On the design of lithographic interferometers and their application

By

Michael E. Walsh

B.A., Physics, Vassar College (1997)
B.E., Engineering, Dartmouth College (1998)
M.S., Electrical Engineering, Massachusetts Institute of Technology (2000)

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

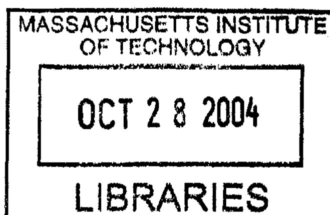
at the
Massachusetts Institute of Technology, September 2004

© 2004 Massachusetts Institute of Technology. All Rights Reserved

Signature of Author ....
Department of Electrical Engineering and Computer Science
September 3, 2004

Certified by ....
Henry I. Smith
Keithley Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Committee on Graduate Students
Department of Electrical Engineering and Computer Science


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 2 8 2004
LIBRARIES

BARKER

On the design of lithographic interferometers and their application

by

Michael E. Walsh

Submitted to the Department of Electrical Engineering and Computer Science on September 3, 2004 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering

ABSTRACT

Interference lithography is presented as an ideal technique for fabricating large-area periodic structures with sub-100nm dimensions. A variety of interferometer designs are discussed and implemented, each of which emphasizes a different attribute. Curvature of the substrate during exposures in a Mach-Zender interferometer is demonstrated as a method for reducing periodicity variations in the printed pattern down to the level of picometers. A robust Lloyd's-mirror interferometer is developed as a simple and flexible lithography system capable of sub-100nm features. Prospects and designs for the use of grating interferometers for the fabrication of structures with 50nm spatial-periodicity are discussed. A novel, integrated, thin-film interferometer exploiting solid immersion is introduced for the fabrication of features below the diffraction-limit. Using 193nm illumination, 45nm dense features are demonstrated. Aspects of sub-wavelength diffraction and thin-film resonance are discussed in relation to the fabrication of structures with sub-50nm periodicity. A selection of applications are discussed which take advantage of the unique capabilities of interference lithography, including patterned magnetic media, DFB lasers and the templated self-assembly of nano-particles. A novel technique for ion-beam etching, applicable to any material system, is shown to improve the patterning of nanoscale magnetic elements.

Thesis Supervisor: Henry I. Smith
Title: Keithley Professor of Electrical engineering and Computer Science

Acknowledgements

This page has perhaps been the most difficult one in the whole document for me to write. I simply cannot thank all the people for all the help that has allowed me to complete this long labor in such a concise form. For those whose remain unnamed, I apologize and am truly grateful.

Of course, I would not be here without my wonderful parents, whose continued support and love have been unwavering. Their guidance and influence have given me the character to stay the course over these six long years at MIT. I don't know to thank my beautiful and patient wife Laura Green, who has too often had to share my attentions with the world of nanotechnology. You are a joy.

It is a bit surprising to look back and see that the classroom has been more of a diversion from my education here than the focus of it. Work in the laboratory and related discussions have been where I learned the most important lessons, and I have been privileged have so many excellent mentors to teach the art and craft of science. My advisor, Professor Henry Smith, is continually inspiring with his passion and energy. His confidence in me and my crazy ideas has given me confidence in myself, and allowed me to pursue things that have seemed impossible at times. Dr. Mark Schattenburg, and Professors Caroline Ross and George Barbasthasis, have been invaluable resources. The staff of the NanoStructures Lab, Mark Mondol, Jim Carter, and Jim Daley, have been a source of technical knowledge and assistance without whom my research would have been impossible. More than simply colleagues, the students I have worked alongside have also been the people from whom I have learned the most, especially Drs. Juan Ferrera, Maya Farhoud, Mike Lim, Tim Savas, Dario Gil, Todd Hastings and Rajesh Menon and Euclid Moon. All of the NSL folks, too numerous to list individually, have been a pleasure to work and learn with, and good friends.

My sincere thanks to everyone.

Contents

CHAPTER 1 INTRODUCTION **13**

1.1 LITHOGRAPHY 13

1.1.1 History *13*

1.1.2 The lithographic process *16*

1.1.3 Interference lithography *18*

1.2 THE SPATIAL THEORY OF IMAGING 20

1.3 THE ROLE OF IL IN NANOLITHOGRAPHY & NANOTECHNOLOGY 21

CHAPTER 2 THE INTERFERENCE OF LIGHT **27**

2.1 INTRODUCTION 27

2.2 FUNDAMENTALS OF INTERFERENCE 28

2.2.1 Maxwell's Equations *28*

2.2.2 Phasor Maxwell Equations *29*

2.2.3 Superposition of two plane-waves *30*

2.3 PERIODICITY OF INTERFERENCE FRINGES 32

2.3.1 The symmetric angle of incidence *32*

2.3.2 Observed fringe periodicity *34*

2.4 VISIBILITY OF INTERFERENCE FRINGES 35

2.4.1 Intensity of the interfering waves *35*

2.4.2 Polarization *38*

2.5 PARTIAL COHERENCE 40

2.5.1 Coherence and spectral bandwidth *40*

2.5.2 The mutual coherence function *41*

2.5.3 Fringe visibility as a function of coherence *42*

2.6 REFERENCES 45

CHAPTER 3 THE MACH-ZEHNDER INTERFEROMETER **47**

3.1 INTRODUCTION 47

3.2 ALIGNMENT TECHNIQUES 49

3.2.1 Reference Plane *49*

3.2.2 Optics positioning via a reference grating. *51*

3.3 SPATIAL PHASE 53

3.3.1 Hyperbolic Phase *53*

3.3.2 Fringe inclination *54*
3.3.3 The importance of assumptions... .... *59*
3.4 MOIRÉ TECHNIQUES FOR MEASURING SPATIAL PHASE 63
3.5 CORRECTION OF HYPERBOLIC SPATIAL-PHASE 66
3.5.1 Plane-waves with lenses *66*
3.5.2 Substrate Curvature *67*
3.6 REDUCING NON-LINEAR SPATIAL-PHASE VIA SUBSTRATE CURVATURE 68
3.6.1 Introduction *68*
3.6.2 Implementing substrate curvature *70*
3.6.3 Bending stresses in single-crystal silicon wafers *71*
3.6.4 Simplification via Symmetry *74*
3.6.5 The spherical surface *76*
3.7 DEMONSTRATION OF SPHERICAL-SURFACE CORRECTION 80
3.7.1 Procedure *80*
3.7.2 Alignment *80*
3.7.3 Zero relative shift *81*
3.7.4 X-direction shift *84*
3.7.5 Y-direction shift *85*
3.8 SURFACE OPTIMIZATION 86
3.8.1 The spherical surface *86*
3.8.2 Polynomial surface description *87*
3.8.3 Optimization via the simplex method *88*
3.8.4 Hyperbolic surfaces *90*
3.8.5 Local Strain Correction *92*
3.9 CONCLUSIONS 96
3.10 REFERENCES 97

CHAPTER 4 THE LLOYD'S-MIRROR INTERFEROMETER 101

4.1 INTRODUCTION 101
4.2 A LLOYD'S-MIRROR INTERFEROMETER FOR LITHOGRAPHY 103
4.3 ALIGNMENT AND FRINGE PERIOD 107
4.3.2 . Angular Alignment of Mirror *110*
4.3.3 Alignment of rotational axis *111*
4.3.4 Practical Alignment and Calibration *113*
4.4 STABILITY 115
4.4.1 Vibrational Stability *115*
4.4.2 Real-time stability monitoring via moiré patterns *118*

4.5 THE MIRROR 120

4.5.1 Phase Distortion *120*

4.5.2 Reflectivity *122*

4.5.3 Mirror Size *123*

4.6 COHERENCE AND CONTRAST 124

4.7 SUMMARY 128

4.8 REFERENCES 130

CHAPTER 5 GRATING INTERFEROMETERS 131

5.1 INTRODUCTION 131

5.1.1 History *131*

5.2 SINGLE-GRATING INTERFEROMETERS 133

5.2.1 Near-field spatial frequency doubling *133*

5.2.2 Zero-Order Diffraction Efficiency *136*

5.2.3 Spatial coherence *137*

5.3 MULTI-GRATING INTERFEROMETERS 140

5.3.1 Two-Grating Interferometer *140*

5.3.2 MIT Achromatic Interference Lithography *142*

5.4 NEXT GENERATION TOOLS 143

5.4.1 Introduction *143*

5.4.2 Short-wavelength photon sources *144*

5.4.3 Material Properties in the VUV *145*

5.4.4 Soft x-ray AIL using membrane parent gratings *148*

5.4.5 Reflection gratings *149*

5.4.6 Liquid Immersion *151*

5.4.7 A novel interferometer designed for immersion *155*

5.5 REFERENCE 159

CHAPTER 6 THE INTEGRATED THIN FILM INTERFEROMETER 165

6.1 INTRODUCTION 165

6.1.1 Concept *165*

6.1.2 Implementation *166*

6.1.3 Fabrication of parent-gratings *167*

6.2 PRACTICAL ADVANTAGES OF THE ITFI 170

6.2.1 Wavelength reduction via solid immersion *170*

6.2.2 Extension of current lithography *172*

6.2.3 Mechanical Stability and Alignment *173*

6.2.4 Source Requirements: Spatial Coherence ... *174*
6.2.5 Source requirements: output power ... *178*
6.3 DIFFRACTION EFFICIENCY ... 180
6.3.1 Introduction ... *180*
6.3.2 Scalar Diffraction Theory ... *180*
6.3.3 Numerical Modeling ... *182*
6.3.4 Finite-difference time-domain simulations (FDTD) ... *187*
6.3.5 The Talbot effect ... *188*
6.4 THIN-FILM STACK DESIGN ... 192
6.4.1 A photoresist for the parent-grating ... *192*
6.4.2 Anti-reflection properties of the thin-film stack ... *195*
6.4.3 Oxide Gapping Layer ... *199*
6.5 EXPERIMENTAL RESULTS ... 200
6.5.1 Design Review ... *200*
6.5.2 360 nm interlayer thickness ... *200*
6.5.3 280nm interlayer thickness ... *205*
6.6 FUTURE DEVELOPMENTS ... 209
6.6.1 Resonant Diffraction ... *209*
6.6.2 A diffractive resonator model ... *210*
6.6.3 High-Index tooth and trench materials ... *212*
6.6.4 Imprinted parent-gratings ... *213*
6.6.5 157nm and 248nm ITFI exposures ... *215*
6.7 CONCLUSION ... 216
6.8 REFERENCES ... 218

CHAPTER 7 APPLICATIONS ... 223

7.1 RESIST PROCESSES FOR INTERFERENCE LITHOGRAPHY ... 223
7.1.1 Higher-order periodicity ... *223*
7.1.2 Anti-Reflection coatings ... *226*
7.1.3 The tri-layer resist process ... *228*
7.1.4 Ultrathin Resist ... *232*
7.1.5 Sensitometry ... *234*
7.2 PATTERNED MAGNETIC MEDIA ... 242
7.2.1 Introduction ... *242*
7.2.2 Fabrication patterned magnetic media using IL ... *244*
7.2.3 Switching behavior of patterned magnetic media ... *245*
7.2.4 Hysterisis loops ... *247*

7.3 TEMPLATED SELF-ASSEMBLY 251
7.3.1 Introduction *251*
7.3.2 Directed assembly of SiGe quantum-dots *252*
7.4 SEMICONDUCTOR NANOCRYSTAL LASERS 258
7.4.1 Introduction *258*
7.5 REFERENCES 262

CHAPTER 8 ETCHING MAGNETIC THIN FILMS 271

8.1 INTRODUCTION 271
8.1.1 Reactive-ion etching *271*
8.2 ION-BEAM ETCHING 273
8.2.1 Etch Selectivity *273*
8.2.2 Faceting *275*
8.2.3 Redeposition *277*
8.3 MASK DESIGN FOR ION MILLING 279
8.3.1 Introduction *279*
8.3.2 Thick polymeric masks *279*
8.3.3 Influence of atomic weight on sputtering yields *281*
8.3.4 Ion milling using Ne^+ and thin tungsten masks *283*
8.3.5 Ion-milling using Xe^+ and photoresist masks *288*
8.3.6 A theory of low-energy sputtering *290*
8.4 REFERENCES 293

APPENDIX A TABULATION OF OPTICAL CONSTANTS 297

Chapter 1
Introduction

1.1 Lithography

1.1.1 History

The modern tools for microlithography are truly wonders of technology, the product of a long history of advances that were, no doubt, equally dazzling in their own day [1]. The engraving of images by hand into materials such as stone, wood, or bone dates back tens of thousands of years. The first transfer of images from one medium to another is credited to the Sumerians roughly 3000 years ago, who engraved images in stone seals that would then be pressed into clay. The Chinese are thought to have been producing prints by rubbing close to 2000 years ago, although the earliest authenticated prints are Japanese from circa 800 A.D. Printing did not arrive in Europe until the 15th century, making its debut in Germany in the form of playing cards printed from woodcuts. Johannes Gutenburg

introduced his famous movable-type printing press in 1440. The use of metal plates for printing soon followed, with the first known example dated 1446. Acid etching of metal printing plates dates to the 16th century, with the work of German artist Albrecht Dürer being among the most famous examples, both then and now [2]. In fact, the first known use of chemical-etching to transfer a pattern was close to 2000 years ago by the Hohokam people in Arizona, although this technology was never disseminated [3].

The art of printmaking was fairly well established by the end of the 18th century, when Alois Senenfelder invented lithography as a way to reduce the high cost of producing and reproducing copperplate engravings. Senenfelder's process involved treating a porous limestone block with a mixture of wax, soap, lampblack, and water. Areas treated with this mixture would repel ink, while the limestone itself absorbed it.

Less than 3 decades later, Frenchman Nicephore Niépce is credited with creating the first permanent photograph, although not entirely out of the blue [4]. Camera obscura - the projection of an image through a small hole into a darkened viewing area - had been known since ancient times. In the 1500's, the camera obscura was greatly improved though the addition of telescope lenses. Similarly, photosensitive chemicals were known as a curiosity for hundreds of years before Niépce's time. Englishman Robert Boyle reported in the 17th century that silver nitrate darkened upon exposure to sunlight. Twenty years or so prior to Niépce's success, Thomas Wedgewood was capturing fleeting images on leather soaked in silver nitrate, although he was unable to stop the exposure and prevent them from completely blackening over time. Niépce himself had been struggling with this problem for more than 10 years before finally solving it in 1826 using bitumen of Judea, a resinous substance which hardened on exposure to light. Niépce subsequently partnered with Louis Daguerre to develop the amazingly high quality photography process known as the daguerreotype, of which many examples still survive.

Although Niépce himself first demonstrated photolithography by etching his images into a metal backingplate, the invention of photolithography is commonly credited to Alphonse Louis Poitevin in 1855. Poitevin developed a method for patterning his limestone lithography blocks photographically [5]. Interestingly, he preferred to call it "chemical printing". Soon after, metal replaced stone as the primary plate material, and spin-coating was introduced as a means of applying a uniform layer of photoresist. Over the next 100 years, a continual string of innovations brought the lithography industry to an advanced state.

The turn towards microlithography began immediately after the invention of the transistor in 1947 [6], however the invention of the integrated circuit by Jack Kilby in 1958 is what truly began the modern semiconductor era [7]. Soon afterwards, Kodak introduced KTFR, the first photoresist designed specifically for microlithography, further enabling the microlithography and semiconductor industries [8]. Up through the 1970's, photomasks were transferred 1:1 (without demagnification) using contact and proximity lithography, ultimately achieving feature sizes of a few microns. Scanning optical projection systems, also 1:1, were first introduced in 1974 by the Perkin Elmer corporation, and quickly replaced proximity lithography as the dominant mode. Reduction step-and-repeat systems were introduced around 1980, although the 1:1 scanning technology continued to be popular throughout the 1980's. Modern semiconductor lithography is generally done using a combination of stepping and scanning, usually with 4:1 reduction.

Parallel to the architectural changes in the optical systems, the wavelength of light used for lithographic exposures has been reduced continually to increase resolution. Beginning at 436nm, the g-line of mercury, high-pressure arc lamps were used down through the 365 nm i-line. Excimer sources such as the 248nm KrF laser were introduced in the early 1980's [9], The most modern tools now use 193 nm ArF lasers. Illumination in the vacuum ultraviolet at 157 nm and the so-called extreme ultraviolet (EUV) illumination at 13 nm have been proposed as possibilities for the next generation of commercial lithography tools.

1.1.2 The lithographic process

The lithographic process can be divided into a few distinct steps. Substrates are chosen and prepared by cleaning and thin-film deposition. These films can be layers that will eventually be part of a functional device, such as metals or semiconductors, or they can be sacrificial layers that enable the process, such as an etch-mask. Next, a pattern is generated on this substrate in a material known as photoresist, shown schematically in Figure 1.1. Generally a polymeric film, photoresists are defined by a change of the dissolution rate in some solvent upon exposure to radiation. The dissolution rate of positive-tone photoresists increases upon exposure, while the dissolution rate of negative-tone resists decreases. Thus, the pattern of the radiation exposure is transformed into a physical structure in the photoresist by the action of the solvent, also called the developer. Although photon exposures are the most common, electrons and ions are also commonly used. Neutral-atom lithography has been demonstrated, but is not widely practiced [10].

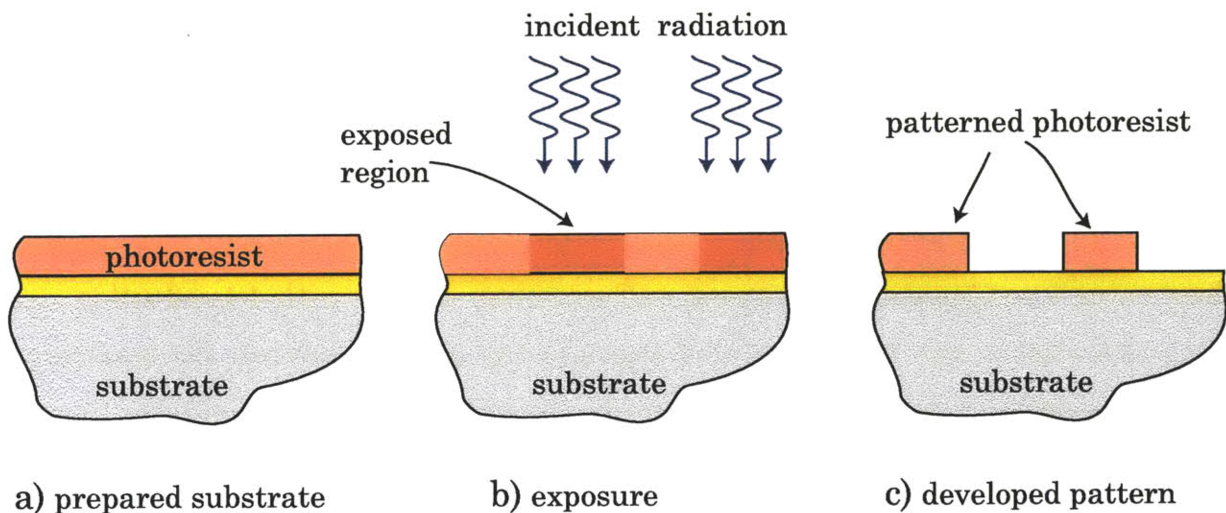


Figure 1.1 Schematic steps of a lithographic exposure. a) a susbtrate (grey) prepared with photoresist (red) over an arbitrary thin-film (yellow). b) the photoresist is exposed to a pattern of radiation. c) shown in the positive-tone, the exposed areas of the resist are developed away to leave a physical structure.

Finally, the photoresist pattern is transferred into a functional material, either through a subtractive process such as etching, or an additive process, such as electroplating or lift-off [11]. This is shown schematically in Figure 1.2. In the lift-off procedure, material is deposited through a sacrificial template, which is then removed leaving a pattern of deposited material which is the inverse image of the template. Chapter 8 of this thesis describes the etching process in more detail in the context of magnetic materials. This cycle can be repeated through many iterations to build complex multi-layer devices using a variety of materials. For example, fabrication of a modern computer processor such as the Pentium 4 chip from Intel, with ~125 million transistors, requires 21 layers of lithography and weeks of processing time [12].

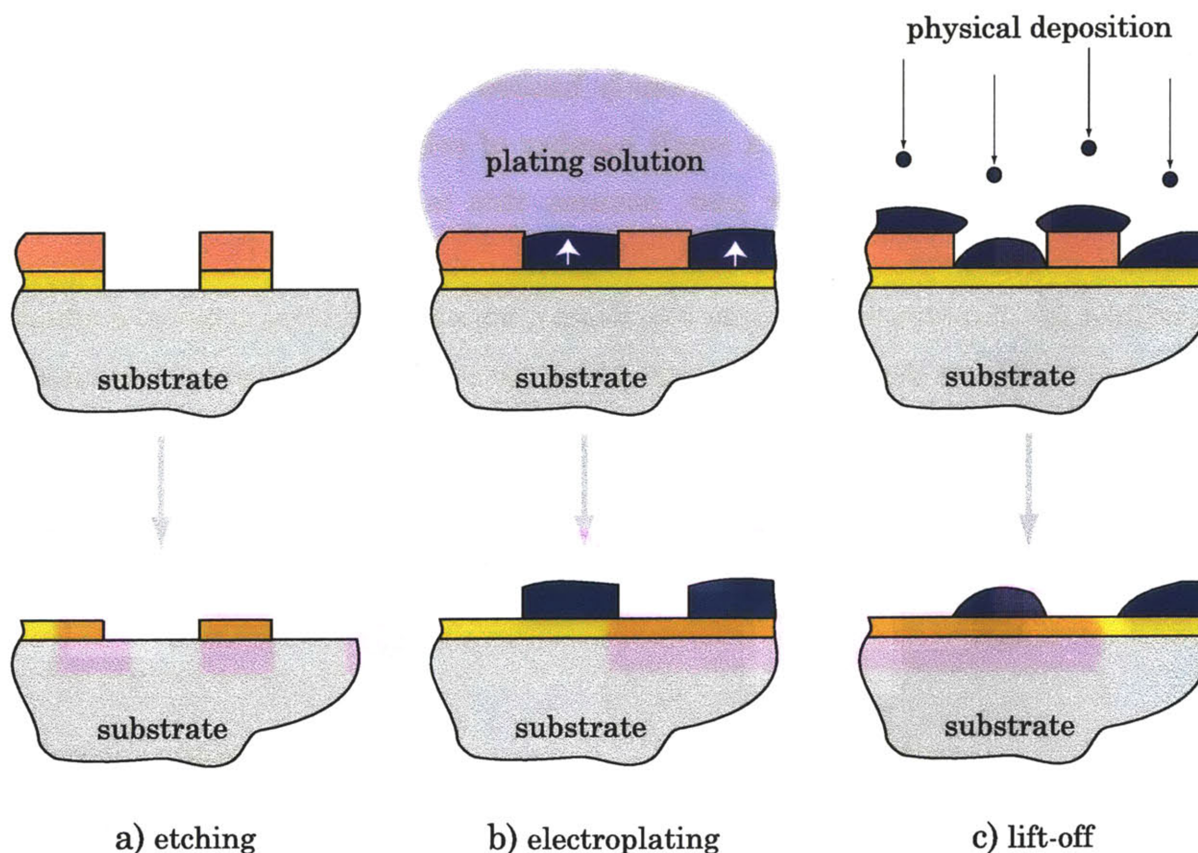


Figure 1.2 Examples of pattern transfer techniques. a) Patternend photoresist is used as a mask for etching into other layers. b) The photoresist is used as a form for electroplating material, inverting the pattern. c) Material deposited over the entire substrate can be "lifted off" leaving an inverse pattern in the substrate.

It is beneficial to consider all of the steps - substrate preparation, lithography, and pattern transfer - as a system when designing a suitable fabrication process. The choice of materials and the physical size of device features, in particular, are critical factors when determining optimal methods of lithography and pattern transfer. The complexity of designing a reliable and repeatable process is not to be underestimated. Thorough coverage of the lithography and pattern transfer processes can be found in a number of excellent references [13,14,15].

1.1.3 Interference lithography

The primary focus of this thesis is the design and implementation of lithographic interferometers, and the fabrication of certain devices using these tools. However, the type of lithography under investigation, known as interference lithography (IL), is inherently extremely limited in its capabilities. IL patterns, formed by the interference of a small number of coherent optical beams, produce only *periodic* patterns. For now, assume this is true - the fundamentals of interference will be discussed in detail in Chapter 2. In fact, the interferometers described in Chapters 3-6 use only two beams, ideally producing a pattern which is periodic in only *one dimension*. Exposing this pattern on a substrate creates a set of parallel lines, often called a grating.

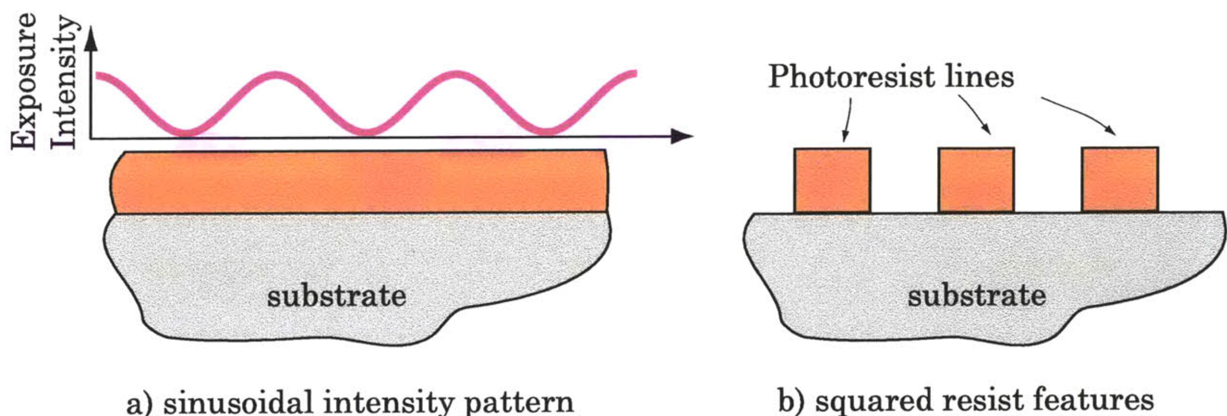


Figure 1.3 The sinusoidal intensity pattern formed by the interference of two coherent beams becomes a periodic array of individual lines in the developed photoresist due to its highly non-linear dissolution rate as a function of dose.

In truth, even the formation of discrete lines, as seen in Figure 1.3, is a result of the non-linear nature of photoresists. The actual intensity pattern formed using interference lithography is sinusoidal; the amplitude and periodicity of the sinusoid are the only lithographic "knobs" that can be adjusted. Arbitrary patterns are impossible using interference lithography. Periodicity in more than one dimension requires either more than two interfering beams or more than one exposure, a strategy explored in Chapter 7.

The range of devices which can be made using IL is limited; a few examples such as patterned magnetic media and distributed-feedback lasers are described in Chapter 7. However, within this admittedly narrow field, interference lithography is usually the best, and often the only lithographic option for periodic devices. Other modes cannot attain the spatial-phase coherence of a truly periodic structure, producing at best only quasi-periodic structures. This is especially relevant to optical devices [16].

Another strength of interference lithography is its ability to make smaller features than any other optical scheme at a given wavelength, rivaling even high-resolution electron-beam lithography. The minimum linewidth (half-pitch) l_{min} in optical lithography is often given in the form of Equation 7.1, where λ is the wavelength, NA is the numerical aperture of the tool, and k_1 is a constant that accounts for a variety of process related details.

$$l_{min} = \frac{k_1 \lambda}{NA} \tag{1.1}$$

Aggressive targets for the most advanced industrial tools are k_1=0.4 and NA=0.75, for a half-pitch of ~0.53λ. However, one must note that such performance requires a multi-million dollar investment in lithography systems and masks, and is out-of-reach for research and university environments. More affordable optical lithography tools tend to be limited to >1 μm resolution. For interference lithography, as will be derived in Chapter 2, k_1=0.25 is guaranteed and an NA of 0.95 is straightforward. The minimum linewidth is $l_{min} \approx 0.26\lambda$, half that of even

state-of-the-art industrial systems. For example, the IL systems described in Chapters 3 and 4 using a HeCd laser at λ=325 nm have demonstrated ~85 nm features. Chapter 5 describes 50 nm lines produced using IL with a 193 nm ArF laser. Chapter 6 introduces IL beyond the diffraction-limit, demonstrating 45 nm features at λ=193 nm, and capable of ~30 nm lines with further research.

1.2 The spatial theory of imaging

Viewed from a different perspective, the ability to create perfect sinusoidal patterns using IL is not a limitation, but a valuable asset. In the application of Fourier theory to imaging, sinusoidal components of different periodicity are the building blocks of an arbitrary image [17]. Just as an audio or communication signal can be decomposed into a spectrum of frequencies, a planar image $u(\vec{r})$ can be decomposed as the sum of independent "spatial frequencies" $\vec{f}$. This is illustrated in Figure 1.4.

$$u(\vec{r}) = \int_0^{\infty}\int_0^{2\pi} U(\vec{f})\, e^{i2\pi\vec{f}\cdot\vec{r}}\, dr d\theta \qquad (1.2)$$

As these spatial frequencies are nothing more than sinusoidal field distributions, we can recognize two-beam interference as producing precisely a single spatial-frequency component, the elementary function for an arbitrary image.

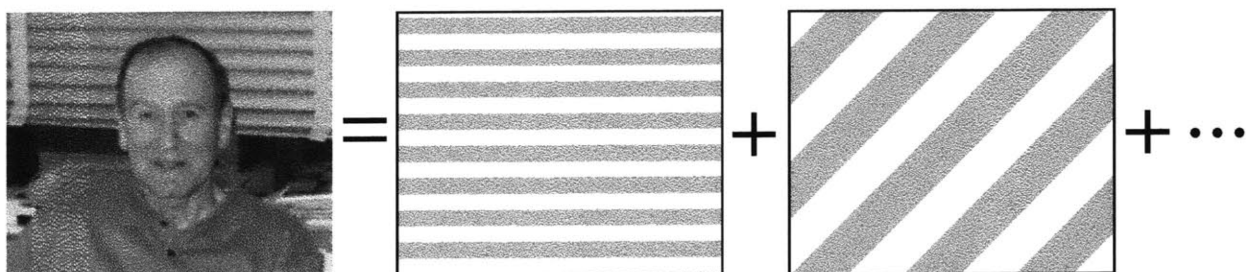


Figure 1.4 The Fourier decomposition of Prof. Smith. His picture can be constructed from a weighted sum of fundamental periodic components known as spatial-frequencies.

This formalism is commonly used to characterize optical imaging systems, such as projection lithography tools, to understand how the frequency content of an image is altered from the original mask. Interference lithography is maskless, using only two well-defined beams, eliminating the need for a transfer function of the tool itself. However, one can abstract this concept and imagine that the lithography process begins *after* the lithography tool. Using the aerial intensity image as the input to the system, the photoresist and pattern transfer steps will each have "transfer functions" which change the input (intensity pattern) into the output, the physical pattern of the final device.

Although the non-linearity of these processes prevents the Fourier formalism from being rigorously applied, it can be useful as a conceptual framework for understanding the lithographic process. Photoresist behavior, as well as process parameters such as etch rates, often have a marked dependence on spatial frequency (feature size). Recognizing and understanding these effects can be extremely difficult using the complex, and often unknown, images obtained via other lithographic tools.

IL offers a way to input single spatial-frequencies, the elementary functions of imaging, so that the lithography process can be learned and understood in a far more general sense than is possible with other techniques. In this regard, despite the limited realm of devices that can be made directly using IL, it is seen to be an invaluable tool for characterizing the lithography processes that underpin the production of *all* devices.

1.3 The Role of IL in Nanolithography & Nanotechnology

The advancement of microlithography over the past 40 years since the invention of the integrated circuit has been so rapid as to have left behind its own name. Lithography is now routinely producing features which are more appropriately measured in nanometers, not micrometers. Nanolithography, and the

blossoming world of nanotechnology in general, can be divided into two rather distinct camps. One camp, in which the work described in this thesis stands, is focused on the "top-down" approach characterized by the lithographic definition of features. Although rarely associated with the nascent nanotechnology boom, in fact semiconductor devices have been made in vast quantities with nanoscale dimensions for decades using this approach.

The alternate camp is defined by a focus on non-lithographic methods often referred to as self-assembly. This approach is in some respects much newer, and in some respects much older than the lithographic one. Creation of materials and molecules with a precise atomic structure has been the domain of chemists for a century or more. Only recently, however, have these techniques been applied towards building *devices* at the molecular scale.

Both approaches hold a great deal of promise, but also harbor a number of limitations. It is the opinion of a growing number of researchers that these two camps are complimentary, and further, that the ultimate realization of the promise of nanotechnology relies on their successful marriage. Lithographic methods, although unlikely to ever match the resolution of self-assembled structures, are the only way to add information content to a system. Self-assembled structures, while capable of molecular or even atomic level precision, are incapable of the high-level organization necessary to interact with other systems. The bridge between the familiar mesoscopic world and the molecular scale of future nanotechnologies must be made lithographically.

This type of research is still in its infancy, but it is clear that interference lithography is well-suited to investigating many fundamental problems of nanotechnology. As shown in Section 1.1.3, feature sizes possible with IL are significantly smaller than most other lithographies, rivaling the finest electron-

beam lithography*. Chapter 7 discusses the use of templates made with IL to induce order in self-assembled systems, such as block-copolymers and epitaxial quantum-dots. Chapter 3 introduces a method for reducing variations in the periodicity of interferometric patters to the level of picometers, (10^{-12} m) which could form the basis of a metrology system necessary to any serious nanotechnology endeavor [18,19]. Thus, IL provides an ideal tool, not just for certain devices, but for investigating current and future lithographic processes. Already, the tools and fabrication processes developed by the author as part of this thesis have led to over 30 published journal articles and conference proceedings, clearly demonstrating the value of the work to the scientific community. A selection of these applications will be discussed in Chapter 7.

* Proximity-X-ray and imprint lithography have been shown capable of *reproducing* features of this scale, although both require a mask and are therefore incapable of *generating* such high-resolution patterns.

1.4 References

[1] Hults, *The Print in the Western World: an Introductory History*. University of Wisconsin Press (1996)

[2] J. Hutchinson, *Albrecht Dürer, a biography*. Princeton University Press, Princeton N.J. (1990)

[3] E.W. Haury, "The Hohokam", *National Geographic*, **131** p.670 (1967)

[4] R. Hirsch, *Seizing the light: a history of photography*. McGraw-Hill, Boston (2000)

[5] W. Weber, *A History of Lithography*, Thames and Hudson, London (1966)

[6] W. Brinkman, D. Haggan and W. Troutman, "A History of the Invention of the Transistor and Where It Will Lead Us". *IEEE Journal of Solid-State Circuits*, **32**(12) (1997)

[7] M. Riordan, L. Hoddeson. *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age*. W.W. Norton & Company, New York (1997)

[8] Eastman Kodak Co., U.S. Patents 2,852,379, 1958, and 2,940,853, (1960).

[9] K. Jain, *Excimer Laser Lithography*, SPIE Press, Bellingham, Washington (1990)

[10] K. K. Berggren, A. Bard, J. L. Wilbur, A. G. Helg, J. D. Gillaspy, J. J. McClelland, S. L. Rolston, W. D. Phillips, M. Prentiss, G. M. Whitesides, "Microlithography by using neutral metastable atoms and self-assembled monolayers,", *Science* **269**, p.1255 (1995).

[11] J.L. Vossen, W. Kern, *Thin Film Processes*. Academic Press, New York (1978)

[12] Website of the Intel corporation, http://www.intel.com/research/

[13] R. Sheats, B.W. Smith, eds., *Microlithography: Science and Technology.* Marcel Dekker (1998).

[14] W. Moreau, *Semiconductor Lithography.* Plenum Press, New York, (1988)

[15] B. Chapman, *Glow Discharge Processes: Sputtering and Plasma Etching.* Wiley, New York (1980)

[16] J. Ferrera, "Nanometer-Scale Placement in Electron-Beam Lithography". Ph.D. Thesis, Massachusetts Institute of Technology, (2000)

[17] J. Goodman, *Introduction to Fourier Optics.* McGraw-Hill, New York (1996)

[18] H.I. Smith, S.D. Hector, M.K. Schattenburg, E.H. Anderson, "A new approach to high-fidelity e-beam and ion-beam lithography using an in situ global-fiducial grid.". *J. Vac. Sci. Technol. B* **9**(6) p.2992 (1991)

[19] M.L. Schattenburg and H.I. Smith, "The critical role of metrology in nanotechnology," *Proc. SPIE Nanostructure Science, Metrology, and Technology* **4608**, p. 116 (2001)

Chapter 2
The Interference of Light

2.1 Introduction

The study of electromagnetic phenomena is a rich and complex field. For the purposes of the discussions in this thesis, only a basic review is required. This chapter will outline the theoretical framework on which our discussion of lithographic interferometers is based. The concept of interference, and the experimental demonstrations of this phenomenon via interferometers, is perhaps the definitive aspect of the wave-theory of electromagnetics. Aspects of interference that are relevant to the design of lithographic interferometers will be the focus. For more detail, full development of the electromagnetic theory can be found in a multiplicity of texts, of which [1] and [2] are examples.

2.2 Fundamentals of interference

2.2.1 Maxwell's Equations

The theoretical basis for understanding all optical phenomena is the set of coupled differential equations known as Maxwell's Equations, which define the relationship between the electric *(E)* and magnetic *(H)* components of an electromagnetic field. These equations are given below (Equations 2.1-2.4) under the assumption of propagation through a linear, isotropic, homogenous, and source-free region. The magnetic permeability μ and electric permittivity ε are physical constants of the propagation medium. The fields in these equations are real functions of both position $\vec{r}$ and time t.

$$\nabla \times \vec{H}(\vec{r},t) = \varepsilon \frac{\partial}{\partial t} \vec{E}(\vec{r},t) \tag{2.1}$$

$$\nabla \times \vec{E}(\vec{r},t) = -\mu \frac{\partial}{\partial t} \vec{H}(\vec{r},t) \tag{2.2}$$

$$\nabla \cdot \vec{E}(\vec{r},t) = 0 \tag{2.3}$$

$$\nabla \cdot \vec{H}(\vec{r},t) = 0 \tag{2.4}$$

The Maxwell equations can be manipulated to form a wave equation in a single variable, known as the Helmholtz equation, given in Equation 2.5. Although only the wave equation for the electric field is shown, an analogous one for the magnetic field can be developed.

$$\nabla^2 \vec{E}(\vec{r},t) - \mu\varepsilon \frac{\partial^2}{\partial t^2} \vec{E}(\vec{r},t) = 0 \tag{2.5}$$

One set of solutions to the Helmoltz equation, shown in Equation 2.6, are sinusoidal functions in space and time known as plane-waves. The wave vector $\vec{k}$ and the frequency ω are constrained by the dispersion relation, Equation 2.7.

$$\vec{E}(\vec{r},t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \tag{2.6}$$

$$\left|\vec{k}\right|^2 = \mu\varepsilon\omega^2 \tag{2.7}$$

2.2.2 *Phasor Maxwell Equations*

Further simplification under the assumption of steady-state fields allows the time-dependence to be removed from the equations. The time-harmonic, or phasor Maxwell equations are shown below, where E and H are now complex functions of only the spatial coordinate $\bar{r}$.

$$\nabla \times \vec{H}(\bar{r}) = -i\omega\varepsilon\vec{E}(\bar{r}) \tag{2.8}$$

$$\nabla \times \vec{E}(\bar{r}) = i\omega\mu\vec{H}(\bar{r}) \tag{2.9}$$

$$\nabla \cdot \vec{E}(\bar{r}) = 0 \tag{2.10}$$

$$\nabla \cdot \vec{H}(\bar{r}) = 0 \tag{2.11}$$

The time dependence of the field quantities can be found at any point using

$$\vec{E}(\bar{r},t) = \mathrm{Re}\left[\vec{E}(\bar{r})e^{-i\omega t}\right]. \tag{2.12}$$

A phasor Helmholtz equation (Equation 2.13) can be found from the phasor Maxwell equations, with plane-wave solutions given by Equation 2.14.

$$\left(\nabla^2 + \omega^2\mu\varepsilon\right)\vec{E}(\bar{r}) = 0 \tag{2.13}$$

$$\vec{E}(\bar{r}) = \vec{E}_0 e^{i\vec{k}\cdot\bar{r}} \tag{2.14}$$

The magnitude of the wave-vector, often referred to as the spatial-frequency, can be rewritten in terms of the wavelength λ, a more familiar optical quantity.

$$k = \left|\vec{k}\right| = \frac{2\pi}{\lambda} \tag{2.15}$$

Rather than the electric field itself, our primary interest the optical intensity. Intensity is the important quantity in a lithographic exposure, and can be formulated in terms of the electric field as

$$I = \frac{c}{4\pi}\sqrt{\frac{\varepsilon}{\mu}}\left\langle \vec{E}(\bar{r},t)^2 \right\rangle. \tag{2.16}$$

In the above equation, c is the speed of light and the pointy brackets represent a time average. Averaging over a time interval that is large compared to the temporal periodicity of the wave, the spatial variations of the intensity distribution

are proportional to the squared phasor field. The proportionality constants are dropped for simplicity.

$$I \approx \vec{E}(\vec{r}) \cdot \vec{E}(\vec{r})^* \qquad (2.17)$$

The plane-wave solutions defined in Equation 2.14 are the building blocks from which we form a theoretical understanding of the interferometers described in the subsequent chapters of this work. It should be noted, though, that the plane-wave is an idealized, non-physical entity. Fortunately, we can approximate an arbitrary wavefront viewed over a small enough region as planar in most cases. Furthermore, application of Fourier techniques allows an arbitrary wavefront to be described as a sum of plane-wave components [3]. Thus, it is sufficient to confine our discussion to the case of plane-waves, which we will assume are monochromatic, with little loss of generality.

2.2.3 Superposition of two plane-waves

A single plane-wave, on its own, is of little interest as its intensity distribution is spatially uniform. However, the intensity distribution of a superposition of two or more coherent plane-waves will have a spatial structure that is non-uniform, a phenomenon known as interference.

For two overlapping plane-waves $\vec{E}_1$ and $\vec{E}_2$, the intensity I_{12} at a given spatial point $\vec{r}$ can also be described by Equation 2.17, except that the sum electric field of the two waves is substituted for the field of a single wave. The phasor descriptions of two plane-waves are given below in equations 2.18 and 2.19, An arbitrary phase difference $\Delta\phi$ between the two waves is included in the description of $\vec{E}_2$.

$$\vec{E}_1(\vec{r}) = \vec{E}_1 e^{j\vec{k}_1 \cdot \vec{r}} \qquad (2.18)$$

$$\vec{E}_2(\vec{r}) = \vec{E}_2 e^{j\vec{k}_2 \cdot \vec{r} + j\Delta\phi} \qquad (2.19)$$

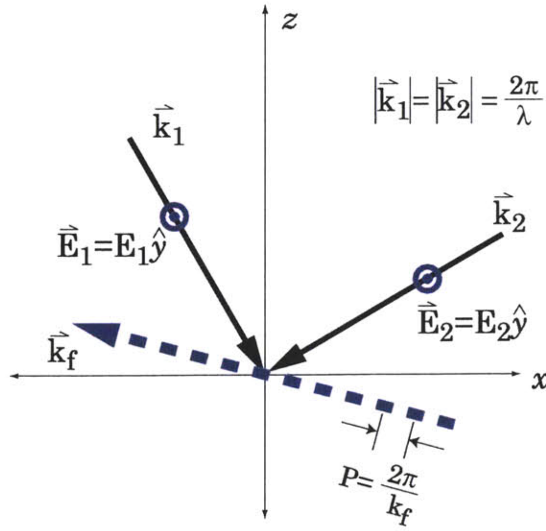


Figure 2.1 The periodicity of fringes formed by the interference of two plane-waves is inversely proportional to the magnitude of the fringe vector $\vec{k}_f$.

The intensity distribution of the superposition of the two waves is equal to the square magnitude of the sum electric field.

$$I_{12} = \left|\vec{E}_1(\vec{r}) + \vec{E}_2(\vec{r})\right|^2 \tag{2.20}$$

$$I_{12} = I_1 + I_2 + \left[\left(\vec{E}_1(\vec{r}) \cdot \vec{E}_2^{\,*}(\vec{r})\right) + \left(\vec{E}_1^{\,*}(\vec{r}) \cdot \vec{E}_2(\vec{r})\right)\right] \tag{2.21}$$

The intensity of the two waves together is seen to be more complex than simply the sum of intensities for the two waves. The bracketed term above is known as the interference term, which modulates the sum intensity of the two waves. Solving the interference term, we see two important features; the spatial dependence is sinusoidal and depends on the orientation of the two electric fields.

$$I_{12} = I_1 + I_2 + 2\left(\vec{E}_1 \cdot \vec{E}_2\right)\cos\left[\left(\vec{k}_1 - \vec{k}_2\right) \cdot \vec{r} + \Delta\phi\right] \tag{2.22}$$

In practical terms, it is easier to write the above equation using only the intensity, which is measurable, rather than the electric field, which is not. Writing the electric fields in terms of their magnitude ($\sqrt{I_1}$ and $\sqrt{I_2}$) and direction ($\hat{e}_1$and $\hat{e}_2$), the general description of the intensity distribution is given in Equation 2.23. This is known as the interference pattern.

$$I_{12} = I_1 + I_2 + 2\sqrt{I_1 I_2}(\hat{e}_1 \cdot \hat{e}_2)\cos\left[\left(\vec{k}_1 - \vec{k}_2\right)\cdot \vec{r} + \Delta\phi\right] \qquad (2.23)$$

The spatial dependence of this equation occurs only in the phase argument of the cosine. For this reason, we often refer to the spatial-phase Φ of the interference pattern.

$$\Phi = \left(\vec{k}_1 - \vec{k}_2\right)\cdot \vec{r} + \Delta\phi \qquad (2.24)$$

Note that the phase difference between the interfering beams, $\Delta\phi$, simply shifts the sinusoidal interference pattern, but does not otherwise affect its spatial dependence. For simplicity, we drop it from the further analysis by setting $\Delta\phi=0$. Consequences of time-dependence in the relative phase of the interfering beams, $\Delta\phi(t)$, will be discussed in the following chapters.

2.3 Periodicity of Interference Fringes

2.3.1 The symmetric angle of incidence

From Equation 2.23, we know that the interference pattern for two plane-waves is sinusoidal. The maxima and minima of this sinusoidal pattern are called *fringes*. For most lithographic applications, the periodicity of these fringes is their salient feature. Let the difference between the wave-vectors of the two interfering beams be called the fringe-vector, $\vec{k}_f$.

$$\vec{k}_f = \vec{k}_1 - \vec{k}_2 = k_{fx}\hat{x} + k_{fy}\hat{y} + k_{fz}\hat{z} \qquad (2.25)$$

If we define the direction of $\vec{r}$ parallel to the fringe vector, then the spatial-phase can be written as a scalar multiplication of the magnitude of the fringe-vector times the position r along this direction, given below in Equation 2.26.

$$\Phi = \left|\vec{k}_f\right| r \qquad (2.26)$$

The spatial periodicity P_f of the interference fringes can now be defined, shown below in Equation 2.27.

$$P_f = \frac{2\pi}{\left|\vec{k}_f\right|} \tag{2.27}$$

Consider the case where the coordinate axes are defined such the fringe vector $\vec{k}_f$, and hence the direction $\hat{r}$, are parallel to one of the unit axes, for example the $\hat{x}$ axis. This is shown in Figure 2.2. Under the monochromatic condition of $|k_1| = |k_2|$, the vector components are defined $k_{1x} = -k_{2x}$, $k_{1y} = k_{2y}$ and $k_{1z} = k_{2z}$. Therefore $\vec{k}_f = \hat{x}2k_{1x}$. Substituting this result, along with Equation 2.15, into Equation 2.27, we get

$$P_f = \frac{\lambda|k_1|}{2k_{1x}}. \tag{2.28}$$

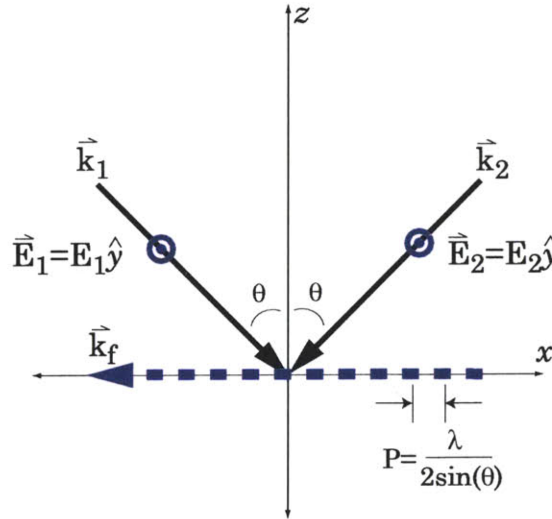


Figure 2.2 The periodicity of the interference pattern is commonly defined in the symmetrical case where bisection of the angle of interference of the beams (2θ) is perpendicular to the plane of observation of the fringes.

Defining the angle between the vectors $\vec{k}_1$ and $\vec{k}_2$ as 2θ, it can be shown that $\sin(\theta) = \frac{k_{1x}}{\left|\vec{k}_1\right|}$. Thus, the periodicity P_f of the interference fringes can be defined in terms of the half-angle between the interfering beams.

$$P_f = \frac{\lambda}{2\sin(\theta)} \tag{2.29}$$

Equation 2.29 is the most common form of the solution for the periodicity of the fringe pattern of two interfering beams. Although the coordinate axes were defined in a specific way to enable the derivation, Equation 2.29 is valid for an arbitrary coordinate system. In an arbitrary coordinate system, however, the fringe vector may not lie along one of the coordinate axis. *In order for Equation 2.29 to be valid, the direction over which the fringe period is measured must be perpendicular to the bisection of the interference angle 2θ.*

2.3.2 Observed fringe periodicity

The derivation of the periodicity given in Equations 2.27 and 2.29 assumed that the plane of observation of the fringes, determined by $\vec{r}$, was parallel to the fringe vector $\vec{k}_f$. However, this is not always true. Relative to the x-axis in Figure 2.1, the fringes form at an angle such that the periodicity of the *observed* fringes is larger. A detail of this situation is shown in Figure 2.3, where the observation plane is canted at an angle to the fringe vector. Recall that the spatial-phase of the fringes is the dot-product of the fringe vector $\vec{k}_f$ and the position $\vec{r}$. In a given observation plane it is beneficial to ensure that these vectors are parallel by defining the projection of the fringe vector into the observation plane. We refer to this projection as the surface vector $\vec{k}_s$, which can be found simply from the angle ϑ between the fringe vector and the observation surface, as shown in Equation 2.30. We can write the periodicity of the observed fringes (P_s) using the surface-vector instead of the fringe-vector. Situations where this distinction is important will be discussed thoroughly in Chapter 3.

$$\vec{k}_s = \vec{k}_f \cos(\vartheta) \qquad (2.30)$$

$$P_s = \frac{2\pi}{|\vec{k}_s|} = \frac{P_f}{\cos(\vartheta)} \qquad (2.31)$$

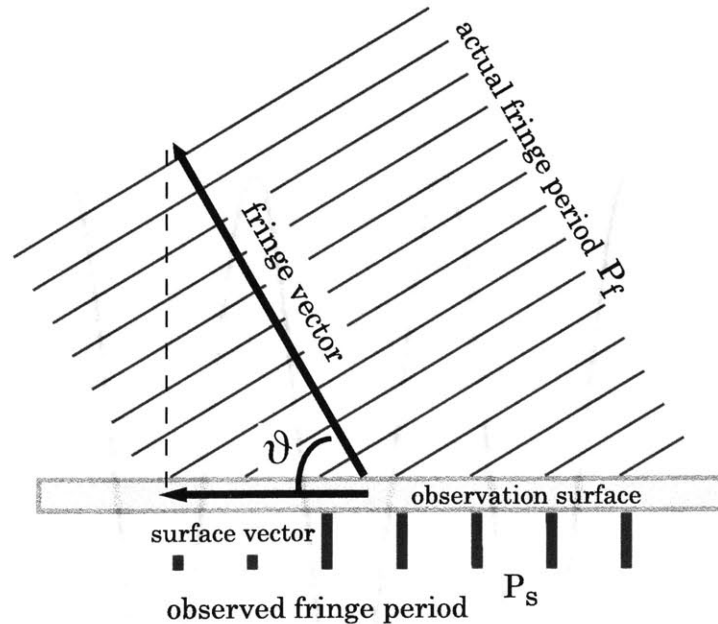


Figure 2.3 The fringe period P_s observed in a plane which is not parallel to the fringe vector will be larger than the actual fringe period P_f. The vector of the observed fringes, $\vec{k}_s$, is proportional to the fringe vector $\vec{k}_f$ by the cosine of the angle between them.

2.4 Visibility of interference fringes

2.4.1 Intensity of the interfering waves

The absolute intensity of an interference pattern is actually a less critical parameter than one might initially imagine in a lithographic exposure. The photosensitive films, called photoresists, which record the pattern are sensitive to the total absorbed energy, known as the exposure dose. There is a reciprocity relationship between the intensity of the aerial image and the exposure time that allows for equivalent results from low-intensity images exposed over long times and from high-intensity images over short exposure times. It is critical, however, that the areas which are nominally unexposed and the areas which are nominally fully exposed can be distinguished. Achieving differentiation between exposed and

unexposed regions is partly the magic of a well-designed photoresist. Of course, it is also necessary for the aerial image itself to have regions of high and low intensity.

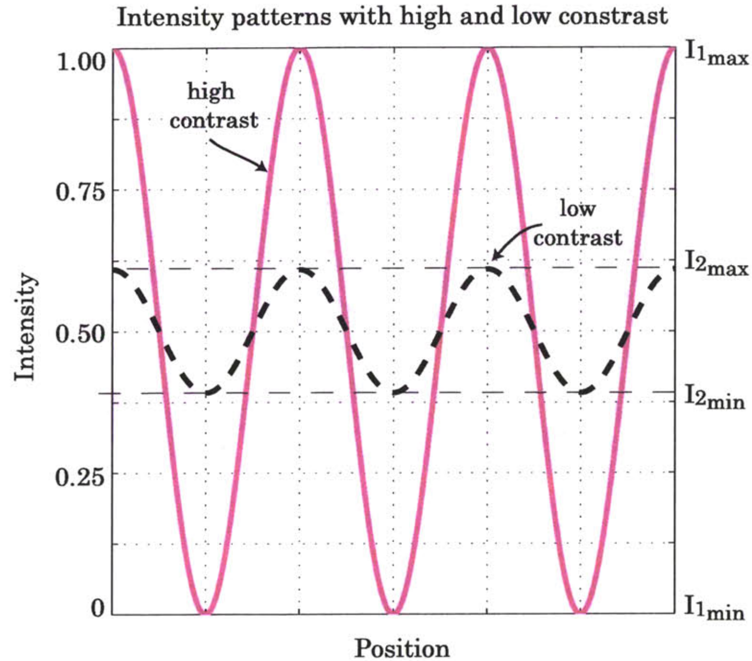


Figure 2.4 The variation in intensity as a function of position is shown for two cases. In the high-constrast (solid line) case, the minima go fully to zero intensity and the peaks to a value equal to twice tha average intensity. In the low-contrast (dashed line) case, the minima are increased and the maxima decreased while the average intensity remains the same.

The ability to distinguish the high- and low-intensity regions can be quantified with a parameter V known as contrast, or visibility, defined in Equation 2.32. Examples of a sinusoidal intensity pattern with high and low contrast are shown in Figure 2.4.

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{2.32}$$

For the interference of two coherent plane-waves, the maximum and minimum intensities shown below in Equations 2.33 and 2.34 can be found by setting the value of the cosine term in Equation 2.23 to (1) or (-1), respectively. Assume for now that the electric fields of the two waves are parallel.

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \tag{2.33}$$

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2} \qquad (2.34)$$

Substituting the above equations back into Equation 2.32, the contrast of the interference pattern of two coherent plane-waves is given in Equation 2.35 in terms of the intensity of the interfering beams.

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} \qquad (2.35)$$

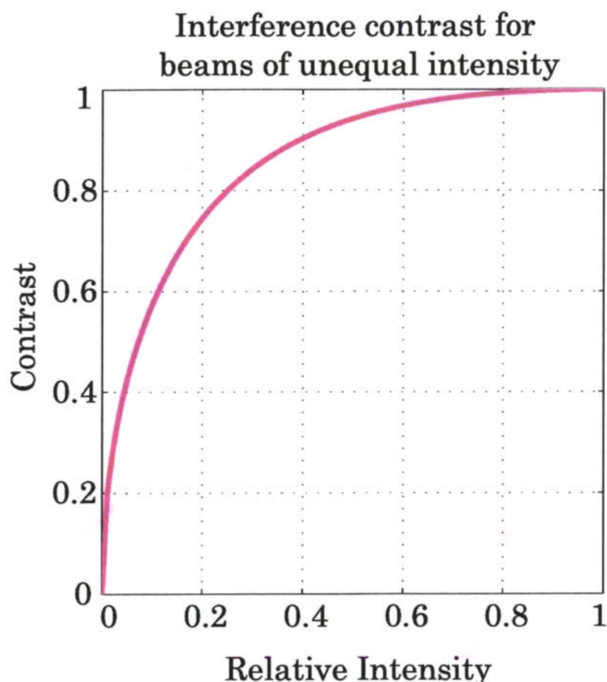


Figure 2.5 Fringe contrast as a function of the relative intensity of the two interfering beams. Relative intensity is defined as the ratio of the less-intense beam to the more-intense beam, for example I_1/I_2.

The contrast is maximized with a value of unity when the intensity of the two beams is equal. Contrast is reduced if either beam is more intense than the other. The visibility as a function of the relative intensity of the interfering beams is shown in Figure 2.5. In practical terms, this curve is very forgiving in that almost perfect contrast can still be achieved for beams with a moderate intensity mismatch.

2.4.2 Polarization

It has been assumed in the previous section that the electric fields of the two interfering waves were parallel. This will occur when the electric field is perpendicular to the plane defined by the two beams, commonly known as the TE (transverse-electric) polarization. This is the optimal configuration for achieving high-contrast interference fringes, as the dot-product $\hat{e}_1 \cdot \hat{e}_2$ in Equation 2.23 is always equal to unity and can be dropped from the analysis. A more complex situation occurs for the TM (transverse-magnetic) polarization, when the electric field vectors are parallel to the plane-of-incidence. A superposition of these two cases is sufficient to describe an arbitrarily polarized beam, as the electric field itself must be orthogonal to the direction of propagation.

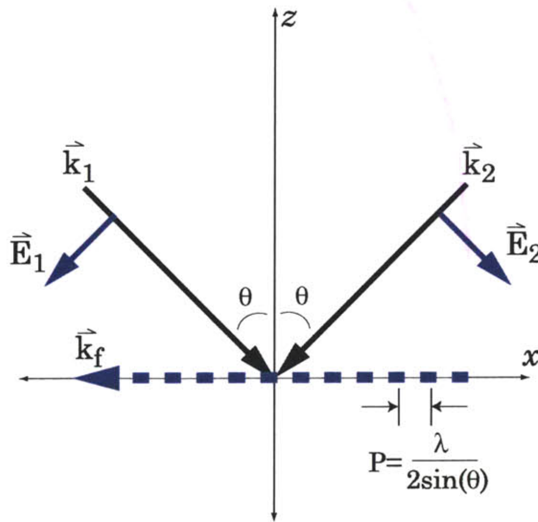


Figure 2.6 Interference of TM polarized plane-waves. The electric field vector is parallel to the plane defined by the interfering beams.

