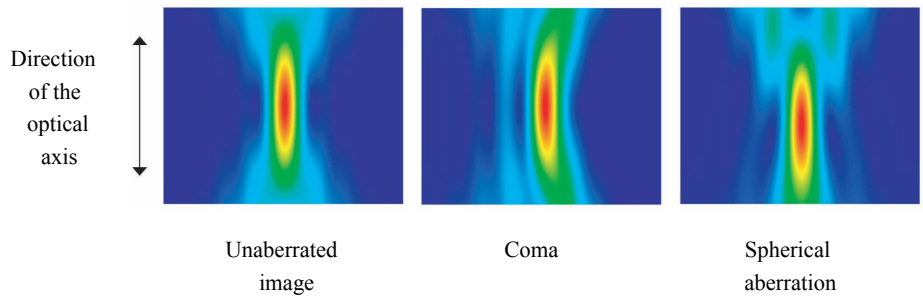


Figure 4.7 Aerial image-intensity contours for a 0.4-μm space on the mask.[10] For an unaberrated lens, the intensity contours have left-right symmetry and are also symmetric across the plane of best focus. Images produced by lenses with coma (Z_7) lose the left-right symmetry, while spherical aberration (Z_9) breaks the symmetry across the plane of best focus. The pictures in this figure were simulated with Solid-C for an i-line tool with $NA = 0.6$ and $\sigma = 0.5$. For the aberrated images, 50 nm were assumed for each aberration (see p. 118).

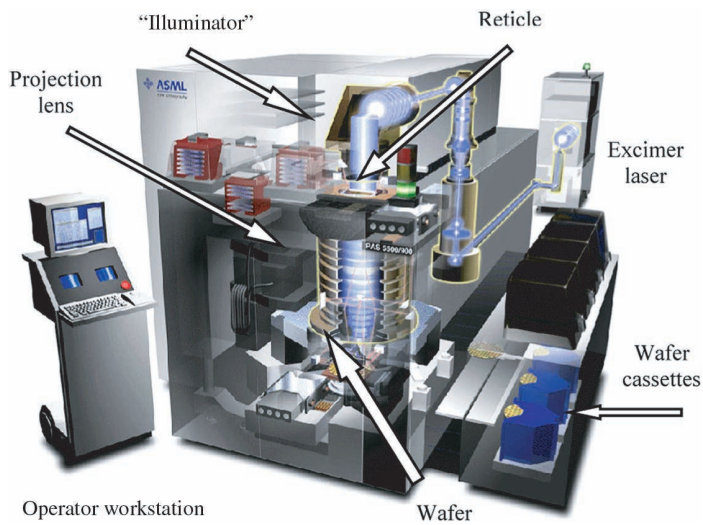


Figure 5.2 ASML PAS5500 wafer stepper (see p. 148).

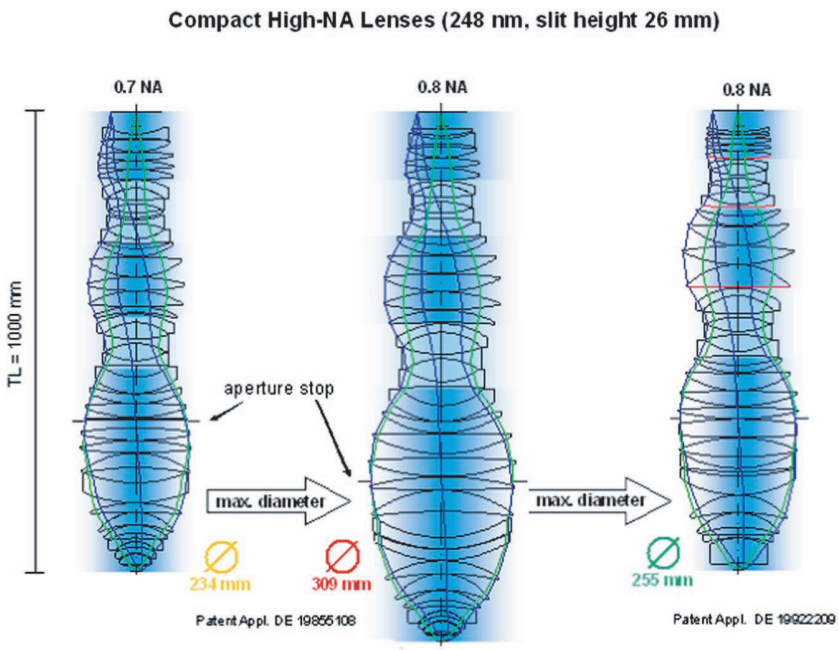


Figure 5.32 The benefit of aspherical lens elements (see p. 185).