Assuming equal intensity TM waves interfering as shown above, Equation 2.23 is simplified as

$$I_{12} = 2I\left(1 + (\hat{e}_1 \cdot \hat{e}_2)\cos\left[\left(\vec{k}_1 - \vec{k}_2\right)\cdot \vec{r}\right]\right). \quad (2.36)$$

Note that when the polarizations are orthogonal, contrast in the interference pattern goes to zero. The intensity is constant, equal to the sum intensity of the two

beams. The dot-product of the polarization vectors, $\hat{e}_1 \cdot \hat{e}_2$, can be expanded in terms of the interference angle θ as shown below.

$$(\hat{e}_1 \cdot \hat{e}_2) = -\cos(2\theta) \qquad (2.37)$$

The values of I_{max} and I_{min} are found by setting the spatial phase argument to $(\pm\pi)$.

$$I_{\max_{TM}} = 2I(1 + abs[\cos(2\theta)]) \qquad (2.38)$$

$$I_{\min_{TM}} = 2I(1 - abs[\cos(2\theta)]) \qquad (2.39)$$

Applying Equation 2.32 to solve for the contrast, we find the result is simply the absolute value of the dot-product of the electric fields, plotted in Figure 2.7.

$$V = abs[\cos(2\theta)] \qquad (2.40)$$

This result should be intuitively correct. For interference angles close to 0° or 180°, the electric fields are approximately parallel or anti-parallel and the interference contrast will be high. Similarly, the beams and electric fields are orthogonal when the interference angle is 90°, and the contrast goes to zero.

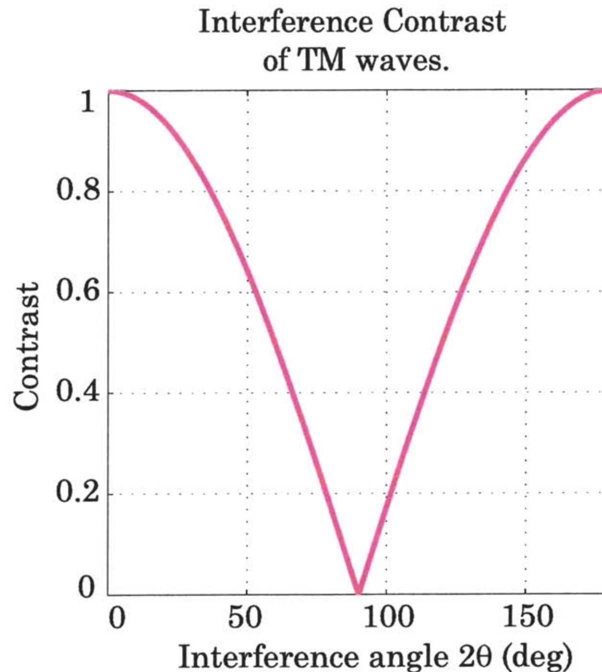


Figure 2.7 Contrast of interference fringes as a function of the interference angle for TM-polarized waves. The contrast goes to zero as the interfering beams become orthogonal. Contrast is maximized for co-propagating (*2θ=0°*) or counter-propagating (*2θ=180°*) beams.

2.5 Partial Coherence

2.5.1 Coherence and spectral bandwidth

The preceding discussion has assumed that the interfering waves were monochromatic and coherent. In general, the term *coherent* is used to describe waves which can interfere with one another. Monochromatic waves of the same frequency from the same source will be perfectly coherent, while monochromatic waves of different frequencies from the same source will be incoherent [4]. We assume throughout this discussion that any interfering waves are emitted from the same source. It is important to realize that the monochromatic assumption also implies that the emitted waves are infinite in extent and duration.

A more physically appropriate description of the emitted waves, however, is that of a monochromatic wavetrain of finite duration τ. For a plane-wave as described in Equation 2.6, we can write the temporal dependence of the electric field at a fixed point as

$$\begin{aligned} E(t) &= \cos(2\pi\nu_0 t) \quad 0 \le t \le \tau \\ &= 0 \qquad\qquad\quad t < 0,\ t > \tau \end{aligned} \tag{2.41}$$

Here we make use of the frequency ν rather than the angular frequency ω used earlier. From Fourier theory [5,6], we know that such a time-domain signal must have a frequency-domain spectrum $u(\nu)$ of finite width. The main lobe of the frequency spectrum will have a width $\Delta\nu$ given as

$$\Delta\nu = \frac{1}{\tau}. \tag{2.42}$$

Thus, the description of a finite duration monochromatic wavetrain is consistent with a description of the emission as non-monochromatic wavetrains of infinite duration over a finite bandwidth. The bandwidth of the emission is seen to be inversely proportional to the duration of the wavetrains. The duration of the wavetrains is referred to as the coherence time $t_c=\tau$. The physical length of the wavetrain, found by multiplying the coherence time by the speed of light c, is known as the coherence length l_c. Simple manipulations show that the coherence length

can also be given in terms of the spread in center wavelength λ_0 and wavelength spread $\Delta\lambda$.

$$l_c = c \cdot t_c = \frac{\lambda_0^{\,2}}{\Delta\lambda} \tag{2.43}$$

The following discussion will attempt to make clear the relevance of the coherence length and coherence time.

2.5.2 The mutual coherence function

The common wisdom of coherence is that two waves will not interfere if the time delay between them is greater than the coherence time. This is useful knowledge, but does not allow us to answer the question of what happens when the time delay is nonzero but less than the coherence time. To understand this, we must revisit the postulates made at the beginning of the chapter. This time, we focus on the complex electric field $E(t)$ at a fixed point due to the broadband emission from a single source. The electric field at this point in space will be a sum of the contributions from wavetrains emitted randomly in time, for example E_1 emitted at time t and E_2 emitted at time $t+\Delta t$.

$$E(t) = E_1(t) + E_2(t+\Delta t) \tag{2.44}$$

The two complex signals comprising $E(t)$ can be given as sums of their frequency components, each with the frequency spectrum $u(\nu)$ given by the Fourier transform of the real field amplitude $\mathrm{E}(t)$.

$$E_1(t) = 2\int_0^\infty u_1(\nu)\, e^{i2\pi\nu t} d\nu \tag{2.45}$$

$$E_2(t) = 2\int_0^\infty u_2(\nu)\, e^{i2\pi\nu(t-\Delta t)} d\nu \tag{2.46}$$

Recall from Equation 2.16 that intensity is given by the inner product of the field and its conjugate.

$$I_{12} = \left\langle E(t)\, E^*(t) \right\rangle \tag{2.47}$$

As was done earlier, we can expand this as

$$I_{12} = I_1 + I_2 + \left\langle E_1(t) E_2^*(t+\Delta t) \right\rangle + \left\langle E_1^*(t) E_2(t+\Delta t) \right\rangle. \quad (2.48)$$

The bracketed terms are conjugates of one another, and therefore we can rewrite their sum as

$$I_{12} = I_1 + I_2 + 2\mathrm{Re}\left\{ \left\langle E_1^*(t) E_2(t+\Delta t) \right\rangle \right\}. \quad (2.49)$$

Let us define a new function, which we call the mutual coherence function $\gamma_{12}(\Delta t)$, as the normalized version of the bracketed term.

$$\gamma_{12}(\Delta t) = \frac{\left\langle E_1^*(t)\, E_2(t+\Delta t) \right\rangle}{\sqrt{I_1 I_2}} \quad (2.50)$$

Using the mutual coherence function, we can write the intensity as

$$I_{12} = I_1 + I_2 + 2\sqrt{I_1 I_2}\, \mathrm{Re}\{\gamma_{12}(\Delta t)\}. \quad (2.51)$$

By separating the mutual coherence function into its magnitude and phase components, where ϕ is the phase argument, the above equation simplifies to

$$I_{12} = I_1 + I_2 + 2\sqrt{I_1 I_2}\, |\gamma_{12}(\Delta t)| \cos(\phi). \quad (2.52)$$

As in the case of the spatial variations, we can extract the maximum and minimum intensity values using the maximum and minimum of the cosine to determine the fringe visibility. Assuming for simplicity that the two beams are of equal intensity, the contrast of the interference pattern is simply

$$V = |\gamma_{12}(\Delta t)|. \quad (2.53)$$

The magnitude of the mutual coherence function is known as the degree of coherence of the interfering waves.

2.5.3 Fringe visibility as a function of coherence.

Recall from Figure 2.5 that the fringe visibility for coherent interference is approximately unity for beams of close to the same intensity. Therefore, in most of the situations we construct, the assumption of equal intensity will be highly accurate. The degree of coherence then plays the dominant role in determining the visibility of the fringes. If we additionally assume that the two wavetrains have an equal frequency spectrum, $u_1(\nu) = u_2(\nu)$, the derivation above is valid for two

separate but otherwise identical wavetrains emitted at different times from the same source. An equally valid interpretation is for the interference of a single wavetrain that has been divided into two parts using some apparatus, e.g. an interferometer, that affects a time delay in one of the beams.

Equation 2.53 shows us that the visibility of the interference fringes we can expect from the interferometer is dependent on the time delay incurred between the two beams. Generally, the time delay of one beam relative to the other is due to a difference in the optical-path-length traversed by the waves. Thus, in the design of practical interferometers, it is important to understand the degradation of the fringe visibility as a function of the path length.

We can rewrite the definition of the mutual coherence function under the assumption that the two waves are of equal intensity and bandwidth, which we recognize as a normalized autocorrelation.

$$\gamma_{12}(\Delta t) = \frac{\left\langle E^*(t)\, E(t+\Delta t)\right\rangle}{\left\langle E^*(t)\, E(t)\right\rangle} \qquad (2.54)$$

Recall that the frequency-domain signal $u(\nu)$ of the time-domain wavetrain $E(t)$ is given by its Fourier transform, which we denote $F\{\ \}$. One of the basic equations of Fourier theory equates the transform of an autocorrelation function to the square of the transform of that signal.

$$F\left\{\left\langle E^*(t)\, E(t+\Delta t)\right\rangle\right\} = \left|u(\nu)\right|^2 \qquad (2.55)$$

Inverting the transform and writing the normalization constant as the intensity, we get the fringe contrast as

$$V(\Delta t) = \frac{F^{-1}\left\{\left|u(\nu)\right|^2\right\}}{I}. \qquad (2.56)$$

The bandwidth of physical sources can generally be well approximated by a Gaussian distribution. This is particularly convenient, since the Gaussian distribution is invariant under the Fourier transform. We can solve for the fringe contrast as a function of time delay Δt as

$$V(\Delta t) = e^{-(\pi\sigma\Delta t)^2}, \qquad (2.57)$$

where σ is the standard deviation of the Gaussian bandwidth. Equation 2.57 can be converted to a more practical form by multiplying the delay time by the speed of light c to form a path length difference Δl. For the slightly conservative definition of σ as half the frequency bandwidth $\Delta\nu$, and applying Equations 2.42 and 2.43 which define the coherence length, l_c. The contrast as function of path length difference is given using the simple formula shown below, and plotted in Figure 2.8.

$$V(\Delta l) = e^{-\left(\frac{\pi\Delta l}{2l_c}\right)^2} \qquad (2.58)$$

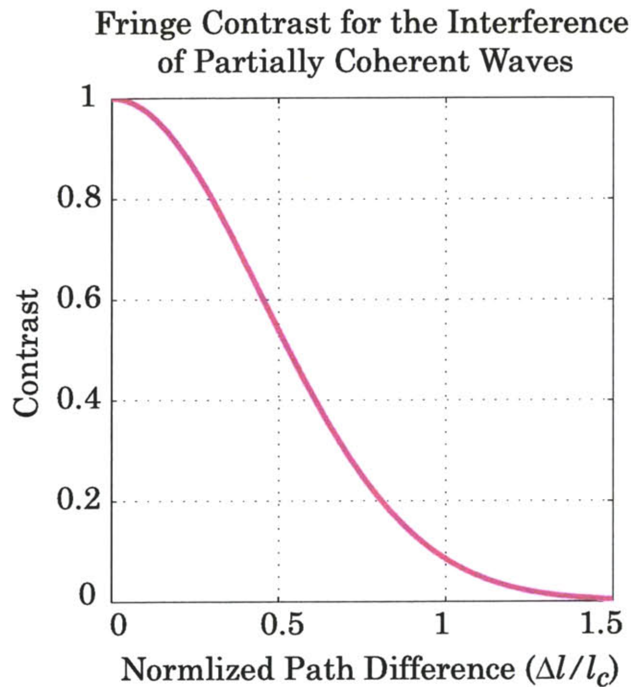


Figure 2.8 Interference contrast as a function of path length difference for a non-monochromatic source. The path length difference is normalized in terms of the coherence length l_c as defined in Equation 2.43.

2.6 References

[1] J.A. Kong, *Electromagnetic Wave Theory*. EMW Publishing, Cambridge MA (1999)

[2] M. Born, E. Wolf, *Principles of Optics*. Cambridge University Press, Cambridge UK (1999)

[3] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, New York (1996)

[4] M. Francon, *Optical Interferometry*. Academic Press, New York (1966)

[5] R.N. Bracewell, *The Fourier Transform and its Applications*. Mcgraw-Hill, New York (1965)

[6] G.B. Arfken, H.J. Weber, *Mathematical Methods for Physicists, 4th ed.*. Academic Press, New York (1995)

Chapter 3
The Mach-Zehnder Interferometer

3.1 Introduction

The most intuitive, although not necessarily the easiest, way to form a set of interference fringes is to simply split a beam in two, and then recombine the two arms. In the classic Mach-Zehnder interferometer [1,2], an additional beamsplitter is used at the intersection of the beams such that the beams will interfere collinearly. However, interference fringes can be observed directly in the area of intersection via the placement of a screen or recording media. For lithography, photoresist is used to record this pattern. This configuration may also be familiar to some as the basis for an offset-reference, or Leith-Upatneiks hologram [3], where the "object" and "reference" beams are identical. For this reason, interference lithography has also been known as holographic lithography.

This scheme was first employed as a lithographic tool in the 1960's [4,5], and has been refined for a wide variety of tasks since then, such as field-emission displays, nano-magnetics and atom diffraction. Tailored to the specific requirements of

lithography, this scheme has been under development here at MIT for around 30 years for the purpose of making extremely fine-pitch diffraction gratings for spectroscopy and metrology [6,7,8,9,10].

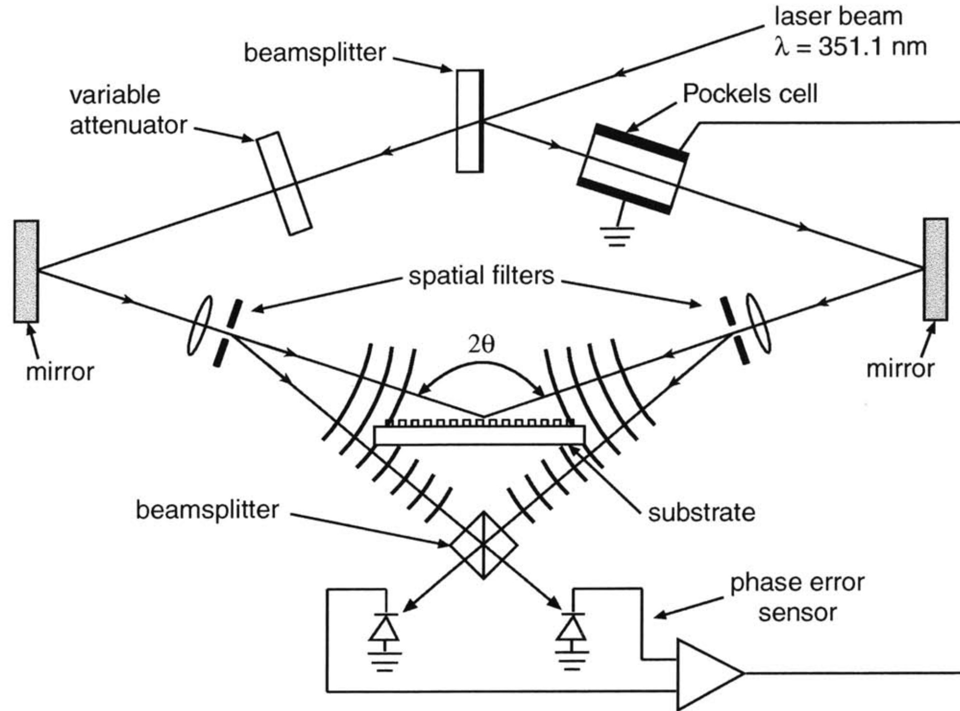


Figure 3.1 Schematic of a Mach-Zehnder style interferometer used in the MIT Space Nanotechnology Laboratory. The laser beam is split into two arms, spatially filtered and expanded. The two arms interfere on a resist coated substrate. The phase sensors and Pockels cell are part of a fringe locking sub-system.

In principle, these systems are extremely simple. However, the closer one looks, the more complex it becomes. A simplified schematic of the system in the MIT Space Nanotechnology Lab is shown in Figure 3.1. A 351.1 nm argon-ion laser beam is split in two arms which are recombined using a set of mirrors. The nominal periodicity P of the resulting grating is dependent on the angle of interference θ and the wavelength λ.

$$P = \frac{\lambda}{2\sin(\theta)} \qquad (3.1)$$

Spatial-filters in each arm serve two functions. One is to expand the beams for dose uniformity over a large area. The other is to remove spatial-frequency noise from beams. Due to the long propagation distances (~1 meter) and the lack of additional

optics after the spatial-filter, the beams interfering at the substrate can be accurately approximated as spherical. Excellent analyses of this are found in the work of Fererra and Chen [11,12]. The spatial position of the interference fringes is determined by the relative phase of the beams, which makes this type of interferometer extremely sensitive to path length differences between the two arms. For this reason, a fringe-locking system is used in which complimentary interference patterns are monitored using a pair of photodiodes as shown in Figure 3.1 [13,14]. Differential changes in the intensity on the photodiodes are converted into a phase error, which is then corrected by a Pockels cell in one of the interferometer arms. This achieves long-term stability of the fringes which would otherwise be impossible. A variable attenuator in the other arm is used to balance the power lost through the Pockels cell for maximum fringe contrast.

3.2 Alignment techniques

3.2.1 Reference Plane

Accurately positioning the point-sources (spatial-filters) with respect to the substrate is critical to controlling the periodicity of the interference fringes. In this section a diffractive alignment technique is described which can aid considerably in the configuration of a Mach-Zehnder type interferometer. Ideally, the positions of the point-sources are symmetrical about an axis normal to the substrate. Proper positioning of the point-sources must be defined, then, in terms of the substrate plane. Therefore, the first step when building, or reconfiguring, this type of system must be to define the angular orientation of the substrate plane. The beam of a HeNe laser, which will be referred to as the reference beam, is used to define the normal to this plane. This beam is also used to define the center position of the substrate. For every substrate used, the specular reflection of the beam can be set back along its incident path. The accuracy $\Delta\phi_{ref}$ to which the angle of the substrate

plane can be defined will be dependent on the path length of the reference beam, l_{ref} and the accuracy to which the reference beam can be placed, Δd_{ref}.

$$\Delta\phi_{ref} = \frac{\Delta d_{ref}}{l_{ref}} \tag{3.2}$$

By using a reference beam that does not expose the photoresist, the angular position of every substrate can be ensured before each exposure. Although this may seem trivial if one has a reliable chuck for mounting the wafer, residue from the spin-coating process often collects on the backside of the wafer, preventing the substrates from lying flat. Figure 3.2 shows the approximate positions of the reference beam and substrate plane for the interferometer described in the rest of this chapter. The path length of the reference beam is ~2 meters, and a conservative estimate of 0.5 mm is used for co-locating the position of the reflected beam by eye. Thus, the substrate surface can be set normal to the beam to within 0.25 mrad. A tighter tolerance could easily be implemented with a more sophisticated method of locating the reflected reference beam.

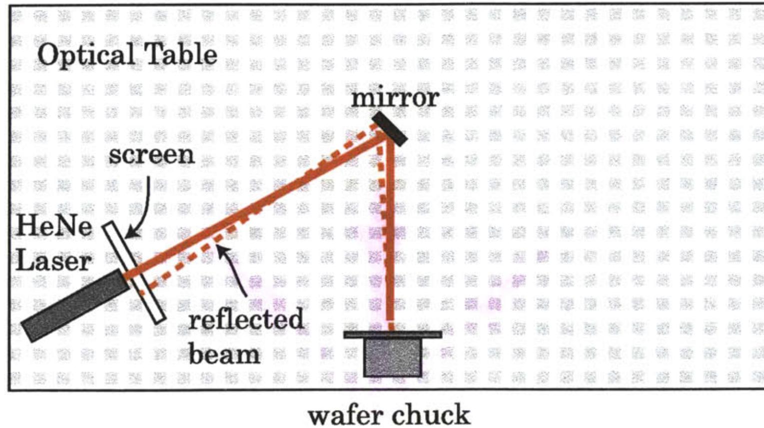
One such method would employ the reference plane in a Michelson interferometer. For this configuration, any angular misalignment of the wafer plane would be observed as interference fringes on the observation screen. The periodicity of the interference fringes would also be given by Equation 1, where the interference angle θ is equal to half the angular misalignment of the stage ($\Delta\phi_{ref}$). Under the small-angle approximation, we can solve Equation 1 for the angular alignment of the wafer plane in terms of the observed fringe period.

$$\Delta\phi_{ref} = \frac{\lambda_{ref}}{P_{observed}} \tag{3.3}$$

By expanding the interfering beams, large period patterns can become visible. If we assume a fringe periodicity as high as 6 cm can be detected, angular misalignment as small as of ~10 μrad can be measured for a HeNe reference beam with wavelength 633 nm. This technique is depicted schematically in Figure 3.2. Although shown in proximity to the wafer for clarity, in reality, the positions of the

beamsplitter and reference plane must be located so as no to occlude the interfering beams during a grating exposure.

Alignment for milliradian accuracy



Alignment for microradian accuracy

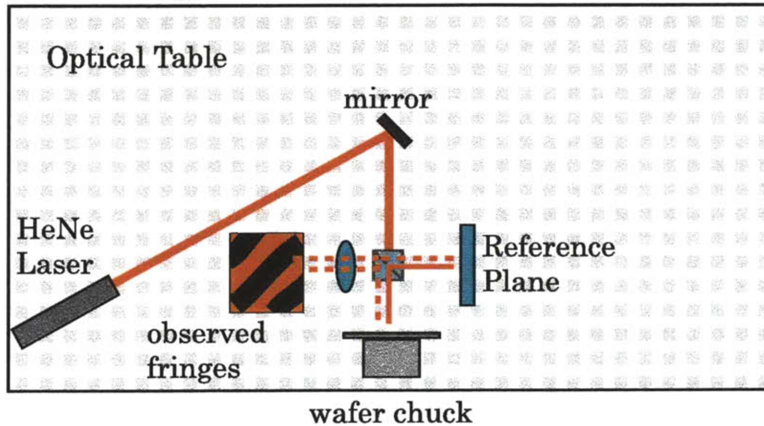


Figure 3.2 Two alignment schemes for the angular position of the wafer plane utilizing a HeNe reference beam. In the simple scheme shown at the top, the reflection of the reference beam is observed on a screen at the laser head. The position of the reflected spot is proportional to the angular misalignment of the wafer (Equation 3.2). The scheme shown in the lower figure observes the interference of the reference beam with its reflection off the wafer. The periodicity of the interference fringes determines the angle of the wafer (Equation 3.3)

3.2.2 Optics positioning via a reference grating.

With the reference plane set, we can now think about positioning the rest of the optics. A simple method for this makes use of a reference grating with a periodicity $2P_0$, exactly equal to twice the periodicity P_0 of the desired grating. This

reference grating can be fabricated with an accuracy approaching 1 part in 10^5 using the Lloyds-mirror interferometer described in Chapter 4. Recall the diffraction equation for the simple case of a normally-incident beam on a grating with period $2P_0$, given in Equation 3.4.

$$\sin(\theta_m) = \frac{m\lambda}{2P_0} \tag{3.4}$$

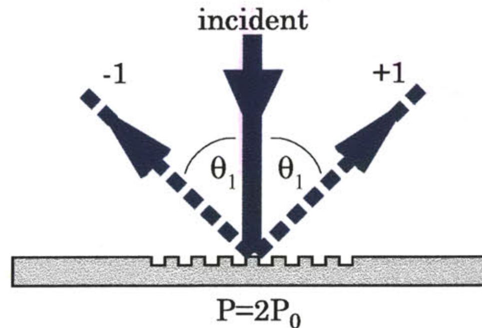


Figure 3.3 First order diffracted beams from a grating of period $P=2P_0$ will be at an angle θ_1 which is precisely the angle of interference necessary to expose a grating of period P_0.

For the first diffraction orders ($m=1$), we can rewrite the equation in the form of Equation 3.1, indicating that θ_m will be exactly the interference angle necessary to form fringes of period P_0. Thus, if the beam at the exposure wavelength is set co-incident with the reference beam, the diffracted orders will trace in reverse the desired path of the interfering beams.

$$P_0 = \frac{\lambda}{2\sin(\theta_1)} \tag{3.5}$$

It is then a simple matter to position the optics such that the beams from each arm are coincident back from the beamsplitter to the laser head, shown in Figure 3.4. This ensures that when the beam propagates in the forward direction, the optics are perfectly aligned. Correct alignment of the forward propagating beams can be checked by viewing the (-1) and (-2) diffracted orders from the reference grating for each arm. The (-1) orders from each arm will diffract coincident with each other and the reference beam, guaranteeing that the incident beams from each arm meet at the center of the wafer, and that each one is at the correct angle. Additionally,

the (-2) order diffracted beam should be coincident with the specularly reflected beam from the other arm. If any of the conditions is not met, the beams are out of alignment and it is easy to discern whether the error is in position or angle of the incident beams.

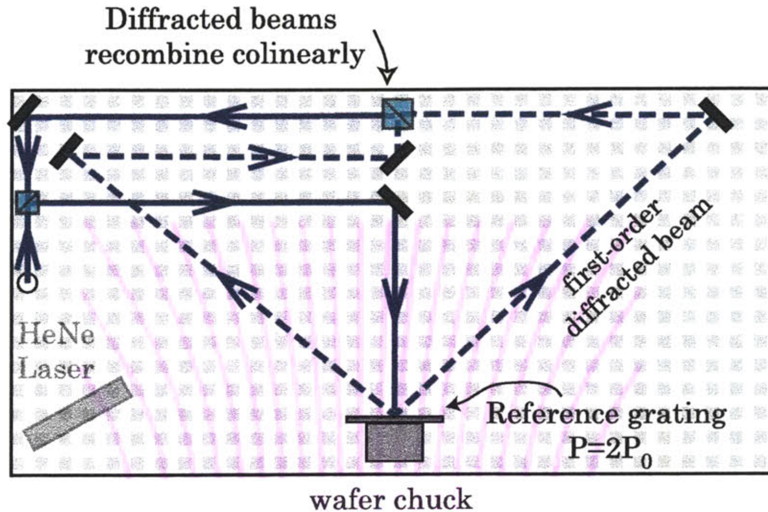


Figure 3.4 Back-diffracted orders from a normally incident beam on a reference grating of period $P=2P_0$ are used to position the optics for the exposure of a grating of period P_0. Mirror positions are set such that the reverse propagating beams are collinear after the beamsplitter and propagate back to the laser head.

3.3 Spatial Phase

3.3.1 Hyperbolic Phase

Another issue, often overlooked, with this interferometer design is that the interfering wavefronts are spherical rather than planar. For non-critical applications, the plane-wave approximation is acceptable if the phase-radius of the spherical wavefronts is large compared to the diameter of the exposure area. For true plane-wave interference, interference fringes will be both straight and parallel. Consequently, fringe periodicity is independent of position, and the spatial-phase of the grating will have a linear progression. In many applications, notably spectroscopy and metrology, the spatial-phase of the grating is extremely

important. The true fringe pattern from interfering spherical beams is a set of confocal hyperboloids [15,16,17,18], shown exaggerated in Figure 3.5. For this reason, the fringes are said to have hyperbolic phase. In simple terms, the period of the grating printed on the wafer matches the desired period only at the center of the wafer, and increases non-linearly away from the center. Why, then, do we use spherical-beam interference instead of plane-wave interference when it is important to print a grating with constant period?

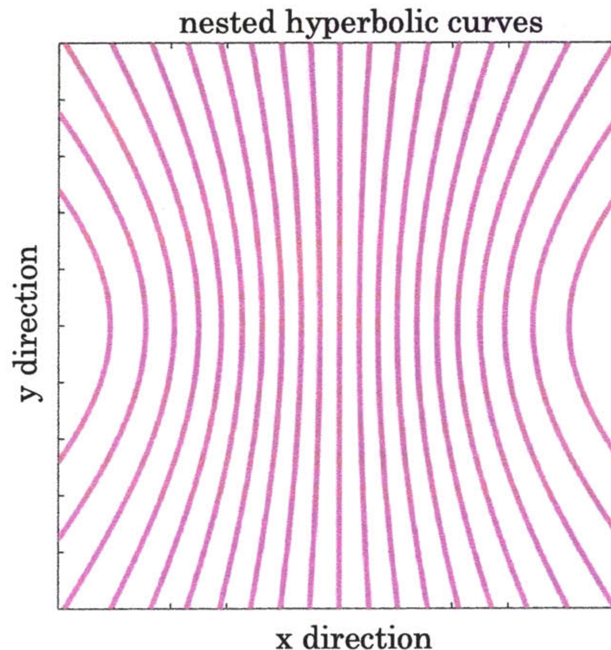


Figure 3.5 An example of the hyperbolic fringe pattern produced by spherical beam interference

The answer to this comes when one considers other figures-of-merit beyond the variation of fringe period. The fringes from spherical beam interference can be made more uniformly and with less error over a large area than the fringes from plane-wave interference. Further discussion of this can be found in Section 3.5.

3.3.2 Fringe inclination

In the next section we will discuss methods for linearizing the spatial phase of gratings made using interference lithography. As a background to this, we must

understand exactly how and why the periodicity of the grating changes. In Chapter 2, the local periodicity P of an interference pattern was presented as the difference in wave-vectors of the interfering beams. Generally, we simplify this in terms of the angle 2θ between the beams to yield Equation 3.1.

For plane-waves, the interference angle 2θ, and thus the period P, are constant. It should be easy to see, however, that the interference angle for spherical waves is spatially dependent. The fringe pattern that results from spherical wave interference has been well studied, one example is a thorough and informative analysis done recently by Ferrera [11]. However, this and other earlier analyses are based on the incorrect assumption that the periodicity P_s of the fringes recorded *on the wafer* is accurately described by the period of the interference fringes, which we will call P_f. To understand this distinction, we must recognize that implicit in Equation 3.1 is the condition that the substrate is normal to the bisection of the interference angle 2θ.

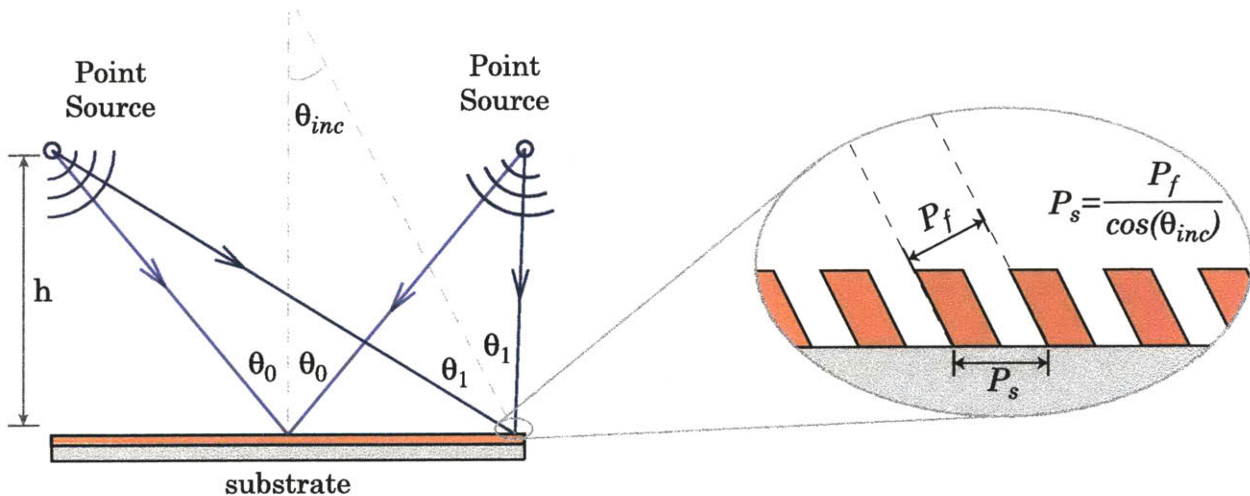


Figure 3.6 Illustration of the change in interference angle $2\theta_1$ and the associated fringe inclination angle θ_{inc} for spherical beam interference on a planar substrate. The interference angle decreases away from the center, while the fringe inclination angle increases. Both effects act to increase the grating period.

Interference fringes always form parallel to the bisection of the interference angle, so under this assumption the fringes themselves are normal to the substrate. However, for spherical beam interference on a flat substrate, this condition is met

only along the plane of symmetry, which we define as the y-z plane at x=0, the center of the substrate. At all other locations, the fringes form off-normal with an inclination angle θ_{inc}. As a result, the period of the interference fringes is modified by the inclination of the fringes when recorded on the substrate.

$$P_s = \frac{P_f}{\cos(\theta_{inc})} \qquad (3.6)$$

Under the assumption that the inclination angles are small, θ_{inc} at a distance x away from the origin can be given simply as the ratio of the substrate position x to twice the height of the point sources h.

$$\tan(\theta_{inc}) = \theta_{inc} = \frac{x}{2h} \qquad (3.7)$$

An extreme but illuminating example of this effect is to consider the pattern recorded at the center of the wafer. In the case shown in Figure 3.6, the wafer lies in the *XY* plane so the fringe period recorded is simply P_s=P_0. However, if the wafer were rotated 90° to lie in the *YZ* plane, only a single fringe would be recorded across the entire substrate, with the surprising value of $P_s=\infty$. In most cases, we can ignore the effects of fringe inclination, just as the variation in interference angle can be ignored in most cases. However, in situations where it is necessary to consider the minor periodicity variations that cause hyperbolic phase, inclusion of the fringe inclination effect is very important.

A more general description for the effect of fringe inclination can be found by returning to the vector notation introduced in Chapter 2. For clarity, a few of the key results are reproduced here. The grating vector of the interference fringes $\vec{k}_f$ is given simply by the difference between the wave-vectors of the interfering beams $\vec{k}_1$ and $\vec{k}_2$ at any point *(x,y,z)* (Eq. 3.8). Equation 3.9 shows the local periodicity of the fringes to be inversely proportional to the grating vector. If the grating vector $\vec{k}_f$ lies in the plane of the substrate, then the vector of the fringes recorded on the substrate $\vec{k}_s$ will exactly equal that of the fringe pattern itself.

$$\vec{k}_f = \vec{k}_1 - \vec{k}_2 \qquad (3.8)$$

$$P = \frac{2\pi}{|\vec{k}|} \qquad (3.9)$$

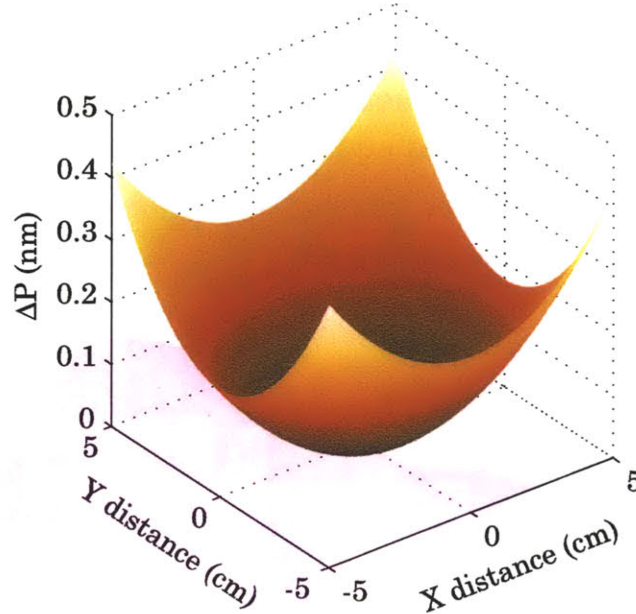


Figure 3.7 The change in periodicity (P-P_0) a grating of nominal period P_0=200nm across a 10 cm diameter area. The source positions for this simulation are at (x,y,z)=(0.848, 0, 0.479), as for the IL system in the MIT Space Nanotechnology Lab. The maximum change is only fraction of a nanometer over the entire wafer, or about 1 part in 10^3.

In general, however, this is not true. The grating vector of the recorded fringes is the projection of the fringe vector into the substrate plane, as given in Equation 3.10. We see that the fringe inclination angle also represents the angle between the fringe vector $\vec{k}_f$ and the substrate plane.

$$|\vec{k}_s| = |\vec{k}_f|\cos(\theta_{inc}) \qquad (3.10)$$

In these terms, the analyses of other authors have assumed, seemingly unintentionally, that $\vec{k}_s = \vec{k}_f$. We can get a sense of the importance of the fringe inclination by looking at the difference between the true grating period on the substrate P_s and the fringe period P_f as a percentage of the total change in period (ΔP) across a wafer, which we will designate ε_{inc}. An example of the total change in periodicity across a wafer is shown in Figure 3.7.

$$\varepsilon_{inc} = \left(\frac{P_s - P_f}{P_s - P_0}\right)\cdot 100 \qquad (3.11)$$

The value of ε_{inc} is shown in Figure 3.8 for a nominal 200 nm period grating where the locations of the point sources, and ΔP, are the same as given in Figure 3.7

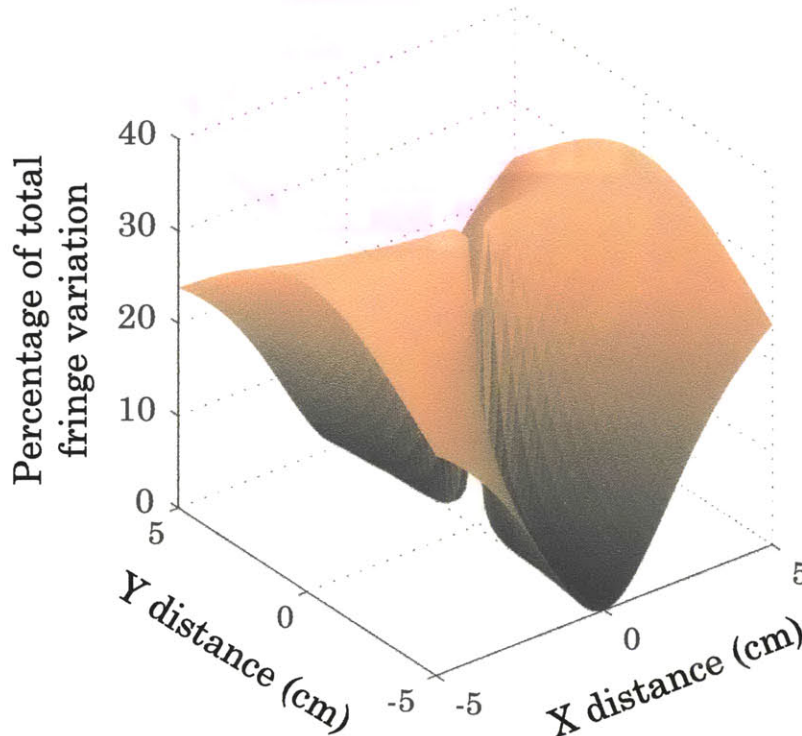


Figure 3.8 The percentage of total period variation across a 10cmx10cm area which can be attributed to the effect of fringe inclination for spherical beam interference recorded on a flat substrate. The point sources are positioned at coordinates *(X,Y,Z)=(0.878,0,0.479)* meters. Notice that fringe inclination has a maximum effect of ~33% along the x-axis and no effect along the y-axis.

The maximum value occurs along the x-axis, where the fringe-inclination contributes as much as 1/3 of the total change in periodicity. Along the y-axis, where the fringes are normal to the substrate (θ_{inc}=0) the inclination obviously does not contribute to the period variation. Over the majority of the wafer area, the fringe inclination is seen to be a significant contributor to the periodicity change.

3.3.3 *The importance of assumptions...*

Based on these results, one would expect that earlier calculations for the change in fringe period would underestimate the actual change in period by about 33%. Surprisingly, calculations in References [11,19] agree with sensitive experimental measurements of the grating period done in [19] to within the error of the measurement, quoted as 40 ppm. How can this be so? It turns out that an additional simplification, $\vec{k}_f \cong \vec{k}_{fx}$ (the fringe vector is well approximated by only its $\hat{x}$ component) is often made for algebraic simplicity on the grounds that $|k_{fx}| >> |k_{fy}|,|k_{fz}|$. Although this condition is true, it is not the correct one to apply in this case.

Recall from Figure 3.7 that the total period variation over a 10 cm wafer is ~1/1000. Therefore, there will be appreciable effects from vector components $\vec{k}_{fy}, \vec{k}_{fz}$ even if their magnitude is only 1/1000th that of $\vec{k}_{fx}$; we are "throwing away the baby with the bath water" by neglecting the small vector components.

We might expect that a more meaningful condition to be one where the $\vec{k}_{fy}$ and $\vec{k}_{fz}$ components are compared to $\Delta\vec{k}_{fx}$, the *change* in $\vec{k}_{fx}$, as shown below.

$$\vec{k}_f \cong \vec{k}_{fx} \quad \text{iff} \quad \left|\vec{k}_0 - \vec{k}_{fx}\right| >> |\vec{k}_{fy}|,|\vec{k}_{fz}| \qquad (3.12)$$

Note that this condition is actually comparing the change in the $\hat{x}$ component to the change in the $\hat{y}$ and $\hat{z}$ components; the fundamental frequency k_0 only has an $\hat{x}$ component and therefore $\Delta\vec{k}_{fy} = \vec{k}_{fy}$ and $\Delta\vec{k}_{fz} = \vec{k}_{fz}$. Figure 3.9 shows a comparison of the magnitude of the $\vec{k}_{fz}$ and $\vec{k}_{fy}$ vectors to the total change $\Delta\vec{k}_f$. We see that the $\hat{y}$ component is fairly small and we can probably justify its elimination. At the same time, though, one could argue that 10% is large enough to merit inclusion. The $\hat{z}$ component, on the other hand, comprises essentially the entirety of the change in the fringe vector except along the y-axis. We have exactly the opposite situation to the expected - $|\vec{k}_{fz}| >> |\Delta\vec{k}_{fx}|,|\vec{k}_{fy}|$. Does this mean that we should neglect the $\hat{x}$-component and only keep the $\vec{k}_{fz}$ component?

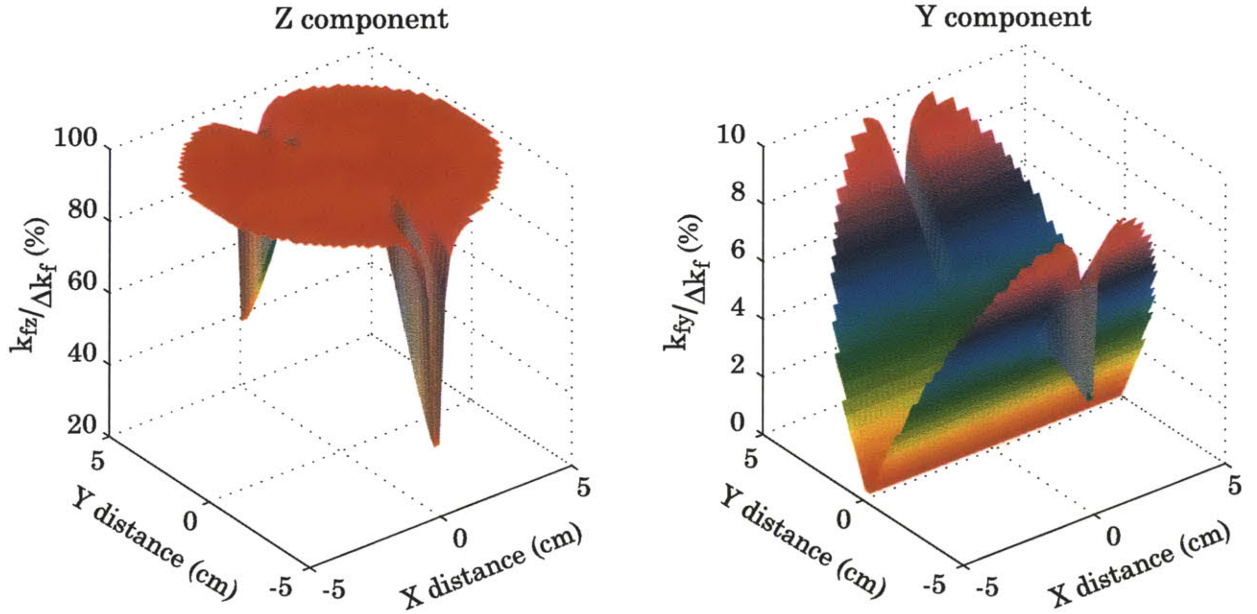


Figure 3.9 Percentage of the change in the fringe vector ($\Delta|\vec{k}_f|$) which can be attributed to its z (left) and y (right) components for a 200 nm period grating. The y component is relatively insignificant, while the z component dominates everywhere except along the y axis.

Consider that the fringe vector at any point can be written as the sum of the fundamental plus an extra term, as shown in Equation 3.13.

$$\vec{k}_f = \vec{k}_0 + \Delta\vec{k}_f \qquad (3.13)$$

Thus, any changes in the periodicity of the fringes will be due solely to the presence of the term $\Delta\vec{k}_f$. Because the fundamental term only has an $\hat{x}$ component, the $\hat{y}$ and $\hat{z}$ components of $\Delta\vec{k}_f$ are exactly those of $\vec{k}_f$. At any given point then, we can write the error term as

$$\Delta\vec{k}_f = (\vec{k}_0 - \vec{k}_{fx}) + \vec{k}_{fy} + \vec{k}_{fz} \qquad (3.14)$$

Because the grating periodicity is determined by the *magnitude* of $\vec{k}_f$, it is not correct to assume that the change in periodicity ΔP is determined simply by the magnitude of $\Delta\vec{k}_f$. The important quantity is the change in magnitude of the fringe vector $\Delta|\vec{k}_f|$, not the magnitude of the change in the fringe vector $|\Delta\vec{k}_f|$.

$$\Delta|\vec{k}_f| = |\vec{k}_0| - |\vec{k}_f| \qquad (3.15)$$

$$\left|\vec{k}_f\right| = \sqrt{\left(k_0 + \Delta k_{fx}\right)^2 + k_{fy}^2 + k_{fz}^2} \tag{3.16}$$

Expanding the parentheses,

$$\left|\vec{k}_f\right| = \sqrt{k_0^2 + \left(2k_0\Delta k_{fx} + \Delta k_{fx}^2 + k_{fy}^2 + k_{fz}^2\right)} \tag{3.17}$$

If we assume that the changes are all small compared to the fundamental, then we can approximate Equation 3.16 by the first two terms of a Taylor series, shown below.

$$\left|\vec{k}_f\right| \approx k_0 + \left(\frac{2k_0\Delta k_{fx} + \Delta k_{fx}^2 + k_{fy}^2 + k_{fz}^2}{k_0}\right) \tag{3.18}$$

By inserting this result back into the Equation 3.15, we get

$$\Delta\left|\vec{k}_f\right| \propto 2k_0\Delta k_{fx} + \Delta k_{fx}^2 + k_{fy}^2 + k_{fz}^2 \tag{3.19}$$

Thus, we should base the condition of whether or not to eliminate the $\hat{z}$ and $\hat{y}$ terms on Equation 3.19: it is not as simple as comparing the components of $\vec{k}_f$ as was done in previous work, nor even as simple as comparing the components of $\Delta\vec{k}_f$ as in Equation 3.12.

$$\vec{k}_f \approx \vec{k}_{fx} \quad \text{iff} \quad \Delta k_{fx}\left(2k_0 + \Delta k_{fx}\right) >> k_{fy}^2, k_{fz}^2 \tag{3.20}$$

In the comparison, the $\hat{x}$-term now has the additional weight of the fundamental, which compensates for the small contribution of Δk_{fx}. Figure 3.10 shows the relative magnitudes of the $\hat{y}$ and $\hat{z}$ terms as percentages of the $\hat{x}$-term. Despite the near dominance of Δk_{fz} in $\Delta\vec{k}_f$, it is seen to contribute only ~1/3 as much to $\Delta\left|\vec{k}_f\right|$ as the $\hat{x}$-term. Nevertheless, this is a significant amount and thus we can conclude that the $\hat{z}$-term should not be wantonly eliminated. The contribution of the $\hat{y}$-components, however, is only a fraction of a percent despite the fact that $\Delta\vec{k}_{fy}$, as seen in Figure 3.9, was fully 10% of $\Delta\vec{k}_f$. The elimination of the $\hat{y}$-component, therefore, can be well justified.

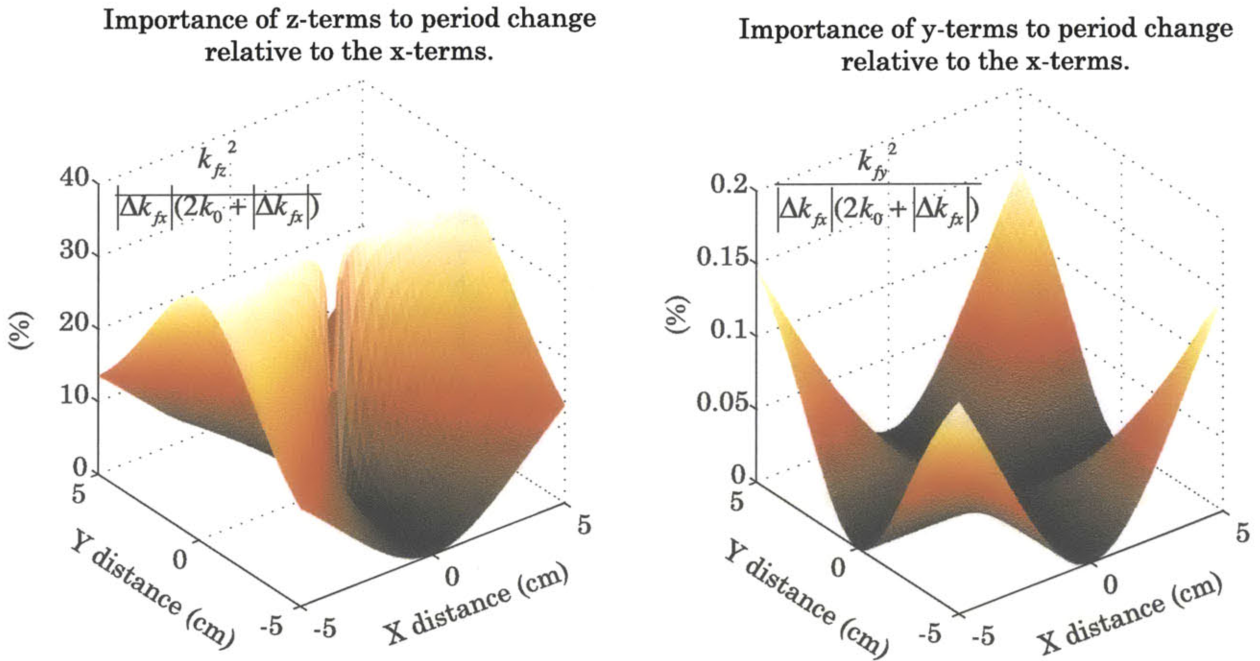


Figure 3.10 The contribution of z (left) and y (right) components of the fringe vector $\vec{k}_f$ to period change (ΔP), shown as a ratio to the x contribution, based on Equation 3.20. The z component is seen to be up to 33% of the value of the x component, indicating that it should not be discarded as an algebraic simplification. The y component, however, contributes only a fraction of a percent and could be eliminated from the analysis without incurring significant error.

Of course, the point of eliminating certain components from the analysis is simplification. However, the simplification is worthless if it is not accurate. Conversely, for increasingly complex conditions to ensure the validity of a simplification, such as that given above in Equation 3.20, one questions the point of actually making the simplification to begin with. It is the opinion of this author that the additional effort necessary to compute the full solution is small compared to the potential for inaccurate results when using a simplified formula.

Of course, now it would appear that the results of prior authors would be way off course, with two solidly incorrect assumptions. Fringe inclination has been ignored, and vital vector components have been eliminated. Comparing Figure 3.10 and Figure 3.6 gives a hint as to the role of the vector simplification - the two figures are identical. On the flat exposure plane, neglecting the $\hat{z}$ component has exactly the effect of projecting $\vec{k}_f$ onto the substrate surface. In other words, the

operation that accounts for fringe inclination is exactly the elimination of the z component of the fringe vector on a flat exposure surface. Thus, the two mistakes cancel one another out and the result is actually highly accurate. The $\hat{y}$ component of the fringe vector is still neglected, however the results above indicate that this is acceptable in terms of the new criteria set forth in Equation 3.20.

It must be emphasized that the accuracy of the earlier calculations is pure luck - if a more rigorous solution had been sought which included the $\bar{k}_{fy}$ and $\bar{k}_{fz}$ components, the result would actually be far more inaccurate. One might wonder at the necessity of exposing the flawed nature of an essentially correct result. The value comes in the application of the methodology to other situations. For other configurations of the substrate plane, as will be discussed in Section 3.6, or asymmetrical placement of the point sources, the fringe inclination will *not* be inadvertently corrected by removing terms from the interference vector $\bar{k}_f$. Specifically, the results in [11] describing the deviations from hyperbolic phase that one would expect from a substrate out of the XY plane will underestimate the true error.

3.4 Moiré Techniques for Measuring Spatial Phase

For a spherical-beam interferometer, we have claimed that the grating lines are not perfectly straight, but are actually a set of confocal hyperboloids. Figure 3.7, which plots the change in periodicity of the fringes over a 10-cm diameter area, gives a sense of the magnitude of the difference between the actual fringe pattern and the ideal linear-phase grating. Only the most dedicated observer would be able to tell that the period had changed at all. Tools generally used to measure linewidth such as the scanning-electron microscope (SEM) would be unable to resolve this level of change. Only in situations where a large number of periods were being averaged would the effect be seen. For instance, if the grating were being used as a metrological standard, where each tooth of the grating was counted

as a specific step size, it could take as little as ~500 periods, or 100 microns, for the measurement to off by a full step. Changes of this magnitude can also affect the response of devices such as Bragg gratings or photonic crystals.

One particularly sensitive technique for observing these variations in periodicity is the use of self-referential moiré images. When a pattern is exposed and developed in photoresist, the resulting physical grating can be placed back into the interferometer. Intuitively, one can consider the intensity image of the interference pattern forming a moiré with the physical grating in the resist. Ideally, these two gratings are the same. Thus, if the resist grating is placed back in exactly the same position as it was exposed, no moiré pattern will be present. However, if the physical grating is moved relative to the intensity image, then the spatial-dependence of the grating period will cause a moiré to form whose pattern is dependent on the direction and magnitude of the relative shift.

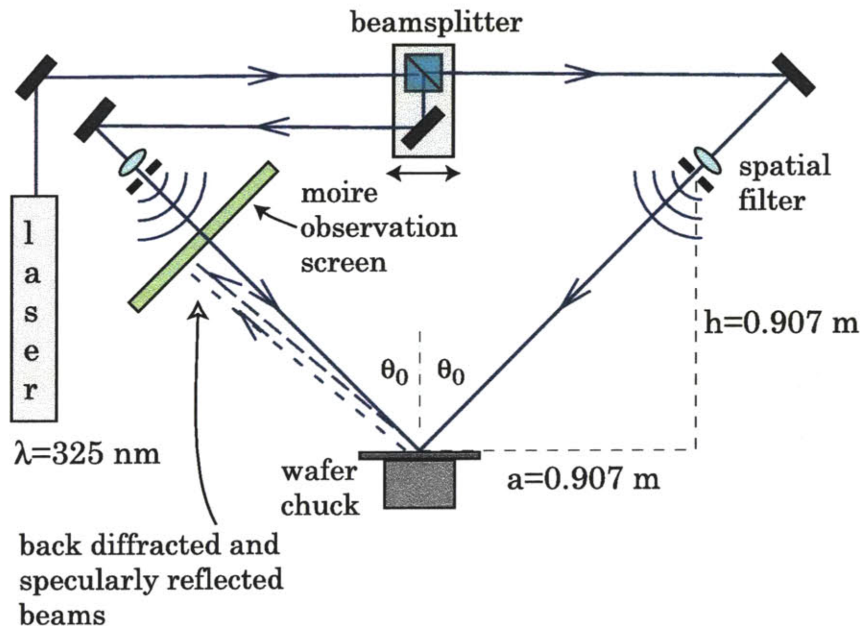


Figure 3.11 Simplified schematic of the IL setup used to observe moiré images .

Ferrera has described a technique for viewing this moiré pattern directly on the wafer by casting a fluorescent polymer over the grating pattern which has been etched into silicon [18]. The same results may be obtained more simply through the

use of a fluorescent photoresist. Modification of a commercial resist with a fluorescent dopant is described in Chapter 6.4.1.

More generally, the moiré patterns can be observed on a screen placed back at the pinhole as shown in Figure 3.11. The pattern viewed on the screen is a result of the back-diffracted beam from one arm interfering with the specularly reflected beam of the other arm. This pattern has been shown to be equivalent to the one formed directly on a flat substrate [11].

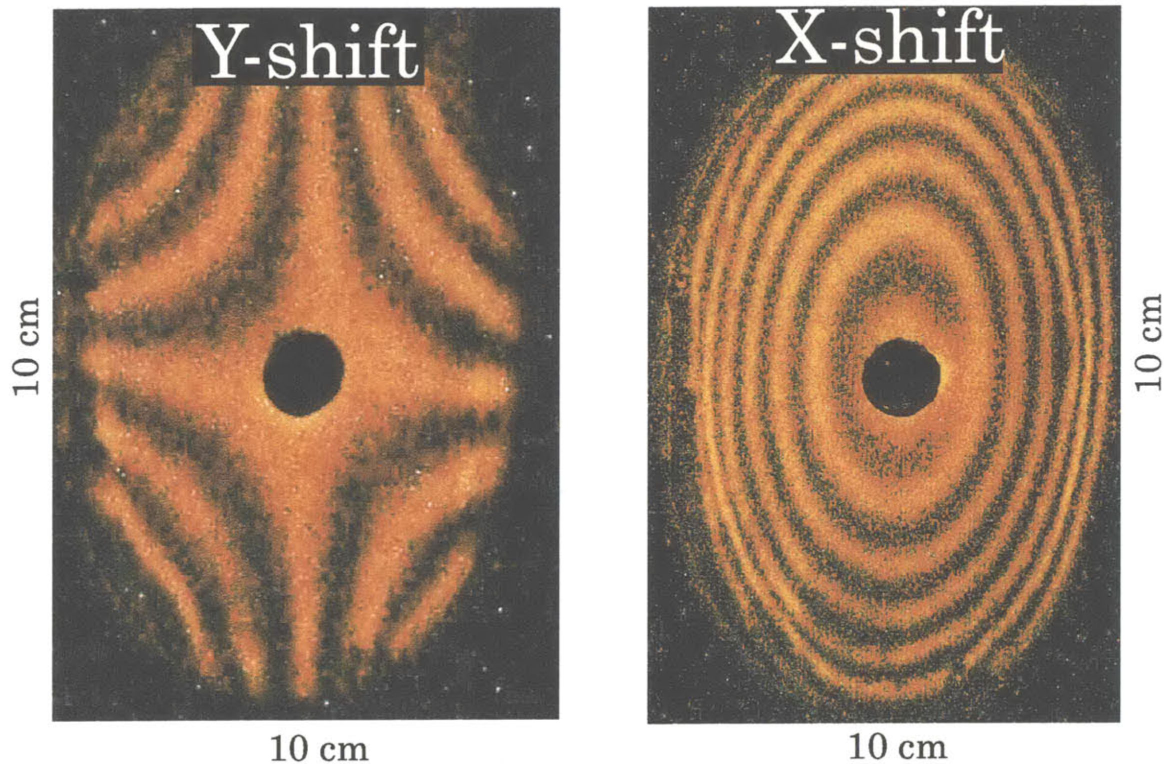


Figure 3.12 Moire images captured with a digital camera verify the hyperbolic phase progression of a grating recorded using spherical beam interference. The y-shift is ~1 cm and the x-shift is ~8mm.

Thus, for a known translation of the two gratings relative to one another, the moiré pattern can be interpreted to reveal the spatial dependence of the gratings. The moiré images shown above in Figure 3.12 display the classic patterns associated with hyperbolic-phase gratings under relative shifts in the x- and y-directions. The

hole in the center of the images is a consequence of the placement of the observation screen in front of the source locations, and is not part of the moiré.

A quantitative measurement of the grating periodicity could be obtained using a technique called holographic phase-shifting interferometry [20], although this has not been attempted as part of this work.

3.5 Correction of Hyperbolic Spatial-Phase

3.5.1 Plane-waves with lenses

In most applications, a grating with hyperbolic spatial-phase is undesirable. Fortunately, the magnitude of the period change is often below the level where it is a primary concern. However, it is almost always true that spatial-phase progressions which are closer to linear are more desirable. The most obvious method to linearize the spatial-phase progression of a grating exposed with the Mach-Zehnder interferometer is to planarize the interfering wavefronts. One way is simply an extension of the current method: use larger radius spherical beams in a region close to the optic axis. This strategy can quickly become impractical as the physical footprint of the optical setup becomes larger.

The use of lenses after the spatial-filter, as shown in Figure 3.13, is another method. Although seemingly desirable, the use of lenses can be detrimental in two respects. First, the lens diameter must be larger than the beam diameter to avoid scattering from the edges of the lens and its mounting hardware. The uniform area of the beam is necessarily limited to some fraction of the lens diameter. Thus, the collimated beam will inevitably have more intensity variation than the uncollimated beam. Although the phase progression in the exposed area will be much closer to linear, the dose-uniformity of the exposed area will be severely compromised. To equal the exposure uniformity of an uncollimated beam, lenses with a diameter on the order of the phase-radius would be necessary. In the current

configurations with a phase-radius ~1 meter, a collimating lens would be prohibitively expensive.

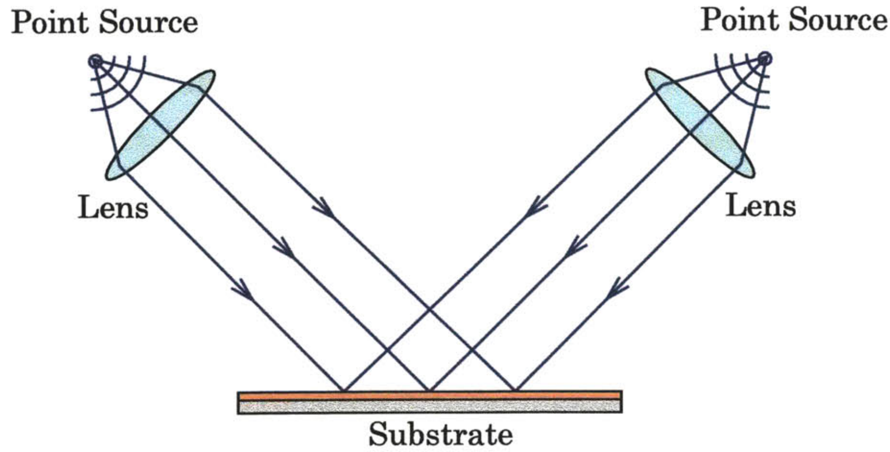


Figure 3.13 Use of collimating lenses in the Mach-Zehnder interferometer allows for exposure of linear-phase gratings. However, wavefronts are prone to distortion from scattering sources such as dust, lens hardware, etc.

If one is willing to accept a smaller exposure area in exchange for a more linear phase progression, there is still a question of the purity of the wavefront. When using uncollimated spherical-beams directly from the spatial-filter, the interfering wavefronts are free of noise and have a known phase profile. By placing a lens in the beam, this purity is lost. Errors in the lens figure and misalignments of the lens itself will lead to a beam which is close to planar, but with some unknown deviations. More importantly, scattering sites such as scratches or the inevitable dust particles on the lens will lead to unpredictable local and global deviations in the linearity of the exposed grating. Despite the drawback of producing gratings with hyperbolic spatial phase, the use of uncollimated spherical beams affords larger exposure area and easily characterized, noise-free exposures.

3.5.2 Substrate Curvature

So far, we have assumed that the ideal exposure surface, and hence the substrate, are planar. However, it was clear in the analysis of Section 3.3 that the position of the substrate plane relative to the interference pattern can have a great

effect on the recorded fringes on that substrate. One could view this as an invitation to alter the substrate plane in order to deliberately change the spatial phase of the grating. In the following section we will show that controlled curvature of the substrate plane can be used to linearize the spatial-phase of gratings made using spherical beam interference [21]. This method is attractive in that it retains all the advantages of the spherical beam, namely high exposure uniformity and noise-free interference, while greatly reducing the non-linear phase components.

3.6 Reducing non-linear spatial-phase via substrate curvature

3.6.1 Introduction

To understand how curvature of the substrate can be used to linearize a grating, let us briefly review the origins of the non-linear phase presented in Section 3.3. One can view the change in periodicity of the grating as arising from the combination of two effects, a change in interference angle θ, and a change in the inclination angle of the fringes. The nominal grating period is found at the center of the exposure field ($x=y=0$), where the interference angle is maximum. Additionally, along the entire y-axis, the interference fringes form normal to the substrate. As one moves away from the center, the interference angle will decrease. The period of the interference fringes, which is inversely proportional to $sin(\theta)$, will increase. Additionally, when the fringes are not normal to the surface of the substrate, the periodicity of the fringes recorded on the substrate will increase as seen in Equation 3.6.

Now consider the effect of increasing the height of the substrate at points away from the origin by bending the substrate as illustrated in Figure 3.14. First, for a given point $(x,y,0)$, increasing the substrate height z will increase the interference angle leading to a decrease in the fringe periodicity. Additionally, the normal to the substrate will now rotate to point towards the x-axis, which will

reduce the effects of the fringe inclination. Thus, the reduction in period associated with increasing z should be able to counter the increase in period associated with increasing x.

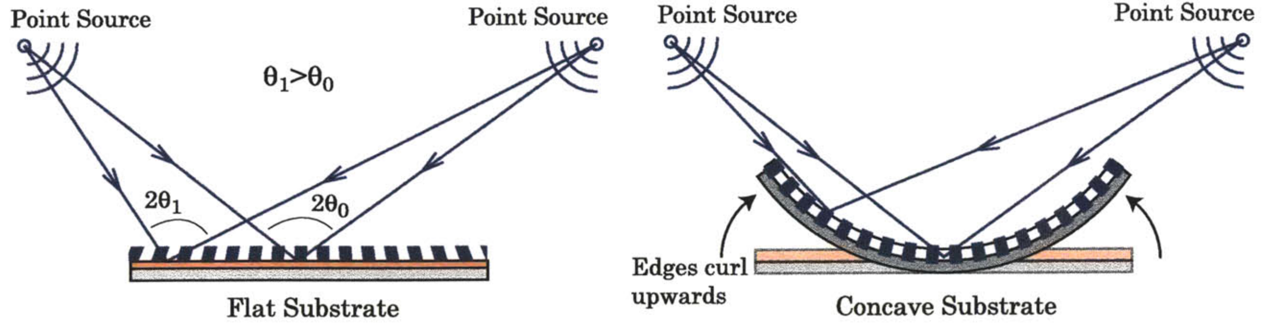


Figure 3.14 Spherical-beam interference on a flat substrate (left) leads to a position dependent fringe periodicity because of changes in the interference angle and an increasing fringe inclination angle. On a concave surface (right) the increase in substrate height compensates for the angle change while the change in slope compensates for the fringe inclination.

The actual fringe periodicity recorded on the substrate (P_s) at any point will be determined by a grating vector $\vec{k}_s$, which is the projection of interference fringe vector $\vec{k}_f$ onto substrate. In the analysis of the flat plane, we defined the projection as a scalar operation in terms of the angle θ_{inc} between the substrate and $\vec{k}_f$. This is an easy and intuitive method that worked because the substrate was known to lie in the XY plane. However, on an arbitrary 3-dimensional surface, defining θ_{inc} requires a careful definition of exactly where the inclination angle is measured from. For such surfaces, the more general vector projection operation is a safer choice.

In vector terms, surfaces are defined by the normal $\hat{n}$. To find the in-plane components of the interference fringe vector $\vec{k}_f$, we can remove the component parallel to the surface normal as seen in Equation 3.21.

$$\vec{k}_s = \vec{k}_f - (\vec{k}_f \cdot \hat{n})\hat{n} \qquad (3.21)$$

The normal vector to a surface at any point p is found using the normalized gradient of the surface function f evaluated at that point, shown in Equation 3.22.

A review of this and other vector manipulations can be found in any text on linear algebra, Reference [22] is one example.

$$\hat{n}(p) = \frac{\nabla f|_p}{\left|\nabla f|_p\right|} \tag{3.22}$$

3.6.2 Implementing substrate curvature

Qualitatively, it seems likely that a curved exposure plane can be used to correct non-linear changes in grating periodicity. However, the entire lithography process flow - beginning with the acquisition of substrates to the function of the device - is designed for flat substrates. Thus, permanent curvature of the substrate is unacceptable. If a planar substrate can be elastically deformed only during the exposure step, then the rest of the process flow can remain unperturbed. The most common substrates used in our laboratory are 10cm-diameter single-crystal silicon wafers. Although quite stiff, the thickness of the wafers (~500 μm) compared to their diameter (10 cm) suggests that they can be temporarily deformed without excessive force. Additionally, the crystalline structure suggests that any deformation will be completely elastic until the wafer breaks.

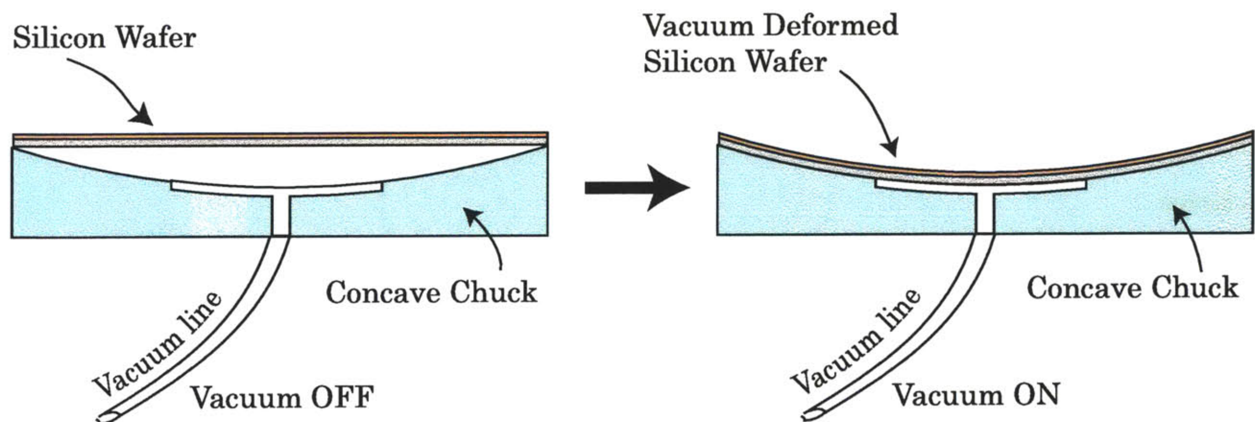


Figure 3.15 Exaggerated depiction of vacuum chucking a silicon wafer onto a concave surface. The actual maximum displacement at the center of a 10 cm wafer is ~1mm.

Temporary substrate curvature can be achieved through the use of a concave vacuum chuck, as illustrated in Figure 3.15. A flat wafer is forced to conform to the curvature of the chuck while vacuum is held, and upon release returns to its original planar shape. Arbitrary surface curvatures are available via diamond-turning for reasonable cost, and many spherical and elliptical shapes are available as off-the-shelf components. Compared to the use of lenses for linearization of spatial-phase, this scheme requires only one item (the chuck) instead of two lenses. Material concerns (uniformity, absorption, etc.) are eliminated as the chuck is not actually a part of the optical path. Furthermore, where lenses much larger than the substrate would be required to achieve high uniformity, the chuck only need to be as large as the substrate itself. Perhaps most importantly, since the chuck is not part of the optical path, this method introduces no additional noise or error into the wavefronts, which remain clean and spherical. Lenses will inevitably have scattering noise from defects and mounting hardware which degrades the fidelity of the exposed grating. The conversion of the interferometer is also very simple; the flat wafer chuck is simply swapped for a concave one. Thus, for very little investment, one can convert an existing IL system to a corrected one. The only concern is exactly what surface curvature is necessary, and whether or not the silicon wafer can handle the additional stress of bending.

3.6.3 Bending stresses in single-crystal silicon wafers

Any advantages of temporarily deforming a silicon wafer during the exposure would obviously be lost if the wafer were to break. To prevent this from occurring, we can develop some guidelines for the acceptable levels of curvature based on a few simple assumptions. We assume that an initially flat substrate of thickness $T=2t$ and diameter D will be deformed into a spherical segment with radius r_d. We also assume that the neutral surface of the deformed wafer is along its mid-line - i.e. when deformed, the middle plane of the substrate remains unstrained, while the top

of the substrate is compressed and the bottom of the substrate is elongated. Tensile stress (along the bottom), rather than the compressive stress (along the top of the wafer), is generally what causes failure in brittle materials such as silicon. We also assume that elastic deformation will occur until the point of catastrophic failure.

In this simple case, we have a condition where the maximum tensile strain is uniform along the lower surface of the wafer due to the spherical deformation. The standard definitions of stress σ (Equation 3.23) and strain e (Equation 3.24) are available in any text on solid mechanics, for example [23].

$$\sigma = Ee \tag{3.23}$$

$$e = \frac{\Delta s}{s_0} \tag{3.24}$$

The strain e is defined as the unitless ratio of the change in length in the deformed state, Δs, to the original length s_0. Due to the symmetry of this simple case, we can reduce the problem to that of a beam bent into a circular segment. The length s of a circular arc is the subtended angle φ multiplied by the radius.

$$s = r_d \varphi \tag{3.25}$$

Thus, the increase in the length of the arc for the elongated bottom-surface (Δs) of the wafer, where tensile strain will be at its peak, will simply be proportional to the change in radius, which is equal to t for a wafer of thickness T=2t. Therefore,

$$e_{\max} = \frac{t}{r_d}. \tag{3.26}$$

The stress along the outside of the wafer is then calculated using Equation 3.23, although it must be modified using Poisson's ratio ν to account for the equal, but orthogonal strain component in the y direction.

$$\sigma_{\max} = \frac{Et}{r_d(1-\nu)} \tag{3.27}$$

If we include a slight margin of safety, and set the maximum stress value equal to the fracture stress of single-crystal silicon, we can solve for the minimum allowable radius of curvature as a function of the thickness of the wafer T.

$$r_{\min} = \frac{2tE}{\sigma_{failure}} = \frac{TE}{\sigma_{failure}} \tag{3.28}$$

Thus, the maximum radius of curvature that can be induced is directly proportional to the wafer thickness by a value determined by the material properties. Common properties for a single-crystal silicon wafer [24] are given in Table 3.1.

Table 3.1

Property	*Symbol*	*Value*
Youngs Modulus	E	170 GPa
Poissons ratio	ν	0.22
Fracture stress	$\sigma_{failure}$	100-1200 MPa
Thickness	T	500 μm
Diameter	D	10 cm

The huge range of fracture-stress values presented in the literature is likely not caused by material variations, but by variations in surface finish. Surface irregularities can cause stress concentrations which, in a brittle material, will cause failure rather than local deformation. Using the values given above, the minimum radius of curvature for a 10cm-diameter silicon wafer can be calculated, and is given below.

$$r_{min} = 7 - 85cm \qquad (3.29)$$

Unfortunately, this range is rather large and prevents us from making an accurate prediction on the tolerable bending radius of a wafer. Intuitively, one would expect that the curvature necessary for spatial phase correction would be of the same order as the phase-radius of the spherical beams. As our interference lithography systems use beam with ~1m phase-radius - greater than the worst-case radius of 85cm - it is unlikely that this level of curvature would cause the wafer to crack. However, it is always possible reduce stress by using a thinner substrate, single crystal silicon substrates are readily available as thin as 100 μm. We can make the general observation that curvatures significantly below 1m radius may be problematic, while curvatures on the order of 1m or greater will likely be successful.

3.6.4 Simplification via Symmetry

Qualitatively, we expect that exposure along a concave surface will reduce the variation in periodicity. It remains, however, to show mathematically that this is the case. Analysis of this problem entails solving for the periodicity *P(x,y,z)* along the surface defined by *Z(x,y)*. For a known surface *Z*, this is a simple process using the technique described in Section 3.3.

1) Solve the k-vectors of the interfering beams $\vec{k}_1$ and $\vec{k}_2$.
2) Subtract $\vec{k}_1$ from $\vec{k}_2$ the to obtain the fringe vector $\vec{k}_f$.
3) Determine the surface normal $\hat{n}$ using the gradient function.
4) Project $\vec{k}_f$ onto the wafer surface to get $\vec{k}_s$.
5) Convert the surface vector $\vec{k}_s$ into grating period.

The current situation, however, is one where the surface curvature is precisely what we want to define. Thus, the approach would be to work backwards by defining the desired periodicity function *P(x,y)*, converting that to the vector $\vec{k}_s$, and then solving for the surface. Analytically, this is an extremely cumbersome approach. For a given desired periodicity, there is no guarantee that a solution even exists.

To reduce the complexity of the problem, we can exploit the symmetry of the interferometer to impose certain constraints on the surface function. Recall that we are assuming the area of exposure is significantly smaller than the phase-radius of the beam, which allows for a highly accurate spherical wave approximation. We have defined our coordinate axis such that the interference pattern is symmetric about the $x=0$ and $y=0$ planes. Less obvious, there is also a rotational axis of symmetry around the line *AB* connecting point sources *A* and *B*. That is to say, for a constant radius from this line, the vector $\vec{k}_f$ will be constant. For surfaces that are rotationally symmetric about the *AB* axis, the recorded interference pattern described by $\vec{k}_s$ will also be constant along the y-direction. This simplifies the task of optimizing our surfaces significantly.

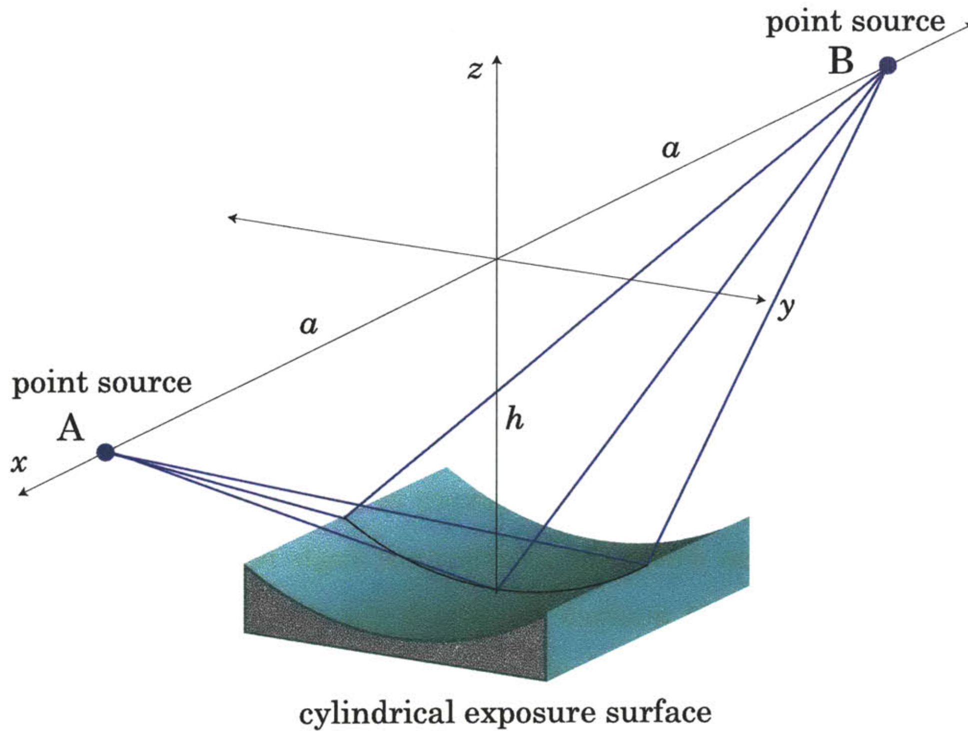


Figure 3.16 A cylindrical exposure surface with circular curvature of radius h in the YZ plane. Surfaces which are rotationally symmetric about the AB axis will ensure constant periodicity in the y direction.

Simply put, circular curvature in the YZ plane ensures that the periodicity of the interference pattern is constant in the y direction. For example, let the exposure surface be a cylindrical section with radius h. as seen in Figure 3.16. The periodicity on the wafer, once it was returned to its planar state, would consist of perfectly parallel lines whose period varied only in the x-direction. The period $P(x,y)$ would be the same as that of the flat surface exposure measured along the x axis, $P(x,0)$. Figure 3.17 shows this effect for an exposure surface that also corrects in the x-direction.

The power of this simple result is two-fold. First, it allows us to *fully correct* variations in periodicity in the y-direction. Second, we can reduce the complexity of the surface solution from 3-dimensional to 2-dimensional. We only need to solve for the $Z(x,0)$ curve, and then rotate it about the AB axis. For a given curve $z(x)$, we can describe the entire 3-dimensional surface $Z(x,y)$ in terms of the circular cross-sections.

$$Z(x,y) = h - \sqrt{(h - z(x))^2 - y^2} \quad (3.30)$$

Recall that h is the z-coordinate of the point sources,

3.6.5 The spherical surface

By imposing rotational symmetry on the exposure surface, we are left with a surface that is truly only a function of the x-dimension, which simplifies the analysis greatly. Recall from Section 3.3 the two effects which act to change the periodicity of the grating: a change in interference angle, and a change in the fringe inclination. We can highlight the importance of considering the fringe inclination by solving for a surface on which the interference angle remains constant. In vector terms, we are solving for a surface for which the magnitude of the interference fringe vector, $|\vec{k}_f|$, remains constant, and equal to its value at the origin.

$$|\vec{k}_f| = \frac{4\pi \sin(\theta_0)}{\lambda} \quad (3.31)$$

Obviously, the condition on the interference angle is that it remains equal to the nominal interference angle at the origin.

$$2\theta = 2\theta_0 \quad (3.32)$$

In this case, it is algebraically simpler to solve for a constant interference angle than for the constant vector magnitude. The interference angle 2θ can be described as the sum of the two components θ_m and θ_p, the incident angles of the two rays. We can rewrite Equation 3.32 in terms of the new angles,

$$2\theta_0 = \theta_p + \theta_m. \quad (3.33)$$

We define the angles $\theta_0, \theta_p, \theta_m$ as

$$\theta_0 = a\tan\left(\frac{a}{h}\right) \quad \theta_p = a\tan\left(\frac{a-x}{h-z}\right) \quad \theta_m = a\tan\left(\frac{a+x}{h-z}\right) \quad (3.34)$$

If we take the tangent of both sides of Equation 3.33 and apply some trigonometric identities, we can get

$$\frac{2\tan(\theta_0)}{1-\tan^2(\theta_0)} = \frac{\tan(\theta_p)+\tan(\theta_m)}{1-\tan(\theta_p)\tan(\theta_m)}. \quad (3.35)$$

Substituting in the identities of Equation 3.34, the *tan(atan(~))* pairs can be cancelled, yielding

$$\frac{2ah}{h^2 - a^2} = \frac{2a(h-z)}{(h-z)^2 - a^2 + x^2}. \tag{3.36}$$

Canceling the common factors, introducing the dummy variable $z'=(h-z)$, and cross multiplying Equation 3.36, we get

$$h\left(z'^2 + x^2 - a^2\right) = z'\left(h^2 - a^2\right). \tag{3.37}$$

Defining a new term b to simplify the notation, we can complete the square on the left-hand side of Equation 3.37.

$$b = \frac{h^2 - a^2}{2h}. \tag{3.38}$$

$$(z'-b)^2 + x^2 = a^2 + b^2. \tag{3.39}$$

The equation above has the form of a circle with radius $r^2=a^2+b^2$ and centered about the point $(x,z)=(0,b)$. However, we would like to rearrange this in terms of the true variable z rather than the dummy variable z'. Additionally, we can simplify the right-hand term by expanding b, such that we can now define the radius of curvature along the x-axis, r_{cx}, in terms of the either the phase-radius of the beam or the coordinates of the point sources.

$$(z-(h-b))^2 + x^2 = r_{cx}^{\ 2}. \tag{3.40}$$

$$r_{cx} = \frac{h^2 + a^2}{2h} = \frac{r_p^{\ 2}}{2h} \tag{3.41}$$

The circular form in Equation 3.39 is unchanged, as is the radius, but the center of the circle is now at the point $[x,z]=[0,(h-b)]$. It is interesting to note how the phase-radius of the beam relates to the radius of curvature of the surface. The ratio of the two can be found by simply from the above equation.

$$\frac{r_{cx}}{r_p} = \frac{r_p}{2h} \tag{3.42}$$

We can solve for the values of a and h such that the two radii are equal by squaring the right-hand side.

$$r_{cx} \geq r_p \;\text{ if }\; a \geq \sqrt{3}\,h \qquad (3.43)$$

Thus, for nominal incident angles *(θ_0)* greater than 60°, the radius of curvature is larger than the phase-radius of the beam. More importantly, perhaps, is the comparison between the surface curvature in the x-direction and the curvature in the y-direction. Recall that the radius in the y-direction is approximately equal to h. The lesser of the x and y radii of curvature will be the one which determines whether or not the wafer breaks when deformed onto the curved wafer chuck.

$$r_{cx} \leq r_{cy} \;\text{ if }\; a \leq h \qquad (3.44)$$

For nominal incident angles (θ_0) less than 45°, the x-direction radius will be less than h, trending towards a minimum at very large periods when $a=0$.

$$r_{\min} = \frac{h}{2} \qquad (3.45)$$

The general result is a section of a football-shaped surface-of-revolution. The interference angle at any point along this surface is guaranteed to be equal to the nominal value of θ_0 at the origin. Interestingly, exactly at θ_0=45°, the surface will be spherical, which provides an easy opportunity to test the theory. High-quality sections of spherical surfaces are readily and inexpensively available as off-the-shelf components in the form of plano-concave lenses or spherical front-surface mirrors. For all other cases, it would be necessary to have a surface with the appropriate curvature fabricated.

Based on the easy availability of spherical concave surfaces for experimental testing described in the next section, we focus now on the particular case of a λ_0=325 nm HeCd laser beam incident at a nominal angle of 45°. The nominal period in this case is P_0=229.8 nm. The position of the point-sources, *[x,y,z]=[+-0.902,0, 0.902]* meters, was set based on the availability of a 3" diameter spherical surface mirror with a radius of 0.902m [25]. Using the algorithm summarized at the beginning of this section, we can now solve for the periodicity of the grating exposed on this spherical surface. Calculations of periodicity on the wafer after it has been returned to its original planar shape are shown in Figure 3.17.

The change in periodicity is clearly reduced from the flat surface exposure, also shown in Figure 3.17. As predicted, the circular curvature of the y-directions has eliminated variations in this direction, yielding grating lines that are now perfectly parallel. The x-direction, persists with a residual, albeit reduced, periodicity variation. This is due to the fact that the spherical surface only compensates for changes in interference angle over the exposure plane, but does *not compensate for the fringe inclination angle.* Although the inclination angles are reduced due to the concave nature of the surface, they are still present. This is an important confirmation that the earlier methods of analysis of grating distortion described in Section 3.3 are flawed. Prior analyses asserted that the constant interference angle on the spherical surface would result in a perfect linear-phase grating.

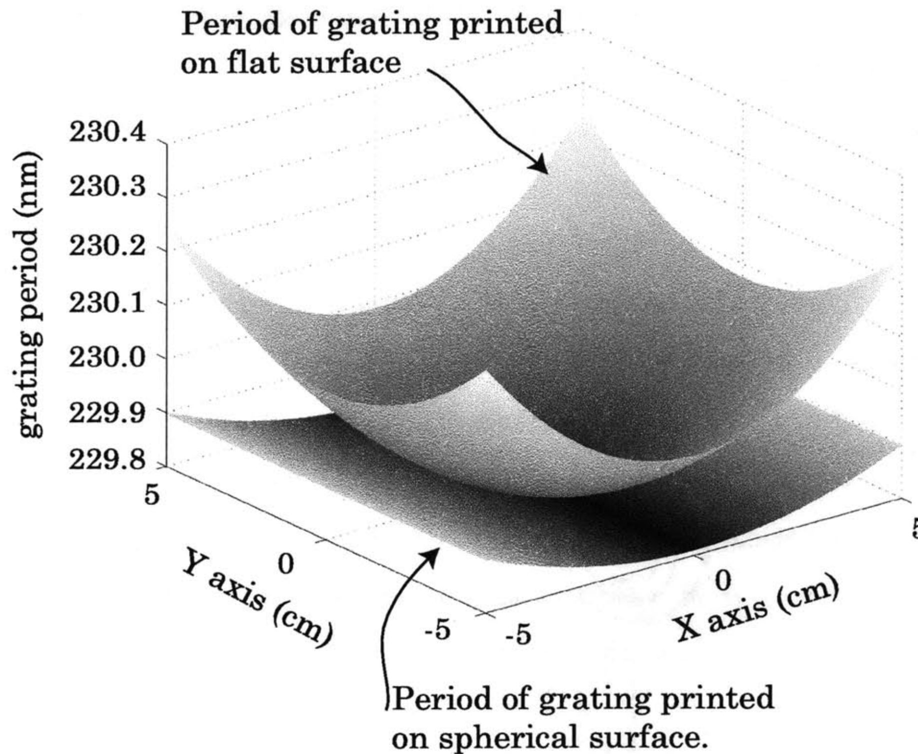


Figure 3.17 Superimposed plots of the grating period printed over a 10cmx10cm area. The upper plot represents a grating with hyperbolic spatial-phase exposed on a flat substrate. The lower plot represents a grating exposed on a spherical concave surface. Notice that the corrected grating shows no variation in the y-direction and a reduced variation in the x-direction.

3.7 Demonstration of spherical-surface correction

3.7.1 Procedure

Careful observation of the scale on Figure 3.17 reveals that the periodicity of the corrected grating only changes by about 1 Ångstrom over the entire wafer. These extremely small variations, while imperceptible to most observations, should be easily detectable using the moiré technique described previously. To experimentally verify the corrections, a spherical front-surface mirror of radius 0.902 meters was converted to a vacuum chuck and mounted in the IL setup shown in Figure 3.11. In rough detail, the procedure used is as follows:

1) A flat Si wafer is deformed under vacuum onto the spherical concave wafer chuck, and a grating is exposed.
2) The wafer is removed from the system and the photoresist image is developed to form a physical grating on the wafer surface.
3) The wafer is re-mounted in the IL system on a *flat* wafer chuck.
4) The moiré pattern between the photoresist grating and the interference fringes is observed and captured using a digital camera.

It should be noted that standard 10cm-diameter, 500μm thickness single-crystal silicon wafers were able to bend and conform to the shape of the chuck without breaking. The worst-case fracture radius (0.85 m) in Section was only marginally different than the 0.9 m radius used in these experiments, so experimental verification of the elasticity of the wafers was a necessary step.

3.7.2 Alignment

To ensure that the observed moiré images can be interpreted properly, the wafer must be placed in exactly the same position when it is re-mounted after development as it is when the initial exposure occurred. This process is complicated

substantially by the fact that the concave wafer chuck must be swapped for the flat wafer chuck between these two steps. Moiré techniques used in the past have enabled the re-alignment of the wafer by assuming that the wafer chuck was stationary between the two steps. Obviously this is not the case, and so the alignment references must be placed on the wafer itself. The alignment procedures described at the beginning of this chapter (Section 3.2) were developed with this problem in mind.

The HeNe reference beam determines the normal to the wafer surface (z-axis) as well as its center, and is used to align the wafer on both chucks. However, it has been shown that the most sensitive parameter, by far, is rotation of the wafer about the z-axis. To determine the angular orientation of the wafer independent of the wafer chuck, a diffractive alignment scheme is used. Prior to the initial exposure, a small section of the center of the wafer is etched with a diffraction grating whose period is large enough to diffract the HeNe reference beam, in this case P=4μm. This wafer is then cleaned thoroughly and spun with a photoresist stack for exposure on the concave surface. Using the reference beam to illuminate this alignment grating once the wafer is mounted in the system, the diffracted orders can be observed and their positions marked. If the wafer is removed for any reason, including when the wafer chucks are swapped, the same orientation can be achieved when the wafer is remounted by simply lining up the diffraction orders from the reference beam with their prior positions. Once the wafer is returned to the appropriate position and orientation, the moiré images can be observed between the resist grating exposed on the spherical surface and the hyperbolic phase intensity pattern.

3.7.3 Zero relative shift

In the common situation where the moiré pattern to be observed is a result of a single pattern interfering with itself, there must be a relative shift between the two patterns to observe an image. When two pattern images are perfectly aligned,

no information is obtained. However, in this case, the two overlapping patterns are different, so we expect to see a characteristic moiré pattern even when there is no relative shift between the two patterns. The expected moiré patterns for any given situation are easy to calculate, in that we can analytically describe both the expected pattern in the photoresist and the known pattern of the intensity image of the interfering beams.

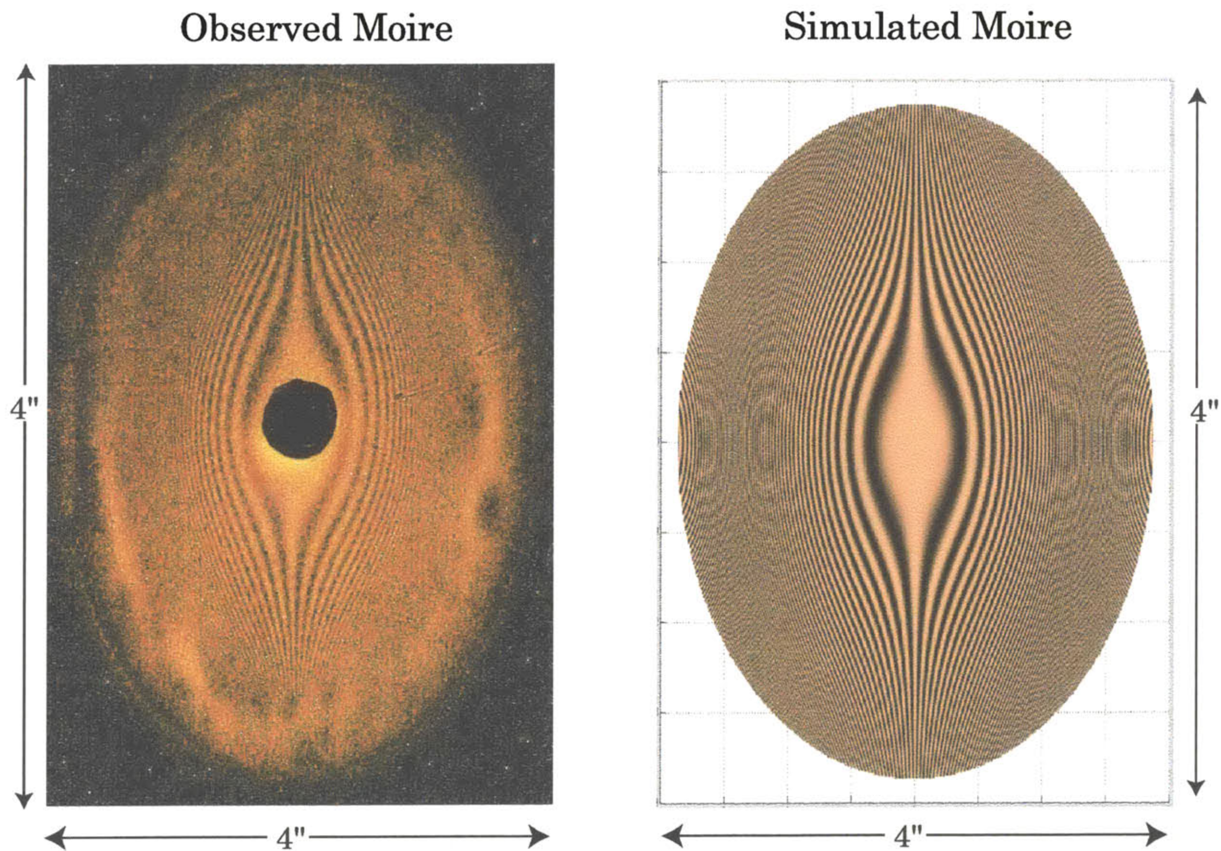


Figure 3.18 Observed and simulated moiré images of the corrected spherical surface grating and the hyperbolic-phase flat-surface grating with no relative shift. The faint elliptical patterns seen along the x-axis of the simulation are artifacts from the image pixellation.

The results shown in Figure 3.12, which prove that the interference fringes on a flat substrate are indeed hyperbolic in their spatial phase progression, are critical to interpreting the results in this section. Thus, we can approach the moiré experiments with knowledge of what the pattern *should* be based on simulations and compare with the actual results.

The most distinctive feature of the interference fringes formed on the spherical surface (Figure 3.17 shows the periodicity of the two overlapping gratings) is that they have been fully corrected in the y-direction. That is, the grating lines should be perfectly straight and parallel, with the period variation only occurring in the x-direction. The variation that does occur in the x-direction is significantly less than the uncorrected case. From these attributes, we can qualitatively describe the moiré pattern that will form between the corrected and uncorrected gratings. In the center of the wafer, where the periodicity of the two gratings is closest, the moiré fringes will be wide. Moving away from the center in the x-direction, the increasing difference in periodicity will cause the fringes to become narrower. This progression will be slowest along the x-axis, and occur more rapidly at greater y-direction distances. Figure 3.18 shows both the simulation and actual moiré images of this situation. One immediately sees that the two images are almost identical, indicating that the spherical surface correction most likely was successful. One obvious difference between the real and simulated moiré images is the presence of the circular hole in the real moiré images. The hole is a necessary consequence of the position of the observation screen at the location of the point sources (Figure 3.11) that allows the forward propagating beam to pass through the screen. However, the quality of the moiré images with zero shift alone does not guarantee that the important structures of the corrected grating have been achieved. For this, shifts in the x-and y-direction are needed.

3.7.4 X-direction shift

The expected results for grating shift in the x-direction should be characteristically similar to the unshifted version, but will lose symmetry. Because the grating lines are straight, but with non-uniform spacing which is symmetrical about the center of the pattern, one side of the pattern will be more closely matched than the un-shifted case, and one less closely matched. This leads to the "bulging" of the fringes seen on the left side of the moirés below, where the periods are more closely matched. On the right side of the images, there is a greater difference in period than the unshifted case and the fringes are "flatter".

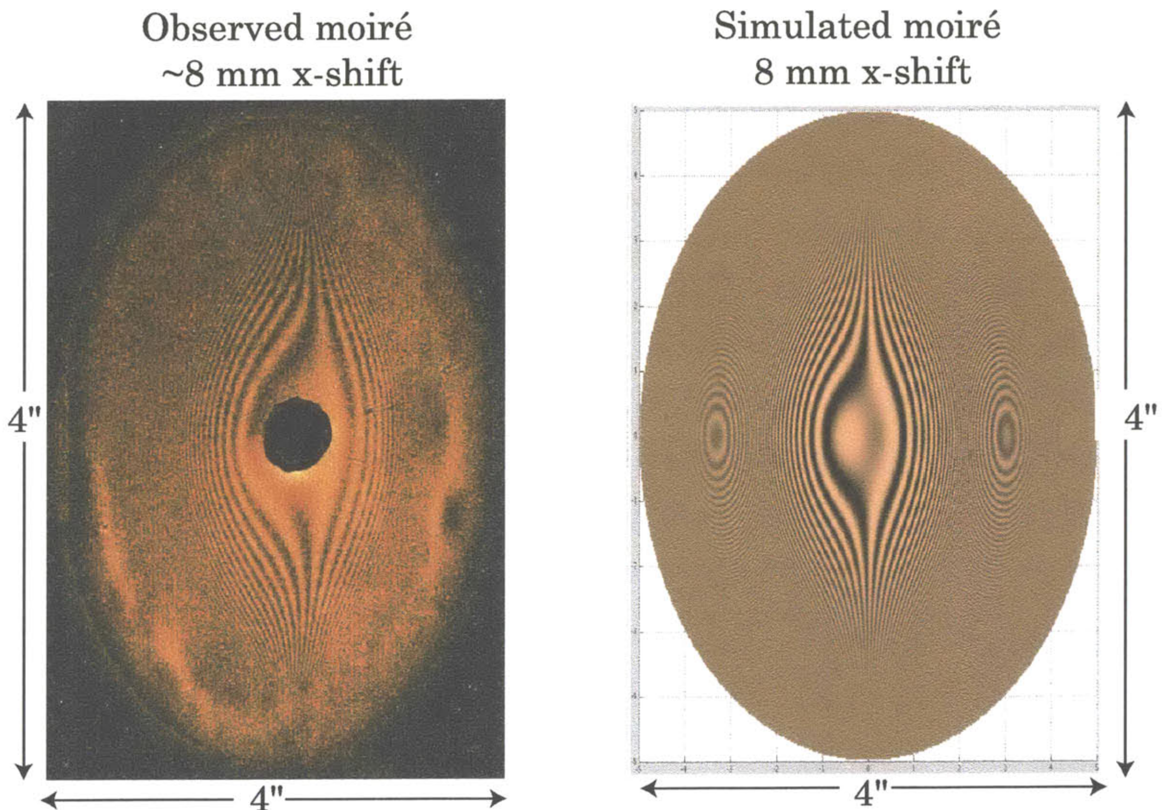


Figure 3.19 **X-Shift** Observed and simulated moiré images of the corrected spherical surface grating and the hyperbolic-phase flat-surface grating with a relative shift of 8 mm in the x-direction. Note the bulging of the fringes on the left-hand side and the flattening on the right-hand side. This is characteristic of a slow nonlinear phase progression.

3.7.5 Y-direction shift

The moiré image between the corrected and uncorrected gratings will be invariant to shifts in the y-direction if the y-direction peiod variation has indeed been eliminated as predicted. In the observed and simulated patterns shown Figure 3.20, the pattern appears to have shifted "down" due to the motion of the substrate itself, not the position of the moiré image. The excellent agreement of the observed and simulated moirés below, and with the unshifted case shown in Figure 3.18 is convincing evidence that any y-direction variation in the periodicity is negligible.

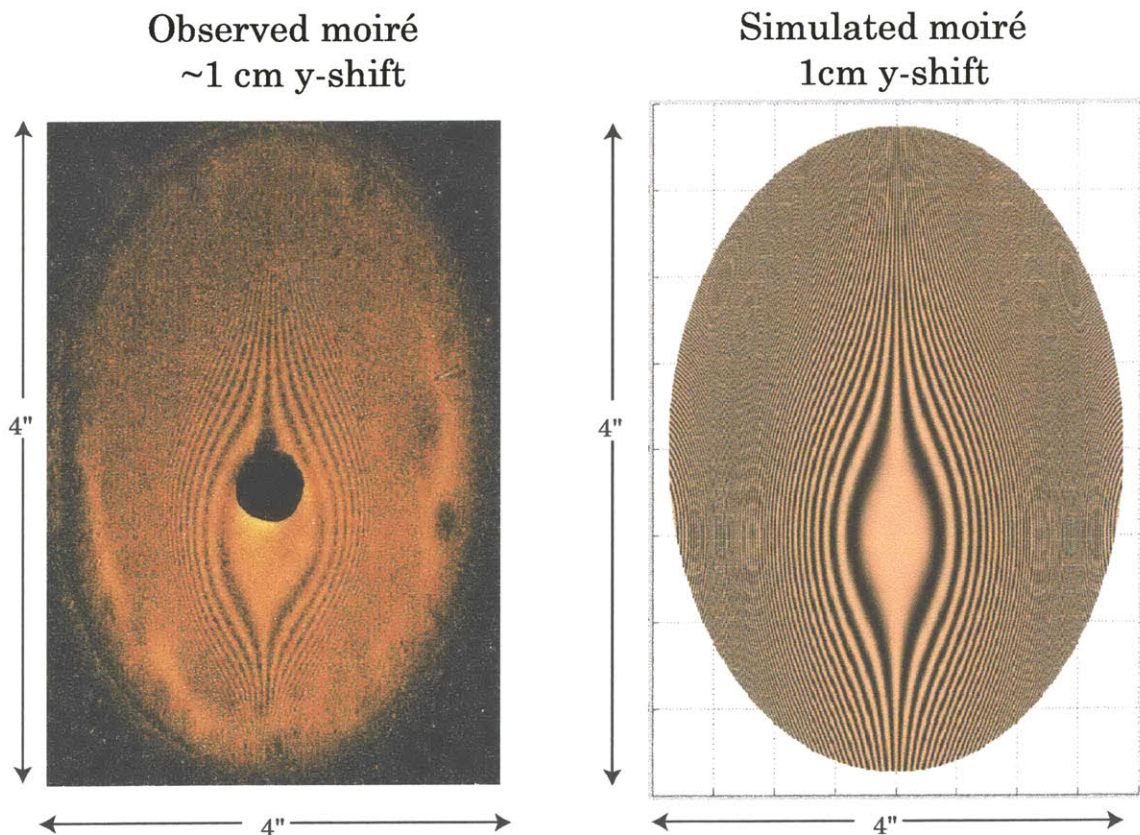


Figure 3.20 **Y-Shift** Observed and simulated moiré images of the corrected spherical surface grating and the hyperbolic-phase flat-surface grating with a relative shift of 1 cm in the y-direction. The moiré pattern should be invariant to y-direction shifts as the spherical surface grating is perfectly corrected in this direction. Compared with the unshifted image in Figure 3.18, the observed moiré above left is in excellent agreement.

In contrast to these patterns are the moirés characteristic of two hyperbolic-phase gratings with a y-direction relative shift shown in Figure 3.12. The hyperbolic image structure characteristic of that case is clearly and significantly different than the one shown below, and would be immediately apparent had the spherical-surface correction not been successful.

3.8 Surface Optimization

3.8.1 The spherical surface

The results of the previous section demonstrate the high level of spatial-phase correction that can be attained with minimal effort by switching from a planar exposure to a spherical concave exposure. No other modifications are necessary - the total cost of the conversion is a few hundred dollars for the purchase of an off-the-shelf spherical mirror blank. However, it should be emphasized that although the spherical surface is the *simplest*, it is not necessarily the *best* surface. In fact, if we consider for a moment the effect of the spherical surface exposure, the nature of the *best* surface becomes more elusive. Recall from Section 3.6.5 that the spherical/elliptical surface was one of constant interference angle, although the fringe inclination was not fully corrected. We can infer from this that a surface on which the fringes all stood normal to the substrate would *not* have a constant interference angle. Similarly, if a surface could be optimized such that the fringe periodicity remained constant, neither the interference angle nor the fringe inclination would individually be optimized. An optimal surface must therefore balance the effects of an increasing interference angle (which decreases the periodicity) with the fringe inclination effect (which effectively increases the periodicity). Unfortunately, there are no guarantees that such a surface exists. Nevertheless, we can likely improve the level of correction through a more complex surface geometry.

3.8.2 Polynomial surface description

The task of optimizing a set of variables for the ideal solution is a problem which is particularly well suited to numerical calculation. This particular problem is made significantly easier because of the geometrical symmetry described in Section 3.6.4. Only the curvature along the x-axis need be optimized, and the rest of the surface is defined via rotation about the AB axis connecting the two point sources (Figure 3.16). A simple surface description for this purpose is the polynomial. A variety of polynomial constructions exist, such as the Taylor or Maclaurin series, which allow any arbitrary shape to be written as a polynomial. For smoothly varying curves such as those necessary for this application, the number of useful terms should be tractably low. A general polynomial description of the curve $z(x)$ is shown below.

$$z(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \ldots \qquad (3.46)$$

At the same time, we must be aware of the practical limitations associated with the deformation of the wafer. For example, in a polynomial surface description the zero and first order terms should be set to zero. The zero-order term in a description of $z(x)$ will effectively change the nominal periodicity of the grating, not change its variation over the wafer. The first-order term, despite any potential value it may have, is equivalent to a rotation of the x and z axes about the y-axis, which will lead to an asymmetry in the exposed grating on opposite sides of the y-axis. If one considers the absolute value of the first-order term, which would eliminate the asymmetry, then the surface has a slope discontinuity at the origin which would be impossible for the wafer to conform to. Thus, the first-order term is undesirable from a practical perspective. Higher order terms should approach the origin smoothly enough for conformal deformation of the substrate, although any odd-order terms would necessarily require the absolute value to preserve symmetry. A general polynomial description of the exposure surface, including the above conditions, is given below in Equation 3.47.

$$z(x) = \left| c_2 x^2 + c_3 x^3 + c_4 x^4 \ldots \right| \qquad (3.47)$$

3.8.3 Optimization via the simplex method

Using the above equation, we can numerically solve for the optimal values of the coefficients c_i for a given polynomial order. The downhill simplex optimization algorithm was used for this process [26,27]. This is an iterative algorithm which begins at a set of initial, "best guess" values for the coefficients. The degree of optimization is determined by a single figure-of-merit, usually corresponding to an error from the optimal value. Thus, optimization is actually defined as the minimization of this error term. The coefficients are modified in each iteration until either an optimal set is reached, or the change in the error term is below a certain bound for each iteration. For a small number of variables (i.e. low polynomial order), this method is easier to apply than more sophisticated algorithms such as the method-of-steepest-descent or the quasi-Newton method [27]. Results of the surface optimization for second and third-order polynomials are shown in Figure 3.21. Also shown are the results for the flat and spherical surfaces. Notice that both of the optimized surfaces reduce the periodicity variation more than the spherical surface. The third-order surface shown reduces the period variation by about another factor of two beyond that of the spherical surface. It is not necessarily the case, however, that increasing the polynomial order will continue to correct the surface in a useful way.

Table 3.2

Polynomial	c_2	c_3	c_4	c_5	c_6
2nd order	0.414				
3rd order	0.569	-3.37			
4th order	0.551	-0.007	-56.04		
5th order	0.551	0.006	-57.02	2.778	
6th order	0.551	0	-56.78	-0.031	6.3

Polynomial coefficients for increasing order are shown above in Table 3.2. The actual values of the coefficients are not important, as they will depend on the geometry of a particular system. The relevant factor is that we can see a trend towards a specific, even-ordered curve. The second-order and third-order surfaces are shown below, but the fourth and higher-order surfaces are not shown as they overlap the third-order too closely. We can infer from this that there is a specific, optimal, curve for which the polynomial is only an approximation. From these results, since a polynomial description can be written for *any* function, we can infer that a surface is not likely to exist on which a linear grating is exposed.

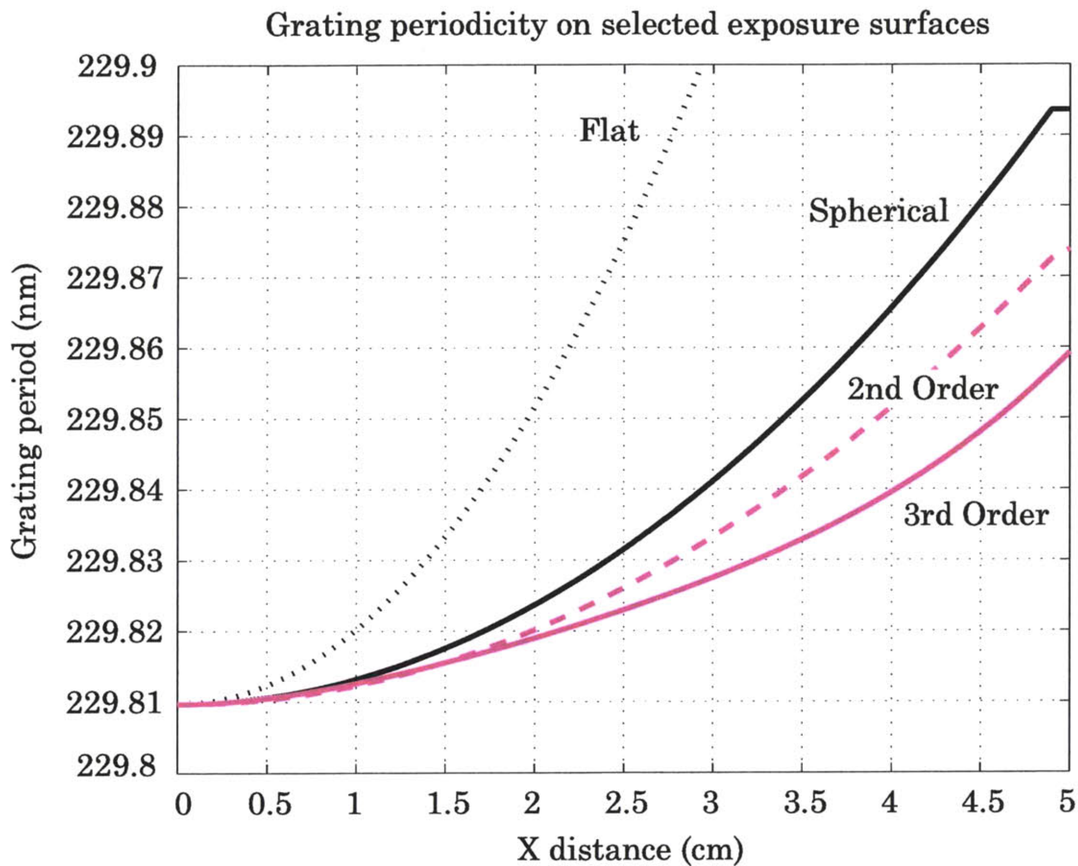


Figure 3.21 Comparison of the grating period exposed on numerically optimized surfaces to the spherical and flat surfaces. The optimized third-order surface shows an improvement of almost a factor of two over the spherical surface.

3.8.4 Hyperbolic surfaces

One can notice from the progression of the polynomial coefficients, given in Table 3.2, that as more orders are included the function behaves as the approximation to a certain "optimal" function. From the presence of only even terms in the series, as well as from the necessary symmetry of the problem, we know that this will be an even function. An educated guess can narrow the field of possible functions, and allow us to compare the Maclaurin series coefficients for a few known even functions to the emerging "optimal" polynomial coefficients. Based on this, we can identify the optimal surface curvature in the x-direction as hyperbolic. This is a satisfying result, given that the progression of the spatial-phase which we are attempting to correct is also hyperbolic. Further work will be necessary to determine if this is simply a coincidence, or is indicative of a deeper relation between the phase progression and the exposure surface. Equation 3.48 gives the general equation of a 2-dimensional hyperbola. Equation 3.49 introduces a shift in the z-coordinate which shifts the minima to $z=0$ at $x=0$, and solves for z. Thus, the optimization parameters are reduced to two, the coefficients a and b, from the arbitrarily large number in the polynomial solution.

$$\frac{z^2}{b^2} - \frac{x^2}{a^2} = 1 \tag{3.48}$$

$$z = \sqrt{b^2 + \frac{b^2}{a^2}x^2} - b \tag{3.49}$$

Recall that the algorithm solves a minimization problem for a scalar function of many variables. Therefore, we must define our problem in such a way that a single number represents the degree of optimization. In this case, we want the grating period to be constant over the exposure surface. Thus, an obvious parameter to minimize is the difference between the actual grating period and the nominal grating period. The important question, however, is how to reduce this error function, which is spatially dependent, to a scalar value. In the previous section on polynomials, we used the sum error of a discrete number of sample

points. This summation is essentially a crude method for evaluating the total non-linear phase accumulation across the substrate. Another method would be to minimize the peak-to-peak variation of the grating period. In fact, the criteria used to determine this error term significantly changes the results of the optimization. Choosing an appropriate minimization criteria is perhaps the most difficult part of implementing numerical optimization.

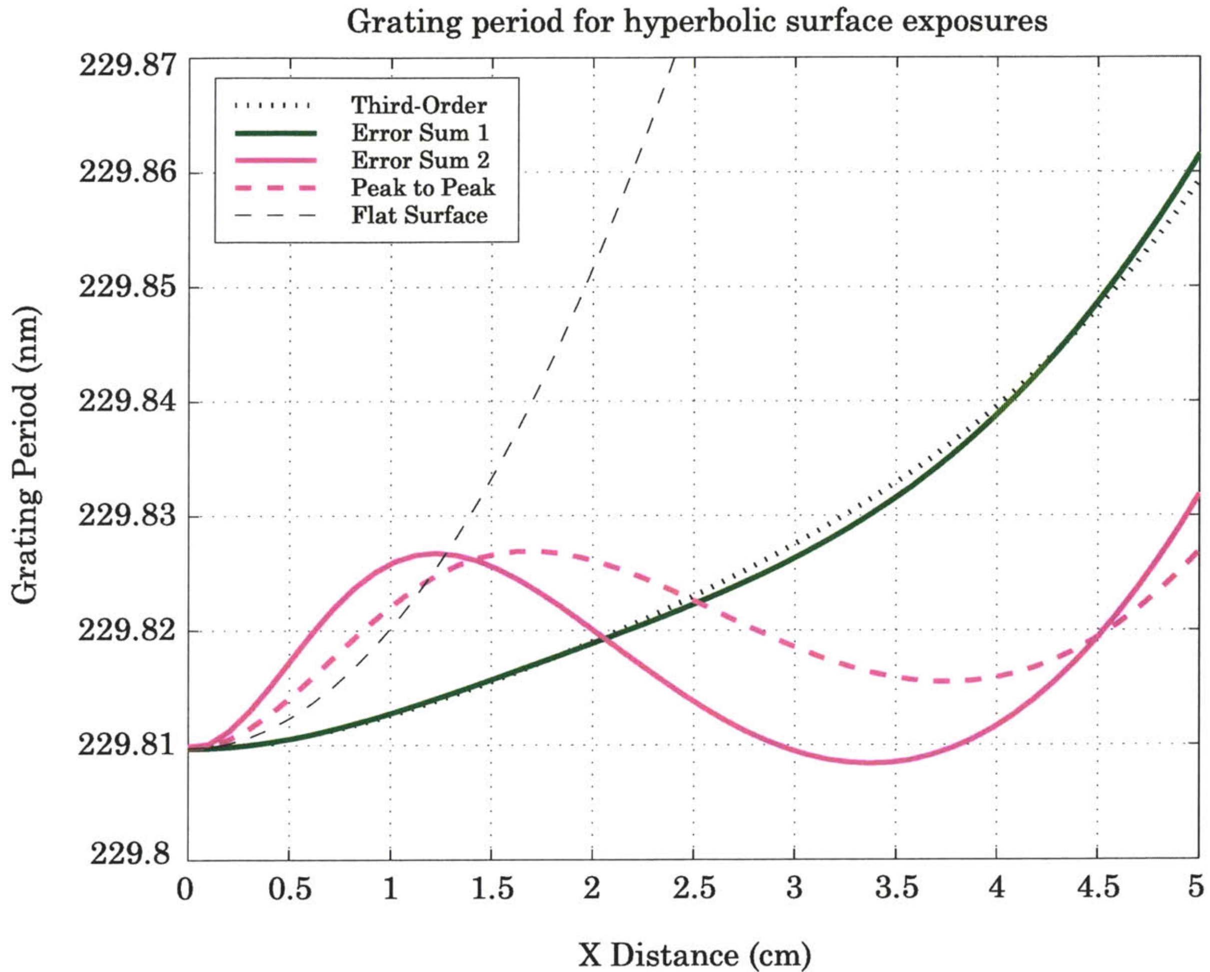


Figure 3.22 Exposed grating period on a selection of optimized hyperbolic surfaces.. The two magenta curves, (Error Sum 2 and Peak to Peak) smaller peak deviations from the nominal grating period, and lower accumulated phase-error, but have the undesirable quality of greater period errors in the preferred region at the center of the substrate. If the hyperbolic surface is constrained at the center of the wafer (Error Sum 1), the optimized curve is likely the function which the optimized polynomials approximate.