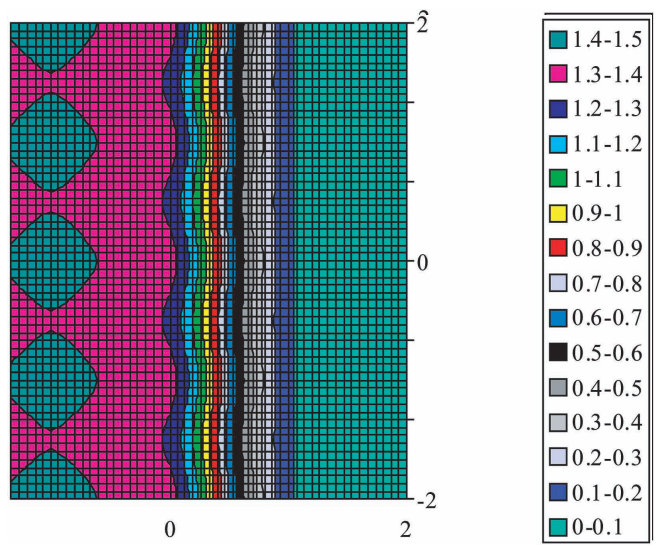


Figure 7.10 Exposure dose contours from Gaussians placed at integer coordinates (x, y), with $x \leq 0$ (see p. 270).

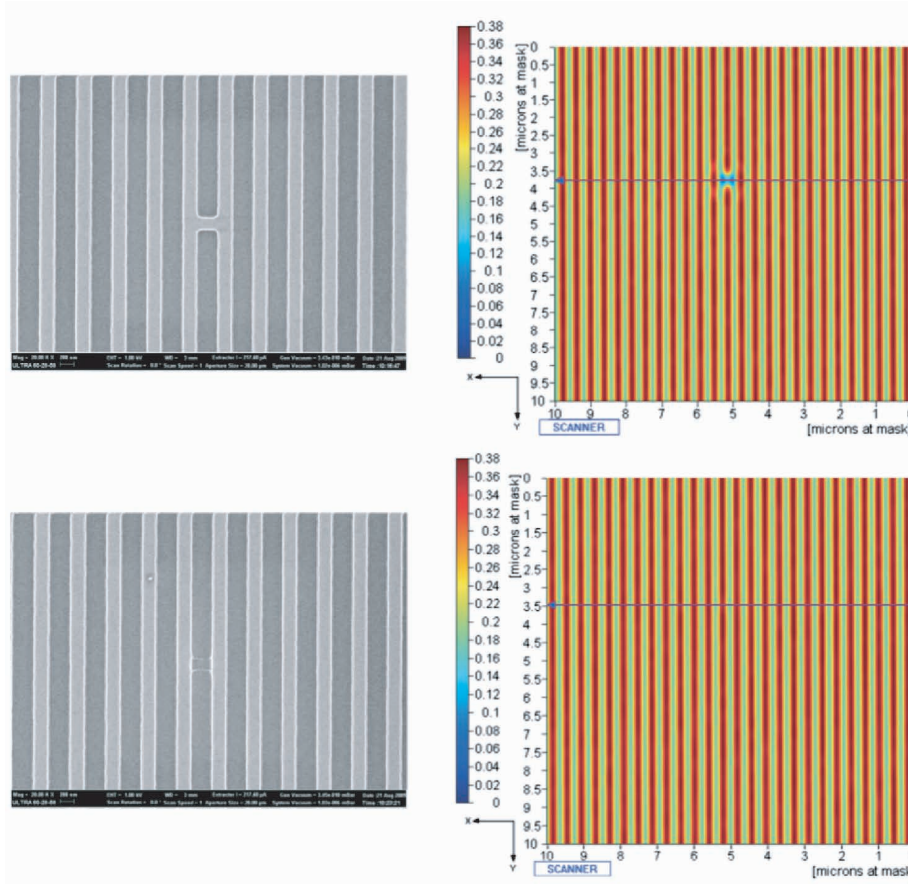


Figure 7.32 (a) A scanning-electron micrograph of a bridging defect on a mask, and (b) the measured aerial image from the mask. (c) The corresponding results after the mask was repaired (see p. 292).

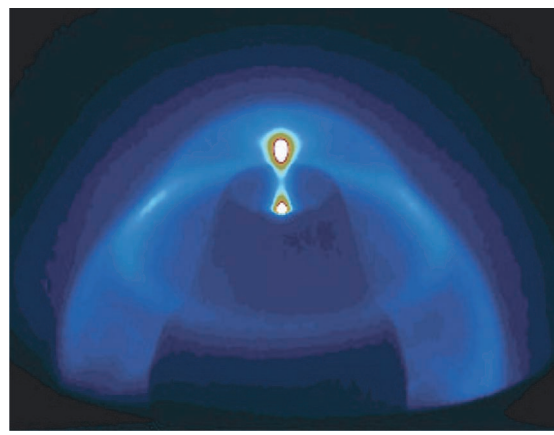


Figure 12.16 An EUV discharge source in operation. (Reproduced courtesy of Cymer, Inc.) (see p. 438).