Figure 3.22 shows the grating period exposed on a selection of hyperbolic surfaces for which different minimization criteria were used. Also shown for comparison are the grating periods for a flat-surface exposure and the third-order polynomial surface. It is worth noticing the similarity of the results for the third-order polynomial surface and the optimal hyperbolic surface under the condition which limits the periodicity to a smoothly increasing function. The two surfaces which are the "best" optimized, under the conditions stated above (error sum and peak-to-peak) show a much more complex pattern of variation. The change in period close to the center of the exposure, which is generally regarded as the ideal region, actually changes faster than the flat surface exposure. It is not obvious that this is an improvement over the smoothly increasing character of the other surfaces. One can conclude that in cases where a complete optimization is impossible, as it seems to be in this case, that the *qualitative* aspects of the minimized function can be as important as the *quantitative* results of the optimization. The determination of which surface is best for a given application may be a subjective one.

3.8.5 Local Strain Correction

The stress/strain relationship for the deformed silicon wafer was examined in Section 3.6.3 for the spherical surface to determine the fracture limits of the wafer. So far, however, we have ignored the possibility that the local strain on the deformed wafer surface will distort the grating periodicity when the substrate returns to its planar state. For the spherical surface, this was not a significant issue since the compressive strain is uniform. When the wafer returns to flat, the surface will uniformly expand as the strain is relieved. Thus, the periodicity of the grating will increase by a scale factor equal to $1+e$, (where e is the strain), but the non-linearity of the periodicity will be unchanged.

On a non-spherical surface, however, the local strain will be spatially dependent. For any concave surface, the top of the wafer will always be under compressive strain during the exposure. The amount of local strain will be

determined by the local radius of curvature R. For areas of greater curvature, the local expansion upon relief of the strain will be greater than areas with less curvature. Whereas the constant radius of curvature of the sphere allowed us to ignore this effect, local strain on the hyperbolic surface must be accounted for in order to accurately predict the periodicity of the grating on the flat substrate. A quick consideration of the hyperbolic curve hints that this may have a beneficial affect. The highest radius of curvature on the hyperbola is at the center, and it asymptotically approaches a straight line with increasing x. Thus, as a wafer bent into a hyperbolic shape is released, the features at the center of the wafer will expand more than the ones at the outside. This behavior is opposite to that of the grating periodicity exposed on this surface, which is lowest at the center and increases towards the edges. It seems possible then, that the two effects could cancel each other out. Cross sections of a hyperbolic surface and the associated strain correction are shown below.

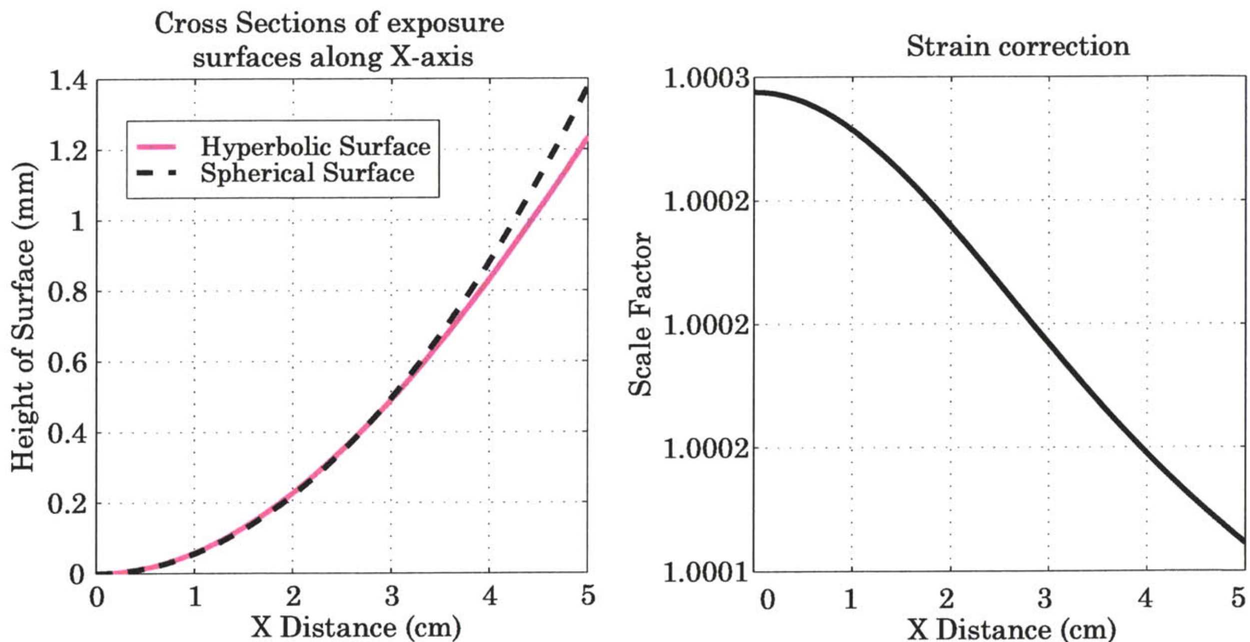


Figure 3.23 The hyperbolic curve optimized for exposure along the x-axis (left) and the local strain correction scale factor $S(x)$, as given in Equation 3.53 (right). For comparison, the surface profile of the spherical surface tested in Section 3.7 is also shown.

Although a full 2-dimensional strain analysis is beyond the scope of this work, we can estimate the local strain along the x-axis via the relatively simple calculation of the local radius of curvature [28], which is given below in terms of the first and second derivatives *(z',z'')* of the hyperbolic function *z(x)* from Equation 3.49.

$$R(x) = \frac{\left[1 + (z'(x))^2\right]^{3/2}}{z''(x)} \tag{3.50}$$

Using the local radius of curvature instead of a constant, we can rewrite Equation 3.26 for the local strain *e(x)* in terms of the wafer thickness *2t*.

$$e(x) = \frac{t}{R(x)} \tag{3.51}$$

The correction to the period after relieving this strain, shown below, will be a scale factor *S(x)* which multiplies $P_s(x)$, the local period of the grating on the surface of the curved grating.

$$P_{flat}(x) = S(x) \cdot P_s(x) \tag{3.52}$$

$$S(x) = 1 + e(x) \tag{3.53}$$

If the optimization algorithm is revised to include not only the optical corrections for the periodicity on the curved surface, but also the strain corrections, the results are greatly improved. Figure 3.24 shows both the period variation on the optimized curved surface and the period error of the same grating compensated for strain relief when returning to flat. The graph on the left shows the data on the scale of the optical correction, which is comparable to the plots shown in Figure 3.22. At this scale, the addition of the strain correction achieves a grating which looks close to perfectly linear. Observing the graph on the right, rescaled to emphasize the strain curve, we can see that the residual period error is less than 1 picometer peak-to-peak over most of the wafer area. The maximum variationin period over the majority of the wafer is ~0.5 pm, or 2 parts-per-million. If the corrected grating were assumed linear and used as a metrology standard, the total accrued error over the entire wafer would be less than 1/2 of a single period. This is

truly an excellent degree of linearization for such a simple method as swapping the planar wafer chuck for a concave one.

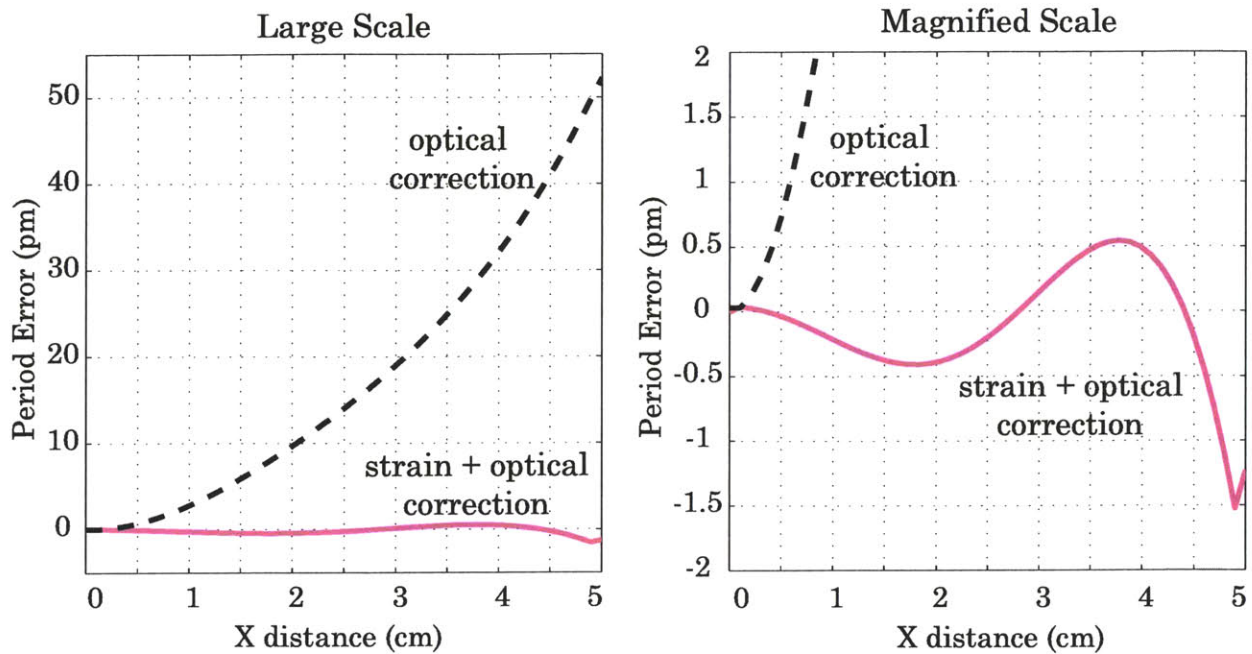


Figure 3.24 Grating period variation along the x-axis for exposures on a hyperbolic surface optimized to include the local strain correction. Both plots depict the same data with different scales on the abscissa. The surface optimized to include the strain correction reduces period variations to less than 1 *pico*meter over almost the entirety of the wafer. This represents an improvement of close to 3 orders of magnitude over the flat surface exposure.

3.9 Conclusions

The Mach-Zehnder style intererometer is a relatively simple and effective way to pattern periodic structures over large areas with feature sizes approaching the diffraction-limit of $\lambda/2$. Spherical-beam interference allows for more exposure uniformity and less noise than plane-wave techniques. Undesirable non-linear spatial-phase terms associated with spherical beam interference can be reduced through deliberate distortions in the substrate plane during the exposure step. Standard 10-cm diameter silicon wafers exhibit enough elasticity to support curvatures on the order of 1 meter radius without fracture. Spherical surface corrections have been demonstrated using a moiré technique and qualitatively match theoretical predictions to a high degree. Further optimization of the exposure surface using numerical methods has been shown to be possible. Hyperbolic surface curvature exhibits the maximum optical correction, which is about a factor of 5 better than the flat surface correction. Inclusion of the local strain effects associated with the wafer relaxing to its planar shape show that the hyperbolic surface exposure can reduce period variations along the x-axis to less than ~1 picometer over a 10 cm wafer, close to 3 orders of magnitude less than the flat surface exposure. Further investigation is necessary to determine the strain effects over the full area of the wafer.

3.10 References

[1] L. Zehnder, *Zeitschr. F. Instrkde,* **11,** p.275 (1891)

[2] L. Mach. *Zeitschr. F. Instrkde,* **12** p.89 (1892)

[3] E.N. Leith, J. Upatnieks, "Wavefront reconstruction and communication theory". *J. Opt. Sci. Am.* **52** p.1123 (1962)

[4] D. Rudolph, G. Schmahl, "Verfahern sur Herstellung von Röntgenlinsen und Beugungsgittern," *Unschau. Wiss.* **67** p.225 (1967)

[5] A. Labeyrie and J. Flamand, "Spectrographic Performance of Holographically made Diffraction Gratings". *Opt. Commun.,* **1** p.5, (1969)

[6] E.H. Andersen, V. Boegli, M.L. Scattenburg, D. Kern, H.I. Smith. "Metrology of electron-beam lithography systems using holographically produced references". *J. Vac. Sci. Technol. B* **9**, p.3606 (1991)

[7] H.I. Smith, *Proc. IEEE* **62** pp.1361 (1974)

[8] M.L. Schattenburg, R.J. Aucoin, R.C. Fleming, I. Plotnik, J. Porter, H.I. Smith, "Fabrication of high-energy x-ray transmission gratings for the Advanced X-ray Astrophysics Facility". *Proc. SPIE* **2280** p.181 (1994)

[9] M.L Scattenburg, *et al.*, "sub-100 nm metrology using interferometrically produced fiducials". *J. Vac. Sci. Technol. B* **17**, p.2692 (1999)

[10] M.L. Schattenburg, "From nanometers to gigaparsecs: The role of nanostructures in unraveling the mysteries of the cosmos". *J. Vac. Sci. Technol. B* **19**, p.2319 (2001)

[11] J. Ferrera, "Nanometer-Scale Placement in Electron-Beam Lithography". Ph. D. Thesis, Massachusetts Institute of Technology, (2000)

[12] C.G. Chen, P.T. Konkola, J. Fererra, R.K. Heilmann, M.L. Schattenburg, "Analysis of vector Gaussian beam propagation and the validity of paraxial and spherical approximations". *J. Opt. Sci. Am. A* **19** p.404, (2002)

[13] L.E. Nilsson H. Ahlén, "Stabilization of the exposing interference pattern in holographic grating production" *Proc. SPIE* **240** p.22 (1980)

[14] H.I. Smith, E.H. Anderson, M.L. Schattenburg, "Holographic Lithography" U.S. Patent 5,142,385 (1992)

[15] W.H. Steel, *Interferometry*, pp. 47, Cambridge University Press, Cambridge, UK (1983)

[16] G. Schmahl, D. Rudolph, "Holographic diffraction gratings," in *Progress in Optics*, E. Wolf ed. North-Holland, Amsterdam. **14** p.195 (1976)

[17] K. Hibino, Z. Hegedus, "Hyperbolic holographic gratings: analysis and interferometric tests", *Appl. Opt.* **33** p.2553 (1994)

[18] J. Fererra. M.L. Schattenburg, H.I. Smith, "Analysis of distortion in interferometric lithography", *J. Vac. Sci. Technol. B* **14**, p.4009 (1996)

[19] D. Dewey, D.N. Humphries, G.Y. McLean, D.A. Moschella, "Laboratory calibration of x-ray transmission diffraction gratings". *Proc. SPIE* **2280**, p.257 (1994)

[20] M. Lim, "Development of X-ray Lithography and Nanofabriation techniques for III-IV Optical Devices", Ph.D. Thesis, Massachusetts Institute of Technology, (2002)

[21] M.E. Walsh, H.I. Smith, "Method for reducing hyperbolic phase in interference lithography". *J. Vac. Sci. Technol. B* **19**(6), p.2347 (2001)

[22] S. Lang, *Introduction to Linear Algebra*, Springer-Verlag, New York (1986)

[23] R.T. Fenner, *Mechanics of Solids*. Blackwell Scientific Publications, Boston, (1989)

[24] C. Wilson, P. Beck, "Fracture Testing of Bulk silicon Microcantilever Beams Subjected to a Side Load". *J. Mem. Sys.,* **5** pp.142 (1996)

[25] Edmund Scientific, Part number 32829

[26] J.A Nelder, R. Mead. *Computer Journal*, **7** pp. 308 (1965)

[27] G. J. Borse, *Numerical Methods with MATLAB*, PWS Publishing, Boston (1997)

[28] Kreysig, E, *Differential Geometry*. Dover, New York (1991)

Chapter 4
The Lloyd's-Mirror Interferometer

4.1 Introduction

The Mach-Zehnder architecture described in Chapter 3 has been implemented at MIT chiefly for the repeatable exposure of the highest quality grating images over a large number of samples - in effect a production tool. A second Mach-Zehnder IL system was built for research purposes such as holographic phase-shifting interferometry and linearization through substrate curvature, discussed in Chapter 3 [1,2]. However, all of the systems fall short when it comes to prototyping, or small batch processes. The versatility of the system is severely limited in that it requires precise alignment, a dedicated set of expensive optics, and a relatively large physical footprint to produce a single grating period. Beyond gratings, there are a number of symmetric, periodic patterns that can be created through multiple exposures. For example, two exposures rotated by 90-degrees to one another produce a square array; two exposures rotated at 60-degrees produce a hexagonal array. Applications that require a multitude of

different periods, such as magnetic studies or templated self-assembly (both discussed in Chapter 7) are difficult to pursue using the traditional Mach-Zehnder style IL setup.

A simpler configuration, known as the Lloyd's-Mirror Interferometer, was introduced in the 1830's by Humphrey Lloyd (1800-1881) [3]. Interestingly, the utility of this design historically is quite different from our current needs. The beginning of the 19th century was a heady time in optics, with the wave theory of light just coming into wide acceptance. Early proponents of the theory, including such heavyweights as Robert Hooke (1635-1703) and Leonhard Euler (1707-1783) were unable to find widespread acceptance of the theory due to the overwhelming influence of Isaac Newtons' (1642-1727) corpuscular theory. In the early 1800's primarily through the work of Thomas Young (1773-1829) and Augustin Fresnel (1788-1827), the tide began to turn. Fresnel's prize-winning explanation of diffraction at the Paris Academy is one of the more dramatic moments in the history of science, although he had assumed light to be a longitudinal vibration and therefore could not explain polarization effects. It was Young, in 1817, who proposed transverse vibrations, and his classic double-slit experiment provided the best experimental evidence to date in support of the new theory. In the wake of this, interferometry was born. One of the simplest early designs, attributed to Lloyd, is shown in Figure 4.1.

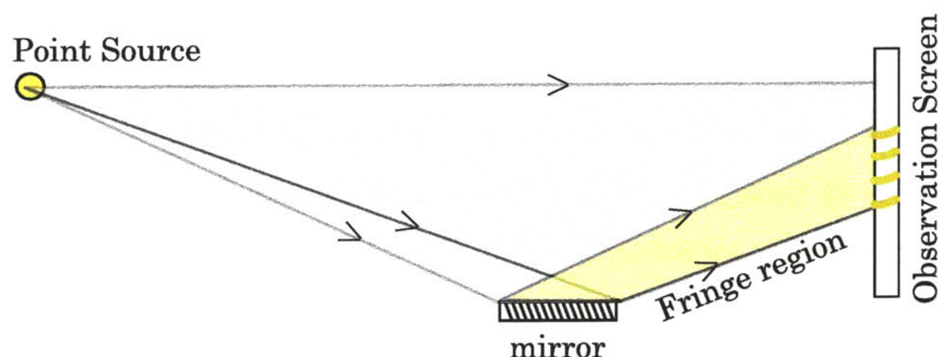


Figure 4.1 Cartoon of the original Lloyd's mirror interferometer. The incident angle on the mirror is depicted as large for clarity, although in practice it would have approached glancing. A large separation between the mirror and screen facilitates glancing angle reflection.

In these early days of interferometry, the challenge was to achieve fringes that were large enough to be seen by eye using a source with poor spatial and temporal coherence. The Lloyd's-mirror was used in such a way that the incident light was at a glancing angle to the mirror, which was placed some distance away from the observation plane. This allowed for a very small interference angle, and hence very large period fringes. Perhaps more important, this also affected only a minimal path-difference between the interfering rays, thus allowing visible fringes with a low-coherence source. Today, however, our interests are just the opposite. The Lloyd's-mirror described here is intended to produce large areas of the finest possible fringes, a result which is only possible with the high-coherence of a modern laser.

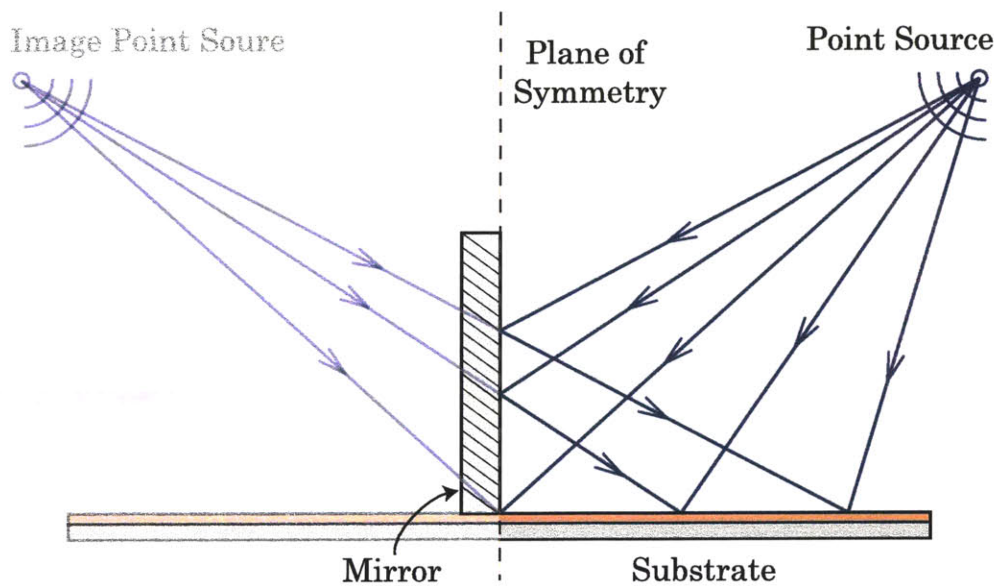


Figure 4.2 The Lloyd's-Mirror Interferometer depicted as equivalent to the Mach-Zehnder style interferometer. The surface of the mirror becomes the plane of symmetry in a Mach-Zehnder, with one image source as one real source.

4.2 A Lloyd's-mirror interferometer for lithography

The practical requirements of a Lloyd's mirror interferometer designed for lithography are, as mentioned, quite different from those of the original configuration. For this reason, it is instructive to present the Lloyd's-mirror in the

context of the familiar Mach-Zehnder configuration. Both systems are designed to produce (nominally) the same result: high-contrast interference fringes of high spatial-frequency over a large exposure area. In fact, as shown in Figure 4.2, the ideal Lloyd's-mirror is optically equivalent to half of a Mach-Zehnder interferometer. Consider the plane of symmetry for the Mach-Zehnder located halfway between the two sources and which bisects the angle of interference. Ideally, this is also normal to the substrate. The system on either side of this plane is a mirror image of the other side. Thus, if indeed we were to place a mirror in this plane, the resulting set of interference fringes would remain unchanged - the image source in the mirror completes an exact replica of the original Mach-Zehnder. This is the Lloyd's mirror. Of course, the region behind the mirror will not be illuminated by either the real or imaginary sources, so only half of the area is exposed. Additionally, the physical differences in implementation do lead to significant practical differences, both beneficial and detrimental. Therefore, rather than being a replacement for the Mach-Zehnder interferometer, the Lloyd's-mirror will be seen to fill a complementary role.

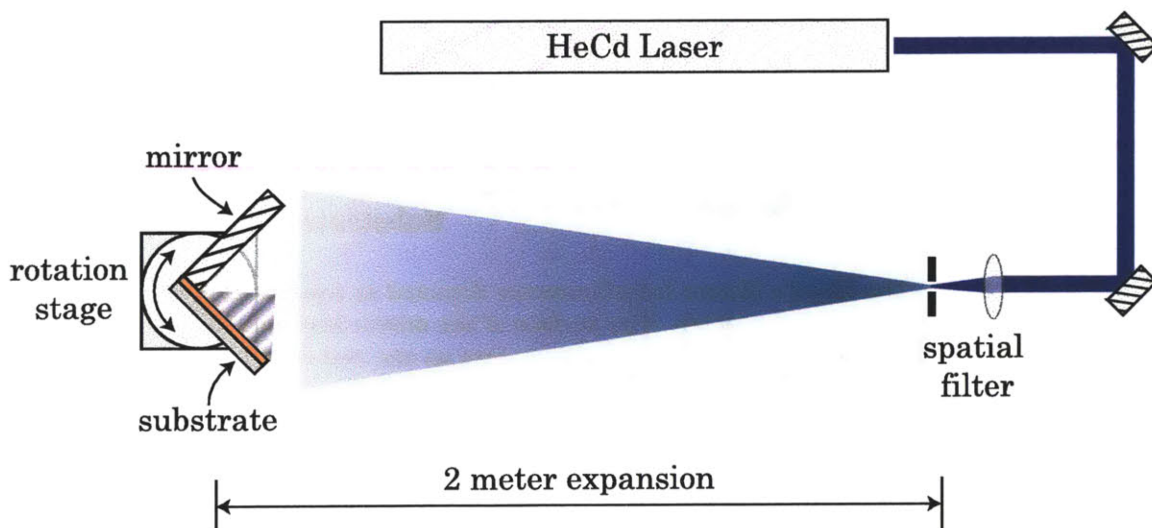


Figure 4.3 A schematic of the Lloyd's mirror Interferometer. The beam from a HeCd laser (λ_0=325 nm) is spatially filtered and allowed to expand for 2 meters. The mirror and substrate are mounted on a rotation stage for easy adjustment of the fringe period.

In simple terms, the Lloyd's-mirror is just a mirror placed normal to the substrate and illuminated with a single beam. A schematic of the implementation of the Lloyd's-mirror is shown in Figure 4.3. The specifics of the design were determined by trying to optimize for high fringe-contrast and exposure areas on the scale of a 10 cm wafer in a highly ergonomic tool. A helium-cadmium (HeCd) laser emitting at λ_0=325 nm in a single TEM00 mode was chosen for the source. The HeCd offers a long (30 cm) coherence length at a mid-UV wavelength in a more robust package and at a lower cost than other options, such as argon-ion or excimer lasers. As explained in Chapter 2, the bandwidth, or temporal-coherence, of the laser determines the maximum exposure area. Although the HeCd has a relatively weak output power, this is of secondary importance. Output power does not limit important parameters such as the minimum pitch of the fringes or the size of the exposed area, and throughput is not an important specification in the research environment.

In any interference scheme, the TE polarization will produce higher contrast fringes than the TM. Thus, the steering mirrors between the laser and the spatial-filter are arranged to ensure that the polarization incident on the substrate is TE (electric field parallel to the substrate). The spatial-filter allows high spatial-frequency noise to be removed from the beam to provide a clean Gaussian profile. The interferometer itself, consisting of a wafer chuck, mirror and rotation stage, is placed approximately 2 meters from the spatial-filter. A large beam expansion is somewhat of a tradeoff, but in this case more is gained than is lost. As a Gaussian beam expands, it changes in three ways. The intensity of the beam decreases, the diameter of the beam increases, and the radius of the phase front increases. Lowering the intensity leads to increased exposure times, but the inherent stability of the system, discussed is Section 4.4, means that this is of little consequence to the patient experimenter. Because of the Gaussian intensity profile, increasing the beam diameter creates a more uniform intensity distribution over the exposed area.

Finally, the increase in radius of the beam diameter means that the beam more closely approximates a plane-wave over the exposure area. Thus, by maximizing the beam expansion, the exposed grating will have a more linear spatial phase and a more uniform linewidth, at the expense of a longer exposure time. The 2-meter expansion used in this system is close to the maximum possible on the available optical table.

The interferometer was designed to accommodate up to a 10 cm diameter substrate. However, to maximize the linearity, contrast, and uniformity over the largest possible area, the wafer is exposed one half at a time, with the other half covered. Recall from Figure 4.2 that the fringe pattern in the Lloyd's mirror will replicate 1/2 of the fringe pattern from a Mach-Zehnder interferometer. Thus, the full fringe pattern can be replicated through separate exposures of the two halves. Although this allows an entire 10 cm diameter substrate to be exposed, the spatial phase of gratings is obviously uncorrelated for the two halves. Spatial phase coherence is limited to the area printed in a single exposure.

To minimize the effects of edge scattering and diffraction, it would be desirable to use a mirror somewhat larger than the exposed sample. Within the coherence limit, a larger mirror will also enable greater exposed areas. However, the commercial availability of large, square, front-surface mirrors is limited. This makes the choice of mirror largely economic, as custom-made mirrors will be made to any specification for increased cost. The mirror currently in use is 10 cm on a side, with a flatness of $\lambda/4$. Aluminum was chosen because of its enhanced UV reflectivity compared to other metals, and for its essentially constant reflectivity over a broad range of angles. A higher reflectivity can be obtained with a dielectric mirror, but the variation in reflectivity with angle can be significant, and would be unacceptable for this application. The effects of the finite mirror size and reflectivity are discussed further in Section 4.5.

4.3 Alignment and fringe period

Although optically equivalent, the obvious practical difference between the Mach-Zehnder interferometer and the Lloyd's-mirror (LM) is that the LM requires only one beam. The power of this simplification is seen most clearly when one considers the work necessary to change the periodicity of the interference fringes. To change the grating period in the Mach-Zehnder, the optics in each arm (mirrors, spatial-filters, ...) must be physically moved and realigned. Methods for this are described in Chapter 3, and it is generally an unpleasant task. Consider the schematic of the Lloyd's-mirror in Figure 4.2; it should be obvious that an arbitrary placement of the real source will always produce a perfectly aligned image source. Perhaps less obvious, if the location of the point-source is fixed in space, its position relative to the substrate, and therefore the fringe period, can be arbitrarily varied by simply rotating the interferometer. In stark contrast to the Mach-Zehnder, no optics need to be moved or aligned. If we assume that the phase-radius of the beam is large enough to be approximated as a plane-wave, then simple trigonometry is sufficient to show that the angle-of-incidence of the reflected beam is always equal to the angle of the directly incident beam, and that the bisection of the interference angle (2θ) is always normal to the substrate.

Figure 4.4 When the mirror is placed normal to the substrate, the simple trigonometry of complementary angles in a right triangle ensures that the angle-of incidence on the substrate for the direct and reflected beams is always equal.

Two conditions must be met to guarantee that this is as simple as it seems. One is that the mirror is truly mounted perpendicular to the substrate. This is a simple procedure, for which a variety of methods exist [4]. The other is that the interferometer axis defined by the intersection of the mirror surface and the substrate surface is the axis of rotation. These issues will be discussed further in the following sections. In practice, the severity of these issues will be largely determined by the build quality of the interferometer. In general, misalignment in these two areas will contribute to errors in the interference angle and therefore error in the period of the exposed grating. For perfect alignment, a rotation of the interferometer stage through a given angle α will change the interference angle θ by exactly α. The ability to accurately set the periodicity of the exposed grating will be dependent on the accuracy and repeatability of the rotation stage. For this reason, a high-quality rotation stage is among the most important parts in the Lloyd's-mirror. Because the effects of misalignment can be reduced to angular errors, we can lump the effects of misalignment and stage accuracy together and consider the effects of total angular error on the grating period. An important distinction is that the alignment errors are systematic, they are repeatable, where the stage error will be largely random. Systematic errors can be eliminated through alignment or calibration, while the stage error will always be present. Thus, we can use the stage error $\Delta\theta$ as an upper bound for the tolerable systematic errors, i.e. we will not notice the effects of the systematic errors if they are kept below the error of the stage. These conditions are derived in the following sections.

The effect of angular error can be understood by considering Equation 4.1 for the periodicity of the exposed grating, first introduced in Chapter 2.

$$P_0 = \frac{\lambda}{2\sin(\theta)} \qquad (4.1)$$

The sensitivity of the grating period to an angular change, given in Equation 4.2 can be found simply from the derivative of Equation 4.1 with respect to θ, shown in Equation 4.3.

$$\Delta P_0 = (\tfrac{dP_0}{d\theta})\Delta\theta \qquad (4.2)$$

$$\frac{dP_0}{d\theta} = \frac{-\lambda\cos(\theta)}{2\sin^2(\theta)} \qquad (4.3)$$

Equation 4.2 can be normalized to give the percentage change in period as a function of angle by simply dividing Equation 4.2 by 4.1.

$$\frac{\Delta P}{P_0} = \frac{\Delta\theta}{\tan(\theta)} \qquad (4.4)$$

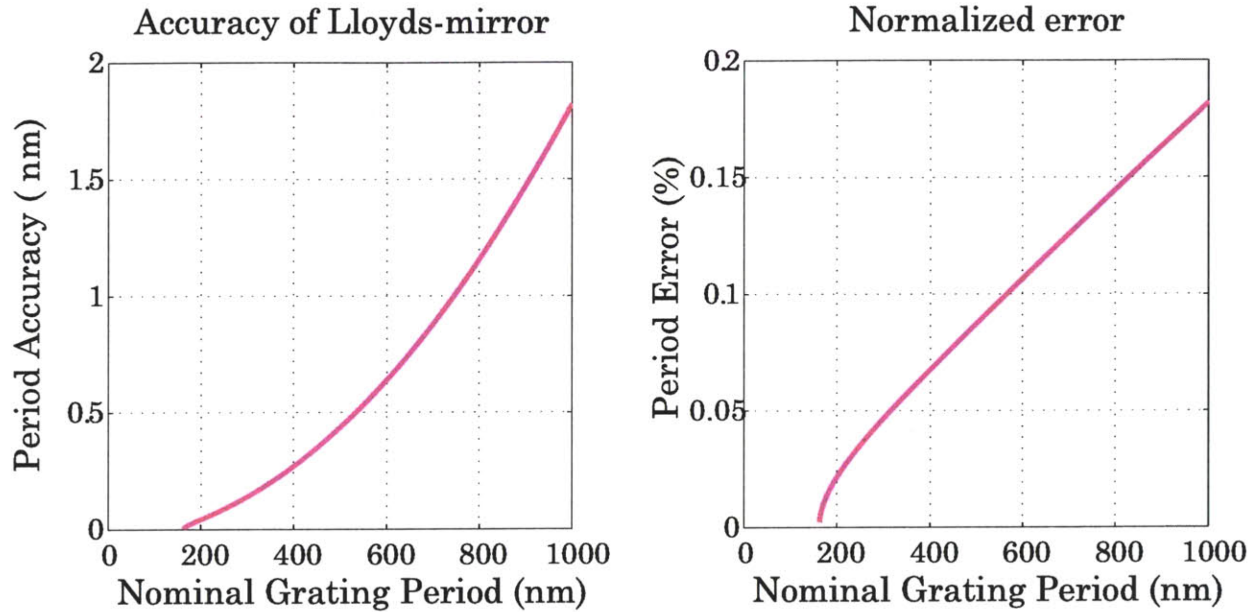


Figure 4.5 The absolute error and normalized error as a function of grating period are shown for the current Lloyd's-mirror configuration. Assume $\Delta\theta$ is equal to the accuracy of the rotation stage is about +- 1 minute of arc, or 0.3 mrad. Notice that at small grating periods, the error in periodicity is significantly less than 1 nm.

The normalized and absolute errors are plotted in Figure 4.5. One could consider it lucky that the smallest period gratings are quite immune to angular error, where the larger period gratings are far more sensitive. In most cases, making smaller feature sizes only exacerbates the troubles observed at larger features. The high sensitivity at large grating periods is actually not particularly troubling, as there are other lithographic options for making features at this scale. The value of interference lithography is generally in its ability to produce small features.

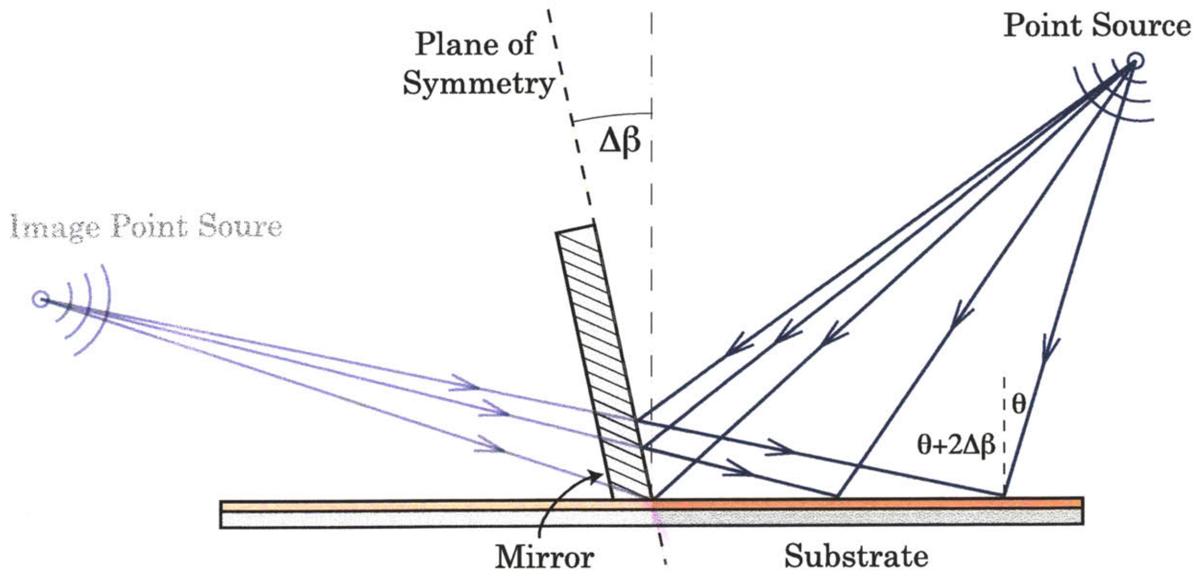


Figure 4.6 Misalignment of the mirror from normal by an angle of Δβ. The image source is no longer symmetrical with respect to the wafer surface, changing the periodicity of the interference fringes and their angle of inclination w.r.t the wafer.

4.3.2. Angular Alignment of Mirror

The easiest way to understand how the mirror angle affects the exposed fringes is to consider what happens to the image source. For mirror angles not equal to 90 degrees, the symmetry will be broken and the image source will be placed in a different position relative to the substrate than the real source. The result of this is a different angle of incidence on the substrate for the reflected beam while the direct beam remains unchanged. If we assume plane-wave illumination, then we have essentially the same situation that arose in the analysis of the hyperbolic phase progression in Chapter 3; the fringes will not form perpendicular to the substrate, and their periodicity will change. The fringe period recorded on the substrate P_s, shown in Equation 4.6 will be the projection of the new fringe period $P_{\theta+\Delta\beta}$, (Equation 4.5) into the substrate plane. For a mirror angle which is equal to $\beta=90^{\circ}+\Delta\beta$, the angle of inclination of the fringes will also be $\Delta\beta$.

$$P_{\theta+\Delta\beta} = \frac{\lambda}{2\sin(\theta+\Delta\beta)} \tag{4.5}$$

$$P_s = \frac{P_{\theta+\Delta\theta}}{\cos(\Delta\beta)} \tag{4.6}$$

Combining the two equations above, we can solve for the the grating period on the substrate, shown below in Equation 4.7.

$$P_s = \frac{\lambda}{2\sin(\theta+\Delta\beta)\cos(\Delta\beta)} \tag{4.7}$$

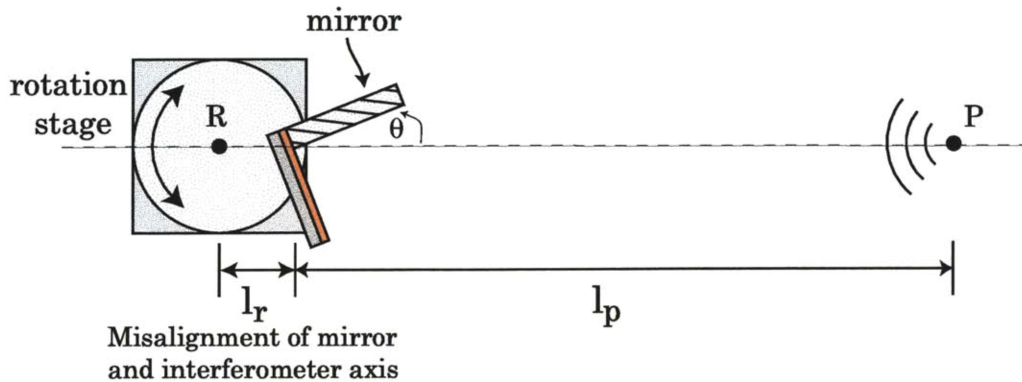
For small deviations $\Delta\beta$ of the mirror from normal, the cosine term can be approximated as unity, leaving the effect of the mirror misalignment as equivalent to a calibration error. That is, the actual interference angle will be slightly different from the expected interference angle by a constant amount. In practical terms, this means that the angle $\Delta\beta$ should be somewhat less than the angular accuracy of the stage $\Delta\theta$, as shown below in Equation 4.8.

$$\Delta\beta < \frac{\Delta\theta}{2} \tag{4.8}$$

4.3.3 Alignment of rotational axis

As mentioned previously, the axis of the interferometer created by the intersection of the plane of the mirror, and the plane of the substrate, which we will call I, should be aligned with the axis of rotation R. In this case, the rotation angle of the stage will exactly equal the interference angle, minus some constant offset. If these axes are not aligned, then there will be an additional error angle α' determined by the separation of I and R and the angle of rotation α. An exaggerated cartoon of this is shown in Figure 4.7.

The presence of this error term is due to the use of point-source illumination. If we were to use a perfect plane-wave, then the axial alignment would be unnecessary. We will simplify the analysis here by assuming that incident light at the interferometer is planar. The direction of the incident wavefronts, however, will be determined by the location of the point source.



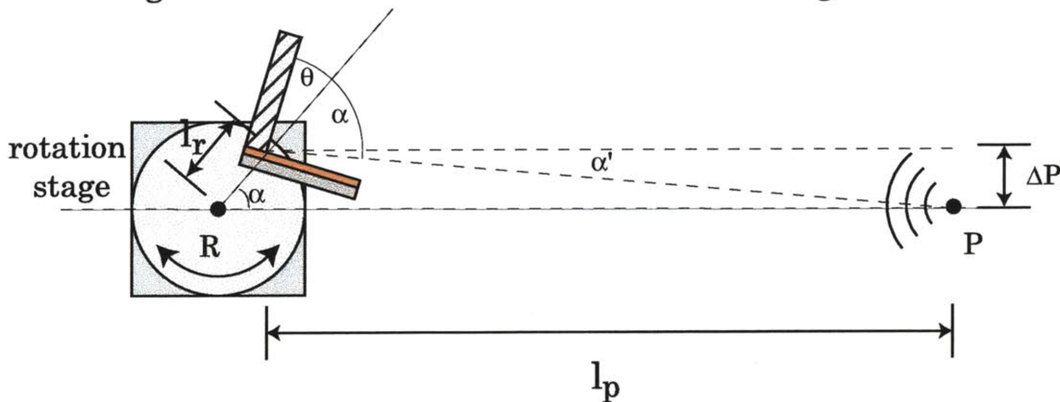


Figure 4.7 When the axes of the interferometer and the rotation stage are misaligned, there will be a lateral shift of the interferometer ΔP as a function of the rotation angle α which creates an error α' in the interference angle.

Assume a set of coordinates where the point source, the intersection axis and the rotational axis are co-linear. In this position, the distance between the intersection axis is l_p, and the distance between P and the intersection and rotational axes is l_r. If l_p is large, we can assume plane-waves incident on the wafer with an angle of θ. From this point, the stage is rotated through an angle α. The location of the interferometer axis has now shifted laterally by a distance ΔP.

$$\Delta P = l_r \sin(\alpha) \tag{4.9}$$

Under true plane-wave illumination, the new angle of incidence of the wafer θ' would simply be $\theta+\alpha$. Because we are using point-source illumination, however, the lateral shift ΔP will induce an additional change in illumination angle α'. For large l_p, we can use the small-angle approximation and neglect the change in axial distance IP due to rotation.

$$\theta' = \theta + \alpha + \alpha' \tag{4.10}$$

$$\sin(\alpha') = \frac{\Delta P}{l_p} \tag{4.11}$$

$$\alpha' \approx \frac{l_r}{l_p}\sin(\alpha) \tag{4.12}$$

Unlike the angular misalignment of the mirror, the axial misalignment will cause an error in the fringe period which varies with the incident angle. Yet we can still apply the same criteria to limiting α' as we did in limiting $\Delta\beta$; it should be smaller than the accuracy of the stage. The maximum value of α' will occur when the sine is unity, leaving the ratio of the misalignment distance to the beam-expansion length as the critical quantity. In the current Lloyd's-mirror setup, the stage accuracy $\Delta\theta$ is ~1 arcmin (0.3 mrad), and the expansion length l_p is ~2 meters. Therefore, the alignment tolerance of the rotation axis is ~1/2 mm, which should be within reach of a well-designed tool.

$$l_r < \frac{l_p \Delta\theta}{2} \tag{4.13}$$

4.3.4 Practical Alignment and Calibration

It has been shown in the past two sections that the critical alignments of the Lloyd's-mirror are the mirror angle and the co-location of the interferometer and rotational axes. The functional definition of the tolerance to these misalignments is related to the accuracy of the rotation stage $\Delta\theta$. If the angular errors due to alignment error are less than the accuracy of the rotation angle, then they will not

be noticeable. The role of the mirror is to automatically align the real and virtual sources, so there is truly very little actual alignment necessary. Under true point-source illumination, the placement of the source relative to the interferometer could be arbitrary. Although it approximates a point source on axis at distances far from the focus, the actual illumination is a Gaussian beam. Thus, it is necessary to align the propagation axis of the beam with the rotation and interferometer axis to ensure both a uniform intensity distribution and the validity of the spherical approximation. This is a simple procedure in the absence of the spatial-filter, where the axis of the interferometer is aligned with the unexpanded beam.

The most important step is actually not an alignment at all, but a calibration of the interference angle in relation to the stage angle. It is extremely unlikely that the zero-point of the stage will coincide with zero incident angle of the beam. The offset between these angles is critical to knowing at what angle to set the rotation stage to achieve a given grating period. This seems to be a simple matter, and in some respects it is. Rough calibration can be found by aligning the back reflection of the laser beam off either the mirror or substrate surfaces. A better calibration would fix the rotation stage at a known angle, and measure the period of an exposed grating. This routine has the benefit of including the effects of an off-normal mirror angle. The trick in this case is to find a measurement technique that is sufficiently accurate. Recall from Figure 4.5 that the error in the stage angle of 1 arcmin correlates to less than 1 nm of absolute error in the grating period in most cases. Thus, we need a method of measuring the periodicity of the grating to less than 1 nm, which is not trivial. One method that works particularly well is to simply count a large number of grating periods, e.g. 10,000, and measure the distance covered. This method has the advantage that extremely small changes in periodicity will add up to measurable distances over a large number of periods. This can be done very precisely by using a laser-interferometer controlled translation stage for distance measuring coupled to a scanning electron microscope (SEM) for counting. This method is described further elsewhere [5], and is accurate to the Ångstrom level.

4.4 Stability

4.4.1 Vibrational Stability

One important consequence of using a single beam in the Lloyd's-mirror versus two beams in the Mach-Zehnder was that it became trivial to align and to change the fringe period. Another consequence of this is vastly improved fringe stability. While the MZ system is very sensitive to vibrations, beam motion, air currents, the LM is almost immune. The physical separation of the beam into two separate arms in the Mach-Zehnder gives rise to this sensitivity. If anything in either arm changes by even a fraction of a wavelength, the interference fringes will shift in position. Obviously, over the duration of a lithographic exposure, fringe motion will lead to a blurring or total loss of the exposed image.

In contrast, the Lloyd's-mirror uses a single beam for most of the optical path. Only in the last few centimeters, when a portion of the beam reflects off the mirror, does the beam split into two arms. A rigid mechanical connection between the mirror and the wafer chuck means that the relative lengths of the two arms is very stable since variations in arm length can only occur if the mirror moves in relation to the substrate. In general, small vibrations of the assembly or wandering of the incoming beam does not cause fringe drift, as the image source created by the mirror will automatically compensate. This negates the need for a phase locking system, a mandatory component of the Mach-Zehnder interferometer. The high inherent stability of the Lloyd's-mirror system enables long exposure times, which in turn enables the use a low-power laser such as the HeCd with a large beam expansion.

Of course, interferometers are never totally immune to environmental disturbances, so it is worth investigation what sort of events might lead to fringe instability. There are three major vibrational modes which can affect the fringes. Shown in Figure 4.8, they are an angular change between the mirror and the substrate, a shifting of the mirror position, and a relative motion between the

interferometer and the point source. Of these, the shift in mirror position is potentially the most damaging. This causes a change in path-length of the reflected beam that shifts the fringes without changing their periodicity. The other two modes, an angular change in the interferometer, and a relative shifting of the point source, are equivalent in that they both lead to a change in periodicity of the fringes. However, the rigid assembly of the interferometer makes this motion extremely unlikely, and we can generally ignore it. Section 4.3 on alignment provides an analysis of these two issues, and we directly apply the results of that investigation.

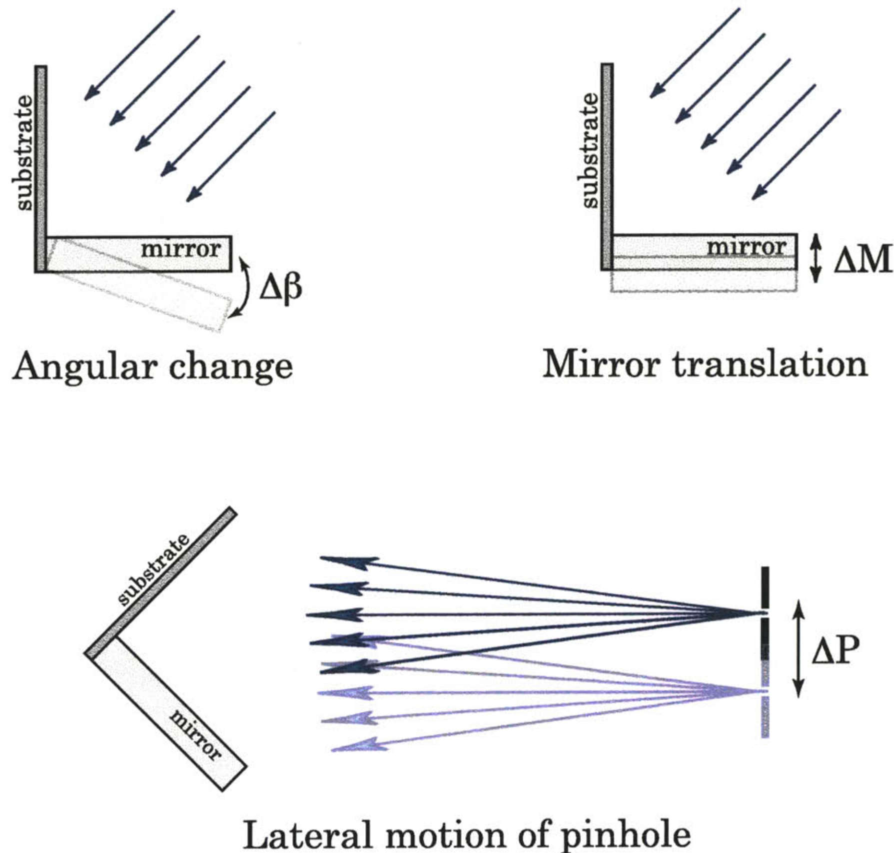


Figure 4.8 Three potentially damaging vibrational modes in the Lloyd's-mirror. An angular change between the mirror and substrate and a lateral shift of the point source will both manifest as changes in the fringe periodicity. A translation of the mirror will change the optical path of the reflected beam, causing the fringes to shift in position.

As with the translational motion of the mirror, the rotational motion of the mirror can be well controlled by proper mounting. That leaves the relative motion of the point-source and the interferometer as the primary cause of vibrational error. Recall from Section 4.3.3 that a relative motion ΔP will cause a change in incident angle α', and therefore a change in the periodicity. Equations 4.7 and 4.2 are rewritten here for convenience.

$$\alpha' = \operatorname{asin}\left(\frac{\Delta P}{l_p}\right) \quad (4.14)$$

$$\frac{\Delta P}{P_0} = \frac{\alpha'}{\tan(\theta)} \quad (4.15)$$

As before, the severity of a small angular change is dependent on the actual angle of incidence. For small angles (large period) the effects will be greatest, while at large angles (small periods) the effect will be minimal. As the periodicity of the fringes shifts, the exposed grating at the nominal period and the shifted period will walk out of phase. One can think of this as a "beating", or the formation of a moiré pattern between the two periodicities. Thus, at some distance l_{beat} from the mirror, destructive interference will cause the exposed image to be entirely lost. The distance over which this happens will be less for large variations, and greater for small variations. In reality, where the angular variation α' fluctuates up to a maximum value α'_{max}, the exposure will be blurred at all distances above that set by α'_{max}. This distance, shown below in Equation 4.16, is easily calculated from the normalized error given in Equation 4.4. For example, if the error $\Delta P/P$ is 1 part in 1000, then it will take 1000 periods for the two to get out-of-phase.

$$l_{beat} = \frac{P^2}{\Delta P} \quad (4.16)$$

If the distance l_{beat} is sufficiently small, it will force a limit on the exposure area. Assume that vibrations in the optical table and mounting hardware allow the position of the pinhole to vary relative to the interferometer by 1μm, and the

distance between the two is 2 meters. If we assume for simplicity that the tangent is unity, then the angle α'=5x10^{-7} rad, will equal the normalized error $\Delta P/P$. The grating period at this angle is 230nm, so the distance l_{beat}=46 cm. If we remember that the coherence length of the laser is only 30 cm, then it is obvious that this amount of relative motion is inconsequential. The relative motion ΔP would have to reach the level of 10μm before l_{beat} became short enough (4.6 cm) to be of concern. This level of motion would only be possible with negligent mounting of the optical hardware.

4.4.2 Real-time stability monitoring via moiré patterns

In short, fringe stability in the Lloyd's-mirror *should* be highly robust in the presence of vibrations and small motions of optical hardware. The proof of this, however, cannot come from assumptions and equations. Assuming that we can execute one good exposure, or use a grating exposed in a different IL system, the fringe stability of the Lloyd's-mirror can be monitored in real-time using the moiré technique described in Chapter 3.

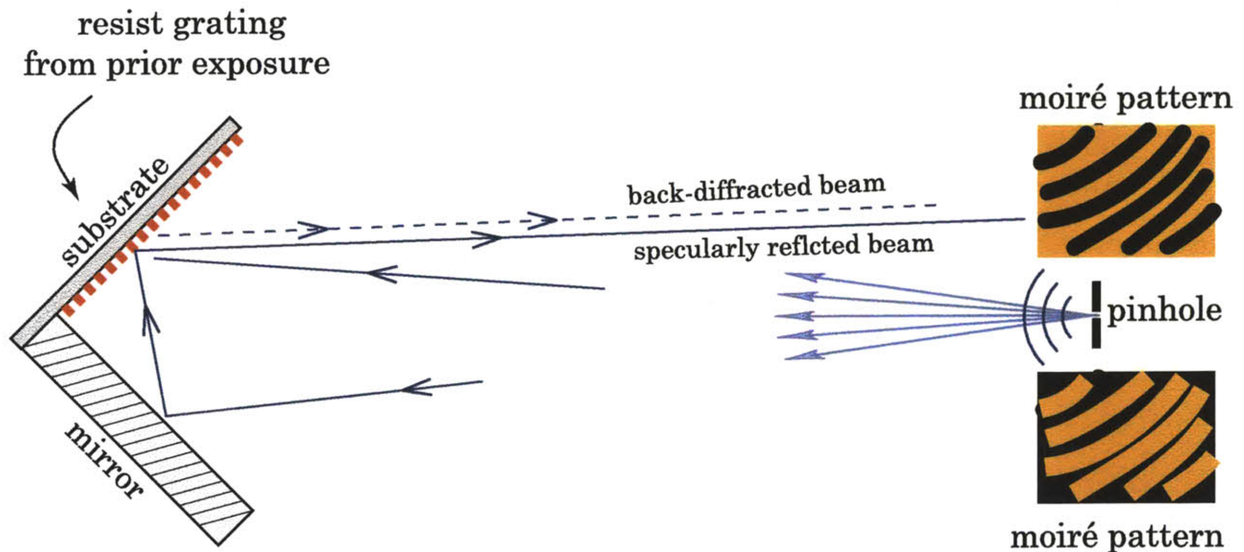


Figure 4.9 Complementary moiré patterns on either side of the pinhole which form due to the interference of specularly reflected and back-diffracted beams. The patterns represent the moiré of the physical grating in resist with the intensity pattern of the intersecting beams. Movement of the observed moiré fringes is indicative of movement of the interference fringes.

Placing an exposed and developed grating back into the interferometer will allow for interference between the fixed physical grating and the potentially shifting intensity pattern. In the proper circumstances, this can be observed directly on the substrate. A more general situation is shown in Figure 4.9, where the same pattern will be displayed on an observation screen placed around the pinhole via the interference of specularly reflected and back-diffracted beams. Notice that this is slightly different from the analogous case in the Mach-Zehnder interferometer. There, the complementary moiré patterns appeared at different pinholes. Here, they appear as mirror-images on either side of the single pinhole. Any motion of the moiré fringes observed on the screen is indicative of unwanted motion in the interference pattern on the substrate. Observation of these fringes is an excellent way to monitor the stability of the interferometer.

Shifting of the interference pattern during the exposure has the effect of blurring; it reduces the contrast of the exposed image. The necessary level of contrast for a good exposure is determined by the choice of photoresist. The interferometer must be stable enough that the interference fringes move by less than $\sim P_0/10$ over the duration of an exposure. A shift of $P_0/2$ in the interference fringes would fully invert the moiré pattern on the screen, so a fringe shift of $P_0/10$ (phase shift of 34°) would be easily visible. Exposure times are typically less than 10 minutes. Moiré fringes of an unperturbed Lloyd's-mirror inside a simple enclosure are observed to be highly stable; *no fringe motion is detected* over times limited by the patience of the observer.

In the absence of an enclosure, however, the moiré fringes are observed to drift in a random fashion on a time scale of a few seconds. The amount of drift is dependent on the distance on the substrate away from the mirror, with more motion occurring further away. Fringes very close to the mirror are highly stable. All of these symptoms, especially the influence of the enclosure, indicate that the cause of the fringe motion is air currents. The use of a single beam in the Lloyd's-mirror system makes it tolerant to air currents which affect the entire beam uniformly. If

the currents are such that one side of the beam is affected differently than the other, then there will be a phase difference between the reflected and directly incident beams. The large diameter and expansion distance (l_p=2m) of the beam create ample opportunity for this to occur. Enclosure of the system suppresses airflow, and eliminates the problem. It is somewhat surprising that a design so robust to vibrations would be crippled by seemingly innocuous air currents. Nevertheless, the use of an enclosure is absolutely necessary for good fringe stability.

4.5 The Mirror

4.5.1 Phase Distortion

So far, the single-beam Lloyd's-mirror has proven to have a number of characteristics which are far more desirable than those of the Mach-Zehnder configuration: it is simple to align, highly robust, and easily reconfigured for different fringe periods. All of the advanced functionality is a direct result of the mirror itself. However, the use of the mirror comes with a penalty, and the price of its simplicity is decreased grating fidelity.

In the Mach-Zehnder interferometer, the last optical element the beams encounter before they interfere is the spatial-filter. With no additional optical elements to alter it, the spherical-wave approximation over the area of a 10cm wafer is accurate enough to be considered exact. Therefore, the exact phase progression of the exposed gratings can be analyzed, as is done in Chapter 3. The Lloyd's-mirror requires an additional optical element in the beam after the spatial-filter: the mirror itself. The mirror becomes a source of phase distortion in the spherical beam, as well as a source of scattering defects. Additionally, imperfect reflectivity makes it impossible to equalize the intensity of the two interfering beams, which reduces fringe contrast. This will be addressed in Section 4.5.2.

Thus, while it is easier to expose gratings with the Lloyd's-mirror, the product is potentially of lower quality.

Imperfections of the mirror can be roughly divided into two categories. The first is flatness. It is impossible to make a mirror perfectly flat. Most mirrors are specified with a flatness value as some fraction of 663 nm, the wavelength of a HeNe laser. The mirror currently in use is specified as $\lambda/4$, meaning that there is a maximum variation of 166 nm between the high and low points of the mirror. The accumulated phase error of the reflected beam will be double this, or $\lambda/2$. Phase error in the beam will directly print as a distortion from the ideal position of the interference fringes. For even a mirror of modest quality, these changes will be slow, and the distortions caused by them insignificant for most users. In general, one would not want to use the Lloyd's-mirror for metrological gratings, although the distortions of the mirror could be mapped to create a distortion map of the grating. Such distortions will be systematic and repeatable as long as the same mirror is used.

The second, and far more damaging, class of defects is local surface defects. These could be scratches, oily fingerprints, dust particles etc. Defects of this kind will act as additional coherent point sources. The recorded interference pattern will no longer be simply a grating formed by two beams, but the sum interference of the two beams plus the many additional point sources on the mirror. Defects large enough to be visible to the human eye will print directly in the exposure due to the loss of reflected power. Of course, in most cases the amount of power in the primary beams will be much larger than the power radiated from these defects. The observed exposure will still be a grating, but the local spatial-phase of the lines will be altered and there may be "rippling" of the lines on a scale larger than the period. It is obvious that great care should be taken to keep the mirror clean. One should always bear in mind, though, that a mirror which *appears* to be clean will always have some defects that will inevitably distort the spatial-phase progression. For many applications, however, the long-range phase of the grating is not important.

4.5.2 Reflectivity

Another aspect of the Lloyd's-mirror design is imperfect mirror reflectivity. Contrast of the interference fringes will be reduced because of unequal intensity in the interfering beams. It is intuitive that the mirror chosen for this application should have high-reflectivity. The wide range of incidence angles used to expose different period fringes also requires that the reflectivity be largely independent of angle. Thus, despite reduced reflectivity, metallic mirrors are preferable to dielectric mirrors, which exhibit high sensitivity to angle. At this wavelength, aluminum is the best reflector, with a minimum value of 0.85. Assuming uniform plane-wave illumination, the intensity of the interfering beams I_1 and I_2 will be scaled simply by the mirror reflectivity R.

$$I_2 = R \cdot I_1 \qquad 4.1$$

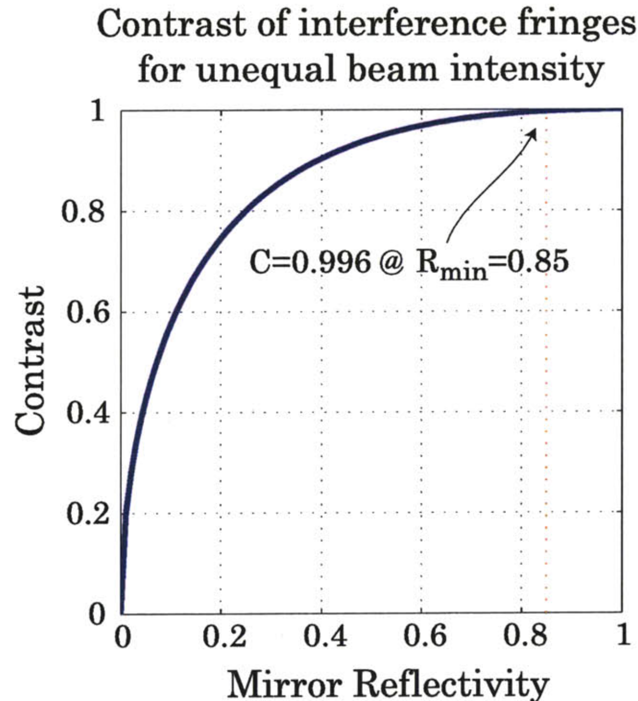


Figure 4.10 Contrast (visibility) of the interference fringes in a Lloyd's-mirror configuration as a function of mirror reflectivity (R). The minimum reflectivity R=0.85 for a commercial UV aluminum mirror is shown, with the corresponding contrast value. TE polarization is assumed.

Figure 4.10 plots the fringe contrast as a function mirror reflectivity. Fringe contrast remains high even for large differences in the intensity of the interfering beams. This forgiving shape is ideal for practical applications in interferometry, where it may be difficult to achieve equal intensity in both beams. For the Lloyd's-mirror, where the minimum relative intensity is 0.85, the contrast remains almost ideal at 0.996. It is interesting to note also that interfering beams of very different intensity can have surprisingly high contrast. This underscores the necessity to prevent sources of coherent scattering, such as mirror defects or dust particles.

4.5.3 Mirror Size

Not only are the optical properties important, but the physical size of the mirror will affect fringe formation as well. For uniformity and linewidth control, it is desirable to have the diameter of the expanded beam be significantly larger than the exposure area. This will likely also mean that the beam is larger than the mirror, and so diffraction or scattering from the edges of the mirror will be a source of distortion in the spatial-phase of the fringe pattern. Use of a mirror large enough to prevent this would be impractical in most cases.

The more present impact of the mirror size is the limiting effect it has on exposure area. Obviously, interference can only occur in regions where there is a reflected beam. The amount of the reflected beam that is intercepted on the substrate will be a function of the incident angle and the mirror size, shown in Figure 4.11. For small-period gratings, the distance l_m will be large enough that the exposed area will be limited by other factors, such as coherence or dose uniformity. The exposed area of large period gratings, however, will be limited by l_m the mirror dimension perpendicular to the interferometer axis. The distance from the mirror over which interference will occur is given in Equation 4.17.

$$l_m = l_w \tan(\theta) \tag{4.17}$$

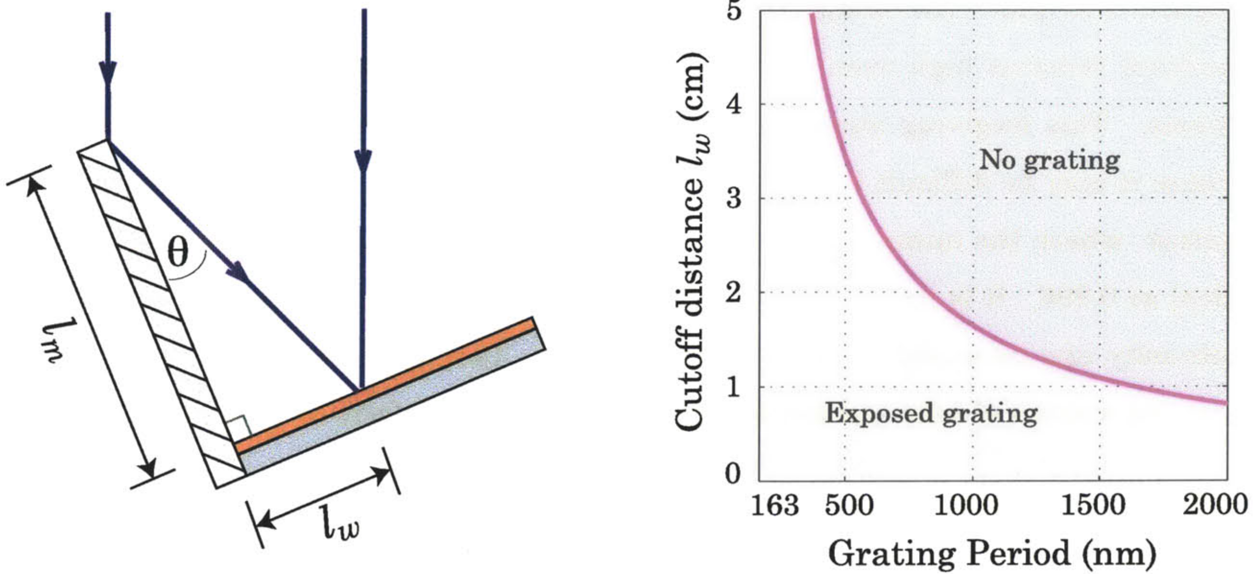


Figure 4.11 Maximum distance lw from the wafer for which interference will occur is a function of the incident angle θ and the grating size. The plot on the right shows lw as a function of the grating period for a 10 cm mirror.

4.6 Coherence and Contrast

In Chapter 2, we described how the visibility of interference fringes decreases as the phase delay between interfering beams increases. This depends on the temporal coherence of the beam, which was defined in terms of the frequency bandwidth of the emitting source. A coherence length l_c was defined as the distance over which the phase delay reduces fringe contrast to e^{-2}, or ~14%. The HeCd laser used with the Lloyd's-mirror interferometer has a bandwidth of ~1 GHz, and therefore a coherence length l_c=~30 cm. The contrast, or visibility *(V)*, of interference fringes as a function of optical path difference (OPD) derived in Chapter 2 is repeated in Equation 4.18, where c is the speed of light, σ is the standard deviation of the source bandwidth, and l is the OPD.

$$V(l) = e^{-\left(\frac{\pi}{c}\right)^2 \sigma^2 l^2} \qquad (4.18)$$

The contrast as a function of OPD for the HeCd laser used in the Lloyd's-mirror is shown in Figure 4.12.

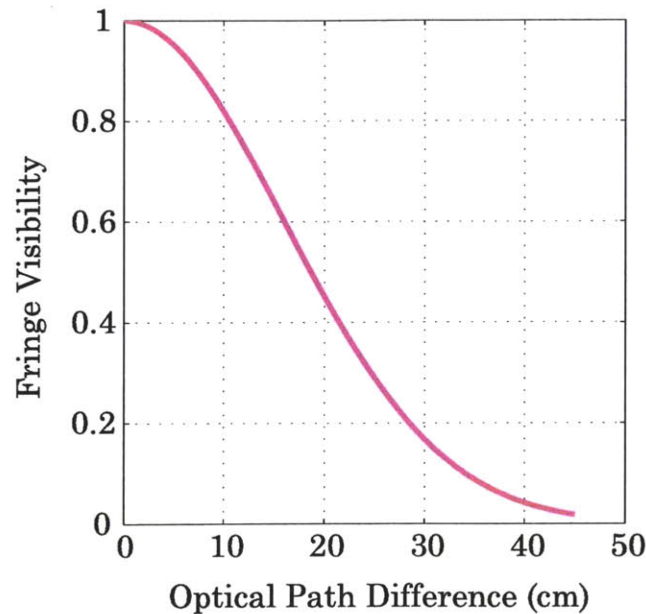


Figure 4.12 Fringe visibility as a function of the optical-path difference for in the Lloyd's-mirror. The source is a HeCd laser, λ_0=325 nm, with a Gaussian bandwidth of 1 GHZ.

Of course, to understand how the reduction of fringe contrast will impact the exposure area we need two more pieces of information. Most importantly, we must know the relationship between the on-wafer distance, measured from the mirror, and the path length difference. Secondly, we must know what level of contrast is necessary to make a good exposure, which is determined by the ability of the photoresist to approximate a binary response. Somewhat confusingly, this is also referred to as contrast.

The geometry of the Lloyd's-mirror configuration determines what the contrast will be on a given area of the substrate. The difference in optical path length between the two arms of the interferometer, i.e. the light directly incident on the substrate and that reflected off the mirror, is nonzero except at the axis between the mirror and the substrate. At positions on the substrate away from the center, the difference in path length increases. Therefore, the contrast of the interference fringes decreases in areas further away from the mirror. The difference in path length also increases as the spatial period of the grating decreases. Figure 4.13 depicts schematically the two contributions to OPD in the Lloyd's-mirror.

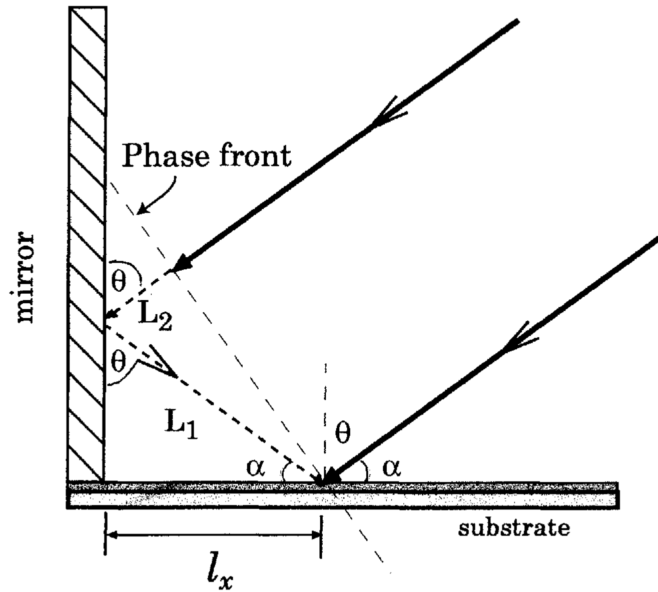


Figure 4.13 The reflected beam accrues additional path length components L_1 and L_2 compared to the directly incident beam. The total OPD is dependant on the incident angle θ and the distance from the interferometer axis l_x

The total OPD in the Lloyd's-mirror can be thought of in two components, L_1 and L_2, which vary with l_x, the distance on the substrate away from the mirror and the angle of incidence θ of the incoming light. For simplicity, we also make use of the complementary angle $\alpha=90^{\circ}-\theta$. The L_1 component defined in Equation 4.19 represents the additional distance that the light must travel in the reflected arm after it hits the mirror.

$$L_1 = \frac{l_x}{\sin(\theta)} \tag{4.19}$$

$$L_2 = L_1 \cos(2\alpha) \tag{4.20}$$

The L_2 component, defined in Equation 4.20, represents the difference between when the phase front reaches the substrate and when arrives at the mirror. This can be positive or negative depending on the incident angle. The optical path difference (OPD) for a given distance l_x is simply the sum of L_1 and L_2.

$$OPD = L_1 + L_2 = \frac{l_x}{\sin(\theta)} \cdot [1 + \cos(2\alpha)] \tag{4.21}$$

Using this measure of the OPD in the Lloyd's-mirror with Equation 4.21, we can get a measure of the fringe contrast at a given point on the substrate for a given period. Figure 4.14 shows a contour plot of the contrast for distances up to 10 cm from the mirror, and periods between 163 nm (the theoretical minimum) and 1000 nm. Two interesting conclusions can be drawn from this graph. The contrast of the fringes will be high even far away from the mirror for large period gratings.

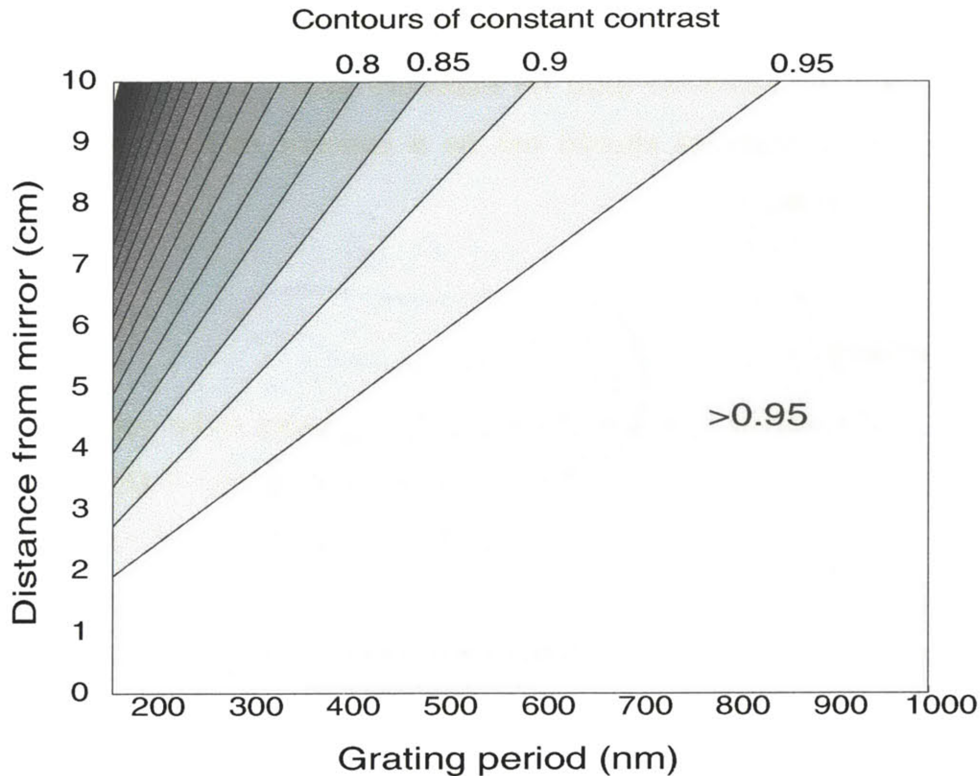


Figure 4.14 Contours of constant contrast in the Lloyd's-mirror. Fine-period gratings lose contrast much more rapidly than large period gratings. In general, the maximum distance from the mirror will be 5 cm for the current LM design, and contrast will be greater than 0.7 for all periods.

This should be intuitively clear, because the light is incident almost normal to the substrate for large periods. The area of exposure for large period gratings will be limited by the size of the mirror rather than contrast of the fringes as seen in Figure 4.11. For grating periods close to the lower limit, fringe contrast drops off rapidly as the distance away from the mirror is increased. As a practical example, consider a 200 nm period grating exposed on a 10 cm diameter wafer. Since the Lloyd's-

mirror only exposes 1/2 of the wafer at a time, the distance from the mirror of the outer edge of the wafer is 5 cm. From Figure 4.14, the fringe contrast of a 200 nm period grating at 5 cm from the mirror is approximately 0.8, which is still quite high. However, if we move to a 165 nm period, close to the theoretical minimum period possible with a 325 nm laser, the fringe contrast drops to about 0.65 at 5 cm from the mirror. This level of contrast may be too low for such high-resolution patterns in certain photoresists. In general, for fine-period gratings the fringe contrast will set the practical limit on exposure area. Under normal conditions, though, the fringe contrast should not be a concern in exposures done on the Lloyd's-mirror system.

4.7 Summary

The Lloyd's-mirror interferometer has been shown to be equivalent to half of a Mach Zehnder interferometer. The enforced symmetry of the system via the creation of an image source leads to high fringe stability and ease of alignment. This also allows for the periodicity of the fringes to be arbitrarily varied with the rotation angle, without otherwise changing the optical configuration. Although the spatial-filter ensures that the illumination is a pure and clear spherical wave, placement of the mirror in the optical path means that the incident waveform will be distorted. Scattering noise, especially from dust, can also be problematic. However, for applications where the spatial-phase coherence of the exposed grating is not important, the Lloyd's-mirror is a near ideal tool. In its current configuration, the Lloyd's-mirror can expose gratings down to ~170 nm in pitch over areas up to $25cm^2$ with a high degree of uniformity. An example micrograph of a 170 nm period grating in photoresist is shown in Figure 4.15. The combination of unique traits of the Lloyd's-mirror has enabled a large variety of applications for which a Mach-Zehnder style interferometer would be too cumbersome. A selection of these are described in Chapter 7.

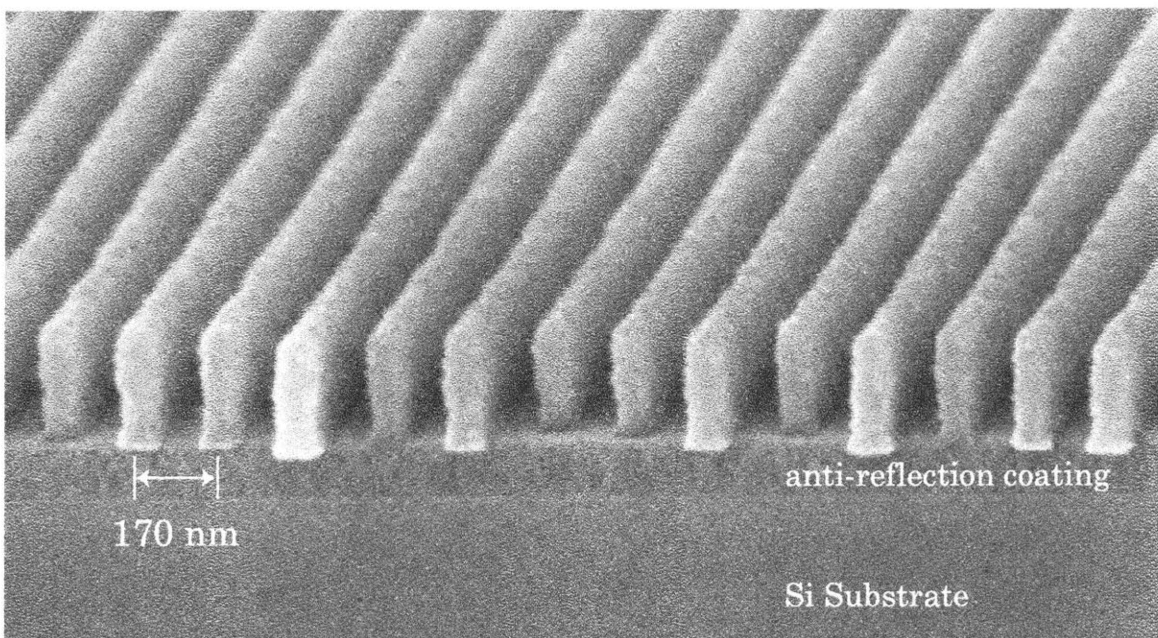


Figure 4.15 Photoresist lines at 170 nm pitch exposed using the Lloyd's-mirror.

4.8 References

[1] M. Lim, J. Fererra, K.P. Pipe, H.I. Smith, "A holographic phase-shifting interferometer technique to measure in-plane distortion". *J. Vac. Sci. Technol. B* **17**(6) 2703 (1999)

[2] M.E. Walsh, H.I. Smith, "Method for reducing hyperbolic phase in interference lithography". *J. Vac. Sci. Technol. B* **19**(6) p.2347 (2001)

[3] S. Tolansky, *Introduction to Interferometry*, Longmans, Green & Co., New York, 1955

[4] D. Malacara, *Optical Shop Testing* 2nd. Ed. John Wiley&Sons, New York, (1978)

[5] J.T Hastings, "Nanometer-precision electron beam lithography with applications in integrated optics", Ph.D. Thesis, Masschusetts Institute of Technology, (2003)

Chapter 5
Grating Interferometers

5.1 Introduction

5.1.1 History

Mach-Zehnder-style interferometers used for lithography, as covered in the last chapter, are a natural extension of the more traditional use of an interferometer as a metrological tool. Most readers will be familiar with the many types of interferometer that can be configured such that the interference fringes can be interpreted to measure some physical parameter. Among the more commonly measured quantities are distance or spatial-position, index of refraction, aerodynamic flow, and the figure of optical elements. Unfortunately, the same design and operational conditions that allow for sensitive measurement can also be problematic, leading to stability problems and the necessity for high source coherence. Early in the twentieth century, grating-based interferometers began to appear which provided the same capabilities while ameliorating many chromatic and stability issues. The use of diffraction gratings as beamsplitting elements in

interferometers dates back to at least 1911 [1]. Ronchi, one of the most influential contributors to the field, began his work in the early 1920's on grating interferometers for the characterization of optical elements [2]. He provides a fascinating and personal account of the early history of these devices in a review article from the 1960's [3]. The work of Weinberg and Wood [4], often credited with the first grating interferometer of a style that would be adapted for lithography, was a direct outgrowth of work by Kraushaar [5] and others [6,7] in the early 1950's, all of whom were investigating aerodynamic flow. After the development of the laser popularized Gabors' new technique of holography [8] it did not take long for grating interferometers, heretofore used only for measurements, to find their roles reversed. Whereas in earlier experiments the fringe pattern was observed to infer the structure of interfering wavefronts, much interest was now focused on how to structure wavefronts for the construction of a specific interference pattern. Interest was rekindled in the historic observations of grating self-imaging by Talbot [9], and the work of Lau, who demonstrated the formation of fringes in grating based interferometers which were both spatially and spectrally incoherent [10].

The secure analytical footing of grating interferometers used for holography and lithography can be credited to the research group of Leith in the 1970's, one of the founding fathers of holography [11,12,13,14]. Through the 1980's, Leith's group [15,16,17] and others [18] continued to develop theory and lithographic applications of the grating interferometer. Following the designs of Hershey and Leith, a number of groups have since developed achromatic interferometers for the exposure of large-area gratings [19,20]. In addition to linear gratings, the achromatic interferometer was demonstrated with crossed gratings (grids) and circular gratings [21,22,23].

Contemporary to these developments, work here at MIT was focused on producing the smallest possible fringe periodicities. Using the near-field spatial-frequency doubling of a single grating, 99nm-period gratings were demonstrated in 1983 [24]. Subsequently, Anderson developed an improved, two-grating

interferometer operating at the 193nm wavelength of an ArF laser for the production of 100nm-period gratings [25]. Further improvements on the design [26,27] culminated with the development of a three-grating interferometer for 100nm-period gratings and grids by Savas with improved stability and exposure area [28,29]. Shown below in Figure 5.1 is an example of a grating exposure from this system.

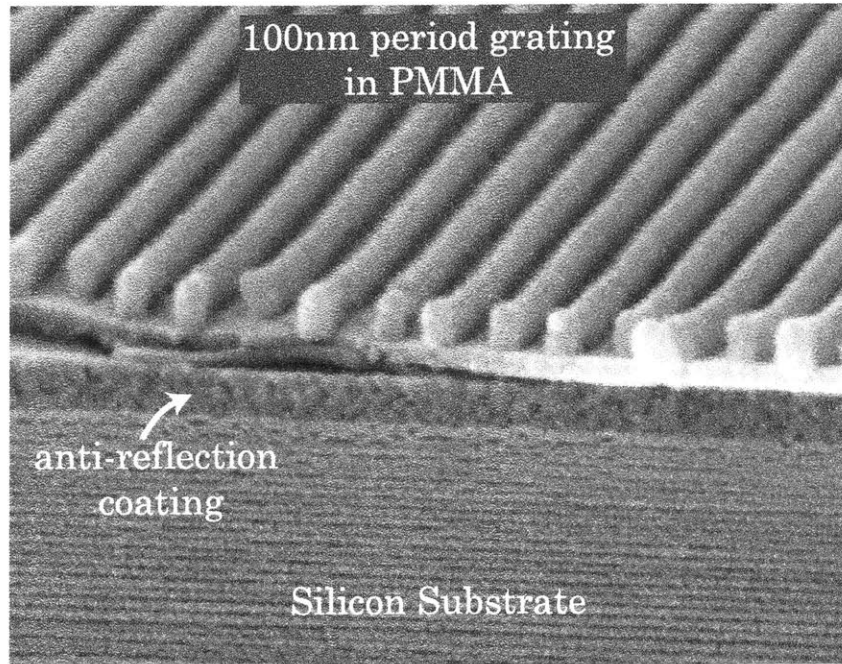


Figure 5.1 A 100nm period grating exposed in PMMA using the three-grating achromatic interference lithography system. An additional layer of anti-reflection coating is used under the PMMA to improve the exposure characteristics.

5.2 Single-grating Interferometers

5.2.1 Near-field spatial frequency doubling

In Chapters 3 and 4, the optical setup has been emphasized more than the light source for its effect on the IL process. However, two parameters of the source are of critical importance. The first is wavelength - recall from Equation 3.1 that the lower limit of the exposed grating period is $\lambda/2$. Therefore, to get a finer-period

structure, a shorter wavelength must be used. The second parameter is the coherence of the light source. The physical size of the grating which can be printed using either the Mach-Zehnder or Lloyds-mirror interferometers, i.e. the number of teeth in the grating, is limited by the temporal coherence of the source. In the ideal case of a perfectly collimated (spatially coherent) source, Equation 5.1 gives the exposed grating width W in terms of the bandwidth of the source, $\Delta\lambda$, where P is the periodicity of the grating.

$$W = P \cdot \frac{\lambda}{\Delta\lambda} \tag{5.1}$$

For instance, the Lloyds-mirror, which uses a 325nm helium-cadmium laser with $\Delta\lambda=3.5\times10^{-4}$ nm, can potentially expose a 200nm-pitch grating up to 18cm wide. To produce gratings with a usefully large area at smaller periods, a source with both high temporal coherence and a short wavelength is necessary, and this is where the problem arises. Short wavelength laser sources currently available at 157 nm, 193nm, and 248nm are all pulsed excimer lasers. Although some line narrowing features are available in more costly models, temporal coherence is notoriously low in all excimer lasers, with $\Delta\lambda=\sim0.5$ nm. An alternate approach based on the diffraction from so-called "parent" gratings has been developed to eliminate the need for coherence.

These interferometers take advantage of the dispersive properties of diffraction gratings: different wavelengths of light are diffracted at different angles in such a way that the periodicity of the interference pattern is the same for all wavelengths of light. Because the wavelength and angle of the interfering beams are coupled by the grating equation, the periodicity of the interference pattern is now dependent on a third parameter, the period of the parent grating. For example, consider a parent grating with period P_0 illuminated with a polychromatic (low temporal coherence) source at normal incidence. The diffraction angles θ_m for a given diffraction order m from the parent-grating are given by Equation 5.2. Solving for the periodicity of fringes P_f formed by first-order ($m=1$) beams

interfering at these diffraction angles (Equation 5.3), we find that the periodicity of the fringes, or "daughter-grating", is equal to 1/2 that of the parent *regardless of the wavelength*. This is known as spatial-frequency doubling (SFD).

$$\sin(\theta_m) = \frac{m\lambda}{P_0} \qquad (5.2)$$

$$P_f = \frac{\lambda}{2\sin(\theta_1)} \qquad (5.3)$$

$$P_f = \frac{P_0}{2} \quad m = 1 \qquad (5.4)$$

This type of achromatic interference is shown in Figure 5.2. The parent grating is optimized such that only first-order beams propagate, and the zero-order beam is eliminated. In the region below the grating where the diffracted beams overlap, a spatially-invariant image at half the pitch of the parent-grating is formed. Of course, the illuminating wavelength does not drop out of the equation entirely; the grating period must be sufficiently large compared to the wavelength such that diffraction can occur. In the limit of Equation 5.2 where $m=1$ and $sin(\theta_m)=1$, the minimum possible grating period for a given wavelength is found to be $P_{f\min}=\lambda/2$.

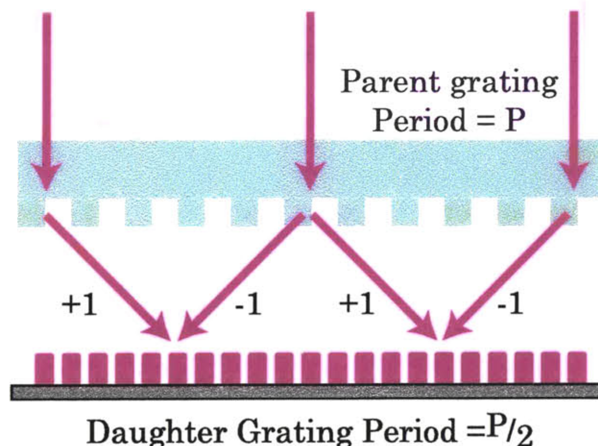


Figure 5.2 Near-field spatial-frequency doubling using first-order diffraction from a parent-grating. The periodicity of the daughter-grating is half that of the parent.

5.2.2 Zero-Order Diffraction Efficiency

It was optimistically assumed for the ideal near-field SFD that the parent-grating has been optimized such that the zero-order is cancelled and only the first-order diffracted beams will propagate. One can theoretically achieve this for parent gratings with a sinusoidal phase profile where the accumulated phase difference between the peak of the tooth and the trough is half a wavelength. Similarly, in square-profile phase gratings one can also achieve this in the case where the duty cycle is 50% and the phase difference between the transmitted light in the tooth and trough is half a wavelength.

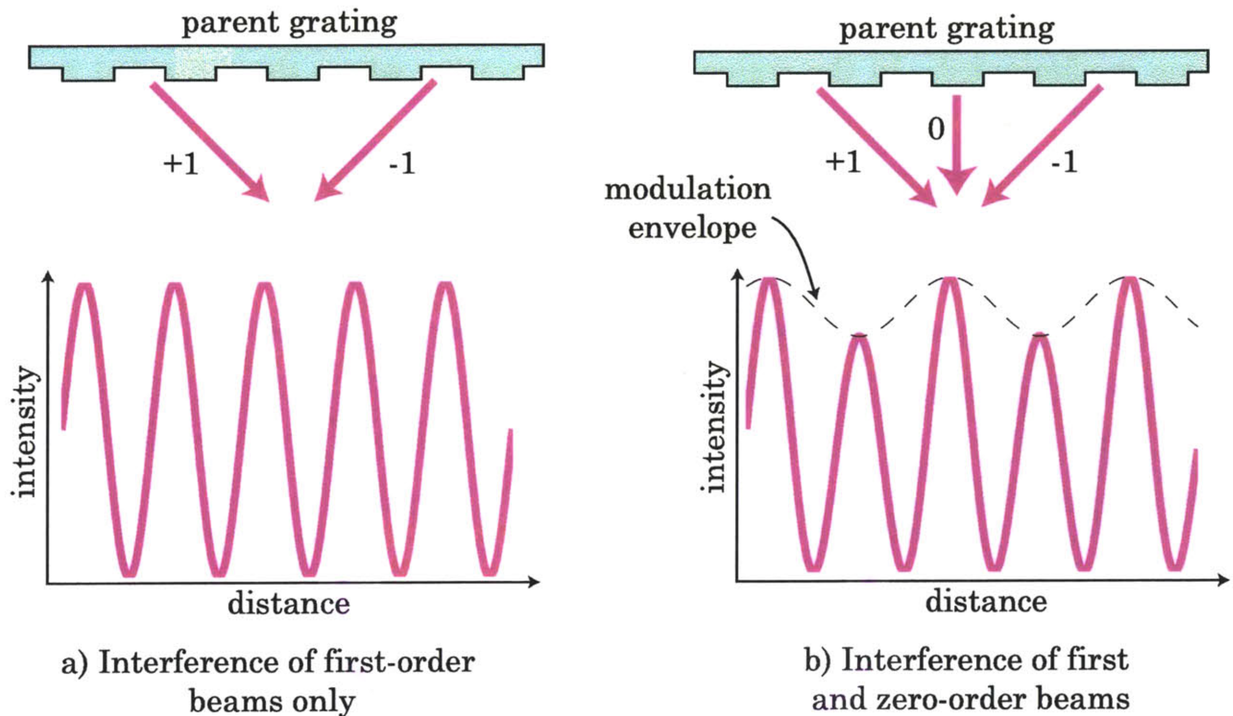


Figure 5.3 a) The interference of only first-order diffracted beams creates a high-contrast fringe pattern with half the periodicity of the parent grating. B) The presence of zero-order beam modulates the frequency-doubled fringe pattern with an envelope at the same periodicity as the parent grating. The amplitude of the modulation envelope depends on the relative intensity of the zero order beam.

Unfortunately, these conclusions are only valid under scalar diffraction theory where the wavelength is much smaller than the period. In the regime which

interests us most, the smallest-period gratings can only be achieved if the periodicity of the parent-grating is approximately equal to the wavelength. Thus, it is unlikely that the zero-order can be fully suppressed in the cases of most interest. A more thorough discussion of the effects of the zero-order beam on the resulting interference pattern can be found in Chapter 6. For now, it will suffice to simply state that the resulting interference pattern will include a modulating envelope at the same frequency as the parent grating, shown in Figure 5.3. The strength of this envelope will be dependent on the relative intensity of the zero-order beam, but will also be dependent on z, the distance below the grating.

5.2.3 Spatial coherence

In the simple model shown in Figure 5.2, we also assumed that the incident light was perfectly collimated. In reality, this will not be true. The short wavelength laser sources desirable for the production of fine-pitch gratings possess limited spatial-coherence in addition to limited temporal-coherence. We can consider the effect of limited spatial-coherence as the presence of light incident at off-normal angles. The effects of this are presented following the discussion by Savas [29]. For simplicity we consider only the first diffracted orders in the absence of the zero-order.

As shown in Figure 5.2, light normally incident on the grating forms interference fringes perpendicular to the plane of the grating. For off-normal angles of incidence, the interference fringes form at different angles relative to the parent grating, i.e. a laterally sheared version of the original pattern. Because of the angle between the on-axis and off-axis fringes, they will progressively become out of phase at increasing distances from the parent grating, shown in Figure 5.4. Thus, the image contrast will degrade as the distance from the parent-grating is increased. The minimum level of contrast in the interference pattern that can be tolerated while still forming a good resist image will determine the depth-of-focus (DOF).

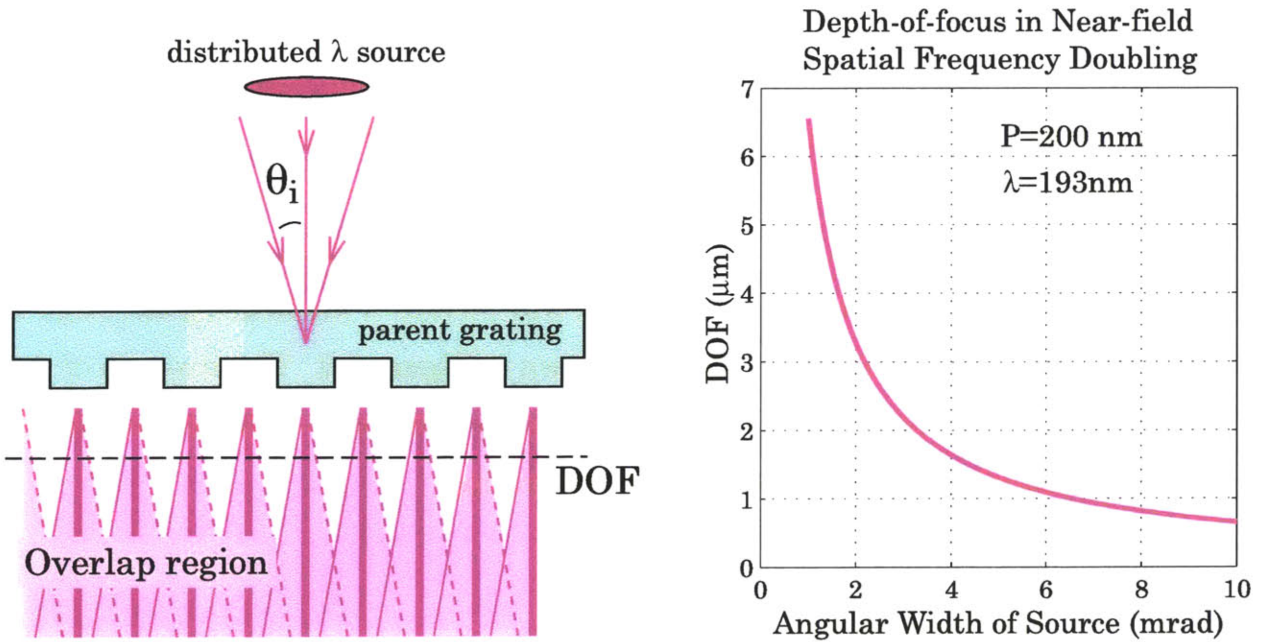


Figure 5.4 The effect of a limited spatial-coherence due to a finite source width for near-field spatial-frequency doubling is to limit the distance from the parent-grating at which high-contrast interference fringes are formed. Off-axis illumination of the parent-grating forms interference fringes at an angle to the normal, leading to fringe overlap at increasing distances from the parent-grating. The depth-of-focus for a 200nm-period parent grating under 193 nm illumination is shown in the plot for a range of angular source widths, calculated from Equation 5.9.

At a given plane located a distance D below the parent-grating, the sheared fringes will be laterally shifted from their nominal position by an amount Δ proportional to D and the tangent of the shear angle $\Delta\theta_0$.

$$\Delta = D\tan(\theta_0) \qquad (5.5)$$

Savas defines the DOF as the point where the shear of the fringes causes a maximum lateral displacement of the daughter-grating of $\Delta=P_2/4$, degrading the contrast of the interference pattern to ~0.63. The required degree of spatial-coherence measured as an angular spread of the source can be found by differentiating the standard diffraction equation (Equation 5.5), where θ_i is the angle of incidence, and θ_0 is the angle of diffraction for normal incidence.

$$\frac{m\lambda}{P_1} = \sin(\theta_i) + \sin(\theta_o) \qquad (5.6)$$

$$\frac{\partial\theta_0}{\partial\theta_i} = \frac{\cos(\theta_i)}{\cos(\theta_0)} \qquad (5.7)$$

If we note that the derivative is maximized when the numerator is unity, and that we will generally be dealing will small variations of the incidence angle about the normal, then the assumption of $cos(\theta_i)=1$ affords both a mathematical simplification and a margin of error useful in engineering practice. Rewriting this as the maximum allowable change in diffraction angle as a function of the spread in incidence angle we get

$$\Delta\theta_0 = \frac{\Delta\theta_i}{\cos(\theta_0)} \qquad (5.8)$$

This change in diffraction angle causes the shear of the interference pattern, and the shear angle is simply $\Delta\theta_0$. Rewriting Equation 5.5 as the DOF for the maximum shear angle, and using the small-angle approximation $tan(\theta)=\theta$, we get

$$DOF = \frac{\Delta}{\tan(\Delta\theta_0)}, \qquad (5.9)$$

$$DOF = \frac{P_2\cos(\theta_0)}{4\Delta\theta_i}. \qquad (5.10)$$

Depth-of-focus is plotted in Figure 5.4 for exposures with λ_0=193 nm and a parent grating with periodicity P_1=200nm. The maximum allowable distance between the parent-grating and the exposure plane is on the order of 2-3 μm for an angular width of ~2-3 mrad, which is characteristic of the 193nm ArF laser used in our lab. Although possible, it is certainly not trivial to reliably position two planar surfaces within less that 3μm without potentially damaging one of the surfaces. Adding to the complexity, silicon wafers often vary in thickness on the micron scale, and the presence of any particles on either surface could foil the required proximity.

5.3 Multi-grating interferometers

5.3.1 Two-Grating Interferometer

Although attractive in its simplistic approach, the practicality of near-field spatial-frequency doubling from a single grating is limited by the necessarily close proximity of the exposure plane to the parent-grating. These issues can be mitigated through the use of additional parent gratings. Figure 5.5 shows a second grating, called the recombiner, placed at a distance d_p behind the first parent grating, which we will refer to as the beamsplitter grating. When these two gratings are of the same period, the *(-2)* diffracted orders from the recombiner grating interfere to form fringes at a distance d_p from the recombiner. By setting the distance d_p such that the diffracted orders are fully separated, a physical stop can be placed to block the zero-order.

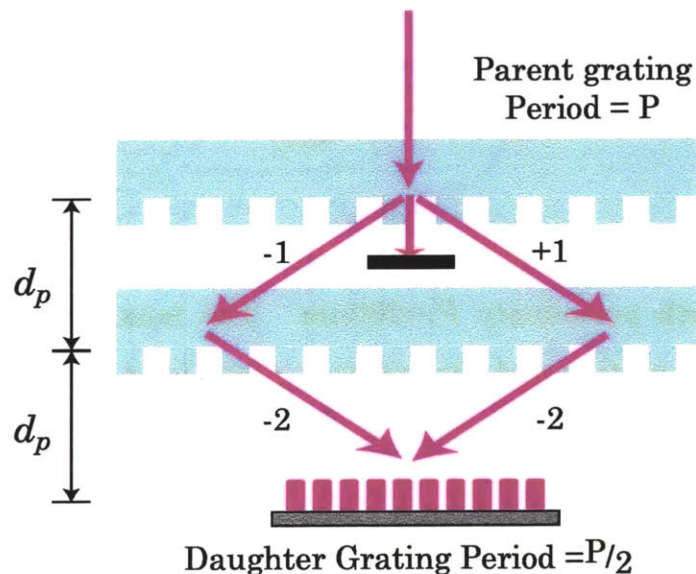


Figure 5.5 A two-grating interferometer uses first-order diffraction from the beamsplitter grating and second-order diffraction from the recombiner grating, with the focal plane formed at a distance d_p equal to the separation of the two gratings. The zero-order beam from the beamsplitter grating can be physically blocked.

Additionally, the location of the exposure plane has now been moved to a macroscopic distance from the parent. The depth-of-focus (*DOF*), although still limited by the spatial-frequency of the source, is improved by a factor of two.

$$DOF = \frac{P_2 \cos(\theta_0)}{2\Delta\theta_i} \qquad (5.11)$$

The multi-grating achromatic interferometer, therefore, is a more practical approach for the production of fine-pitch gratings. The two-grating interferometer described above was implemented at MIT for the production of 100 nm period gratings by Yen *et al.* [26]. In practice, it was found that the area of the exposed grating was limited by the requirement that the beamsplitter and recombiner gratings must be matched in periodicity to better than ~1 part in 10^5 [29]. If we consider that the parent gratings used in this type of interferometer will be made using a Mach-Zehnder type IL scheme, as described in Chapter 3, then we must account for the hyperbolic spatial-phase of the parent-gratings. In Chapter 3, we showed that the variation in periodicity of the parent-gratings could be as high as a few parts in 10^4. Thus, despite the ability to make large parent-grating using IL, the actual usable area is limited by spatial-phase non-linearities.

5.3.2 MIT Achromatic Interference Lithography

To circumvent the limitations of the two-grating design, the recombiner grating was split into two separate gratings. Substantial work by Savas has brought this design to an advanced state. Pictured in Figure 5.6, this system is known as achromatic interference lithography (AIL). The potential exposure area can be as large as the full parent grating. The gratings can be positioned such that the local periodicity is more closely matched in each of the arms, allowing for more of the area of the parent gratings to contribute. An example of the 100 nm period gratings produced by this system is shown in Figure 5.1.

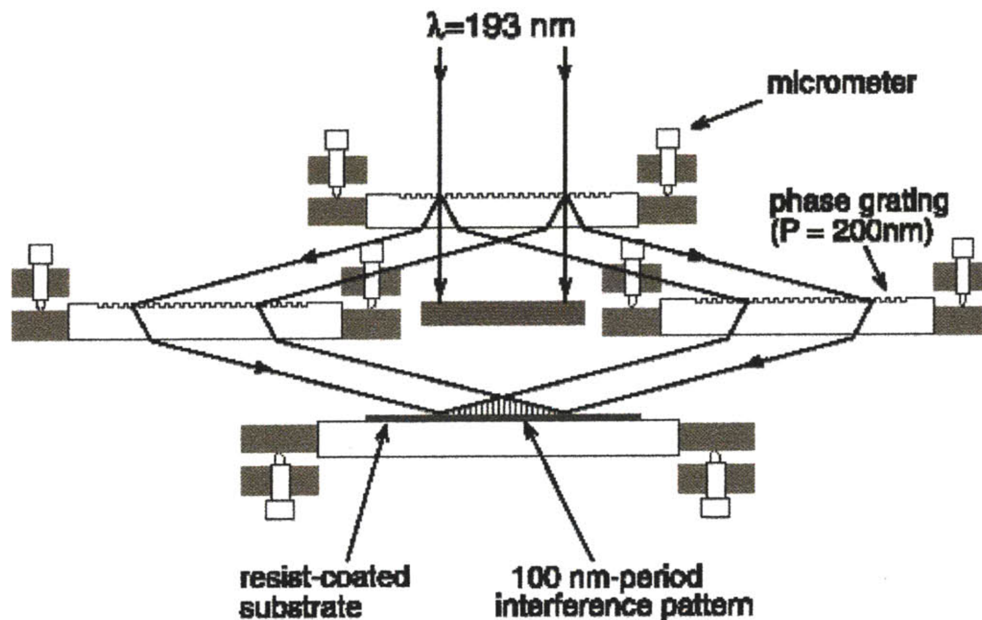


Figure 5.6 Schematic of the MIT achromatic interference lithography (AIL) tool. An ArF laser provides 193nm light which illuminates a set of 3 matched parent gratings of period P_1=200nm. The interfernce pattern recorded on the substrate has a period of P_2=100nm.

5.4 Next Generation Tools

5.4.1 Introduction

The nature of the grating interferometer, using a set of parent-gratings to make daughter-gratings with half the original period, invites bootstrapping from the current generation to a finer-period one. This begs the question of whether it is possible to use the 100nm-period gratings from the current AIL tool to make 50nm-period gratings. There are a number of compelling reasons for trying, some of which will be discussed in Chapter 7 on Applications. The interest in moving down another generation using 100nm-period parent-gratings is not something new. A proposed solution was published more than 10 years ago based on soft x-ray synchrotron radiation at λ=14nm to double 100nm-period parent gratings made in gold on a silicon membrane [30]. A similar approach was demonstrated shortly afterward by Anderson at the Lawrence Berkeley National Laboratory with unconvincing results [31]. Good results based on this approach have proved elusive. Very recently, this scheme has been resurrected by Nealey at the University of Wisconsin and Solak at the Paul Scherrer Institute to produce gratings with periodicity down to ~50 nm, using an undulator attached to a synchrotron [32,33]. Like Anderson's attempt 10 years ago, the recent results are based on a single, segmented parent grating, shown in Figure 5.7. However, this design requires much higher spatial coherence than the previously described interferometers. Shear of the interference fringes occurs in exactly the same manner as shown in Figure 5.4, but in this case the imaging occurs in the far-field rather than the near-field. The necessity of a large, highly-coherent beam makes the production of fine-period gratings using this technique impossible without a synchrotron source. The inability to-date to make *large-area*, sub-100 nm period grating indicates that there are still substantial challenges to this seemingly straightforward process.

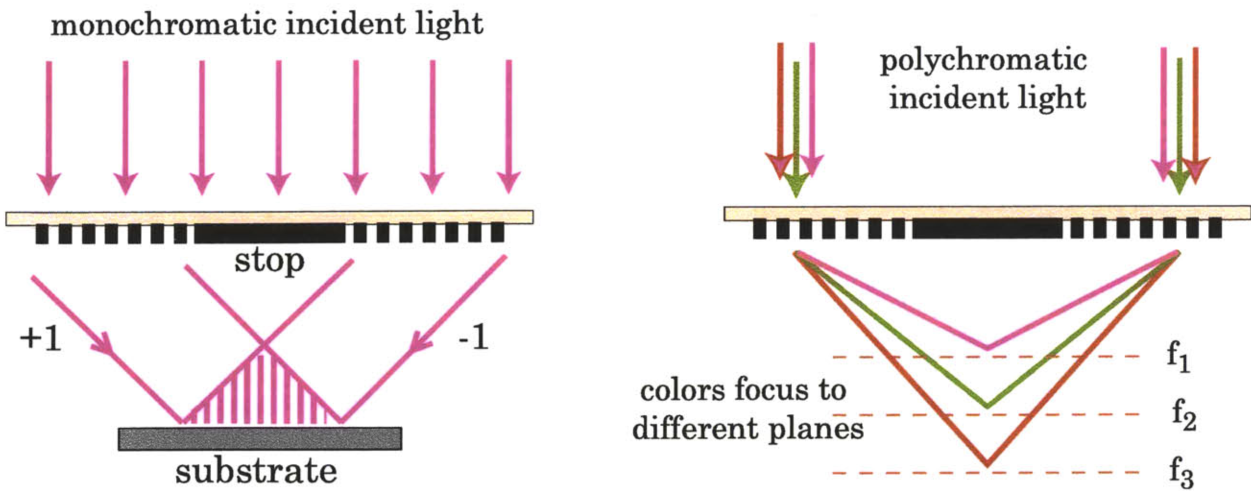


Figure 5.7 Spatial-frequency doubling from a single segmented parent grating. The zero-order stop in the center of the field moves the image plane to the far-field, requiring much higher-spatial coherence than other schemes.

5.4.2 Short-wavelength photon sources

The most fundamental requirement for any grating interferometer is that the wavelength must be smaller than the period of the parent-grating to support diffraction. For example, in order to use parent-gratings with a period P_1=100 nm, it is necessary to use a wavelength less than 100nm. This is where the problems begin. Photon sources with wavelengths less than 100nm have an unappealing combination of poor spatial coherence, low power, high cost, and lack of commercial availability. The lack of a suitable source is the most commonly stated reason for the failure to-date to double the current generation of 100nm-period gratings. However, this is largely an issue of money and available technology. It is not a coincidence that those groups to demonstrate ~50 nm periods so far - Anderson Nealy and Solak, have access to synchrotrons with undulators on dedicated lithography beamlines. Although it is not impossible to locate other sub-100 nm photon sources, their specifications are far from ideal, lacking in both output power and spatial coherence.

5.4.3 Material Properties in the VUV

While the main problem with acquiring an appropriate source is economic, a more fundamental problem is the optical properties of materials at the necessary wavelength. Absorption in all materials tends to increase as wavelengths decrease through the VUV band [34]. The name itself, VUV, is the acronym for vacuum ultra-violet, indicates that even atmospheric gasses are absorbing. The lowest wavelength for which refractive optics are possible would probably be 121 nm, from a discharge source emitting at the Lyman-α of hydrogen [35,36]. This is at or very near the cutoff for the highest grades of fluorinated crystals, such as LiF, CaF_2, or MgF_2, and all of these present certain practical challenges. For example, all of these crystals are hygroscopic, leading to difficulty in even the simplest chores such as cleaning. LiF will absorb water from the atmosphere leading to an irreversible chemical reaction which will alter optical properties [37], as well as being mechanically soft to the point where optical polishing is impossible. All of these crystals are prone to radiation damage over time which can radically alter their absorption characteristics. Cutoff wavelengths and solubility in water for a selection of materials are given in the following table [38,39].

Table 5.5.1 Transmission cutoff for Selected Materials

Material	*Cutoff λ (nm)*	*H_2O Solubility (g/100cm^3)*
LiF	104.5	0.27
MgF_2	113.0	0.0076
CaF_2	122.0	0.0016
Sapphire	142.5	Insoluble
Quartz	145.5	Insoluble
Fused silica	152.5	insoluble

Below these cutoffs, refractive optics are generally impossible. Moreover, from wavelengths of roughly 110 to 20 nm, the reflective properties of materials are also very poor. High material absorption limits the reflectivity of even a well-designed multilayer mirror [34,40]. A few quasi-free-electron materials - aluminum, silicon, and beryllium - have sufficiently low plasma-frequencies that they can exhibit low absorption and high reflectivity at wavelengths down to ~60 nm. However, these materials must be remain free of oxide layers which form spontaneously even in minute concentrations of oxygen. It is generally considered impossible for these conditions to be maintained in a terrestrial laboratory, let alone a loosely controlled university environment. In the soft x-ray band below ~20 nm, absorbtions decrease but the refractive index of most materials tends towards unity, limiting index contrast in multilayers. Thus, impractically high numbers of layers can be necessary to achieve good reflectivity.

Separate from the limitations on optics, the availability of materials for lithography, specifically photoresists, is also a challenge at VUV wavelengths. One necessary property of photoresists, not widely recognized, is transparency. To ensure uniform imaging throughout the thickness of the film, absorption should be limited. A common guideline for the minimum transmission through a resist film is 40% [44]. As photon energies increase into the VUV band, the polymeric materials used for photoresists tend to follow the absorption of carbon. Figure 5.8 shows the absorption coefficient α for PMMA, a common DUV and x-ray photoresist, for the wavelength range of 60-130 nm [41]. It should be noted that optical data for all materials is extremely limited in this band, and that reported results can vary by a factor of 2 or more [42,43,44]. The absorption coefficient α ranges from ~60-130 per micron. This can be converted to a maximum film thickness using the following equation, where T is the transmission of the film.

$$thickness = \frac{-\log_{10}(T)}{\alpha} \qquad (5.12)$$

Assuming a transmission of 40%, we find that the maximum thickness of the resist must be below 5nm for the wavelengths between 60 and 100 nm.

Lithography using conventional processes with resists of this thickness is impractical. Recent work on the exposure and resist properties of self-assembled monolayers may enable imaging in ultra-thin layers [45,46,47], however this is still in its infancy. Extrapolating based on the absorption spectrum of carbon [48], the absorption of polymeric resists will be unacceptably high except at wavelengths below 10-15 nm. The successful use of PMMA for high-aspect ratio exposures using C_K radiation at λ=4.5 nm is testament that resist processes are possible using soft x-rays [49,50,51].

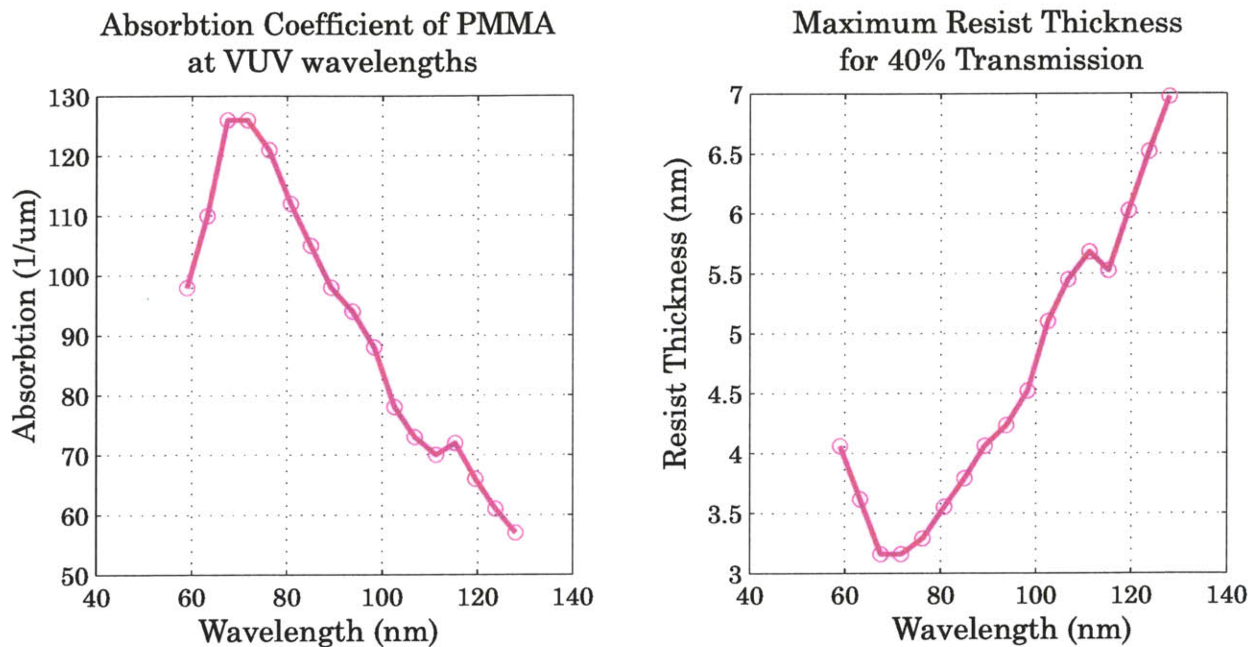


Figure 5.8 Absorption data from [41] for PMMA at VUV wavelengths. The maximum thickness of a photoresist layer which retains 40% transparency is shown in the right-hand plot. For wavelengths 20nm<λ<100nm, which would be necessary to make 50 nm period gratings, resist thicknesses are limited to less than 5nm.

In summary, due to material properties, the optical design for interferometers operating at sub-100nm wavelengths is extremely challenging. The extraordinary development efforts towards the proposed EUV lithography tools at 13 nm are evidence of the difficulties in this area. Large teams of people with vast sums of money have been stymied by even simple optical elements such as mirrors and masks. One can infer that development of even a simple lithography tool

operating at wavelengths between EUV and ~100nm, where material absorptions are much greater, would virtually impossible.

5.4.4 Soft x-ray AIL using membrane parent gratings

One solution, mentioned at the beginning of this section, is the use of gold gratings in thin membranes at a soft x-ray wavelength. There is a significant body of work which has shown that lithographic systems can be successfully developed in this regime. Although it may seem to be a logical step because it builds on existing x-ray technology, and membrane diffraction gratings have been successfully fabricated in the past for applications such as atom diffraction [29], ultimately I do not consider it to be the right choice. One important point is that the demonstrations to date have been for the spatial-frequency doubling of a single parent-grating over a small exposure area [31]. To move these results beyond a curiosity and into the realm of useful, large-area gratings, one must necessarily move to a three-grating design akin to AIL (Section 5.3).

While the relatively high transmission of thin membranes allows the architecture of the system to remain essentially unchanged, their mechanical properties and low diffraction efficiency will ultimately compromise the utility of the system. Large membranes of a few square centimeters have been demonstrated as x-ray masks with acceptably low distortion for that application. However, the crucial difference is that three membranes, not one, are necessary for this application. It is not clear that the combination of distortions in three separate membranes will be small enough, or stable enough, for interferometry. Savas gives the required period match for the three-grating configuration with a 27 nm photon as 2 parts in 10^4. Using the holographic phase-shifting technique described in Chapter 3, Lim has measured the periodicity changes in SiN_x x-ray masks at about 2 parts in 10^3, an order-of-magnitude greater than the tolerance of the lithographic interferometer in which they would be used [52]. Although this could potentially be

reduced through modification of the fabrication process, it is not guaranteed. Additionally, the membranes proposed for this application would be ~1 μm thick and millimeters to centimeters in diameter, making them prone to vibration. Depending on the nature of the vibrations, the position of the interference fringes can potentially also shift on the substrate during the exposure.

One solution to this problem would be to limit the size of the membrane to increase its mechanical strength. With smaller membranes, however, the utility of the system would suffer. A primary advantage of interference lithography is the capacity for small features over a large exposure area. With membranes (and exposure areas) of millimeter dimensions or less, this advantage is effectively lost. Additionally, the range of possible experiments will be limited to those which require only a small area of grating. Of course, these arguments are academic, in that there is no suitable soft x-ray source readily available to pursue this path.

5.4.5 Reflection gratings

An analogous interferometer using reflection gratings rather than transmission gratings, shown in Figure 5.9, can potentially solve some of the problematic issues associated with membranes. First and foremost, the stability of bulk substrates on which the reflection gratings reside allows for large parent gratings, which in turn allows for large exposure areas. A secondary advantage may come with the throughput of the interferometer. Because reflection gratings are essentially phase objects, diffraction efficiencies potentially be significantly higher than from the amplitude gratings of the transmission scheme. The reflection scheme also has the advantage of functionality independent of wavelength. Unlike transmissive optics, which may be opaque at some wavelengths, the reflection optics will always function, although the diffraction efficiency and absolute reflectivity will certainly will vary with wavelength.

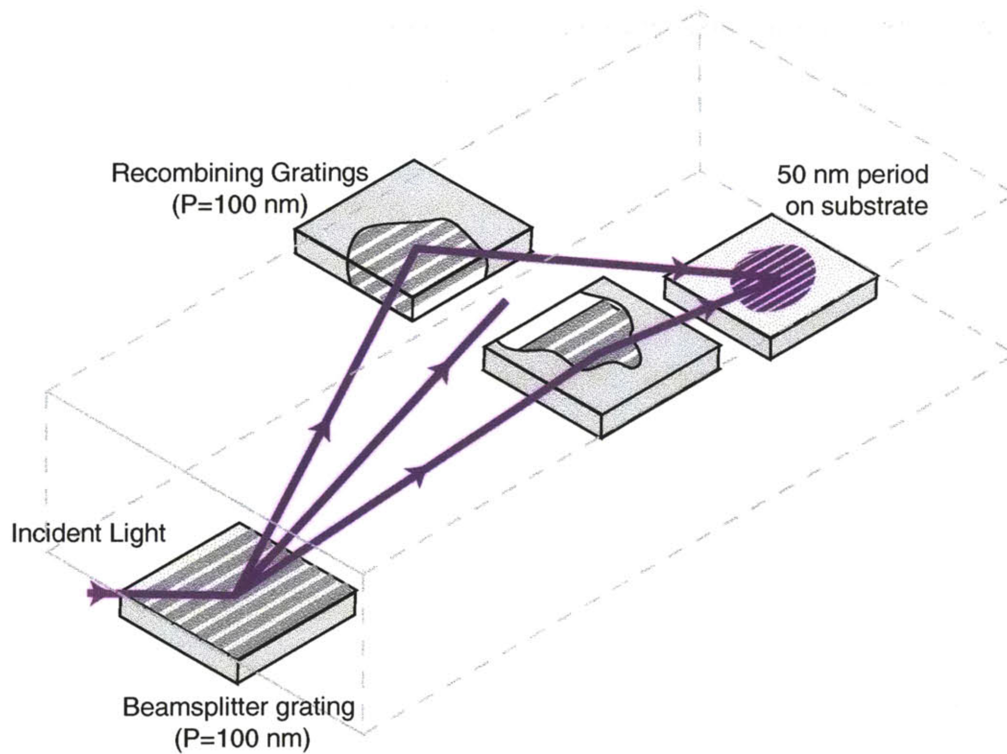


Figure 5.9 A three-grating interferometer made with reflection gratings. The beamsplitter grating and the substrate are coplanar, as are the recombiner gratings.

In either the reflection or transmission scheme, however, the problem of low output power from the source is compounded by the limited diffraction efficiencies at each of the gratings, and it would not be unreasonable to suffer two to three orders-of-magnitude loss of power at the substrate. The membrane grating proposed in [30] actually loses close to 6 orders-of-magnitude between the source output and the incident power on the wafer. Additionally, polarized light is necessary for high contrast in the exposed image. Polarizing technology for sub-100 nm light is largely undeveloped and the few options available, such as reflections at the Brewster angle, will incur significant losses. The practical limit to the length of an exposure, outside of the patience of the user, would be determined by the stability of the interferometer. One point of reference would be the time of a proximity x-ray lithography exposure here at the MIT NanoStructures Lab using the electron bombardment sources emitting at ~1 nm. Using only a single

membrane transmission mask, exposures in PMMA commonly last about 1 day. If one imagines the increase in exposure time for this source used with a grating interferometer, with limited diffraction efficiency and absorptions through a second membrane, it would be necessary to measure exposure times in months or years. Even with a significantly more powerful source, it is likely that the interferometer must be designed for stability over a timescale measured in hours, which is not trivial, especially when compounded with the already difficult material issues.

5.4.6 Liquid Immersion

Resolution enhancement in microscopy using liquid immersion has been well-known since the nineteenth century. One can reduce the effective wavelength of the photon by a factor equal to the refractive index n of the immersion fluid. Although immersion for lithography has been proposed in the past [53,54,55], only recently has this concept been gained acceptance as a viable method [56,57].

$$\lambda_{effective} = \frac{\lambda_{vacuum}}{n} \tag{5.13}$$

In the context of a lithographic grating interferometer, one could potentially see this same benefit, although it would manifest slightly differently. In microscopy or conventional projection lithography, the increase in resolution associated with the decrease in effective wavelength can be observed immediately upon introduction of a high-index medium between the optics and the substrate. The achromaticity of the grating interferometer eliminates this benefit. What is gained, however, is the ability to diffract propagating orders from parent gratings whose periodicity is less than the vacuum wavelength of the incident light, whereas ordinarily these orders would be evanescent (at normal incidence). The minimum possible pitch of the parent gratings is reduced to $P_{min}=\lambda/n$. Thus, to see the benefits of immersion in a grating interferometer, it is necessary to design and build a new tool, one which would not support diffraction in the absence of the immersion fluid.

Of course, the potential if this technique is limited by the index of the immersion fluid - higher index fluids allow for smaller features. In order for immersion to be a viable option, the fluid must be transparent, as well as providing a high enough index to realize gains that are beyond what is attainable through simpler means. With these considerations in mind, the most viable combination of photon and fluid for immersion lithography is the use of pure, clean, deionized water at λ=193 nm [58]. Critical optical parameters are given below in Table 5.5.2. For example, the effective wavelength for water immersion at λ_0=193 nm is 131nm, low enough to make it attractive compared to the next available laser wavelength at λ_0=157 nm. Immersion at λ_0=157nm is topic of research, although absorption and other material issues at this wavelength are currently unresolved problems [56]. Using Equation 5.12, the propagation distance through which a high transmission will occur can be calculated. It is found that 90% of the incident radiation can be retained for about a centimeter of propagation through water, which should be sufficient for a well-designed interferometer.

Table 5.5.2 Optical data for H_20 at λ=193 nm [59]

λ_0	n_{water}	$\lambda_{effective}$
193 nm	1.47	131nm

Absorption α	*Depth (T=95%)*
0.036 cm^{-1}	6 mm

The present three-grating AIL design, developed for use in air, is not ideal for immersion. The entire apparatus including the ancillary alignment functions, would need to be submersed. One important consequence of this is the potentially large immersed path-length that is required for large area exposures. The minimum distance between the splitter and recombiner gratings must be large enough that the first-order diffracted beams from the beamsplitter grating land in a corresponding position on the recombiner gratings. The distance that the diffracted beam must travel is inversely proportional to the sine of the diffraction angle θ_0. The same distance must be traversed by the beam between the recombiner grating

and the exposure plane. The total immersed path length for a current three grating interferometer is given below, where W is the width of the parent grating.

$$L_{immersed} = \frac{2W}{\sin(\theta_0)} \qquad (5.14)$$

For the case where the diffraction angle is large, as in the case of parent gratings near the diffraction limit, the immersed path length can be well approximated as simply twice the width of the parent grating. It is obvious, that for even modestly sized parent gratings, the immersed optical path, and therefore absorption, can become quite large. This, in turn, can lead to greater problems than simply reduced throughput. Heating of the immersion fluid due to the absorbed optical energy will cause changes in the refractive index of the material, in turn leading to local refraction of the beams. Over the course of an exposure a few minutes long, convective currents would create random index variations in the two arms of the interferometer. Although further investigation of this phenomenon is necessary, probable outcomes are that achromaticity will be compromised and the focal plane will drift, ultimately reducing or ruining contrast in the exposed image.

As an example, consider the amount of optical power absorbed into the immersion fluid for an interferometer similar to the current AIL. Assume that the width W of the parent grating is a modest 2.5 cm, for a minimum immersed path of ~5cm. From Equation 5.12, the transmission of this amount of water can be calculated as ~66%. Thus, the exposure for a resist such as PMMA would increase by 50% over the current time in air. A 180-second exposure with ~2 Watts of optical power incident on the interferometer is representative of the current exposure conditions. Under immersion, this would increase to 270 seconds for the same incident power. The total incident energy would be ~540 Joules, of which 33% would be absorbed in the water. Thus, we estimate 180 Joules of energy would be absorbed by the water. The specific heat of water at room temperature is ~4.2 J/g/°K [39], so this absorbed energy would be sufficient to increase the temperature of about 43 mL of water by 1 degree. It is important to note that the absorbed energy increases non-linearly with the immersed path length when we consider that

the incident dose on the substrate must remain the same for a given photoresist. Equation 5.15 shows that the incident dose *D* on the interferometer is inversely proportional to the transmission of the immersion fluid, where *Do* is the incident dose for 100% transmission. The amount of energy absorbed in the fluid, shown in Equation 5.16, is proportional to the incident dose *D*, where the transmission can be given in terms of the width *W* of the parent grating as in Equation 5.17. A plot of the energy absorption normalized to incident dose for a range of parent-grating widths is shown in Figure 5.11.

$$D = \frac{D_0}{T} \tag{5.15}$$

$$E_{absorbed} = \frac{D_0(1-T)}{T} \tag{5.16}$$

$$T = 10^{\frac{-2\alpha W}{\sin(\theta_0)}} \tag{5.17}$$

It is clear that there are only two parameters which can be adjusted to minimize the energy absorbed by the immersion fluid. Either the width of the parent grating can be reduced, or the dose *Do* an be reduced. Not accounted for in the above equations, reduction in the width of the parent grating has the effect reducing the exposed area of the daughter grating. The incident dose will scale linearly with the exposure area. Further reduction on the value of *Do* can either come from improvements in the efficiency of the interferometer or greater photospeed in the resist. Resists designed for use in immersion at 193 nm are currently under development by all major manufacturers. However, it is unclear whether these products will have the necessary resolution and roughness properties to be practical alternatives to PMMA at sub-50nm linewidths.

5.4.7 A novel interferometer designed for immersion

It becomes obvious that the absorption of the immersion liquid ultimately will limit the width of the parent grating that can be used, and thus the size of the exposed gratings. Recall from Section 5.3.2 that one of the reasons for using a three-grating interferometer over the two-grating interferometer was the increase in usable area. However, it is likely that under liquid-immersion conditions, the usable area of the parent-grating will not be limited by the period-match requirement, as in the "gas-immersion" design. Rather, fluid absorption will limit the maximum area of the exposure, rendering the additional complexity of the three-grating design unmerited. Rewriting Equation 5.17, we can solve for the maximum width of the parent grating at a given level of transmission, shown in Equation 5.18, assuming $\sin(\theta_0)=1$.

$$W_{\max} = \frac{\log_{10}(T)}{-2\alpha} \qquad (5.18)$$

To achieve a transmission greater than 90%, the maximum width of the parent grating can only be ~6 mm. With this in mind, a simplified interferometer can be designed with an emphasis on simplicity.

The configuration shown in Figure 5.10 is functionally the same as the three grating interferometer shown previously. However, using a hybrid of reflection and transmission parent gratings, the whole system can be integrated into a single thick substrate. Fabrication of this arrangement can be accomplished easily from a single phase-grating. Using a masked evaporation of aluminum or chromium, the center 1/3 of the grating area can be made reflective. Further, alignment consists only of positioning the substrate at the focal plane.

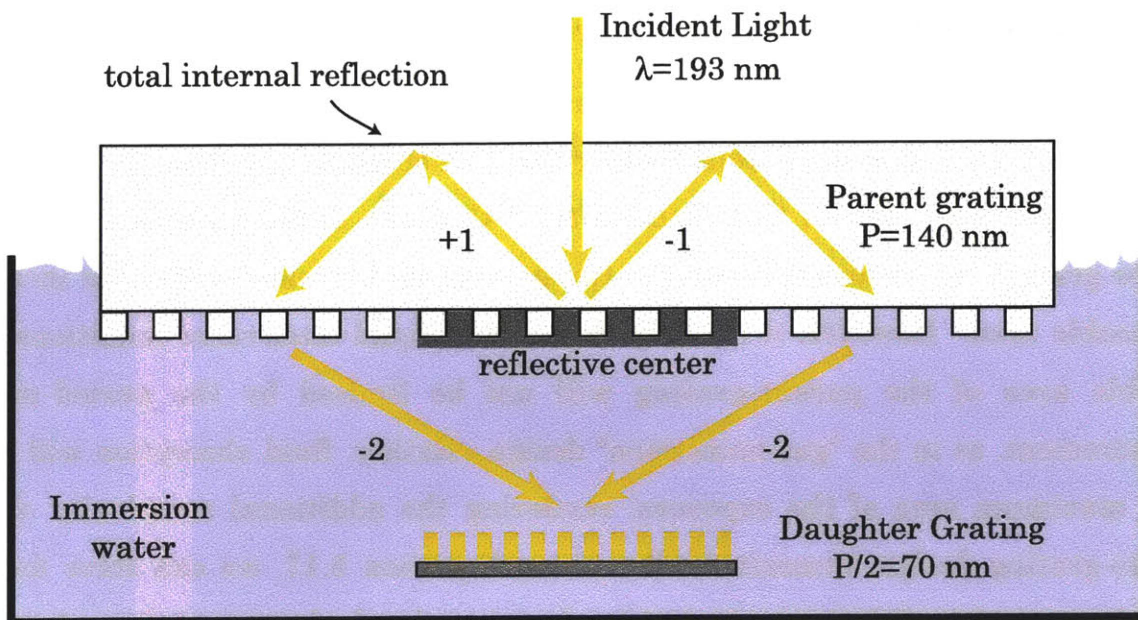


Figure 5.10 A three-grating-interferometer wrapped into a single monolithic block of fused silica. A single grating is used, with the center portion coated to act as a reflection grating. Total-internal-reflection off the top surface of the block redirects the diffracted beams to the outer sections of the grating. The (-2) order diffracted beams from the outer sections of the grating interfere to form a grating with half the period of the parent. Using water immersion and 193 nm light, 70nm-period gratings are possible.

A primary advantage inherent to this scheme is the stability and accuracy of the grating positions and alignment. The splitter and recombiner sections are inherently in perfect alignment and it is impossible for misalignment to occur. Further, only the lower section of the parent grating optic and the substrate must in contact with the immersion fluid, as opposed to three gratings and the substrate in the traditional three-grating arrangement. Because the path from the beamsplitter grating to the recombiner sections occurs entirely within the optic, the immersed optical path is also reduced by half. Further, the reduction in the number of partially reflecting surfaces, the backside of each parent grating, is reduced from 3 to 1, thus improving overall efficiency. The zero-order beam from the beamsplitter grating is automatically directed away from the substrate.

The size of the parent gratings is limited in this design in the same way as the two grating interferometer, by period-matching requirement. In Chapter 3 it was shown that the periodicity of a grating exposed with spherical beam interference, the method that would be used to produce the parent grating, increases nonlinearly for distances away from the center of the exposure. In the monolithic design, the largest period of the reflective beamsplitter section would be shorter than the shortest period in the transmissive recombiner section. For the interferometer to function, the maximum difference in period between the shortest period at the center of the beamsplitter section and the longest period at the outside of the of the recombiner section must not differ by more than the period-matching tolerance.

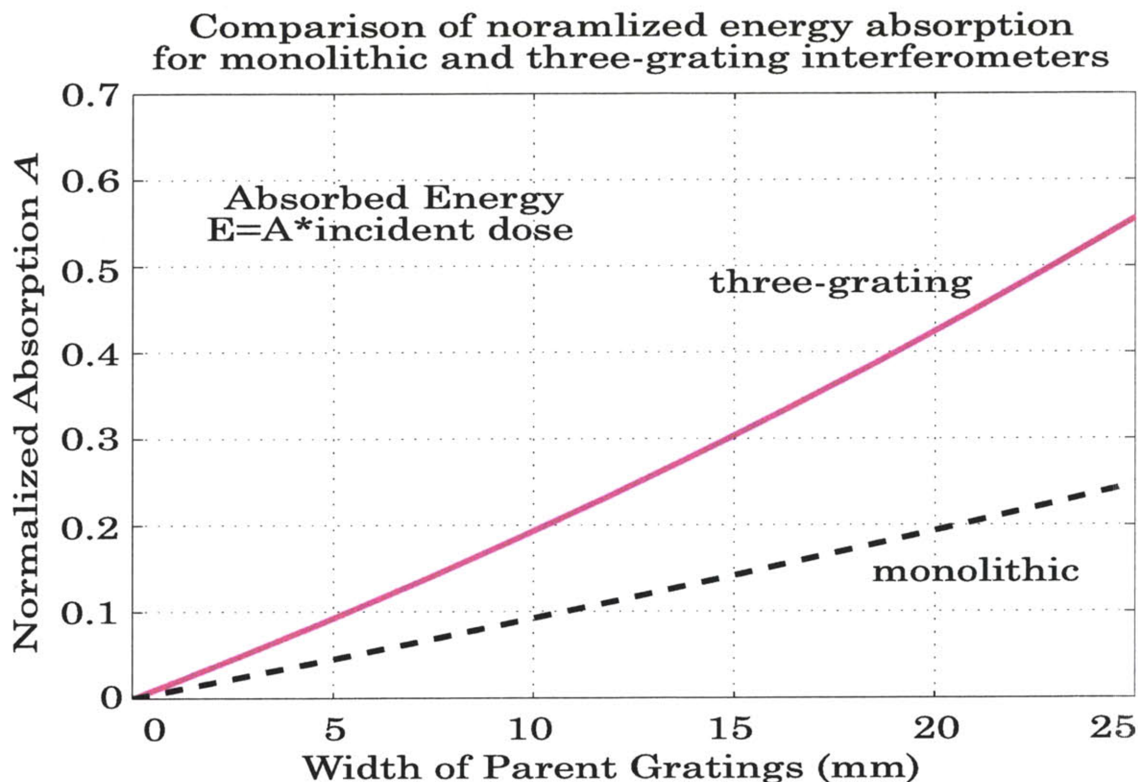


Figure 5.11 The energy absorbed by the immersion fluid normalized to the incident dose as a function of the width of the parent grating. Calculations assume water immersion for λ_0=193 nm and a parent grating of period P_I=140nm. For a given energy absorption, the monolithic interferometer can support parent gratings at least twice as wide as the three-grating interferometer.

Thus, it becomes difficult to make large-area parent-gratings for this monolithic system. Methods for reducing period variations, discussed in the Chapter 3, could greatly benefit this design. However, we have shown that for immersion lithography, large-area parent gratings are undesireable due to the heating caused by absorbed optical energy. By limiting the total width of the parent grating to ~3 cm, each section would only be ~1 cm in width. Absorption in the immersion fluid would be minimized, and a high degree of period matching can be achieved over this distance. In fact, despite the limited potential for very-large-area gratings from this design, the actual potential grating widths can be greater from the monolithic interferometer than from the three-grating interferometer for a given allowable level of energy absorption due to the improved efficiency. Figure 5.11 compares the normalized energy absorption A for the monolithic and three-grating interferometers. The normalized absorption is found simply by dividing Equation 5.16 by the dose.

5.5 Reference

[1] C. Barus, Carnegie Inst. Wash. Publ. **149** [Part 1 (1911), Part 2 (1912), Part 3 (1914)]

[2] V. Ronchi, Ann. Scuola Normale Superiore di Pisa, **15** (1923)

[3] V. Ronchi, "Forty Years of History of a Grating Interferometer", *App. Opt.* **3** p.437 (1964)

[4] F.J. Weinberg, N.B. Wood, "Interferometer based on Four Diffraction Gratings". *J. Sci. Instrum.* **36** 227 (1959)

[5] R. Kraushaar, "A Diffraction Grating Interferometer", *J. Opt. Sci. Am.* **40** p.480 (1950)

[6] J.R. Sterrett, J.R. Erwin, "Investigation of a Diffraction-Grating Interferometer for use in Aerodynamic Research". *NACA TN2827* (1952)

[7] A.R. Maddox, R.C. Binder, "The Use of an Improved Diffraction Grating Interferometer". *App. Opt.* **8** p.2191 (1969)

[8] D. Gabor, "A new microscope principle", *Nature* 161 (1948)

[9] H.F. Talbot, "Facts Relating to Optical Science IV". *Philos Mag.* **9** (1836)

[10] E. Lau, "Beugungserscheinung an Doppelrastern". *Ann. Phys.* **6** p.417 (1948)

[11] E.N. Leith and B.J. Chang, "Space Invariant Holography with Quasi-Coherent Light". *App. Opt.* **12** p.1957 (1973)

[12] B.J. Chang. R. Alferness, E.N. Leith, " Space-invariant achromatic grating interferometers: theory". *App.Opt.* **14** p.1592 (1975)

[13] E.N. Leith , G.J. Swanson "Achromatic interferometers for white-light optical processing and holography" *App.Opt.* **19** p.638 (1980)

[14] G.J. Swanson, "Broad Source Fringes in Grating and Conventional Interferometers" *J. Opt. Sci. Am. A* **1** p.1147 (1984)

[15] E.N. Leith, G. Swanson, S. Leon, "Construction of Diffractive Optical Elements in Non-Coherent Light". *Proc. SPIE* **53** p.3 (1984)

[16] H. Chen, R. Hershey, E.N. Leith, "Sawtooth profile fringes with a two-grating interferometer". *App. Opt.* **27** p.1193 (1988)

[17] R. Hershey, E. Leith, "Grating interferometers for producing large-area holographic gratings". *App. Opt.* **29**(7) p.937 (1990)

[18] D. Post, K. Patorski, P. Ning, "Compact grating interferometer for producing photoresist gratings with incoherent light" *App. Opt.* **26**(6) p.1190 (1987)

[19] K. Hibino, Z.S. Hegedus, "Aberration compensation in a grating interferometer". *App. Opt.* **33**(19) p.4156 (1994)

[20] I.N. Ross, C.J. Hooker, P.Dombi, "Efficient generation of large diffraction gratings with a grating interferometer". *App. Opt.* **40**(34) p.6153 (2001)

[21] E. Leith, H. Chen, G. Collins, K. Scholten, G. Swanson, J. Upatneiks, "Interferometric construction of circular gratings". *App. Opt.* **19** (1980)

[22] Y.S. Cheng, "Fringe formation with a cross-grating interferometer". *App. Opt.* **25** p.4185 (1986)

[23] Y.S. Cheng, "Analysis of interference pattern in a cross-grating interferometer". *App. Opt.* **27** p.3025 (1988)

[24] A.M. Hawryluk, H.I. Smith, D.J. Ehrlich, "Deep-UV spatial frequency doubling by combining multilayer mirrors with diffraction gratings". *J. Vac. Sci. Technol.B* **1**(14) p.1200 (1983)

[25] E.H. Anderson, "Fabrication and Electromagnetic Application of Periodic Nanostructures". Ph.D thesis, Massachusetts Institute of Technology (1988)

[26] A. Yen, E.H. Anderson, R.A. Ghanbari, M.L. Schattenburg, H.I. Smith, "Achromatic holographic configuration for 100-nm period lithography". *App. Opt.* **31**(22) p.4540 (1992)

[27] T.A. Savas, S. N. Shah, M.L. Schattenburg, J.M. Carter, H.I. Smith, "Achromatic interferometric lithography for 100-nm-period gratings and grids" *J. Vac. Sci. Technol.B* **13**(6) p.2732 (1995)

[28] T.A Savas, M.L. Schattenburg, J.M. Carter, H.I. Smith, "Large-area achromatic interferometric lithography for 100 nm period gratings and grids". *J. Vac. Sci. Technol.B* **14**(6) p.4167 (1996)

[29] T.A. Savas "Achromatic Interference Lithography". Ph.D. Thesis, Massachusetts Institute of Technology (2003)

[30] A. Yen, M.L. Schattenburg, H.I. Smith, "Proposed method for fabricating 50-nm period gratings by achromatic holographic lithography.". *App. Opt.* **31**(16) p.2972 (1992)

[31] M. Wei, D.T. Attwood, T.K. Gustafson, E.H. Anderson, " Patterning a 50-nm period grating using soft x-ray spatial frequency multiplication". *J. Vac. Sci. Technol.B* **12**(6) p.3648 (1994)

[32] H.H. Solak, C. David, J. Gobrecht, V. Golovnika, F. Cerrina, S.O. Kim, P.F. Nealey, *Microelectron. Eng.* **67-68** p.56 (2003)

[33] V. Golovnika, F. Cerrina, P.F. Nealey, J.W. Taylor, H.H. Solak, C. David, J. Gobrecht, "Exploring the limits of the ultimate resolution of positive-tone

chemically amplified resists: 26 nm dense lines using extreme ultraviolet interference lithography". *J. Vac. Sci. Technol.B* **22** p.99 (2004)

[34] J.A Samson, D.L. Ederer, *Vacuum Ultraviolet Spectroscopy*, Academic Press, San Diego, (2000)

[35] J. Yan, A. El-Dakrouri, M. Larousi, M.C. Gupta, "121.6 nm radiation source for advanced lithography". *J. Vac. Sci. Technol.B* **20**(6) p.2574 (2002)

[36] J. Yan, M.C. Gupta, "High-power 121.6 nm radiation source". *J. Vac. Sci. Technol.B* **21**(6) p.2839 (2003)

[37] D.A. Patterson, W.H. Vaughn, *J. Opt. Sci. Am.* **53** p.851 (1963)

[38] W.R. Hunter, S.A. Malo, *J. Phys. Chem. Solids.* **30** p. 2739 (1969)

[39] D.R. Lide *ed., Handbook of Chemistry and Physics, 73rd Edition.* CRC Press, Cleveland , (1992)

[40] Multilayer database at www.crxo.lbl.gov/multilayer/survey.html

[41] I.E. Ferincz, Cs. Toth, J.F. Young, "Imaging characteristics of poly(methyl methacrylate) at vacuum ultraviolet wavelengths". *J. Vac. Sci. Technol.B* **15**(4) p.828 (1997)

[42] K. Kudo *et al., Jpn. J. Appl. Phys. 1*, **29** p.2572 (1990)

[43] K. Kudo *et al., Jpn. J. Appl. Phys. 1*, **31** p.2572 (1992)

[44] V. Liberman, M. Rothschild, P.G. Murphy, S.T. Palmacci, "Prospects for photolithography at 121 nm". *J. Vac. Sci. Technol.B* **20**(6) p.2588 (2002)

[45] P.M St.John, H.G. Craighead, "Moloayers of fluorinated silanes as electron-beam resists" *J. Vac. Sci. Technol. B* **14** p.69 (1996)

[46] M.J. Lercel, C.S. Whelan, H.G. Craighead, K. Seshadri, D.L. Allara, "High-resolution silicon patterning with self-assembled monolayer resists". *J. Vac. Sci. Technol.B* **14** p.4085 (1996)

[47] X.M Yang, R.D. Peters, T.K. Kim, P.F. Nealy, "Patterning of self-assembled monolayers with lateral dimensions of 0.15 μm using advanced lithography". *J. Vac. Sci. Technol. B* **17**(6) p.3203 (1999)

[48] F.R. Powell, P.W. Vedder J.F. Lindblom, S.T. Powell, "Title". *Opt. Eng.* **29** p.614 (1990)

[49] A.C. Warren, I. Plotnik, E.H. Anderson, M.L. Schattenburg, D.A. Antoniadis, H.I. Smith, "Fabrication of sub-100nm linewidth periodic structures for study of quantum effects from interference and confinement in Si inversion layers" *J. Vac. Sci. Technol. B* **4**(1) p.365 (1986)

[50] H.I. Smith, M.L. Schattenburg, "X-ray lithography, from 500 to 30nm: X-ray nanolithography". *IBM J. Res. Dev.* **37**(3) p.319 (1993)

[51] H.I.Smith, M.L. Schattenburg, S.D. Hector, J. Fererra, E.E. Moon, I.Y. Yang, M. Burkhardt, "X-ray nanolithography: Extension of the limits of the lithographic process". *Microelectronic Eng.* **32** p.143 (1996)

[52] M. Lim, "Development of x-ray Lithography and Nanofabrication Techniques for III-V Optical Devices". Ph.D. Thesis, Massachusetts Institute of Technology (2002)

[53] C.V. Shank, R.V. Schmidt, "Technique for producing 0.1 μm periodic structures". *App. Phys. Lett.* **23** p.154 (1973)

[54] B.J. Lin, "The future of sub-half-micron optical lithography" *Microelectron. Eng.* **6** p.31 (1987)

[55] H. Kawata, J. Carter, A.Yen, H.I. Smith, "Optical projection lithography using lenses with NA greater than 1" *Microelectron. Eng.* **9** p.31 (1989)

[56] M. Switkes, M. Rothschild, "Immersion Lithography at 157 nm", *J. Vac. Sci. Technol.B* **19**(6) p.2353 (2001)

[57] M. Switkes, M, Rothschild, *J. Microlith, Microfab., Microsyst.* **1** p.225 (2002)

[58] M. Switkes, R.R. Kunz, R.F. Sinta, M. Rothschild, P.M. Gallagher, V.J. Krukonis, K. Williams, "Immersion Liquids for Lithography in the Deep Ultraviolet", *Proc. SPIE* ***5040*** *p.690 (2003)*

[59] M.R. Querry, D.M. Wieliczka, D.J Segelstein, in *Handbook of Optical Constants of Solids,* E.D. Palik ed., Academic Press, Boston (1991)

Chapter 6
The Integrated Thin Film Interferometer

6.1 Introduction

6.1.1 Concept

In this chapter, a novel concept for an achromatic interferometer will be presented which can circumvent many of the issues that currently inhibit the fabrication of sub-100 nm period gratings. Recall from Chapter 5 the simplest scheme for near-field spatial-frequency doubling (SFD) utilizing a single parent-grating, shown again in Figure 6.1b. In the region under the parent-grating before the diffracted orders separate into distinct beams, they will interfere to form an image whose spatial-frequency is twice that of the parent-grating. It was argued in the last chapter that this scheme is difficult to implement for two reasons. First, finite spatial-coherence of the exposure source limits the separation of the exposure-plane and the parent-grating to a few microns. Second, it can be difficult to fully suppress the zero-order diffracted beam with common fabrication techniques. To circumvent these problems, the three-grating AIL interferometer was developed. If

we assume, however, that a properly optimized grating could be fabricated and placed appropriately close to the exposure-plane, then the simplicity of near-field SFD from a single parent-grating becomes very attractive. The Integrated Thin-Film Interferometer (ITFI) is a new approach to near-field spatial-frequency doubling where the entire system - parent-grating, gapping mechanism, and substrate - are all integrated onto a *single substrate with no external optics*.

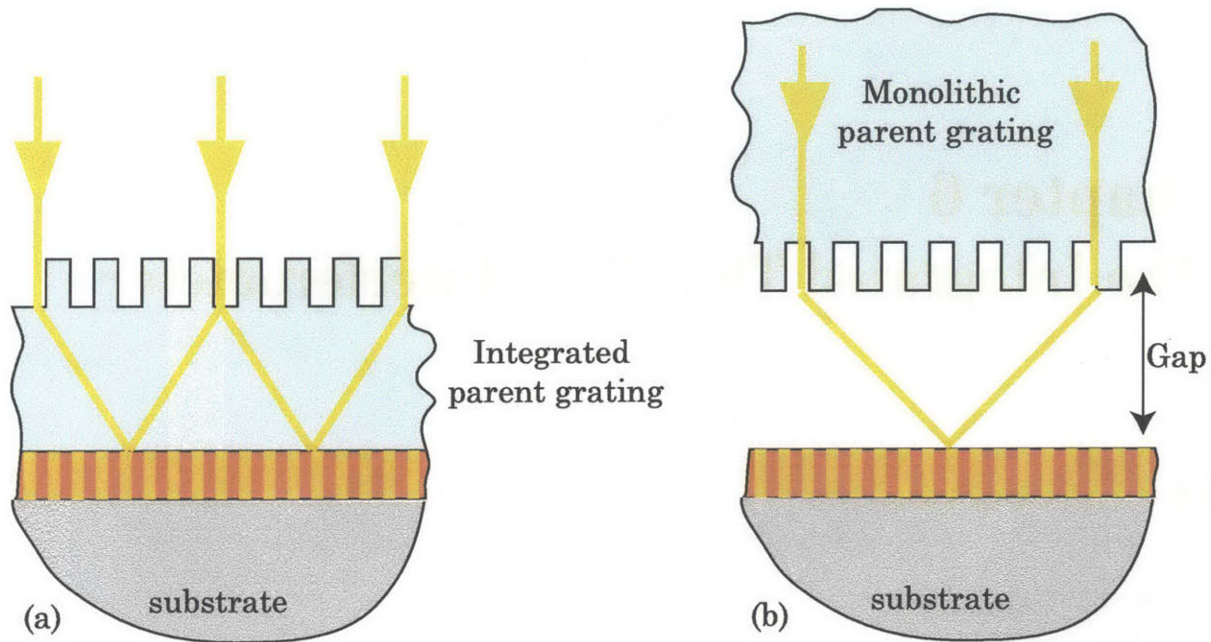


Figure 6.1 (a) Near-field SFD with the parent-grating fabricated in a thin-film as part of the resist stack. (b) SFD as previously practiced with a monolithic parent-grating brought into proximity with the substrate.

6.1.2 Implementation

Figure 6.1(a) shows the ITFI concept of a single-use parent-grating fabricated in intimate contact with the wafer as part of the thin-film resist stack. Assume that the parent-grating is optimized such that only first-order diffraction occurs, and that the fabrication of the parent-grating does not affect the underlying photoresist. A single broad beam, or a narrow scanned beam, illuminates the parent-grating at normal incidence to expose the daughter-grating. After a single exposure, the

parent-grating is washed away with a wet chemistry. The frequency-doubled image in the photoresist can now be developed to produce a daughter-grating with half the spatial-period of the parent. Although the *function* is exactly the same as the near-field SFD with a separate monolithic parent-grating (Figure 6.1(b)), the *differences in implementation* create numerous advantages that significantly raise the potential for this technique.

6.1.3 Fabrication of parent-gratings

The ITFI is essentially a specially designed thin-film stack, and therefore its implementation should, in theory, be no more complicated than spinning and flood-exposing a wafer. The simplicity of this implementation, however, comes at the price of a complex design. The idealized cartoon in Figure 6.1(a) shows us how the ITFI scheme is intended to operate, without concern for the gritty details. It is worthwhile to mentally segregate the fabrication process for the parent-grating from that of the daughter-grating when considering what attributes are desirable. Foremost, the process for the parent-grating must not perturb the photoresist in which the daughter-grating will be exposed, either chemically, mechanically, or optically. Second, it must have high resolution and repeatablity - capable of ~50-100 nm features with precise control of parameters such as duty-cycle and tooth-height. Finally, it must be simple enough that it can be performed on a per-wafer basis without difficulty. We can dismiss immediately the possibility of plasma-etching the parent-gratings, as this would violate all of the above conditions. Notably, ionic bombardment and the broad-band emission of a plasma discharge would likely compromise the underlying photoresist layer.

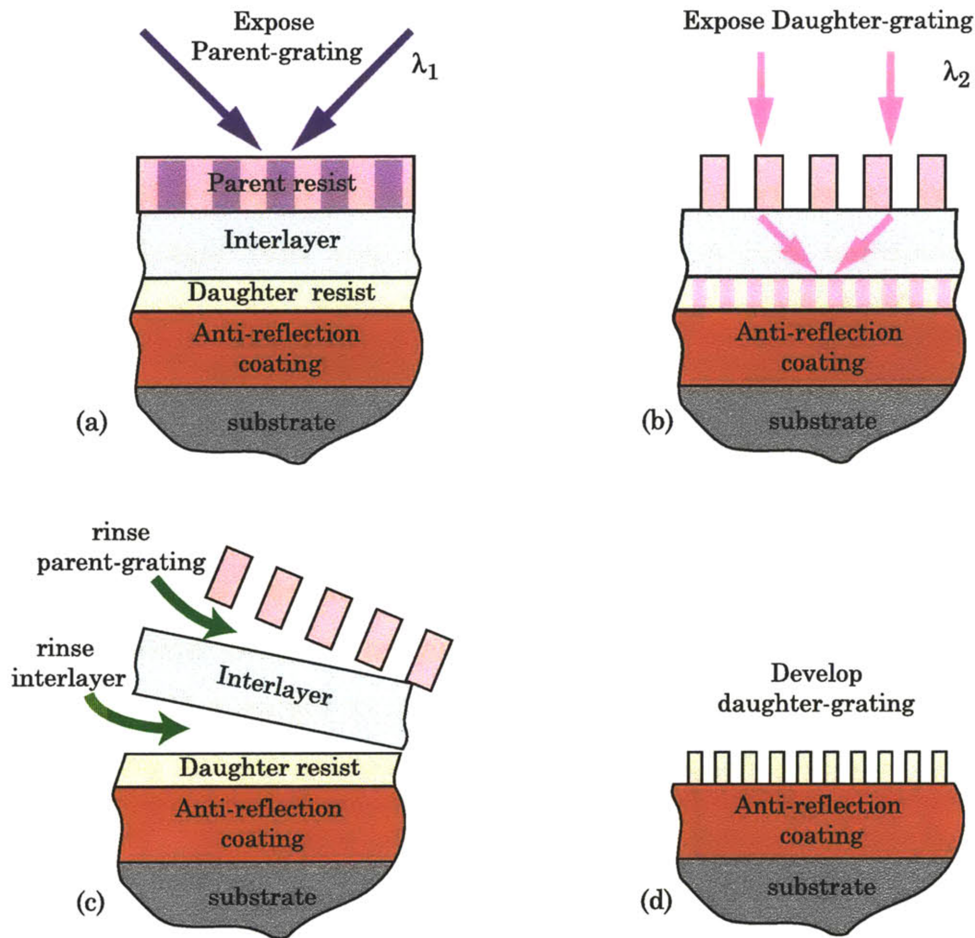


Figure 6.2 An integrated thin-film interferometer using photoresist as parent-grating. Near-field spatial-frequency-doubling from a parent photoresist grating is captured in a second layer of resist which is insensitive to the exposure of the parent. The gapping material and the interlayer are removed using a wet etch prior to developing the photoresist for the daughter-grating.

An intuitively simple approach is to use a photoresist pattern directly as the parent-grating, which could be made using any of the interference lithography techniques described in Chapters 3-5. This offers highly accurate and precise control of all the important parameters of the parent-grating. Periodicity is set by the exposure geometry, tooth-height is set by the spin speed of the resist, and the duty-cycle can be accurately cotrolled by the exposure time. The subtle challenge of this approach is that now two photoresist layers are required, one for the parent-grating and one for the daughter-grating, and the lithography of the parent-grating

must not affect the resist for the daughter-grating. Fortunately, most DUV resist materials are insensitive to mid-UV radiation, so this is not as challenging as it might first appear. For example, PMMA, is a high-resolution resist which can be used for exposures at λ=193nm or 157nm, but is wholly insensitive to light above ~250 nm. If a mid-UV exposure, such as at λ_0=325nm using the Lloyd's-mirror, is used for the parent-grating, then the energy of the photons will be insufficient to interact with the PMMA resist for the daughter-grating.

An example process is shown in Figure 6.2. The wafer is prepared as for a regular IL exposure, with a few additional layers, including PMMA for the daughter-grating. A detailed discussion of resist processes for IL is given in Chapter 7. The parent-grating is exposed and developed in the standard fashion at λ_0=325 nm. The wafer is then flood exposed with λ_0=193 nm to expose the daughter-grating. The parent-grating is re-developed and dissolves completely as result of the flood-exposure. The oxide gapping-layer can be removed with a dip in dilute hydrofluoric acid (HF). Finally, the PMMA is developed to create the daughter-grating. Thus, the process for making a frequency-doubled grating with the ITFI requires only a few more steps than a standard IL exposure.

The two-wavelength exposure system allows for the entire process of fabricating both the parent and daughter-gratings to be only slightly more complex than the process for other IL gratings. Fabricating the parent-grating is identical to the exposure process for gratings made with current IL techniques. Exposure of the daughter-grating is trivial; an unaligned flood-exposure followed by a three non-critical "development" steps. The first two steps remove the parent-grating and gapping layer, and the third develops the resist image of the daughter-grating.

6.2 Practical Advantages of the ITFI

6.2.1 Wavelength reduction via solid immersion

Perhaps the most compelling advantage of using the ITFI is the ability to expose significantly shorter grating periods using a given wavelength than is possible with any of the other IL schemes. Recall that the minimum period at a given wavelength is $\lambda/2$. For λ_0=193 nm, the shortest possible period is theoretically 96.5 nm. As discussed in the previous chapter, there is a notable dearth of commercially available sources emitting at λ_0<193 nm, limiting further reduction of feature size by simply shortening the wavelength. The combination of a number of attributes of the ITFI neatly sidesteps this limitation. In Section 6.2.5, it will be shown how the range of possible exposure sources for the ITFI is greatly extended by reducing the power and spatial-coherence requirements necessary for current achromatic schemes. Even better, utilizing sources below 193nm or 157nm may actually be unnecessary.

A critical difference between the ITFI and other IL schemes is that light diffracted from the parent-grating is always propagating in a high-refractive-index medium. The effect of this is that diffracted orders that would ordinarily be evanescent (i.e. all orders above $m=0$ for parent-gratings with $P<\lambda_0$) are allowed to propagate in the ITFI. Although none of the layers are actually liquid, this effect is the same as for the liquid-immersion discussed in Chapter 5. One could consider the ITFI as *solid-immersion lithography*, where the effective wavelength λ_{eff} of the light is equal to the vacuum wavelength λ_0 divided by the index of the propagation material. The index of the interlayer material (n_i) between the parent-grating and the photoresist (resist index n_r) generally determines the effective wavelength of the system, shown in Figure 6.3. Compared to liquid-immersion, the index of solid materials can potentially be higher, and the absorption of the immersion medium is far less critical. The propagation distance in the ITFI is only a few hundred

nanometers, and many problems caused by absorptive heating of a liquid medium, such as convection, are circumvented.

$$\lambda_{effective} = \frac{\lambda_0}{n_i} \tag{6.1}$$

$$P_{min} = \frac{\lambda_0}{2n_i} = \frac{\lambda_{effective}}{2}, \quad n_i < n_r \tag{6.2}$$

A lower effective wavelength means that the minimum spatial-period possible with a given source decreases. Using the same ArF laser currently used on the AIL (λ_0=193nm) and a hydrogen-silsesquioxane (HSQ) interlayer whose refractive index is 1.534, the effective exposure wavelength is λ_{eff}=126 nm and the minimum period is ~63 nm. Using a λ_0=157nm F_2 laser, the same HSQ interlayer material has an index n_i=1.7, allowing the exposure of 46nm-period daughter-gratings.

It should be noted that the refractive-index of the photoresist can also be a limiting factor. If the index of the resist is less than index of the interlayer, diffracted orders allowed in the interlayer may not be allowed in the resist. In more familiar terms, this is simply total-internal-reflection at a high-index to low-index transition, as seen in Figure 6.3. The critical angle will equal the diffraction angle from a parent-grating with period λ_0/n_r. The minimum period in this case is

$$P_{min} = \frac{\lambda_0}{2n_r}, \quad n_i \geq n_r \tag{6.3}$$

It is obvious that the optical constants of the photoresist, interlayer, and other materials in the thin-film stack are critical to designing the ITFI. A tabulation of complex refractive indicies for materials relevant to interference lithography can be found in Appendix A.

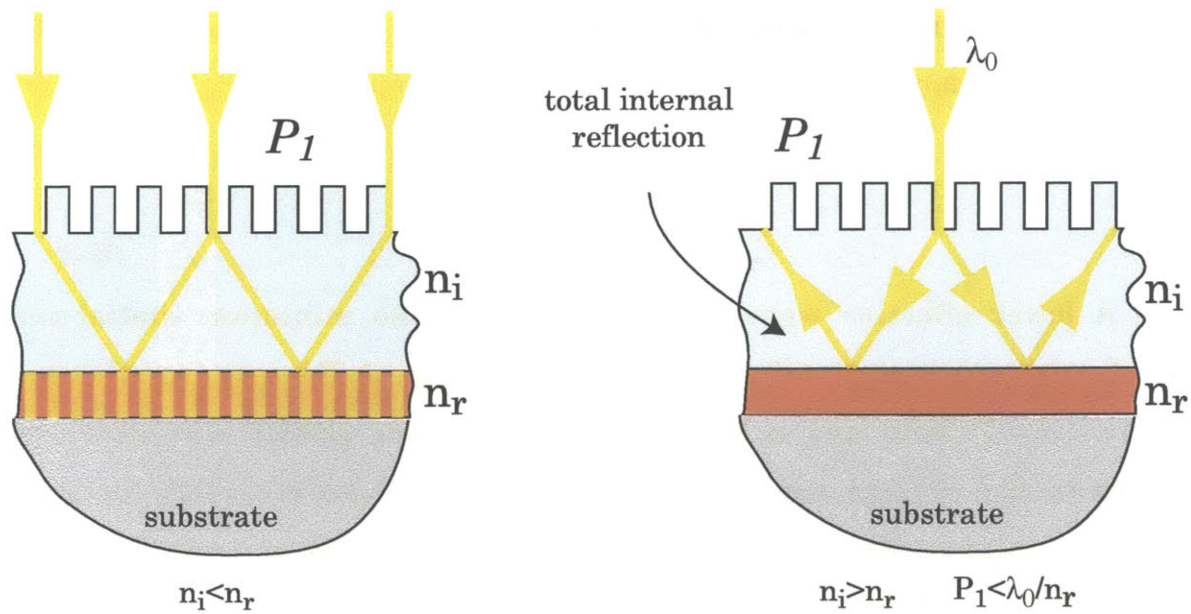


Figure 6.3 A) Minimum period of $P_{min}=\lambda/2n_i$ determined by the index of the interlayer material for the case where $n_i<n_r$. B) For the case where $n_i>n_r$, total internal reflection at the interlayer/resist boundary limits the minimum period to $P_{min}=\lambda/2n_r$.

6.2.2 Extension of current lithography

In contrast to other proposed IL schemes for similar feature sizes, including liquid immersion, the ITFI requires no additional investment or equipment beyond what is currently in use here in the NanoStructures Lab at MIT. In this sense, it is a "free" technology. The benefits of working with materials characterized and optimized for use at 193 nm (or 157nm), rather than uncharacterized materials at a shorter wavelength, should not be underestimated. Furthermore, unlike the fixed parent-gratings of other achromatic schemes, ITFI is not restricted to exposures at only one grating period. The single-use nature of the ITFI parent-grating affords a flexibility not possible with other designs. For any parent-grating that can be produced, a frequency-doubled grating over the same area can be made with an ITFI exposure.

One example of this is the fabrication of parent-gratings in the Lloyd's-mirror, which is capable of exposing gratings of arbitrary period down to P_I=170 nm. Frequency-doubling these parent-gratings using the ITFI can allow for *arbitrary* grating periods down to 85nm. Currently, the only comparable option is 100nm-period gratings from the AIL. Another exciting possibility is to use parent-gratings from the new Nanoruler scanning-beam interference lithography system developed in the Space Nanotechnology Lab here at MIT [1]. The Nanoruler is capable of writing gratings with exquisite spatial-phase control over substrates as large as 30cm, potentially at periods as low as 200 nm†. Using the ITFI, these gratings could be converted to 100nm-period gratings over an entire 30 cm wafer.

It is interesting to note that neither of these cases exploits the capacity of the ITFI for the smallest possible features. Rather, the value comes from a unique ability to extend the specific qualities of current IL tools to smaller spatial-periods. From this perspective, one could consider the ITFI not as a stand-alone technology, but as an extension for *any* of the current IL tools. Development of the ITFI does not just open the door to the long-sought 50nm-period grating, it can allow the Lloyd's-mirror to print spatial-periods down to 85nm and the Nanoruler to print 100nm-period gratings as big as a dinner plate, all without building or buying any new equipment. In fact, this extension of capabilities can potentially be generalized to other lithographies producing arbitrary features.

6.2.3 Mechanical Stability and Alignment

In addition to wavelength reduction, the ITFI obviates the standard concerns in interferometry of alignment and mechanical stability. Both part of the same thin film stack, it is impossible for the parent-grating and substrate to move relative to one another. Lateral position of the exposure pattern in the photoresist is

† Currently, the Nanoruler writes only 400nm-period gratings, although development is underway for shorter periods.

determined by solely by the position of the parent-grating (which is fixed), so it is impossible for the pattern to blur during exposure even if the entire substrate moves within the illuminating beam.

Additionally, parallelizing the parent-grating with the substrate occurs by default due to the uniformity of spin-coating. If we assume the gapping layer spun over a standard 10cm-diameter substrate has a thickness variation of 10nm from one end to the other, the parent-grating will be parallel to the substrate within 0.1 μrad, Compare this to the parallelism obtained with a simple Fabry-Perot cavity or Fizeau style interferometer operating with a HeNe laser at λ=633 nm. For 1 fringe visible across the wafer surface, the difference could be as large as $\lambda/2$=~316 nm, or parallel to within ~3 μrad. Thus, without any special effort, parallelism in the ITFI will be at least an order-of-magnitude better than in other schemes.

6.2.4 Source Requirements: Spatial Coherence

Defining the parent-grating as well as the gap using spun films also neatly solves the depth-of-focus issue caused by the finite spatial-coherence of the source in other achromatic schemes. Recall from Section 5.2.3 that for spatially-incoherent light, the interference fringes shear by an angle $\Delta\theta_o$ which is a function of the spread in incident angles $\Delta\theta_i$. At planes below the parent-grating, the positions of the interference maxima shift by a distance Δ which is proportional to the distance between the observation plane and the parent-grating. Contrast (V) as a function of the shift distance Δ (given a fraction of the image period) is plotted in Figure 6.4. A shift of $\Delta=P_2/4$, ($V=0.63$), is assumed to be the maximum tolerable value. For exposures with λ_0=193 nm, the maximum distance D between a 200nm-period parent-grating (P_1) and the exposure plane was on the order of 3 μm.

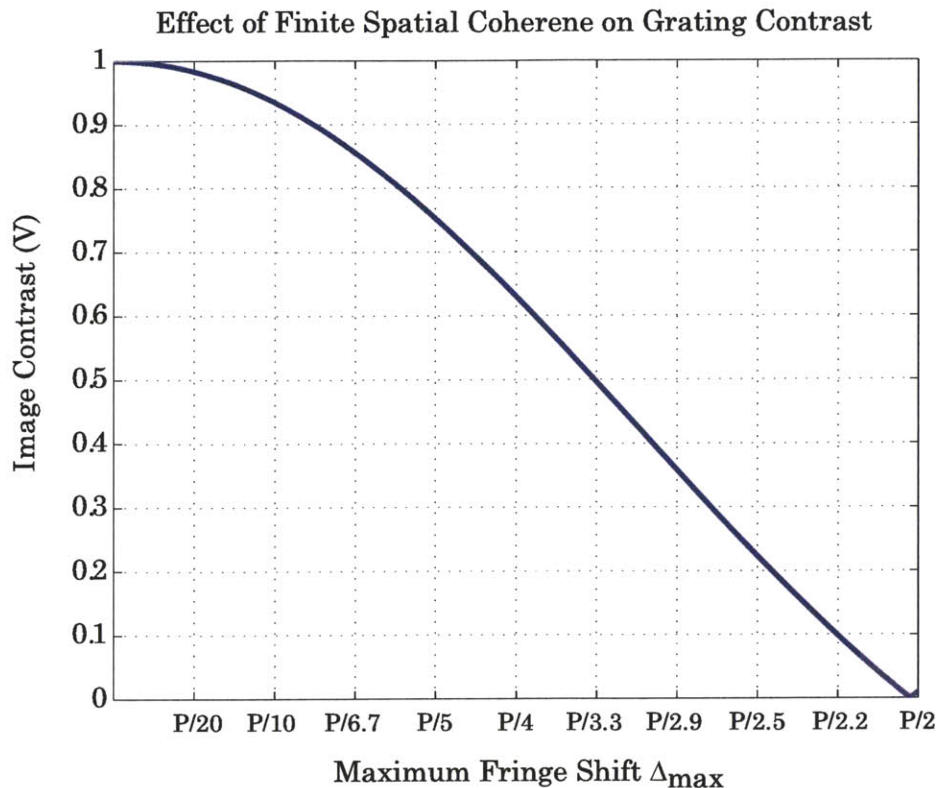


Figure 6.4 The contrast of the interference pattern shown as a function of the maximum fringe shift. The fringe shift is shown as a fraction of the period of the interference pattern.

Although 3μm may seem like a very small depth of focus when trying to bring together two separate objects, it is in fact quite a large depth of focus in the context of the ITFI. Spin coating can reliably achieve films as thin as tens of nanometers, and films 100-200nm-thick are commonplace in a resist stack for ~100 nm features. Defining the separation of the exposure plane from the parent-grating becomes as simple as fixing a spin-speed, no additional alignment is necessary. In light of this, referring to the separation D between the parent-grating and the exposure-plane as "depth-of-focus" is misleading; this distance is fixed and no focusing can occur.

For either the standard near-field spatial-frequency doubling, or the 3-grating AIL scheme the problem could be summarized as: "For the given spatial-coherence of the source, what image contrast am I willing to tolerate to achieve a practical depth of focus?". In the ITFI, however, the parameter D is fixed and can

be made (almost) arbitrarily small, such that it is no longer a limiting factor. Nanometer-level control of the parent-grating-to-substrate gap allows the use of sources with less spatial-coherence while improving the image contrast of the exposure when compared to other grating interferometers .

The maximum tolerable image shift Δ_{max} is defined in Equation 6.4 simply as proportional to the gap D by assuming that the change in diffraction angle is small,

$$\Delta_{\max} = D\Delta\theta_0 \tag{6.4}$$

In conjunction with Equation 5.8, we get

$$\Delta_{\max} = \frac{D \cdot \Delta\theta_i}{\sqrt{1 - \left(\frac{\lambda_{eff}}{P_1}\right)^2}}. \tag{6.5}$$

We can rewrite Equation 6.5 in terms of the maximum allowable spatial coherence, given as an angular spread of the source,

$$\Delta\theta_i = \frac{P_1}{8D}\sqrt{1 - \left(\frac{\lambda_e}{P_1}\right)^2}. \tag{6.6}$$

Looking at this relationship we see that when the wavelength is somewhat smaller than the period of the parent-grating, the diffraction angles are small and the required spatial coherence is a simple linear function of the gap. However, when diffraction angles are high, i.e as P_1 approaches λ_{eff}, the tolerable spatial-coherence decreases.

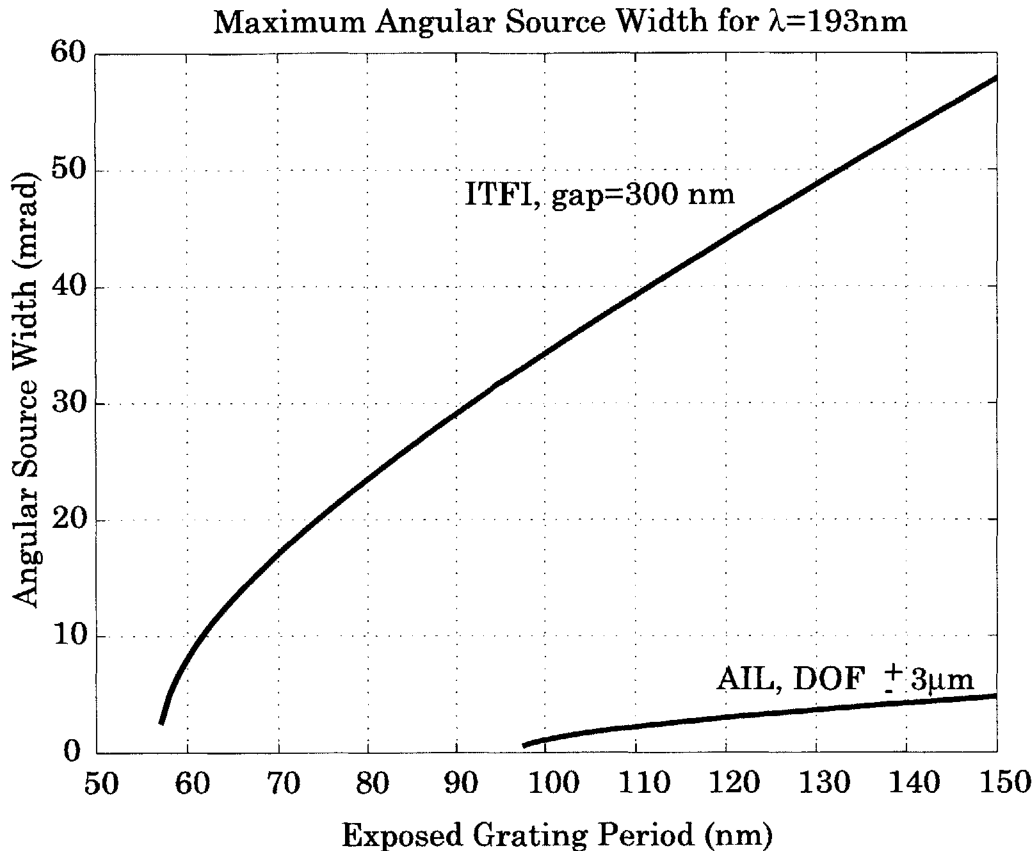


Figure 6.5 Required spatial coherence measured as the angular spread of the source as a function of the grating period for exposures using a 193nm source. The ITFI plot assumes an interlayer gap of 300 nm and an interlayer index of 1.7. The plot for the current AIL system assume a depth of focus +- 3μm.

Figure 6.5 plots Equation 6.6 for the ITFI and the AIL described in Chapter 5. The advantage of the ITFI is twofold; the gap D is an-order of-magnitude less and the high-index medium reduces diffraction angles via a shorter effective wavelength. The angular spread of the source can be about a factor of 30 worse for the ITFI than the AIL at the same grating period. Alternately, once can consider that the image contrast of the daughter-grating in the ITFI will be greater for exposures using the same source. Tbe solid immersion effect is also seen in the

plot, allowing the 193nm ArF laser used for ITFI to expose grating periods down to ~60nm‡.

6.2.5 Source requirements: output power

The availability of sources for use in an AIL configuration was limited not just by spatial coherence, but also by the generally low output power of short-wavelength discharge sources. As discussed in Chapter 5, a number of factors conspire to reduce power transmitted through the AIL system to the wafer. For example, Savas reports that the gratings used in the current AIL diffract ~27% into each first-order beam, and ~57% into the second-order beams [2]. Couple this with ~40% reflection at the air/glass interface for a 70 degree incidence, and the power efficiency through the interferometer is only 8%, ignoring a few other minor factors. The stability of the interference fringes depends on the stability of the relative positioning of the parent-gratings. Nanometer-scale shifts in the position of any of the parent-gratings during the exposure will cause the interference pattern to move, thus blurring the image. For this reason, a high-power source is desirable to minimize exposure times.

The ITFI eliminates this need. The intrinsic stability of the interferometer, discussed in Section 6.2.3, means that exposure times can be arbitrarily long without concern for fringe stability. Limits to the exposure time will only be set by the patience of the user. The tolerance of the ITFI to low spatial-coherence also means that a given extended source can be placed closer to the ITFI than to an AIL system, thereby capturing a greater percentage of the output power. Through the combination of these unique advantages, the restrictions on available exposure sources which are present for AIL systems are largely eliminated for the ITFI.

‡ The ITFI interlayer material is assumed to be PMMA in this plot, with a refractive index n=1.7.

It should be noted, however, that the elimination of requirements on the physical properties of the source, i.e. spatial-coherence and power, does not automatically transform a source at an arbitrary wavelength into a viable option for lithography. Material properties, notably transparency, at the chosen wavelength must be such that lithographic exposures are possible. Although the parent-grating in the ITFI is only on the order of 100 nm thick, it is necessary that the material be transparent for suppression of the zero-order. Absorption of the interlayer, although it reduces power throughput efficiency, does not compromise the image of the daughter-grating.

Therefore, the ITFI runs afoul of the same material issues that plague any photon-based lithographic tool as wavelengths decrease further into the vacuum-ultraviolet. Transparency in polymer materials, both organic and inorganic, drops rapidly below 157 nm. It becomes obvious, as discussed in Chapter 5, that material properties are a more critical limiting-factor than source availability when determining the shortest possible exposure wavelength for lithography. Recall from Figure 5.8 that the thickness of polymer films is limited to less than 7nm for λ_0<~130nm. In practical terms, it would be both unwise and unnecessary to use wavelengths below λ_0=157 nm for ITFI exposures. Using HSQ as a gap material of (refractive index n=1.7 at λ_0=157 nm), the effective exposure wavelength is ~93 nm. The minimum linewidth is $\lambda_{eff}/4$=23 nm, and could in fact be smaller for low-duty-cycle patterns. It is expected that other limits of the lithographic process, such as photoresist resolution, line-edge roughness, or pattern transfer, would halt the continued reduction in feature-size before it became necessary to move beyond the 157nm-wavelength .

6.3 Diffraction Efficiency

6.3.1 Introduction

So far, it has been assumed that all transmitted light is in the first-order beams. As we mentioned in the discussion of near-field SFD in Chapter 5, achieving this ideal will probably not be possible. In fact, difficulty in sufficiently suppressing the zero-order beam was cited as one reason why near-field SFD with a monolithic parent-grating was dismissed as impractical. Thus, it is crucial to a successful ITFI design to understand what attributes of the parent-grating are necessary for minimizing the zero-order. Additionally, methods for achieving good results even in the presence of a finite zero-order contribution will be discussed. It is expected that if an ideal design can be found, the performance of the actual parent-grating will very closely match the predicted diffraction efficiencies. Tooth height and width can be controlled with nanometer-level precision and accuracy using spin-coating and interference lithography§. Fabrication of the parent-gratings in this fashion should allow for much better control of the grating parameters than is possible with other means.

6.3.2 Scalar Diffraction Theory

In examining these issues, we will see that what appears to be an almost trivial optical system actually harbors some very subtle and non-intuitive physical phenomena. This is, in part, due to the scale of features in the ITFI being on par with, or smaller than, the optical wavelength. Intuition is based on the observable results of experiments that match our understanding of mathematical equations which describe the process. Unfortunately, neither the equations nor the experiments will generally apply in this case. Direct observations of diffraction

§ Refer to Section 7.1 for a discussion on linewidth control in IL exposures.

invariably happen in the far-field regime. Familiar analytical solutions for diffraction problems use the scalar approximations of Maxwell's equations, which assume the wavelength is significantly shorter than the feature size and the propagation distance is large.

For comparison, the diffraction efficiency η for a given order m predicted by the scalar theory is given in Equation 6.7 [3]. The duty-cycle f, and tooth height h are simple physical parameters, while optical indicies and incident-angle are wrapped up in the parameter $\Delta\phi$, which is the phase difference between the tooth and trench areas. In Equation 6.8, α and β are the propagation angles in the tooth and trench, respectively.

$$\eta_m = 4f^2 \mathrm{sinc}^2(mf)\cdot\cos^2\left(\tfrac{h\pi}{\lambda_0}\Delta\phi + \tfrac{m\pi}{h}\right) \qquad (6.7)$$

$$\Delta\phi = \tfrac{2\pi h}{\lambda_0}\left(n_2\cos(\beta) - n_1\cos(\alpha)\right) \qquad (6.8)$$

It is interesting to note that the all of the physical parameters of the grating, except the duty-cycle, are only included in the cosine term of Equation 6.7. The efficiency will be optimized in a given order when the cosine term is unity, and minimized when the cosine is null. This can be simplified by considering only the relevant situation of normally incident light, and minimizing the zero-order. Note that this condition, expressed in Equation 6.9, simultaneously maximizes first-order diffraction. The optimal tooth height for suppressing the zero-order is given in Equation 6.10, where Δn is the difference between the tooth index n_2 and the trench index n_1.

$$\frac{h\pi\Delta\phi}{\lambda_0} = \frac{\pi}{2} \qquad (6.9)$$

$$h_{optimal} = \frac{\lambda_0}{2\Delta n} \qquad (6.10)$$

The condition for minimizing the zero-order beam, then is that the light transmitted through the tooth will be exactly out of phase with the light transmitted through the trench

Examinations of the discrepancies between scalar theory and full-vector solutions for square-tooth gratings can be found in references [3,4]. For situations

where the feature size is less than $\sim 15\lambda_0$, error in the predictions of scalar theory becomes significant (>5%). In the most interesting cases of the ITFI, where the periodicity of the parent-grating (P_1) is actually *less* than the vacuum wavelength λ_0, accurate analytical solutions for diffraction phenomena are impossible and numerical methods must be used.

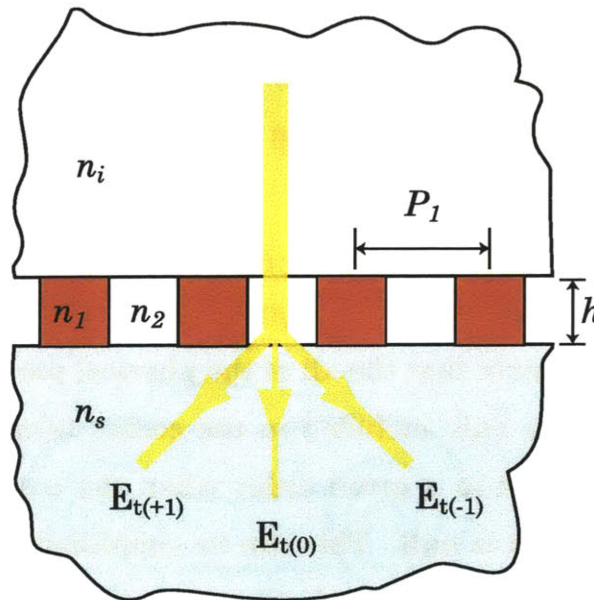


Figure 6.6 The general model used for calculation of diffraction efficiency, both in the numerical simulations and in the equations of scalar diffraction efficiency. Grating teeth are assumed to be rectangular in profile. Infinite half-planes are assumed for the incident and substrate media. The refractive index of the grating tooth and trench are defined separately. Only first-order diffraction is shown, although this is not a necessary condition.

6.3.3 Numerical Modeling

The realm of the ITFI -diffraction from sub-wavelength features over wavelength-scale propagation distances- is only accessible through rigorous numerical simulation. The coupled-mode technique was used to solve Maxwell's equations for the diffraction efficiency of a square-tooth grating [5,6,7] using a

hybrid algorithm developed by Anderson [8]. The simplified physical model used to calculate the diffraction efficiency of the parent-grating is shown in Figure 6.6.

Grating teeth are assumed to be perfectly rectangular, and the top and bottom planes are assumed semi-infinite. The grating tooth and trench materials are independently defined and can have complex refractive indicies. In general, normal-incidence is also assumed, and the desired result is suppression of the zero-order transmission. The periodicity of the parent-grating, unless otherwise stated, will be P_1=180 nm, and the wavelength λ_0=193 nm. The refractive index of the grating material, unless otherwise stated, is $n_1=1.72-0.03i$. A resist developed specifically for use with the ITFI is described in Section 6.4.1. The material between the grating teeth, also called the trench, is air ($n_2=1$).

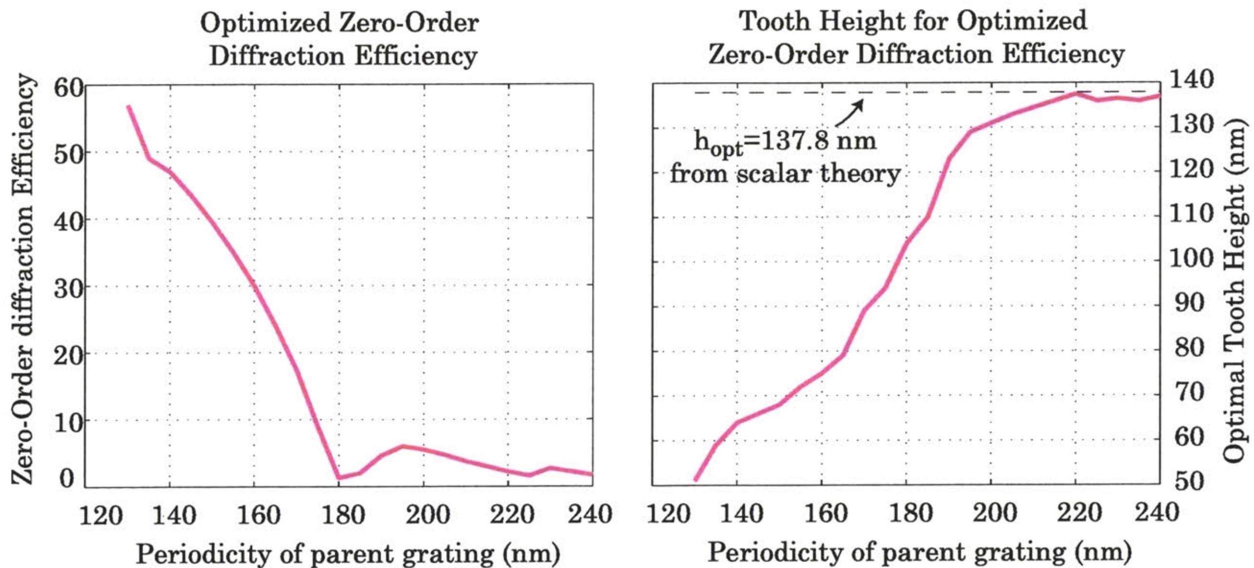


Figure 6.7 The diffraction zero-order efficiency as a function of grating period for the optimal duty-cycle and tooth-height (left). The optimal tooth height as a function of period is shown in the plot at the right.

Figure 6.7(a) shows the minimized zero-order diffraction-efficiency for an optimized grating as a function of periodicity. For parent-gratings whose period is significantly below the vacuum wavelength (λ_0<193nm), the zero-order contribution rises rapidly. Additionally, the optimal tooth height, shown in Figure 6.7(b) decreases for finer-period parent-gratings.

Considering the corresponding high zero-order efficiency, the "optimal" tooth height is actually not very well optimized. The tooth height predicted by scalar-diffraction theory, shown as a dotted line across the top of Figure 6.7(b), is such that light transmitted through the tooth and through the trench are exactly out-of-phase. We see that the true optimal height is well matched with the scalar prediction in the same region that the zero-order diffraction efficiency can be successfully cancelled, $\lambda_0 > P_1$. Intuitively, it is not surprising that a tooth which is too short to affect a π phase-shift will be unable to fully cancel the zero-order. It is not clear, however, why the shorter tooth is a more optimal condition for sub-wavelength gratings, and why the "minimized" zero-order efficiency is actually so high.

Although more investigation is necessary to fully understand this effect, it is likely that the limiting condition is the depth of the trench, rather than the height of the tooth. While light in the tooth has an effective wavelength λ_0/n_1=112nm, light in the trench remains at the original, longer, wavelength (λ_0=193nm). The trench, which is narrower than half a wavelength, cannot support propagation. Therefore, light in the trench will be increasingly attenuated at greater depths. The optimal conditions seen in Figure 6.7 represent a compromise between the ability of the tooth to phase-shift and the ability of the trench to transmit through an aperture in which propagation is not allowed.

This interpretation fits well with the effect of tooth index on diffraction efficiency, shown in Figure 6.8. The plot on the left (a) shows the minimized zero-order efficiency as a function of refractive index for a 180nm-period grating. The suppression of the zero-order is highly correlated to index, with high-index materials being preferable. The plot on the right, (b), shows the tooth height of the optimized grating as a function of index as a solid line, and the tooth-height necessary for a π-phase shift as a dotted line. For higher-index materials, the phase-shift of the tooth in the optimized grating approaches π, and the diffraction efficiency of the zero-order drops dramatically. If the trench-depth (in air) is indeed

the limiting factor, then it makes sense that a higher-index tooth is beneficial. The more rapid phase-shifting of the high-index material requires less propagation through the trench, and therefore less attenuation occurs.

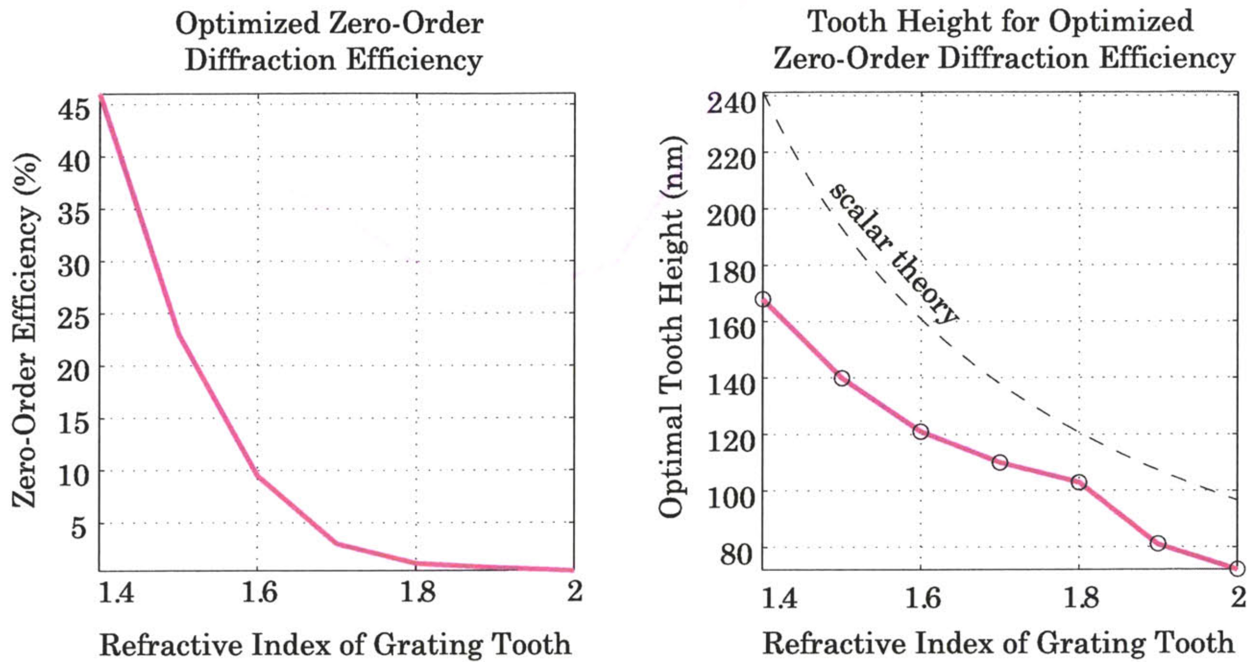


Figure 6.8 (a) Diffraction efficiency as a function of tooth index for a 180nm-period grating illuminated with λ_0=193nm. (b) the optimal tooth height as a function of tooth index for the same situation. Optimal tooth height is seen to be consistently lower than the predictions of scalar diffraction theory, shown by the dotted line.

Figure 6.9 shows the zero-order efficiency as a function of duty-cycle for a 180nm-period parent-grating made in resist. The optimal duty-cycle is found to be around 33%, significantly less than the 50% optimal duty-cycle predicted by scalar theory. This also makes sense in the context of the sub-wavelength grating trench. Trench width increases for decreasing duty-cycle, potentially becoming wide enough to again support propagation. Additionally, attenuation which does occur is offset by the greater area intercepted by a wider trench. Thus, the lower duty-cycle is a method to equalize the power transmitted through the tooth and trench. This is, in fact, advantageous from a practical perspective. It will be shown in Chapter 7 that low duty-cycle lines have a wider process window than lines with a 50% duty-cycle.

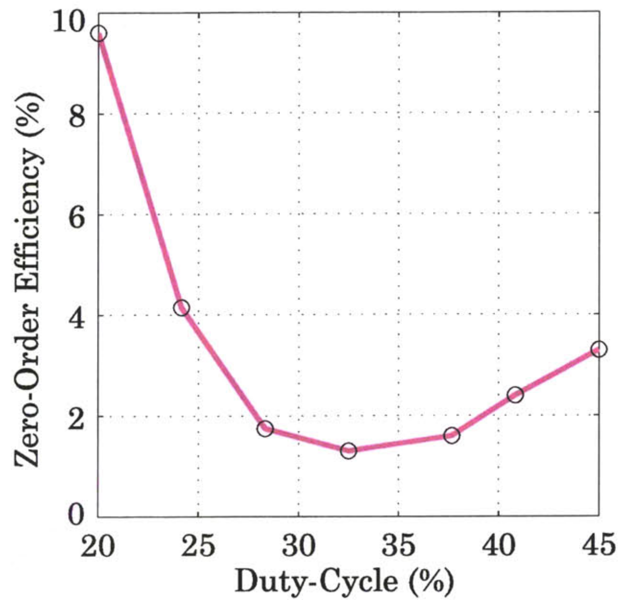


Figure 6.9 Diffraction-efficiency as a function of duty-cycle for a 180nm-period grating with a tooth-height of 103nm. The grating tooth index is *N=1.72-0.03i*. The minimum zero-order efficiency is found at a lower duty-cycle, ~33%, than would be expected from scalar theory predictions.

Strategies for further suppressing the zero-order, especially for parent-gratings much smaller than the wavelength, will be discussed in Section 6.6. The parameters for the optimized 180nm-period grating are summarized in Table 6.1. For now, assume that the first-order efficiency of the 180nm-period parent-grating is sufficiently high that we can shift our focus to the to a demonstration of the feasibility of the ITFI concept.

Table 6.1

Parent-grating Parameters	
Period	180 nm
Index	1.73-0.03i
Height	103nm
Duty-Cycle	33%
Zero-Order	1.3%

6.3.4 Finite-difference time-domain simulations (FDTD)

Coupled-mode calculation of the diffraction efficiencies is extremely useful for determining which grating parameters minimize the zero-order contribution. Generally given as a real number representing the fraction of transmitted power in a certain order, the efficiency is actually the square-modulus of a complex electric-field coefficient that carries both amplitude and phase. By propagating these electric-fields through an arbitrary distance z, the field at any plane beneath the grating can be determined. However, building a picture of the intensity distribution as a function of depth below the grating in this fashion is rather tedious. A complementary simulation technique, the finite-difference time-domain method [9], while not suited for optimizaton of the grating geometry, is a convenient way to visualize the diffraction pattern from a known grating. The software used was a public-domain FDTD package called TEMPEST available from the University of California, Berkeley [10,11].

Figure 6.10 shows an FDTD simulation of the intensity pattern beneath the optimized parent-grating (**Error! Reference source not found.**) in a semi-infinite high-index medium. The medium is assumed to be the spin-on-glass HSQ (hydrogen-silsesquioxane), with an index of $n_s=1.534$. The outline of the parent-grating is shown in pink. It is clear that the desired result, a pattern with twice the spatial-frequency of the parent-grating, is achieved for all depths. It is also clear, however, that there is periodicity in the vertical direction caused by the presence of the zero-order. A cross section of the intensity at any plane will show the frequency-doubled image with a modulation envelope at the period of the parent-grating. Recall that this was depicted schematically in Figure 5.3.

Interestingly, at certain depths indicated by the dotted lines, amplitude of the modulation envelope goes to zero. In these planes, a perfect frequency-doubled image is formed. If the resist layer for imaging the daughter-grating is placed in one of these planes, the deleterious effects of the zero-order can be overcome. Thus, the thickness of the gapping layer between the parent- and daughter-gratings can

be tuned such that excellent frequency-doubled daughter-gratings can be exposed despite the uncompensated zero-order diffraction.

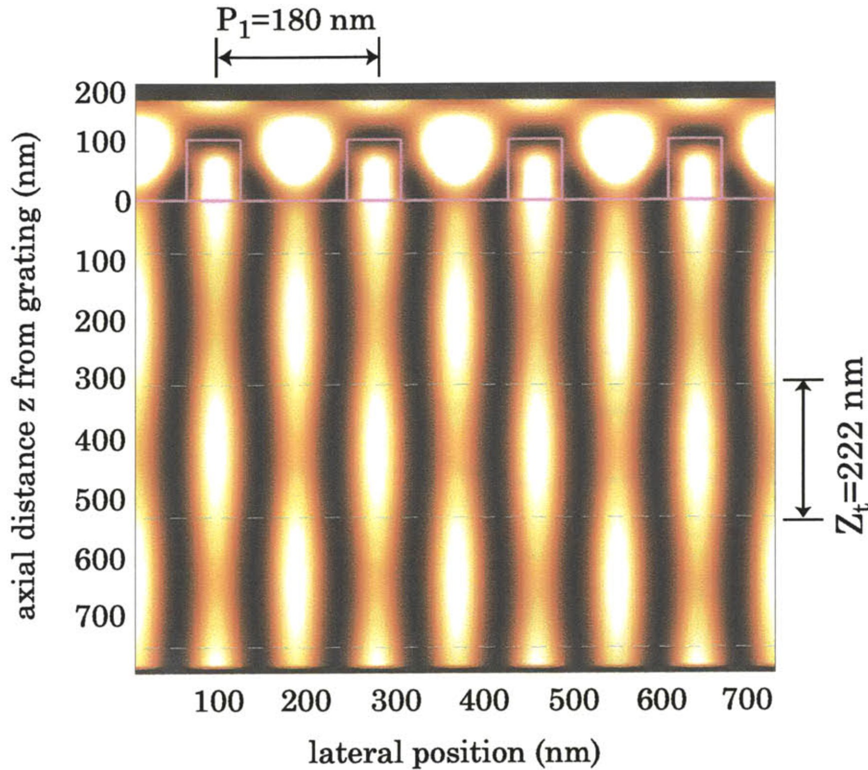


Figure 6.10 FDTD simulation of the diffracted intensity from the parent-grating shown outlined in pink. The grating period is 180 nm, with a 33% duty-cycle and a tooth height of 102 nm. TE polarized, 193nm illumination is used. The Talbot sub-planes, separated by z_t=222nm, are shown by the dotted lines.

6.3.5 The Talbot effect

The formation of high-contrast frequency-doubled images at certain planes is characteristic of the self-imaging of gratings first reported by Talbot in 1836, known as the Talbot effect [12]. More generally, this phenomenon is just one example of the self-imaging described by Montgomery for all periodic structures [13,14]. The common description of the Talbot effect is formulated for amplitude gratings from the Fourier-transform description of the Fresnel diffraction equation [15]. At axial distances which are integer multiples of z_t (known as the Talbot distance), a perfect

image of the grating is formed [16]. Frequency-doubled images, known as sub-images, form at planes halfway between the self-images.

$$z_t = \frac{nP^2}{\lambda} \tag{6.11}$$

The self-imaging of phase gratings, however, is subtly different than that of amplitude gratings. The self-images of a phase object are themselves only phase objects, and have no intensity contrast. Intensity images of a phase parent-grating form in the sub-planes, while a frequency-doubled image *does not form at any plane* [17]. This is very different than the ideal ITFI, however, where a frequency-doubled intensity image is formed at *all planes below the grating.*

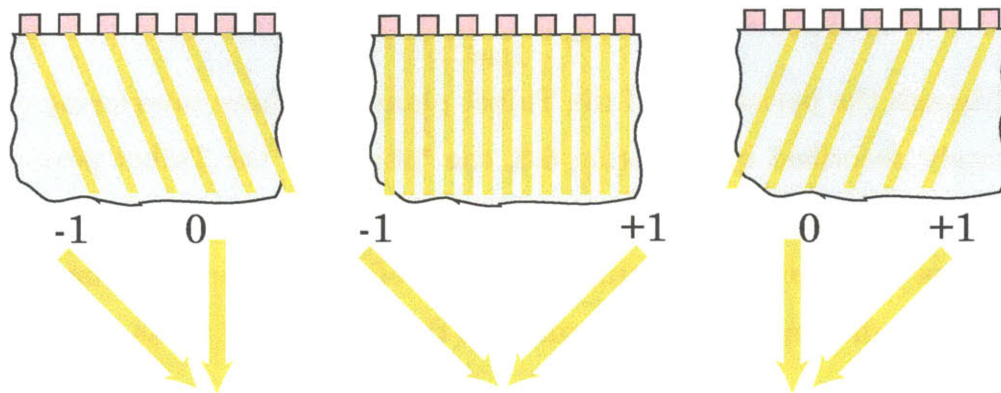


Figure 6.11 Intensity patterns formed by the interference of separate pairs of orders in the case of only first- and zero-order diffraction. The inclination of the fringes for first and zero orders means that the lateral intensity distribution varies as a function of the axial distance.

The dramatic discrepancy of these predictions can be ascribed to the Fresnel equation, on which the theory of Talbot images is based, which is not valid for wavelength-scale structures or large diffraction angles. It has been reported that predictions of the Talbot effect based on the Fresnel equation will not be valid for ratios of $P_1/\lambda_0<20$ [18].

For the case of only three orders ($m=-1,0,+1$), an intuitive (if not rigorously correct) understanding of vertical periodicity can be found by imagining the overlap of three sets of standing waves, depicted in Figure 6.11. Note that the fringes

formed by the interference of the first- and zero- orders shift laterally as a function of the axial distance z. The inclination θ_{inc} of these fringes with respect to the plane of the parent-grating is equal $\theta_1/2$, half the diffraction angle of the first-order beam. The planes where the maxima of these fringes are in phase will be the planes in which the modulation envelope, seen in Figure 6.10, is maximized. The distance z_t between these planes, which we will also call the Talbot distance for convenience, is given in Equation 6.8. For the 180nm-period parent-grating, the classical Talbot distance given by Equation 6.7 is z_t=257nm, somewhat larger than the distance observed in Figure 6.10 and Figure 6.12. The actual distance observed in these simulations, z_t=222nm, is accurately predicted by Equation 6.8

$$z_t = \frac{P_1}{2\tan(\frac{\theta_1}{2})} \qquad (6.12)$$

Although the periodicity of the Talbot planes for three interfering orders can be accurately described by Equation 6.8, the "phase" of the Talbot periodicity, i.e. actual location of the planes, is not easily characterized. While the locations of the Talbot planes for amplitude-gratings are well defined, grating parameters such as tooth index, height, and duty-cycle will shift the locations of these planes for phase gratings.

A depiction of this phenomenon is shown in Figure 6.12. The amplitude of the modulation envelope is shown for a number of different cases**: parent-gratings with low zero-order efficiency are shown on the left, and gratings with high zero-order efficiency are shown on the right. This distinction is made only to avoid the confusion of too many curves on one axis.

** The modulation envelope can be easily determined from an FDTD simulation (e.g. Figure 6.10) by comparing the intensity along vertical stripes at the peak and null positions of the frequency-doubled image.

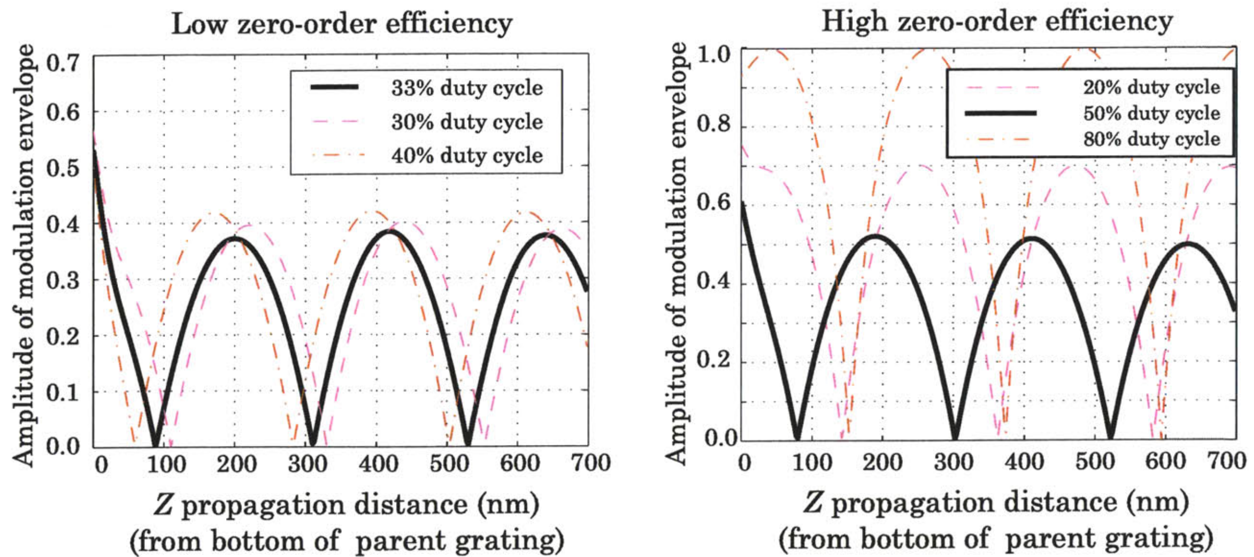


Figure 6.12 The amplitude of the modulation envelope as a function of the propagation depth z in HSQ for selected duty-cycles. The parent-grating has a period P_l=180 nm, and the grating tooth has a height of 103 nm. Notice that the null points for the different curves occur at different distances. from the parent-grating The plot on the left shows values for D.C. which have low zero-order efficiency, D.C. values on the right have high zero order efficiency. The periodicity of the envelope is 222nm, as given by Equation 6.8.

Distilling a plot such as Figure 6.10 down to a single curve can help illuminate some very important practical issues. Recall that the null-planes of the modulation envelope were proposed as a good location for imaging the daughter-grating. The location of these planes, as well as their necessary placement accuracy, can be clearly seen in the above plots. Even the small zero-order contribution of the optimized parent-grating, with $E_0=1.3\%$, is seen to have a strong modulating envelope (the solid line in Figure 6.12(a)). The "depth-of-focus" collapses towards a single plane for parent-gratings with high-zero-order efficiency, such as the 80% duty-cycle grating shown in Figure 6.12(b). Although not shown in the above plots, the *contrast* of the frequency doubled image in the null-planes of the envelope is also reduced with increasing zero-order. Thus, the concerns of contrast and depth-of-focus, which are influenced by spatial coherence in other achromatic schemes, are brought back to the ITFI by the presence of the zero-order.

The nature of the DOF, however, is somewhat different than in other schemes. The thicknesses of the interlayer and the daughter-resist, which determine the gap between the parent-grating and the exposure plane of the daughter-grating, are set with nanometer precision and accuracy using spin-coating. No adjustment or focusing can occur. Instead, the DOF refers to allowable thickness of the photoresist layer for which the modulation envelope is sufficiently small. Comparing the plots in Figure 6.12, it is clear that the window around the null positions for which the envelope is below a given amplitude is reduced for higher zero-order contributions. Tolerable zero-order contributions will be determined by one of two factors; either a minimum required contrast level or a minimum required photoresist thickness for the exposure of the daughter-grating.

Two other points are worth noting. One is that the "phase" of the envelope changes with different duty-cycles. In fact, all the grating parameters, i.e. tooth height or duty-cycle, can change the position of the optimal imaging planes without changing their periodicity. Thus, control over the parent-grating geometry is necessary for placement of the imaging plane as well as suppression of the zero-order. Additionally, in the region closest to the parent-grating, the regular periodicity seen at larger distances breaks down due to the effects evanescent orders. The effect of the evanescent fields disappears at propagation distances $z>$ ~50 nm. This can be interpreted as a lower limit to the thickness of the HSQ gapping layer.

6.4 Thin-Film Stack Design

6.4.1 A photoresist for the parent-grating

One subtlety to the approach outlined in Section 6.1.3 is that the material of the parent-grating, in addition to its imaging requirements as a mid-UV photoresist, must also function as a DUV diffractive optical element. To optimize the first-order diffraction efficiency from the parent-grating, we want it to act as a

phase-modulating element and not an amplitude-modulating one. Therefore it must be highly transparent to the λ_0=193nm photons used for exposing the daughter-grating.

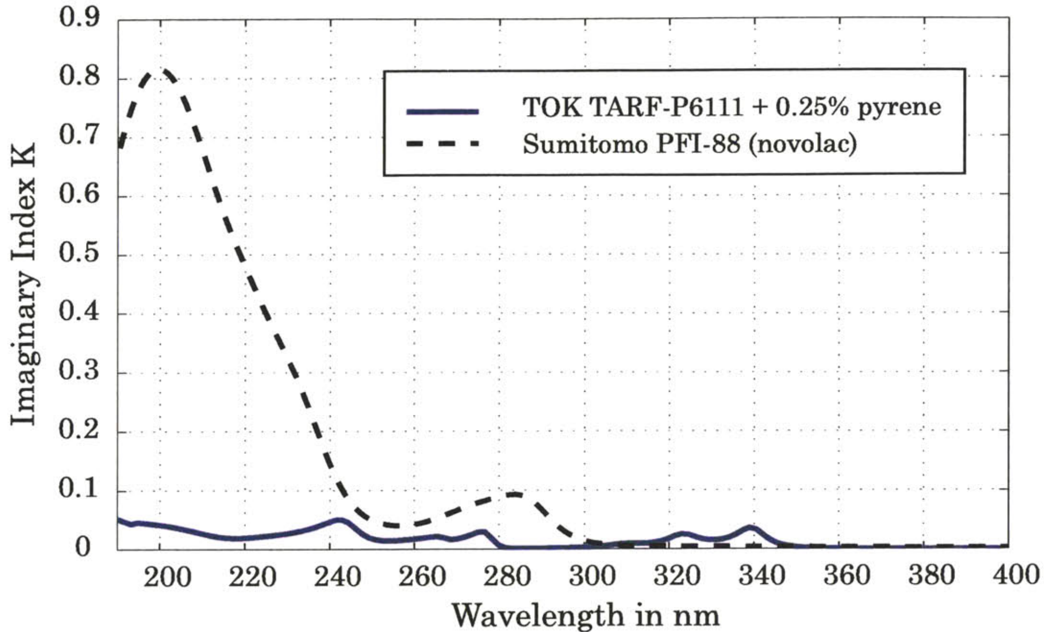


Figure 6.13 Absorption as a function of wavelength is plotted for (solid) Sumitomo PFI-88 A2, a representative i-line novolac resist, and (dotted) TOK TARF-P6111, a newly developed methacrylate resist for 193nm which was doped with pyrene to increase sensitivity at λ=325nm Lloyd's-mirror exposures. Absorption values were measured using spectroscopic ellipsometry.

This requirement is not trivial, as all commercial photoresists developed for mid-UV use novolac (cresol-formaldehyde) as a base resin. Absorbance of the novolac polymers, plotted in Figure 6.13, increases rapidly below 300 nm, to the point of being as opaque as many metals at λ_0=193nm. Although not shown, photoresists designed for λ_0=248nm exposure, based on poly(hydroxystyrene), are also too absorbing at λ_0=193 nm. Newly developed resists based on methacrylate polymers for λ_0=193 nm have the required degree of transparency, but like PMMA they are insensitive to light above ~250 nm [19,20]. Thus, no commercially available product meets the necessary requirements.

The solution to this problem was to enhance the mid-UV sensitivity of a commercial 193nm resist††, which offers the necessary transparency, without significantly increasing its absorption. The sensitivity of chemically-amplified resists is determined by the absorption spectrum of a component known as the photoacid-generator (PAG). Onium salts, the class of PAG's used in most commercial resists [21], have been shown exhibit efficient energy transfer via electrons with polycyclic aromatic compounds such as anthracene and its derivatives [22]. By absorbing longer-wavelength photons and transferring their energy to the PAG, these compounds can be used as sensitizers for chemically-amplified resists [23,24].

Pyrene‡‡ was chosen as a sensitizer based on its strong absorption peaks at 320nm and 335nm, and reduced absorption at 193nm compared to other polycyclic aromatic hydrocarbons [25]. It was found that concentrations of 0.25% -0.5% pyrene by weight in the base photoresist were sufficient to allow exposure of the sensitized resist at 325nm with doses comparable to a standard i-line resist. The absorption at 193nm, plotted as a solid line in Figure 6.13, increased only slightly over its initial value.

A major difficulty in the use of chemically-amplified photoresists is their extreme sensitivity to environmental and process conditions. Minute amounts of airborne contaminants, as well as surface contamination of the substrate, are enough to impede or prevent the catalytic reaction which leads to a solubility change in the exposed regions of the resist [26,27]. To prevent contamination, laboratories and industrial facilities which use CA resists are heavily filtered for organic vapors as well as particulate matter. As the Nanostructures Lab here at MIT is not filtered for organics, significant process development was required to overcome the contamination and achieve high-quality resist structures at high-

†† TARF-P6111 from Tokyo Ohka Kogyo (TOK), a new, positive, chemically-amplified resist with a methacrylate polymer base.

‡‡ Pyrene: C_6H_{10} CAS# 129-00-0

resolution for use as parent-gratings in the ITFI. A comparison of the type of resist profiles observed using an unoptimized (a) and optimized (b) resist process are shown in Figure 6.14. The formation of an insoluble surface layer, as well as "footing" (widening of the line at its base), both visible in the micrograph on the left, are indicative of a contaminated resist. Using optimized baking steps before and after the exposure, a modified developer, and a protective overcoat during the exposure, resist profiles such as those shown in Figure 6.14(b) were achieved. Two different overcoats were used interchangeably, poly-acrylic acid (PAA) and a flouropolymer from TOK called TSP3A. Neither has yet proven superior.

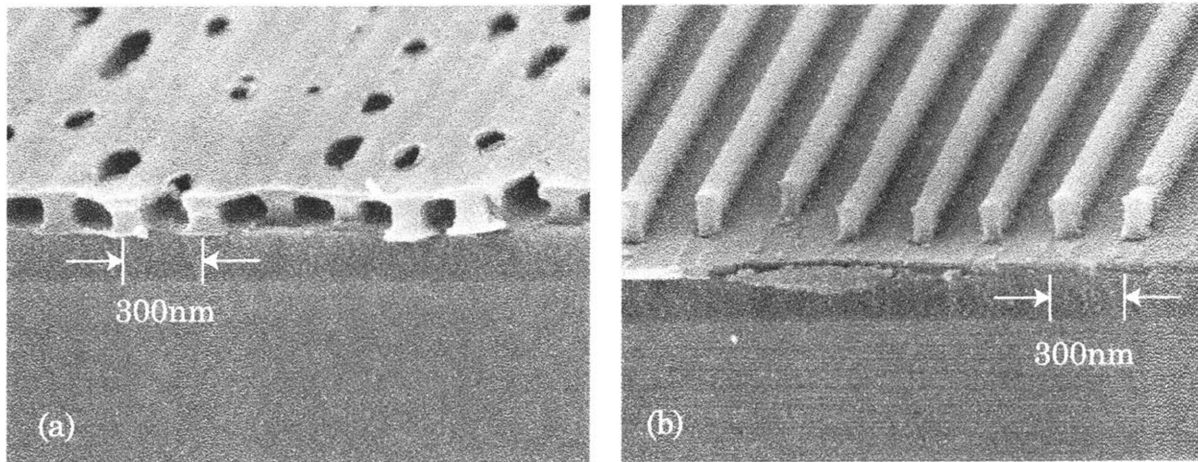


Figure 6.14 (a) an example of resist structures in the pyrene-sensitized resist (TOK TARF-P6111 + 0.25% Pyrene/wt) using standard bake and development conditions. (b) Example of the same resist using optimized bake and development conditions at a ~33% duty-cycle.

6.4.2 Anti-reflection properties of the thin-film stack

The design methodology illustrated below, while detailed for the specific case of 90nm-period daughter-gratings, will be generally applicable to other periodicities. Readers unfamiliar with wafer-processing techniques may find it valuable may to first read Section 7.2 in Chapter 7, which discusses resist processing. Anti-reflection coatings (ARC) under the photoresist, which function through a

combination of thin-film interference and absorption, are used in the optical lithography of fine features to reduce unwanted interference effects. We can begin the design of the ITFI thin-film stack by considering only the exposure of the daughter-grating. A tri-layer (see Section 7.1.3) is used to minimize back-reflections into the daughter-resist, which includes an absorbing oxide material over a standard organic ARC§§. A schematic of the thin-film stack used to model the reflectivity is shown in Figure 6.15a. The imaging layer for the daughter-grating is PMMA, and the incident medium (the interlayer between the PMMA and the parent-grating) is HSQ. Calculations for the reflection coefficient at the interface of a layer of PMMA the DUO are shown in Figure 6.15b. The angle of incidence is set at θ_i=0.77rad, the angle of first-order diffraction inside the HSQ interlayer from a parent-grating with periodicity P_I=180 nm. Complex refractive-index values for all relevant materials presented can be found in Appendix A.

Shaded regions in the reflectivity plot correspond to acceptable combinations of the two ARC layers, with the darker regions having lower reflectivity. If the thickness of the DUO193 interlayer is held between 15-20 nm, there is a very large range of acceptable thicknesses for the XHRi layer.

§§ Rather than the standard interlayer material of evaporated SiO_2, a new spin-on-glass available from Honeywell was used, marketed under the trade-name DUO193. DUO is an organo-siloxane polymer similar to HSQ with additional dyes attached to the molecule such that it absorbs at λ_0=193 nm. Thus, it can provide the absorptive properties of an anti-reflection coating as well as the high etch-selectivity of an inorganic oxide. The organic ARC is XHRi from Brewer Science Inc..

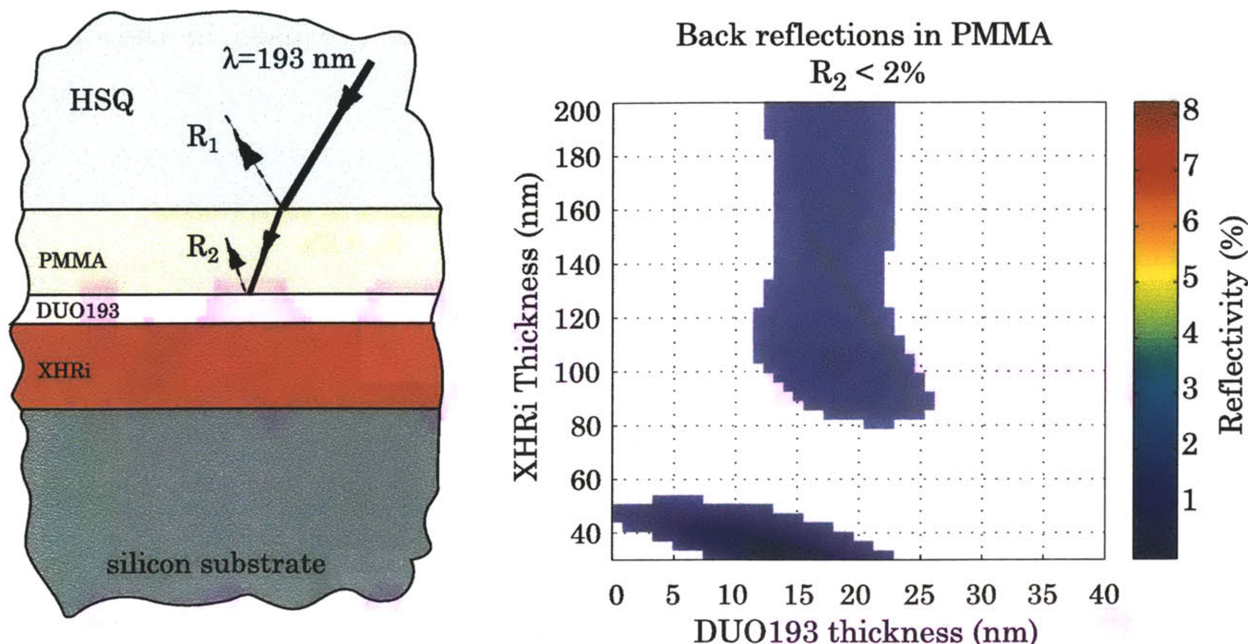


Figure 6.15 A) Schematic of the thin film stack simulated to determine the thickness of the XHRi and DUO193 layers which minimize the reflection coefficient R_2 during the daughter-grating exposure in the PMMA. B) Reflection coefficient R_2 shown for variable ARC thicknesses for of λ_0=193 nm illumination with an incident angle 0.77 rad in HSQ (n_{HSQ}=1.534). Data above 2% is removed from the plot to visualize the process window.

This is a useful property when we consider that this same thin-film stack must also be optimized to minimize back-reflections during the exposure of the parent-grating, which occurs at a different wavelength and incident angle. Reflectivity minimization is necessary in the exposure of the parent-grating to form grating teeth with square profiles. To help minimize back-reflections during both exposures, at least one of the remaining layers must be eliminated as a variable. One important criteria when defining the thickness of the PMMA layer is to ensure that the aspect-ratio (resist height divided by linewidth) of the developed pattern will be below 2:1. As a rule-of-thumb, this prevents mechanical deformation of the resist due to surface-tension during development (see Section 7.1.4). If we fix the thickness of the PMMA film at 75 nm, another simulation can be run to determine the best thicknesses of the XHRi layer and the HSQ interlayer for the parent-

grating exposure. Recall that any XHRi thickness above 80 nm was acceptable for exposure of the daughter-grating.

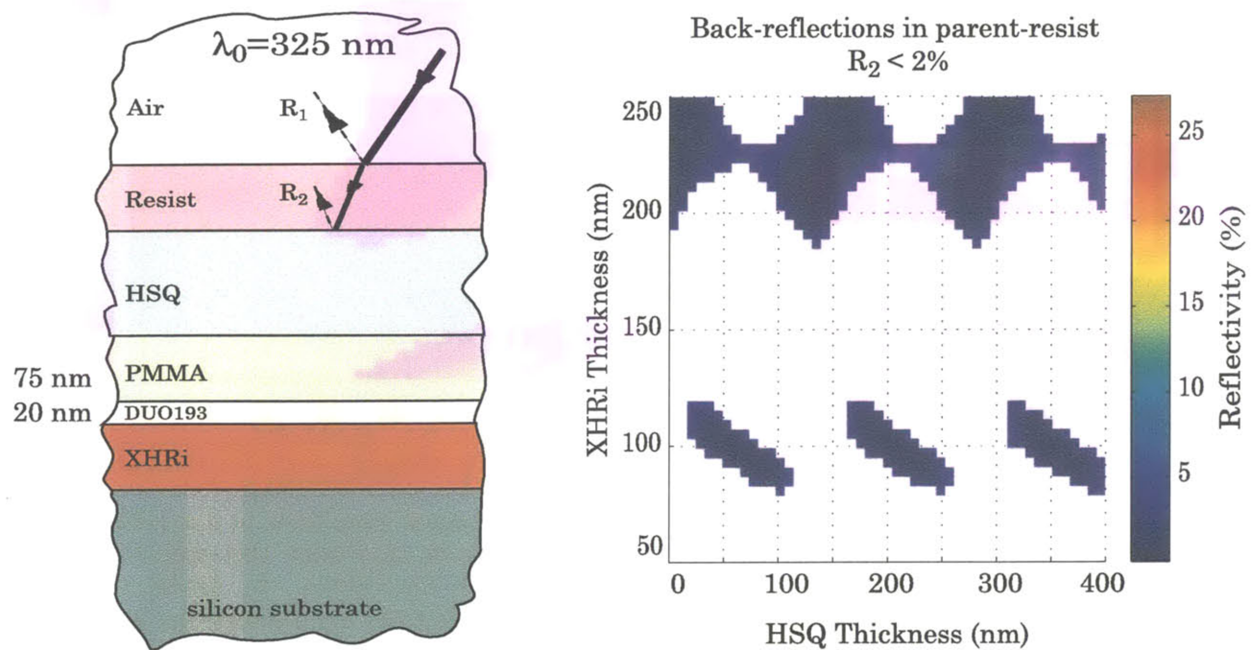


Figure 6.16 Schematic and simulations of the back reflections into the photoresist during the exposure of a 180 nm period parent-grating. The illumination is λ_0=325 nm at an incident angle of 1.13 rad. The reflection coefficient as a function of thickness for both the XHRi anti-reflection layer and the HSQ gapping layer is shown only for values of R_2<2% to approximate the process window.

The data in Figure 6.16 indicates that there is a lot of latitude available for choosing the thickness of both layers, the XHRi (ARC) and the HSQ (interlayer), although we know from Section 6.3.5 that the thickness of the interlayer should not be chosen arbitrarily. Recall from Section 6.3.3 that thickness of the top photoresist layer, 103nm, is determined by the tooth-height of the parent-grating which minimizes zero-order diffraction.

6.4.3 Oxide Gapping Layer

It was found that spin-coating of the HSQ and development of the parent-grating work better if the HSQ film is chemically isolated between two thin layers of evaporated SiO_2. The evaporated oxide between the PMMA and HSQ films prevents mixing of the two layers, since the developer for PMMA (methyl-isobutyl ketone, MIBK) is also used as the casting solvent for HSQ. The evaporated oxide between the parent-grating and the HSQ is used to prevent the developer of the parent-grating (aqueous TMAH***) from attacking the HSQ.

It would be desirable to use only evaporated oxide, or only HSQ, however, neither material alone can meet all the necessary requirements. The evaporated oxide is chemically compatible with the other layers, but the deposition has an unacceptably large thickness error, up to ~30%†††. The thickness of the spun HSQ can be controlled very precisely, but it is incompatible with the PMMA and top resists. Thus, the sandwich is used: evaporated oxide layers are used to isolate the HSQ, but are very thin to minimize deposition error. It is likely that with additional research, a spin-on glass other than HSQ can be found which uses a different casting-solvent and is resistant to alkaline developers. Unless otherwise noted, however, the oxide sandwich will be used for the gapping layer in all ITFI designs described in this work.

*** Alkaline solutions are used as developers for almost all commercial photoresists. In this case , MFCD-26 from Shipley was used, a 0.26N solution TMAH (tetra-methyl ammonium hydroxide) in H_2O.

††† Although the thickness control of e-beam evaporated films *should* be very good, this is not the case for the particular tool used in these experiments.

6.5 Experimental Results

6.5.1 Design Review

At this stage, the methodology for designing the thin-film stack is well defined. The two-wavelength exposure system allows for the entire process of fabricating both the parent- and daughter-gratings to be only slightly more work than the process for other IL gratings. Fabricating the parent-grating is identical to the exposure process for gratings made with current IL techniques. Exposure of the daughter-grating is trivial; an unaligned flood-exposure followed by a three non-critical "development" steps. The first two steps remove the parent-grating and gapping layer, and the third develops the resist image of the daughter-grating.

Based on the results of numerical calculations, ITFI wafers were prepared with a variety of different HSQ spacer thicknesses. Results will be shown comparing wafers that were prepared with interlayer thicknesses of 360nm and 260nm. According to the plot in Figure 6.12, these thicknessses should place the daughter resist close to the worst and best exposure planes, respectively, based on the strength of the modulation envelope.

6.5.2 360 nm interlayer thickness

If the sum of both evaporated oxide layers is ~20 nm, an HSQ layer spun 340nm-thick will give a total interlayer thickness of 360nm. The reflectivity plot shown in Figure 6.16 is used to find the thickness of the ARC layer which is optimized for this interlayer thickness. The full thin-film stack for this design is shown in Figure 6.17. The 180nm-period parent-grating is exposed using 325 nm illumination, while the 90nm-period daughter-grating is exposed with 193nm illumination.

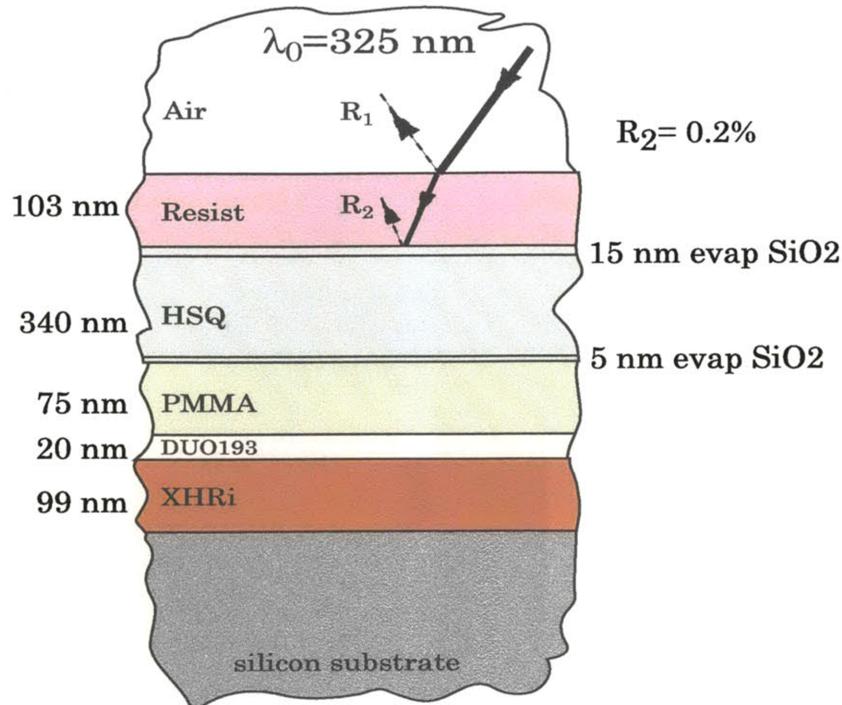


Figure 6.17 Thin-film stack for ITFI. The thickness of the oxide/HSQ gapping layer is designed such that the exposure of the daughter-grating in the PMMA film occurs at the Talbot sub-image plane predicted by paraxial diffraction theory, z_t=386 nm for λ_0=193 nm.

Parent-gratings with 31% and 36% duty-cycle are shown in Figure 6.18. All exposures were performed using the Lloyd's-mirror interferometer and the duty-cycle was varied by changing the exposure dose.

In both micrographs the residual effects of resist contamination are clear: significant line-edge roughness and a slight "T" shape in the resist profile‡‡‡. It is unclear why the excellent profiles achieved in this resist shown in Figure 6.14 were not achieved in these experiments. A number of months separated the initial mixture of the sensitized resist and the ITFI exposures shown below. It is possible that either the base resist, the pyrene sensitizer, or the mixture of the two has a limited shelf life. The base resist, which is itself an experimental product, was ~6

‡‡‡ Refer back to Section 6.4.1. Sources of contamination in the laboratory include ammonia, HMDS, an adhesion promoter used with i-line resists, and NMP, a common organic solvent. Both the NMP and the ammonia are generally used at elevated temperatures.

months old when these experiments were conducted. Additionally, the resist was thinned from the original formulation to achieve the optimal thickness of 103nm, which may have altered the concentrations of PAG and pyrene in the spun film.

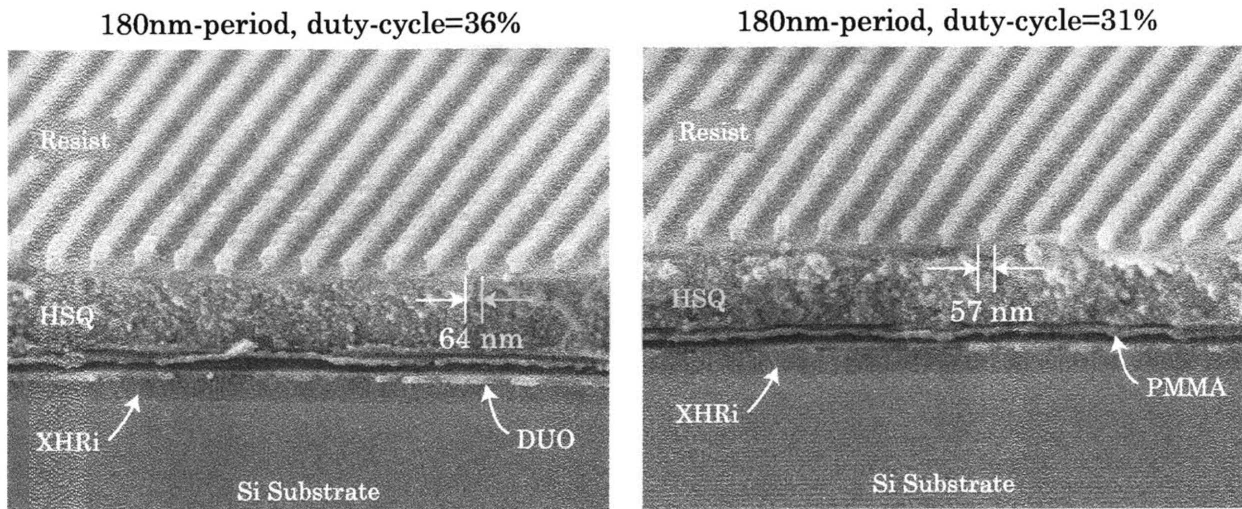


Figure 6.18 ITFI thin-film stack with 180 nm period gratings. The HSQ gapping layer, PMMA, and ARC layers are all clearly visible. The optimum duty-cycle of 31% is shown on the right, and a slightly detuned value of 36% shown on the left. Line roughness is caused by resist contamination.

The 180nm-period parent-gratings are flood exposed with an ArF laser at 193nm. Reflections from a Brewster window are used to select only the TE-polarization from the unpolarized laser. The beam is scanned during the exposure to enhance uniformity and increase the exposed area. After the 193nm exposure, parent-gratings are removed by a second development in an aqueous alkaline solution, and the HSQ/ oxide layers are stripped in 100:1 H_20:HF solution. This step is easily timed by noticing that oxide is hydrophilic, while PMMA is generally hydrophobic. As a result, when the oxide interlayer is fully stripped, the substrate can be removed dry from the HF solution. The daughter-gratings of the 36% duty-cycle parent-grating, exposed and developed in PMMA, are shown in Figure 6.19§§§.

§§§ PMMA development procedure. 60 second immersion in 3:2 isopropyl alchohol (IPA): methyl-iso(butyl) ketone (MIBK). Rinse with IPA and blow dry with nitrogen.

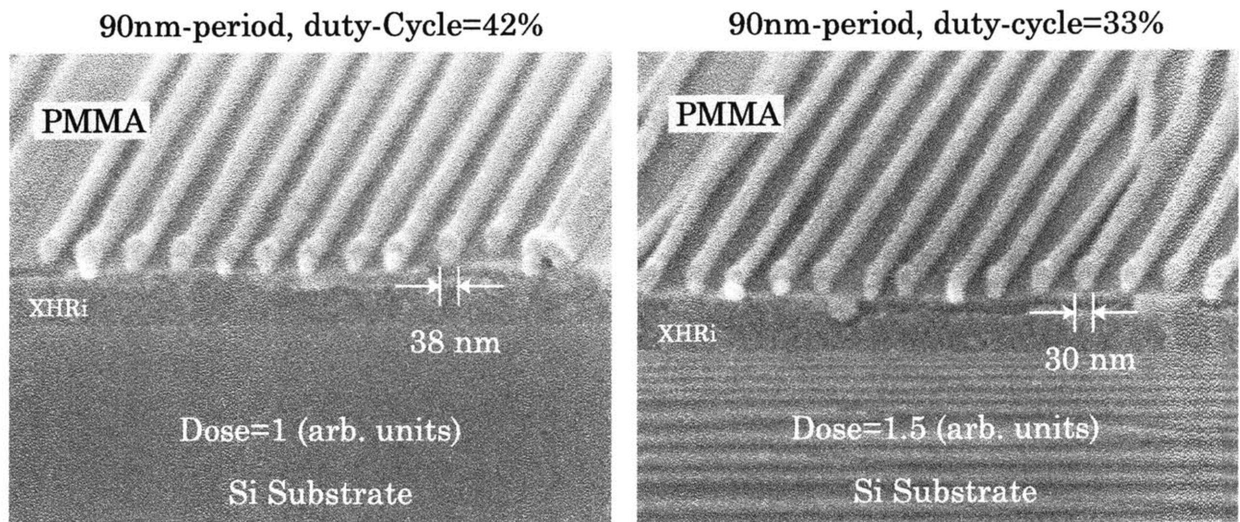


Figure 6.19 90nm-period daughter-gratings from 36% duty-cycle parent-gratings for two different doses. The lower dose (left) produces 38 nm lines with fewer adhesion defects than the higher dose shown on the right. Line-shifting and adhesion loss occur due to surface tension during development and drying.

Making an exception for the adhesion defects, these results are surprisingly good. The slow change in linewidth with increasing dose indicates indicates high-contrast in the exposure image. Especially important is the smoothness of the lines when compared to the roughness of the parent-gratings. It is likely that this is due to a combination of two effects. First, the roughness of the parent-gratings is significantly smaller than the effective wavelength, and may not significantly affect the diffraction pattern outside the range of evanescent orders. Second, the image formed at any location in the PMMA is a sum of diffracted fields from many lines in the parent-grating, thus random defects in the teeth of the parent-grating will be averaged out in the daughter-grating. Adhesion of the PMMA lines seems to be more problematic at the higher dose. One possibility is that there is a reaction between the MIBK:IPA developer and the underlying DUO193. This can potentially be resolved with a reduced development time, or a thinner PMMA layer to reduce the effect of capillary forces during development.

Shown in Figure 6.20, are daughter gratings from the 31% duty-cycle parent, exposed at slightly lower doses. The increased linewidth of the reduced-dose exposures improves the line adhesion as expected. The lowest exposure dose, with almost an even 1:1 line:space ratio, showed only minor mechanical deformation. Despite the fact that the strength of the modulation envelope at the fundamental frequency is expected to be quite high in this case (see Figure 6.12) we see little evidence of it in these exposures.

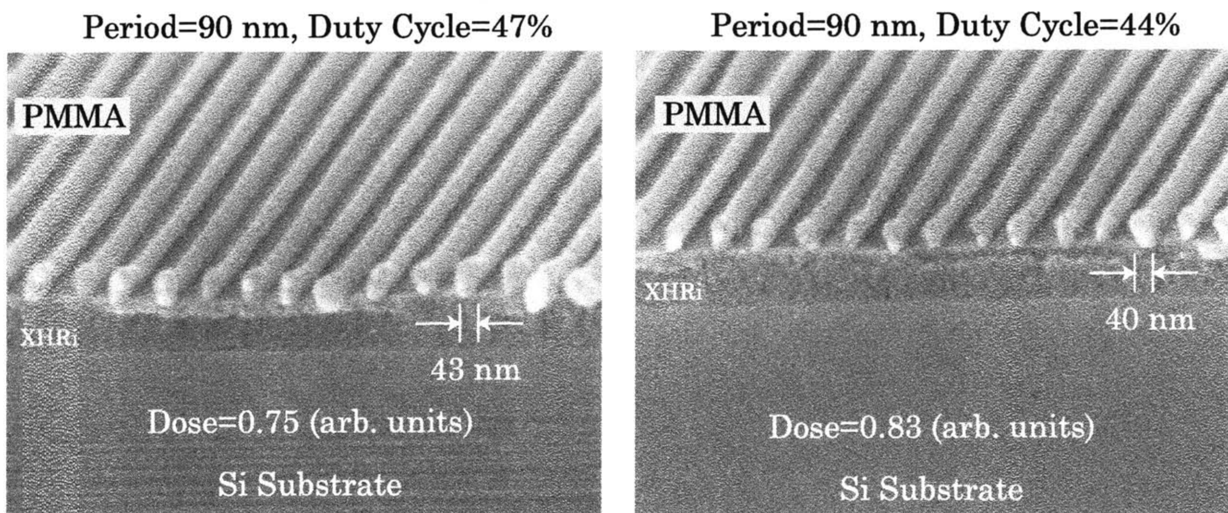


Figure 6.20 90 nm period daughter-gratings from 31% D.C parent-gratings for two different doses. In both cases, line adhesion is improved, although deformation of the lines continues to illustrate the effect of surface tension.

Duty-cycles of the parent-gratings used for these exposures were very close to the optimum values. Zero-order diffraction efficiency for the two examples above is close to 2%, compared to the optimal value of 1.3%. It is expected that the increased zero-order contribution from a parent-grating with a duty-cycle significantly off optimum would show more structure at the periodicity of the parent-grating. Shown in Figure 6.21 is a parent-grating with much larger duty-cycle, 45%, which produces a daughter-grating with a noticeably different structure. The effect of the increased modulation envelope is to shift the line centers closer to the low-intensity

region. This manifests as the bi-periodic structure shown below, which trend toward single lines at the periodicity of the parent-grating in the situation of very high zero-order contribution.

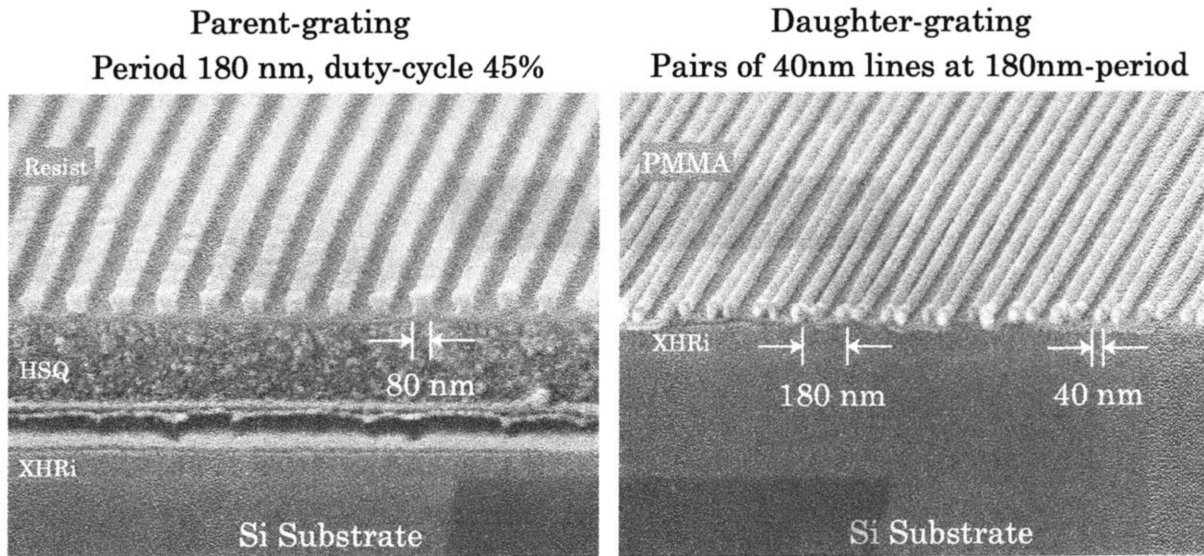


Figure 6.21 Parent and daughter-gratings in the case of increased zero-order efficiency. The duty-cycle of the parent-grating is significantly higher than optimal, yielding ~4% zero-order efficiency. Line-pairs in the daughter-grating are centered at a 180 nm period, and have pulled together in many spots due to surface tension.

6.5.3 280nm interlayer thickness

Numerical simulations using both the FDTD method and coupled-wave theory are in excellent agreement, and predict that the optimal plane for exposure of the daughter-grating would be at 310nm below the parent. As before, the interlayer thickness is chosen to place the optimal exposure-plane close to the middle of the PMMA layer. Based on the success of the exposures shown in Figure 6.20, in a plane predicted to be close to the peak of the modulation envelope, we should expect excecllent results when the proper distance is used. The thin film-stack for this configuration is shown below is Figure 6.22. Based on the reflectivity calculations shown in Figure 6.16, the thickness of the XHRi (ARC) was changed

along with the HSQ thickness to minimize back-reflections during exposure of the parent-grating. Otherwise, the design of the film-stack is unchanged.

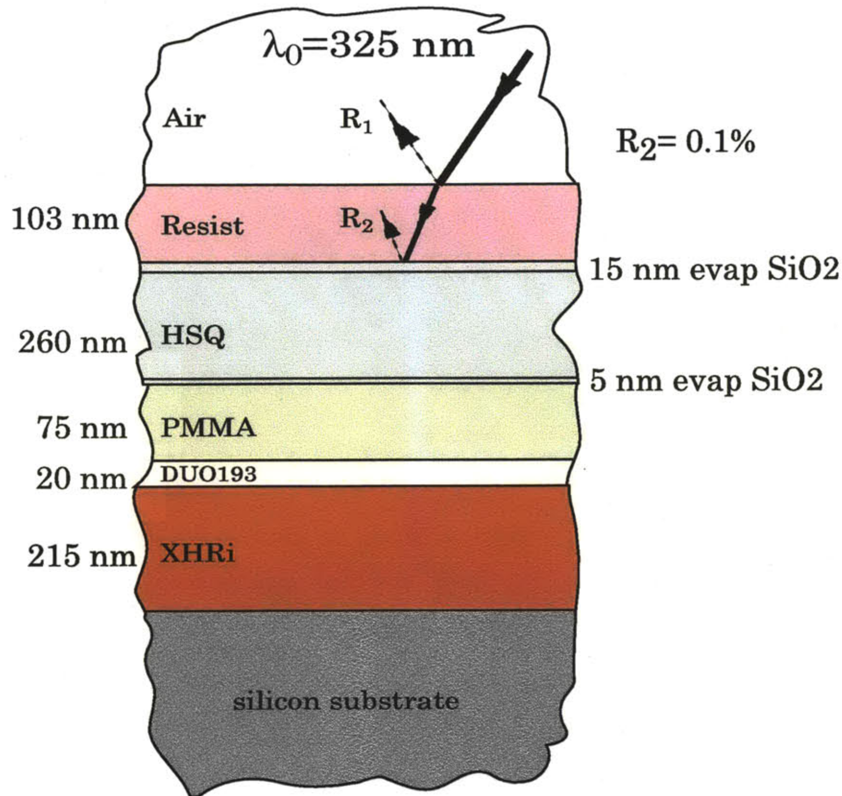
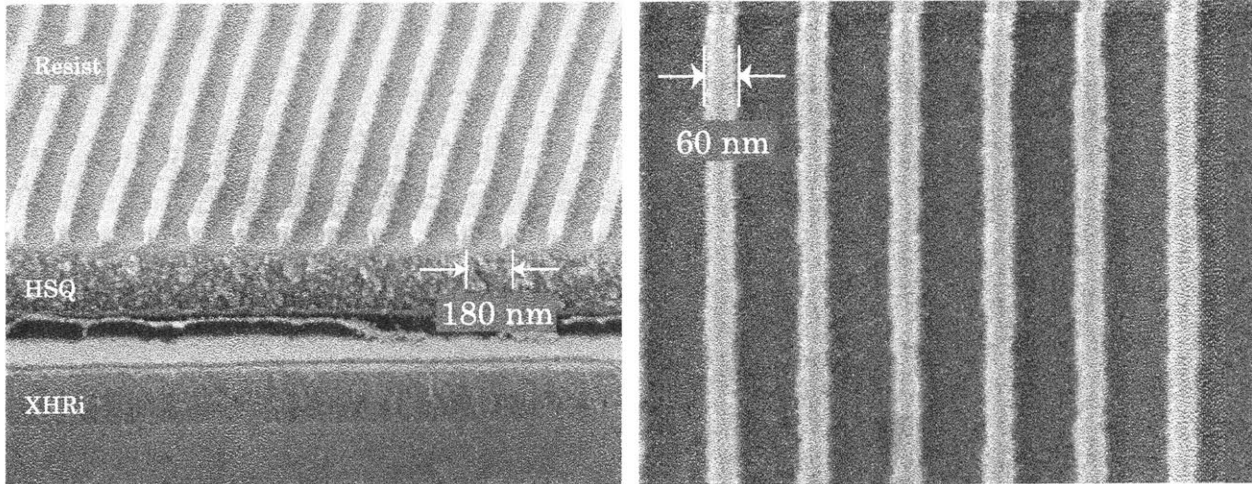


Figure 6.22 ITFI film-stack designed to place the optimal "true" sub-image plane for exposure of the daughter-grating (z_t=310 nm) near the center of the PMMA layer. In this case it will be slightly below center. Plots in Figure 6.15 and Figure 6.16 show the reflectivity calculations for this stack.

The procedure for exposure and development of the daughter-gratings was identical to the previous experiments. Although only a few micrographs are shown here, a wide variety of parent-gratings and exposure conditions were tested using this interlayer thickness. Typical results for this stack are shown in Figure 6.23.



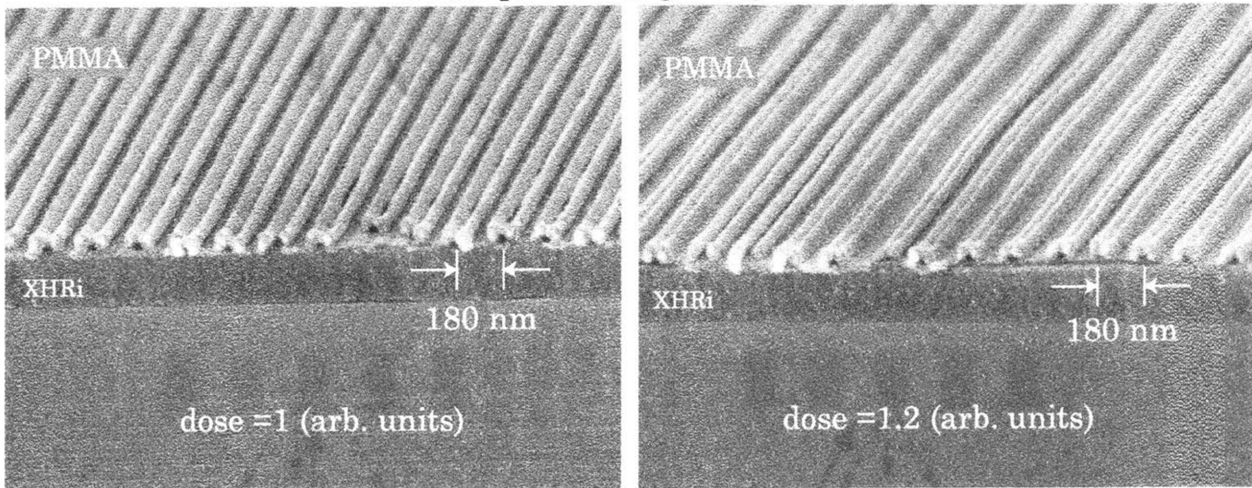


Figure 6.23 Representative examples of parent and daughter-gratings for a gap layer of 310 nm. Daughter-gratings are shown for two dose values. The low dose (left) shows full clearing of the PMMA only at 180 nm period. The higher dose (right) is fully cleared between all lines but the close proximity lines pull together.

It is immediately apparent that the daughter-gratings are similar to the high order case shown in Figure 6.22 than they are to the good results seen in Figure 6.20. Pairs of lines spaced at the period of the parent grating are indicative of an exposure plane with a strong modulation envelope. This is the opposite of what was expected, in that the exposure-plane of the PMMA was shown to be near a *minimum* of the modulation envelope for this interlayer thickness. In situations

such as this, where experimental results differ markedly from the theoretical predictions, the theory merits re-evaluation. In the next section, we investigate why such a disconnect between prediction and experiment has occurred. It turns out that the answer is rather subtle - and may actually be the answer to a problem which was introduced, but tactfully avoided, in Section 6.3.

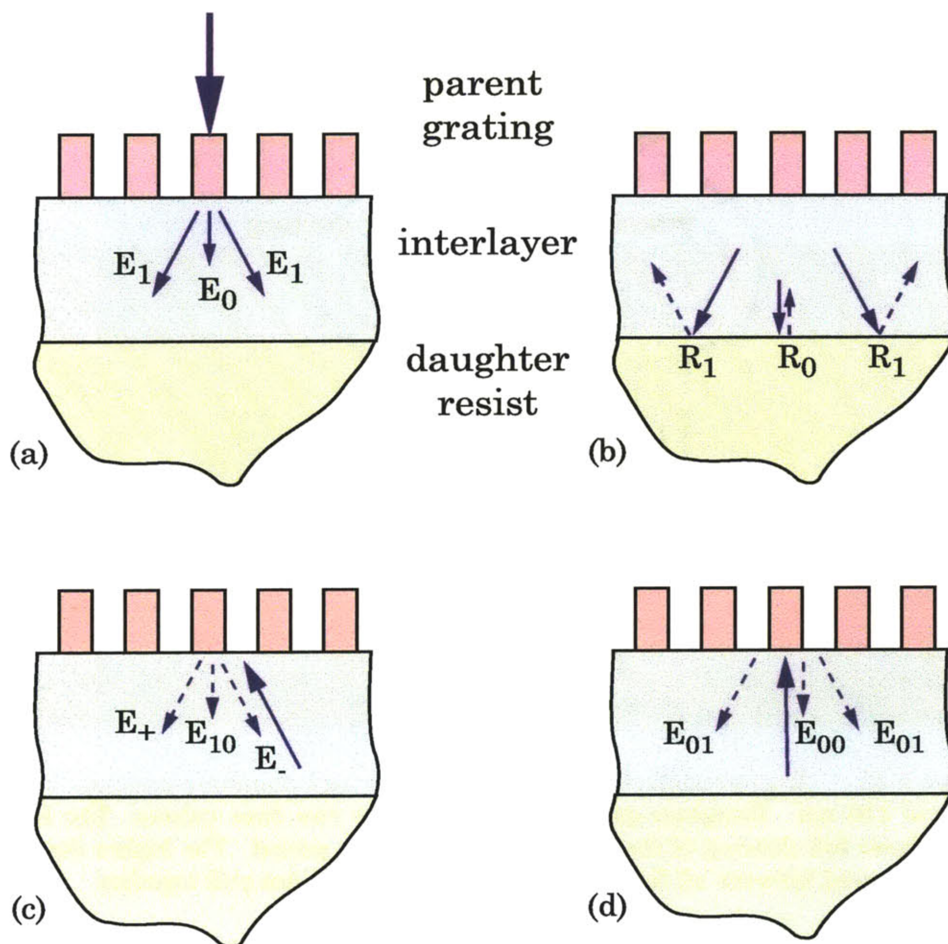


Figure 6.24 The interlayer material forma a resonant cavity through which the diffracted orders can coupled light into one another. (a) Incident light is diffracted into only the first orders. (b) First-order beams are reflected at the lower interface of the interlayer. (c) The first-order beam now incident on the parent-grating from the other direction, produces three more diffracted orders, including the zero-order which was not initially present. (d) The zero-order beam goes through the same process, and itself is re-diffracted back into all three orders.

6.6 Future developments

6.6.1 Resonant Diffraction

The reason that the exposures performed in the optimal imaging plane predicted by the numerical simulations performed poorly is not that the simulations were incorrect, but that they were the wrong simulations. The coupled-wave calculations used to optimize the diffraction-efficiency only consider the parent-grating itself, placed between two semi-infinite spaces, while the reality is more complicated.

Rather than a half-space, the interlayer is more aptly described as a resonant cavity. The parent-grating forms one reflector, and the stack of dielectric materials up through the PMMA forms the other. Light initially diffracted into one of the three orders (-1,0,+1) gets coupled into each of the other orders during subsequent passes through the cavity. For example, consider an ideal parent-grating that diffracts into only the first-order, as shown in Figure 6.24(a). After reflection from the surface of the PMMA, the orders are now incident on the parent-grating off-axis from *inside* the interlayer material. When they re-diffract off the parent from this direction, each of the initial first order beams produces diffracted components in all three orders (Figure 6.24(c)). Thus, although there was initially no zero-order component, one is produced on secondary passes through the cavity. Therefore, optimization of diffraction efficiencies without considering resonant effects will not accurately predict the behavior of the system.

This type of resonant diffraction is an interesting field, in which there is only a small amount of published work. A few papers present the diffractive element inside a high-Q Fabry-Perot cavity [28,29]. The ITFI is slightly different in that it uses the diffractive element as one end of the resonator and is very lossy. A structure of similar geometry has been studied recently for use as a spectral filter of reflected light [30,31,32]. Further investigation into the resonance properties of the ITFI will likely be of interest well outside the immediate lithographic community.

6.6.2 A diffractive resonator model

Although the resonance effect has been presented as detrimental to the diffraction efficiency of an ideal parent-grating, a more optimistic mind will view it as an opportunity to minimize the zero-order from a non-ideal parent-grating. This requires a counter-intuitive leap to realize that the "optimal" parent-grating may not actually minimize the zero-order by itself. Instead, the "first-pass" zero-order is cancelled by the "second-pass" diffraction of first-order beams into the zero-order. That is, the thickness of the interlayer is chosen such that the light re-diffracted into the zero-order is out-of phase-with initial zero-order component.

Determining this condition is not as simple as setting the interlayer thickness to $1/4\lambda_{eff}$. Phase-shifts will be induced in the diffracted orders, as well as on reflection from the interlayer/PMMA interface. As discussed in Section 6.3, numerical calculations for the diffraction efficiency are required because periodicity of the parent-grating is less than the wavelength. An excellent future project for someone continuing this work would be to develop a simulation tool which would optimize the both the resonator thickness and the grating geometry to absolutely minimize the transmitted zero-order.

Assuming that all relevant diffraction efficiencies are known, a matrix equation for the resonator can be developed. Note that the diffraction efficiencies and reflection coefficients will have both magnitude and phase components. Figure 6.24 identifies all of the individual diffraction efficiencies relevant to this discussion. Assume normal incidence, such that the efficiency of (+1) and (-1) orders will be equal. We can construct a 3-dimensional vector quantity that represents the complex electric-field amplitudes for the three forward-propagating diffracted orders inside the resonator. The initial values will be equal to the diffraction efficiencies of the parent-grating in the absence of the resonant cavity.

$$\overline{E}_{init} = \left[E_{-1}, E_0, E_{+1}\right] \qquad (6.13)$$

The orders will traverse the cavity acquiring a phase delay, Φ_1 for the first-orders, and Φ_0 for the zero-order, proportional to the thickness of the cavity d. The first-order diffraction angle is θ_1.

$$\Phi_1 = e^{i2\pi\left(d/\lambda_{eff}\cos(\theta_1)\right)}$$
$$\Phi_0 = e^{i2\pi\left(d/\lambda_{eff}\right)} \quad (6.14)$$

The orders will then reflect off of the lower interface of the interlayer with reflection coefficients R_1 and R_0, and travel back across the cavity acquiring a second phase delay. The trip across the resonator and back can be described by the matrix $\overline{\overline{P}}$.

$$\overline{\overline{P}} = \begin{bmatrix} R_1\Phi_1^2 & 0 & 0 \\ 0 & R_0\Phi_0^2 & 0 \\ 0 & 0 & R_1\Phi_1^2 \end{bmatrix} \quad (6.15)$$

Each of the orders, upon arrival back at the parent-grating, will itself be back-diffracted into all three orders,. For example, the (+1)-order incident on the parent-grating will reflect back into the (-1)-direction with efficiency E_{+1-1}, couple into the zero-order (E_{+10}), and continue along the same direction (E_{+1+1}). This is represented by the matrix $\overline{\overline{D}}$, which can be simplified using the symmetry of the (+1) and (-1) orders.

$$\overline{\overline{D}} = \begin{bmatrix} E_{+1+1} & E_{+10} & E_{+1-1} \\ E_{0-1} & E_{00} & E_{0-1} \\ E_{-1+1} & E_{-10} & E_{-1-1} \end{bmatrix} = \begin{bmatrix} E_{+} & E_{10} & E_{-} \\ E_{01} & E_{00} & E_{01} \\ E_{-} & E_{10} & E_{+} \end{bmatrix} \quad (6.16)$$

The fields after one pass through the resonator can be described as

$$\overline{E} = \overline{\overline{D}}\,\overline{\overline{P}}\,\overline{E}_{init} = \overline{\overline{M}}\,\overline{E}_{init}, \quad (6.17)$$

where $\overline{\overline{M}} = \overline{\overline{D}}\,\overline{\overline{P}}$. After subsequent passes through the resonator, the fields will continue to undergo the same operation $\overline{\overline{M}}$. $\overline{\overline{I}}$ is the identity matrix.

$$\overline{E} = \left[\overline{\overline{I}} + \overline{\overline{M}} + \left(\overline{\overline{M}}\right)^2 + \left(\overline{\overline{M}}\right)^3 \ldots\right]\overline{E}_{init} \quad (6.18)$$

Similar to a standard Fabry-Perot resonator, the geometric series can be written in closed form, as shown below.

$$\overline{E} = \left[\overline{\overline{I}} - \overline{\overline{M}}\right]^{inverse} \quad (6.19)$$

Mathematical restrictions on $\overline{\overline{M}}$ will always be met for real-life diffraction efficiencies and reflection coefficients, which must be less than unity.

This matrix formalism can be used to solve for the actual diffraction efficiencies of a given system, accounting for the resonance effects. The efficiencies as a function of the interlayer thickness can be solved, allowing the optimal exposure plane to be determined. The drawback to this, however, its that the diffraction efficiencies and the reflection coefficients must already be known; no indication is given as to whether a different parent-grating will perform better.

It is likely that the optimal interlayer thickness will be determined by the phase of the diffraction efficiencies and reflection coefficients, rather than their magnitudes. If we assume an interlayer thickness such that the optimal phase relationship is found, Equation 6.19 can be recast in terms of only the magnitudes of the diffraction efficiencies and reflection coefficients.

From a design point of view, the magnitude of the reflections at the lower surface would be the least flexible parameter, determined by the film thicknesses of the anti-reflection stack and the PMMA imaging layer. For the given reflection coefficients, one could solve for the optimal diffraction efficiency magnitudes. Using these optimum values, the physical parameters of the parent-grating could be determined, such as its tooth-height and duty-cycle. Once the parameters of the parent-grating are fixed, the optimal thickness of the interlayer can be determined using the complex diffraction efficiencies and reflection coefficients.

6.6.3 High-Index tooth and trench materials

The ITFI will, in general, be a low-Q cavity. That is, only a few passes of the light will occur before it leaks out of the resonator. Given this, it is unlikely that resonance effects can be exploited to significantly reduce a large initial zero-order component. Rather, the magnitude of the residual zero-order which cannot be suppressed, even with an optimized grating, can be driven towards zero through the use of resonance effects. Recall from Section 6.3.3 that the zero-order diffraction

efficiency could be as high a ~50% for parent-gratings with periodicity significantly shorter than the wavelength. If the ITFI is to fully exploit the high-index propagation medium, the zero-order from such parent-gratings must be made significantly lower.

The reason that sub-wavelength gratings exhibit such low first-order diffraction efficiency is that the space between the grating teeth was filled with air. Without the benefit of a high-index medium to reduce the effective wavelength, propagation in these regions is forbidden. A simple solution to this problem is to create a parent-grating where the tooth and trench are both high-index materials. This could be achieved with little additional effort by spin coating another material over the top of a resist-grating made using IL. For example water-soluble polymers such as polyvinyl alcohol (PVA, $n=1.70$) or polyacrylic acid (PAA, $n=1.69$) could be spun over the patterned resist, before exposure of the daughter-grating. Both of these materials are transparent at 193nm and have a refractive-index higher than that of an oxide interlayer ($n=1.534$), The same aqueous alkaline developer that currently strips the parent-grating would also remove a water-soluble overcoat.

6.6.4 Imprinted parent-gratings

The ITFI, implemented for 193nm exposures as described so far, was the most direct way to demonstrate a proof-of-concept for spatial-frequency doubling using solid-immersion. Readily available techniques and materials (with the exception of the pyrene-sensitized chemically-amplified resist) were used to minimize process development.

It would be advantageous, however, to reduce the number of process steps and improve process latitude by minimizing the number of necessary layers. By changing the method of fabrication for the parent-grating, a process which much more closely approximates the ideal schematic in Figure 6.2 can be developed. Imprint lithography has been shown to be capable of replicating nanometer-scale

features from master template with excellent fidelity in a wide variety of materials [33,34,35].

Fabrication of ITFI parent-gratings using imprint lithography could allow for significant process simplifications. For example, the thin film stack must currently be designed to minimize back-reflections during the exposure of both the parent-grating and the daughter-grating. As the two exposures occur at different wavelengths, in different parts of the stack, and at different angles, this optimization is not trivial. By imprinting the parent-grating, only reflections during the exposure of the daughter-grating need be considered. Even this can potentially be eliminated if the PMMA layer for imaging the daughter-grating becomes thin enough that an anti-reflection coating is unnecessary. The use of ultrathin resists without an ARC will be discussed further in Chapter 7.

Currently, the interlayer material is actually a sandwich of evaporated oxide, spun HSQ, and evaporated oxide. Recall that the top evaporated oxide is necessary to prevent the developer of the parent-grating from attacking the HSQ. An imprinted parent-grating, with no developer, would not require these oxide buffer layers. Further, the parent-grating could be stamped directly into the interlayer material. Although imprint lithography in HSQ has been demonstrated [36], it would be beneficial to eliminate the HF dip currently required to strip the HSQ and oxide. Imprinting the parent-grating directly into an interlayer material which is water soluble is an especially attractive possibility.

Perhaps the most difficult challenge so far has been the development and use of a photoresist for the parent-grating which has both the necessary resolution and high transparency to 193nm light. Looking towards ITFI exposures at 157nm, this problem will only get worse. By imprinting the parent-grating, the resist problem can be solved. Material choices for the parent-grating are widened significantly, in that the material needs only to be transparent, but does not require the additional chemical functionality of an advanced photoresist. Further, diffraction efficiencies are likely to be closer to the ideal calculated values, assuming that a good imprint

mold can be fabricated. Critical features such as the tooth height and width of the parent-grating will be highly repeatable. It is clear from the parent-gratings shown in Figure 6.18 andFigure 6.23 that resist contamination currently prevents the teeth of the parent-grating from being ideally square, or even of uniform height and width.

Combining all of the simplifications afforded by an imprinted parent-grating, the ITFI thin-film stack could be reduced from its current 8 or 9 layers to a much more tractable 3 or 4. The ITFI stack using an imprinted parent-grating consists of an ARC and resist for the exposure of the daughter-grating, and an interlayer into which the parent-grating is stamped. Even the inclusion of a high-index overcoat to improve diffraction efficiency would only bring the total to 4 layers. Implementing this scheme, however, will require a significant investment in equipment and time as there are currently no imprint lithography tools, or a knowledge-base of imprint techniques, here at MIT.

6.6.5 157nm and 248nm ITFI exposures.

Lithography at 157nm, which seemed to be a strong possibility even just a few years ago [37], has recently fallen out of favor. Unlike most other lithography schemes, however, the ITFI requires essentially no optics external to the susbtrate, and therefore can operate without difficulty at λ_0=157nm. Most of the materials which were used in the ITFI demonstration at 193nm could also be used at 157nm. The spin-on-glass interlayer material HSQ is highly transparent and has an index of $n=1.705$, yielding an effective wavelength of λ_{eff}=92nm. Using the ITFI at 157nm, the effective wavelength is short enough to frequency-double $P_1=100nm$ parent-gratings made using the AIL. Although its absorption is significantly higher than at 193nm, PMMA films up to 70nm thick should be transparent enough for the exposure of good resist profiles [38], although at ~20nm linewidths, it is likely that resist thicknesses will need to be less than ~40nm to prevent mechanical collapse.

In some situations, it may also make sense to use a longer wavelengthfor ITFI exposures. For example, standard i-line photoresists based on novolac, which are extremely absorbing at λ_0=193nm, are actually almost ideal for use at λ_0=248nm. The refractive index is close to 2, while the absorption is still quite low. Thus, standard, i-line resists, which have none of the environmental problems of chemically-amplified resists, could be used to make high-diffraction-efficiency parent-gratings for 248nm ITFI exposures. The proof-of-concept 90nm-period gratings, exposed using 193nm, would actually have been easier to implement and a more impressive demonstration of the solid-immersion effect if they were exposed with λ_0=248nm. The HSQ interlayer, with a refractive index at 248nm of $n=1.467$, would allow for daughter-gratings down to P_2=~$85nm$ to be exposed. For certain applications, where feature-size is not the most important quality, 248nm ITFI may be an attractive option. For example, the frequency-doubling of any parent-grating made using the Lloyd's-mirror is possible at this wavelength. Also, for applications such as the frequency-doubling of gratings produces using scanning-beam interference lithography [1], where the parent-gratings are of 400nm period, it is unnecessary to use the shortest possible wavelength.

6.7 Conclusion

A new concept for lithography, the integrated thin-film interferometer, has been proposed. Taking advantage of diffraction inside a high-index, solid-immersion medium, spatial-periods below the diffraction limit in air can be successfully imaged. As a proof-of-concept, 90nm-period gratings were fabricated with this technique using 193nm illumination. The diffraction efficiency of sub-wavelength structures was investigated to facilitate the development of ITFI processes for 50nm period gratings. The ITFI as a diffractive resonator was introduced as a way to understand unexpected optical behavior, although exploiting resonance properties has the potential to further optimize diffraction efficiency

beyond what is possible by considering only the geometry of the parent grating. Through additional efforts developing a complete theoretical understanding of the system, as well as materials and process development, the ITFI provides a viable and attractive route to achieving dense ~20nm features using optical lithography.

6.8 References

[1] C.G. Chen, P.T. Konkola, R.K. Heilmann, C. Joo, M.L. Schattenburg, "Nanometer-accurate grating fabrication with scanning-beam interference lithography", *Proc. SPIE* **4936** p.126 (2003)

[2] T.A. Savas, "Achromatic Interferometric Lithogrpahy", P.h. D. Thesis, Massachusetts Institute of Technology (2003)

[3] D. Pommet, M.G. Moharam, E.B. Grann, "Limits of scalar diffraction theory for diffractive phase elements". *J. Opt. Sci. Am. A* **11**(6) p.1827 (1994)

[4] D. Gremaux, N.C. Gallagher. "Limits of scalar diffraction theory for conducting gratings", *Appl. Opt.* **32**(11) p.1948 (1993)

[5] K. Knop, "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves". *J. Opt Sci. Amer.* **68** p.1206 (1978)

[6] F.G. Kaspar, "Diffraction by thick, periodically stratified gratings with complex dielectric constants". *J. Opt Sci. Amer.* **63** p.37 (1973)

[7] R.S. Chu, J.A. Kong, "Modal theory of spatially periodic media". *IEEE Trans. Microwave Theory Tech.* **23** p.123 (1977)

[8] E.H. Anderson, "Fabrication and Electromagnetic Applications of Periodic Nanostructures." Ph.D. Thesis, MIT, (1988)

[9] K.S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media". *IEEE Trans. Ant. Prop.* **14** p.302 (1994)

[10] R. Guerrieri, K.H. Tadros, J. Gamelin, A.R. Neureuther, "Massively parallel algorithms for scattering in optical lithography". *IEEE Trans. On CAD* **10**(9) (1991)

[11] A. Wong, "Rigorous three-dimensional time-domain electromagnetic simulations" Ph. D Thesis, University of California, Berkeley (1994)

[12] H.F. Talbot, "Facts Relating to Optical Science IV". *Philos. Mag.* **9** 401 (1836)

[13] J.T. Winthrop, C.R. Worthington, "Theory of Fresnel images I. Plane periodic objects in monochromatic light" *J. Opt. Sci. Am.* **55**(4), p.373 (1965)

[14] W.D. Montgomery, "Self-imaging objects of infinite aperture". *J. Opt. Sci. Am.* **57**(6) p.772 (1967)

[15] J. Goodman, *Introduction to Fourier Optics.* McGraw-Hill Co., New York (1996)

[16] S. Teng. L. Liu, J. Zu, L. Luan, D. Liu, "Uniform theory of the Talbot effect with partially coherent illumination" *J. Opt. Sci. Am. A* **20**(9) p.1747 (2003)

[17] V. Arrizon, J. Ojeda-Castañeda, "Irradiaance at Fresnel planes of a phase grating". *J. Opt. Sci. Am. A* **9**(10) p.1801 (1992)

[18] P.M. Mejia, R. Martinez-Herrero, "Diffraction by one-dimensional Ronchi grids: on the validity of the Talbot effect". *J. Opt. Sci. Am. A* **8**(2) p.266 (1991)

[19] R.R. Kunz, R.D. Allen, W.D. Hinsberg, G.M. Wallraff, "Acid-catalyzed single-layer resists for ArF lithography". *Proc. SPIE **1925*** p.167 (19xx)

[20] R.D. Allen, G.M. Wallraff, D.C. Hofer, R.R. Kunz, "Photoresists for 193-nm lithography", *IBM J. Res. Dev.* **41**(1) p.95 (1997)

[21] S.P. Pappas, "Photogeneration of Acid: Part 6 - A Review of Basic Principles for Resist Imaging Applications" *J. Imaging Tech.* **11(**4) p.146 (1985)

[22] S.P. Pappas, J.H. Jilek, L.R. Gatechair, "Photogeneration of Cationic Initiators by Donor-Acceptor Interactions - 7. Ag(I), Cu(II) and Onium Salts as Acceptors" *J. Imaging Sci.* **30**(5) p.218 (1986)

[23] H. Ito, C.G. Willson, J. Frechet, *United States Patent #4,491,628* (1985)

[24] R.D. Allen, G.M. Wallraff, W.D. Hinsberg, L.L. Simpson, "High performance acrylic polymers for chemically amplified photoresist applications" *J. Vac. Sci. Technol. B* **9**(6) p.3357 (1991)

[25] I.B. Berlman, *Handbook of the Fluorescence Spectra of Aromatic Molecules*, Academic Press (1971)

[26] S.A. MacDonald, et al.. "Airborne Chemical contamination of a Chemically Amplified Resist". *Proc. SPIE* **1466** p.2 (1991)

[27] W.D. Hinsberg et al. "Quantitaion of Airborne Chemical Contamination of Chemically Amplified Resists Using Radiochemical Analysis". *Proc. SPIE* **1672** p.24 (1992)

[28] L. Escoubas, F. Flory, F. Lemarchand, E. Drourard, L. Roux, S. Tisserand, G. Albrand. "Fabry-Perot multilayers for enhancing the diffraction efficiency of ion-implanted gratings". *Appl. Opt.* **40**(10) p.1587 (2001)

[29] A.Sinha, G.Barbastathis, "Resonant Holography". *Opt Lett.* **27**(6) p.385 (2002)

[30] D. Rosenblatt, A. Sharon, A. Friesem, "Resonant grating waveguide structures". *IEEE J. Quantum Electronics* **33**(11) p.2038 (1997)

[31] D.K. Jacob, S.C. Dunn, M.G. Moharam, "Design Considerations for narrow-band dielectric resonant grating reflection filters of finite length". *J. Opt. Sci. Am. A* **17**(7) p.1241 (2000)

[32] D.K. Jacob, S.C. Dunn, M.G. Moharam, "Normally incident resonant grating reflection filters for efficient narrow-band spectral filtering of finite beams". *J. Opt. Sci. Am. A* **17**(7) p.1241 (2000)

[33] S.Y. Chou, P. Krauss, P.J. Renstorm, "Nanoimprint Lithography". *J. Vac. Sci. Technol. B* **14** p.4129 (1996)

[34] S.Y. Chou, P.R. Krauss, W. Zhang, L. Guo, L. Zhuang, "Sub-10nm imprint lithography and applications" *J. Vac. Sci. Technol. B* **15**(6) p.2897 (1997)

[35] M. Colburn *et al.*, "Step and Flash Imprint Lithography: A New Approach to High-Resolution Patterning". *Proc. SPIE* **3676** p 379 (1999)

[36] S. Matsui *et al.*, "Room-temperature nanoimprint and nanotransfer printing using hydrogen silsesquioxane". *J. Vac. Sci. Technol. B* **21**(2) p.688 (2003)

[37] T.M. Bloomstein, M.W. Horn, M. Rothschild, R.R. Kunz, S.T. Palmacci, R.B. Goodman, "". *J. Vac. Sci. Technol. B* **15** p.2112 (1997)

[38] R.R. Kunz, T.M. Bloomstein, D.E. Hardy, R.B. Goodman, D.K. Downs, J.E. Curtin, "Outlook for 157nm resist design". *J. Vac. Sci. Technol. B* **17**(6) p.3267 (1999)

Chapter 7
Applications

7.1 Resist processes for interference lithography

7.1.1 Higher-order periodicity

It was stated in Chapter 1 that patterns produced by interference lithography are necessarily periodic. Attempts to hybridize the resolution capabilities of interference lithography with the patterning of arbitrary geometries have met with limited success [1]. The gratings produced by the two-beam interferometers described in Chapters 3-6, with only one-dimensional periodicity, are the simplest periodic structure that one can create. Through the use of three or more interfering beams, periodicity in all three spatial dimensions can be created. Multi-beam interferometers similar in design to the Mach-Zehnder and AIL interferometers can be designed to produce structures with multi-dimensional periodicity [2,3,4,5]. However, for more than two beams, the complexity of the interferometer increases dramatically such that the benefit is often outweighed by the complexity. In most

cases, sacrifices must be made, either in exposure area, image contrast, or image quality in order to make such a system functionally robust. For example, the depth-of focus for a two-beam interferometer is essentially infinite; the same grating pattern is produced everywhere that the beams overlap. As few as three beams, however, will create periodicity not just within the desired exposure plane, but also through the depth of the field. The recorded pattern becomes dependent on the vertical position of the substrate, and the depth-of-focus can be small as 10's of nanometers.

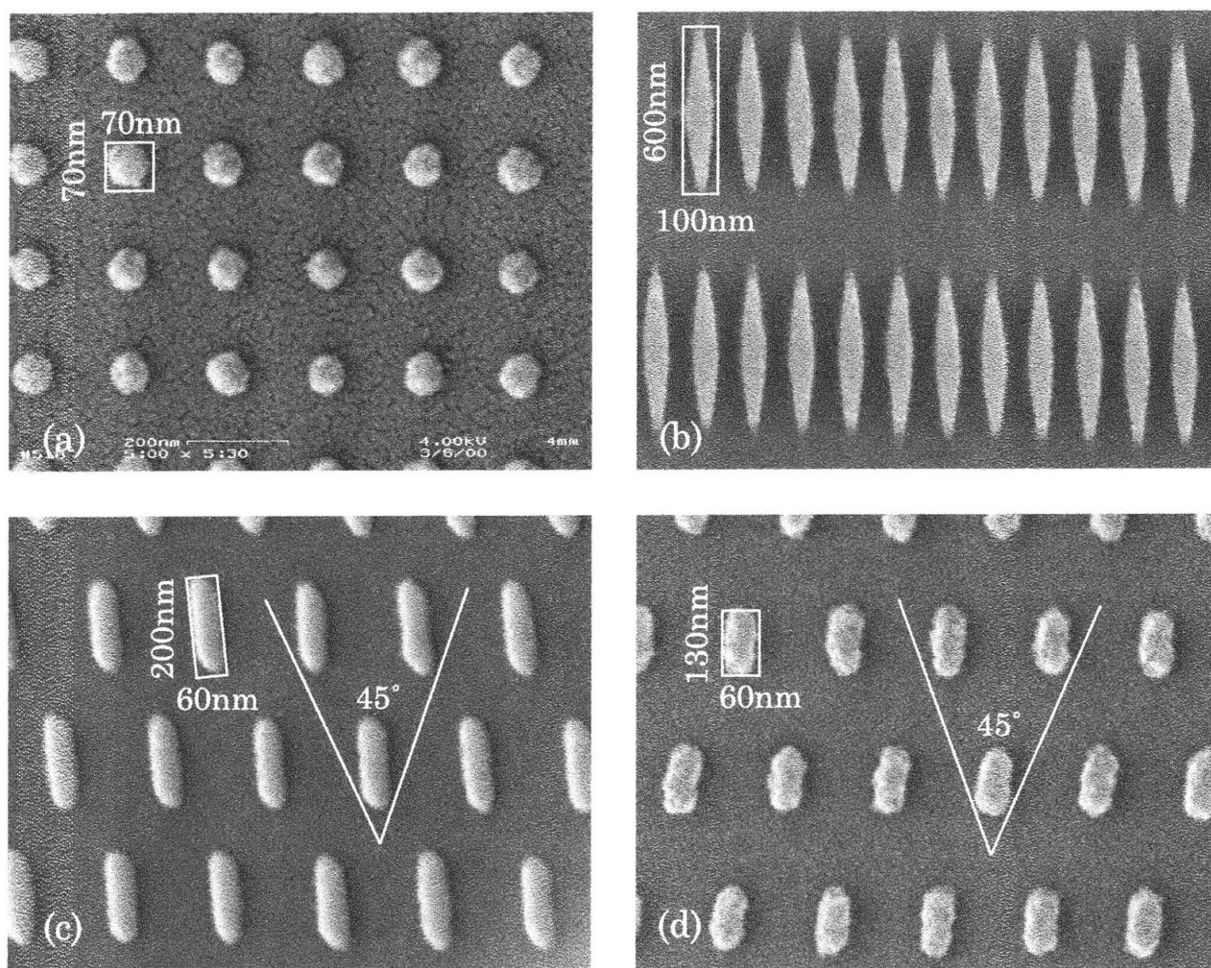


Figure 7.1 (a) Square lattice created through two exposures rotated at 90° to one another. (b) A rectangular lattice created by different periodicities in the two directions. (c) Rotation of 45° between the two exposures creates highly elongated structures. (d) Modulation of the relative dose for exposures at 45° changes the aspect ratio of the structure.

An alternate method is to build 2-dimensional periodic structures as a superposition of shifted and rotated gratings. For example, an array of points on a square lattice can be created using two exposures rotated at 90° to one another, as seen in Figure 7.1(a). This technique works well for two exposures, but becomes rapidly less desirable as the number of exposures increases. One reason is that the image is created by a summation of intensities rather than a summation of fields, as would be the case in multi-beam interference. Image contrast becomes increasingly poor over a larger number of incoherent exposures, whereas control of the polarization and phase can create a similar pattern with much higher image contrast using multiple beams. Additionally, spatial-phase registration is necessary in a series of three or more exposures. For these reasons, applications described here are limited to either a single grating exposure or two exposures rotated azimuthally to form an array.

The patterned magnetic media described in Section 7.2 are one particular case where a variety of techniques have been employed to create arrays of different cell width and spacing. Three basic options are available when using the two-exposure method. First, a square lattice can be created with different periodicity in the two directions. Figure 7.1(b) shows a rectangular array with a periodicity of 1000nm in one direction and 200nm in the other. Second, the angle of rotation between the two exposures can be varied anywhere between 0 and 90°. While an angle of 90° produces the array of round posts seen in Figure 7.1(a), exposures at a smaller angle will produce elongated structures, as seen in Figure 7.1(c). Finally, one can expose a lattice with equal periodicity in both directions, but adjust the relative dose of the two exposures. Comparing Figure 7.1(c) and (d), the rotation of the second exposure is 45° in each case but the relative dose is different, leading to the features in (c) being much longer. Of course, these methods are not mutually exclusive, one can easily imagine using a combination of all three.

7.1.2 Anti-Reflection coatings

Interference lithography, as has been emphasized repeatedly throughout this work, is the interference of two beams to form a standing-wave which is recorded in photoresist. In practice, the situation is more accurately described as the creation of *two* standing waves from the interference of *four* beams. Reflections of the primary beams at the lower interface of the photoresist film will lead to the formation of a standing wave perpendicular to the plane of the film. Thus, the intensity profile in the photoresist will vary through the thickness of the film as well as across the surface of the wafer. The periodicity of the vertical standing wave can be derived similarly to that of the lateral standing wave as

$$P_{vertical} = \frac{\lambda_0}{2n \cdot \cos(\theta)}, \tag{7.1}$$

where λ_0 is the wavelength of the light, n is the refractive index of the photoresist, and θ is the incident angle of the beams. The effect of this unwanted standing wave is to "scallop" the sidewalls of the photoresist features, seen in Figure 7.2, potentially leading to the loss of linewidth control and poor pattern transfer. In more pronounced cases, the narrow waist can cut entirely through the resist line. Minimal reflectivity at the resist/substrate interface R_2 is ideal in all cases, although the effects will be less significant at larger feature sizes.

One approach commonly used in the semiconductor industry quantifies the severity of the standing wave in terms of a "swing ratio" SR derived from a simplified consideration of the resist as a resonant cavity [6].

$$SR = 4\sqrt{R_1 R_2}\, e^{-\alpha t}, \tag{7.2}$$

where R_1 is the reflectivity at the top resist interface, R_2 is the reflectivity at the lower resist interface, α is the resist absorbance, and t is the film thickness. This type of characterization is useful in that it forces one to consider the top-surface reflection, as well as the resist thickness and absorption as important parameters. However, it is misleading in that a null of the swing-ratio does not guarantee a

minimal standing wave. Consider the top-surface reflection $R_1=0$ and $R_2=1$; $SR=0$ but the contrast of the vertical standing wave would actually be maximized. Therefore, it is usually more effective to minimize R_2 directly.

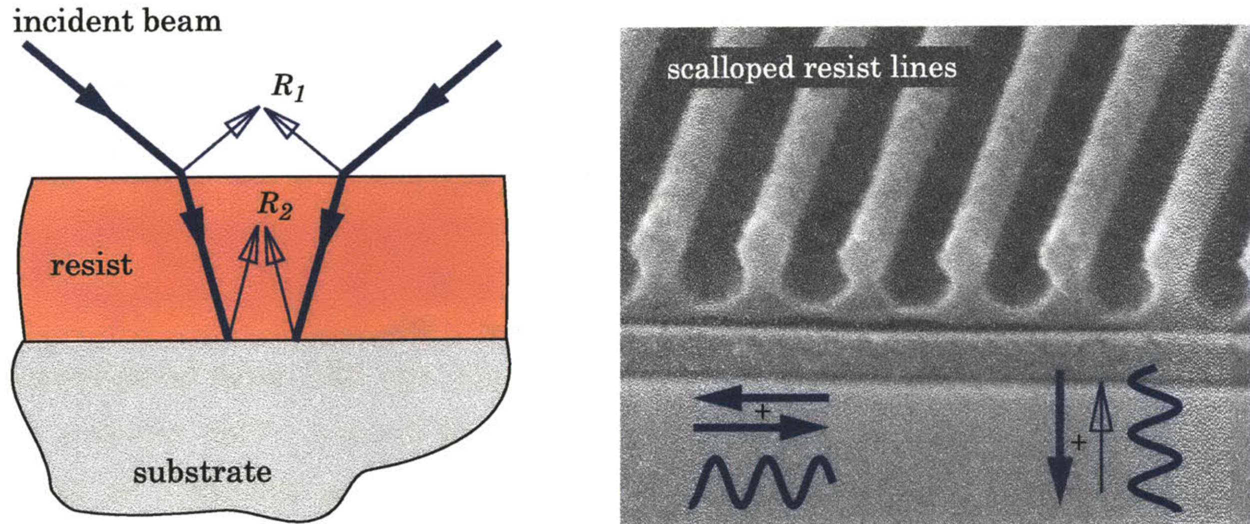


Figure 7.2 Light reflected from the substrate interferes with the incident light to create an unwanted vertical standing wave which degrades resist profiles. The micrograph shows a 200nm-period grating exposed in photoresist over an improperly optimized anti-reflection layer.

The problem of the vertical standing-wave has long been recognized in optical lithography. At large feature sizes (>>1μm), one could afford to disregard its effects without significantly impacting process latitude. However, for sub-micron feature sizes, the vertical standing-wave is a much more serious problem. The original solution relied on a post-exposure bake (PEB) to smooth out the standing wave via diffusion of photoacids within the resist. One can consider the ratio of the periodicity of the vertical standing wave to that of the desired grating period as a rough guideline for the applicability of this approach. If $P_{vertical}$ is a significant fraction of the of the linewidth, diffusion during the PEB will also noticeably blur the desired resist pattern.

The use of anti-reflection coatings (ARC) between the resist and substrate was proposed in the early 1980's, although it was not until the 1990's that this

became widely accepted [7,8]. ARC's function through a combination of thin-film-interference and absorption, requiring optimization of both the film thickness and complex refractive index ($N=n-ik$) [9,10]. In practice, the optical properties of a commercial ARC layer are determined by the manufacturer such that reflectivity is minimized by adjusting only the film thickness. Solving for the reflectivity at the lower interface of the photoresist requires treating the substrate, device layers, and ARC film together as a thin-film stack - essentially a dielectric mirror. A computerized algorithm for this optimization was developed and is described elsewhere [10].

As the reflectivity of the stack can show a strong angular dependence as well as sensitivity to the optical properties of each layer, the ARC must be optimized individually for different grating periodicities and devices. This requires an accurate knowledge of both the optical properties and film thickness for every component layer. While the optical properties of some materials can be found in published references [11,12], others must be measured directly by spectroscopic elliposmetry [13]. A tabulation of the optical constants for most materials relevant to interference lithography is given in Appendix A.

7.1.3 The tri-layer resist process

It has been found that a thin interlayer of an inorganic material, generally SiO_2, between the resist and ARC improves linewidth control and sidewall profiles when transferring the photoresist pattern into the underlying films [14]. These three layers are collectively known as a tri-layer resist stack, and form the basis of a general resist process for interference lithography. Specific parameters such as the type and thickness of ARC, and the choice between positive and negative photoresist are determined by the nature of the device being fabricated. Positive resists are generally used for low-duty-cycle lines or arrays of posts, while negative photoresist is used for high-duty-cycle lines or arrays of holes [15]. Examples are shown in Figure 7.3.

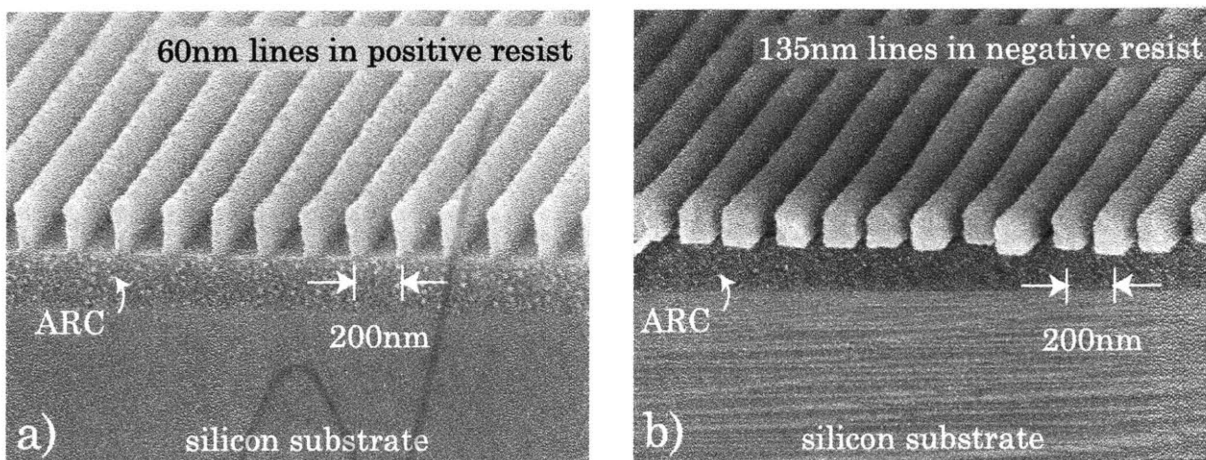


Figure 7.3 200nm-period gratings. a) low duty-cycle lines in positive resist. b) high duty-cycle lines in negative resist.

For exposures at λ=325 nm using the Lloyd's mirror or Mach-Zender interferometers, commercial i-line (365 nm) photoresists based on the novolac-diazonapthoquinone (DNQ) chemistry are used [16]. For AIL and ITFI exposures at λ=193 nm, polymethyl-methacrylate (PMMA) is used****. Commercial photoresists are available for lithography at this wavelength, but they rely on a chemically-amplified chemistry [17,18,19] which is extremely sensitive to environmental contamination and process conditions [20,21]. Although commercial resists exhibit higher photospeed and etch-resistance, these products cannot match the high resolution and low line-edge-roughness of PMMA [22].

**** A commercial, chemically-amplified photoresist designed for λ=193 nm was modified for use in the ITFI process, discussed in Chapter 6, although the imaging exposure occurs in PMMA.

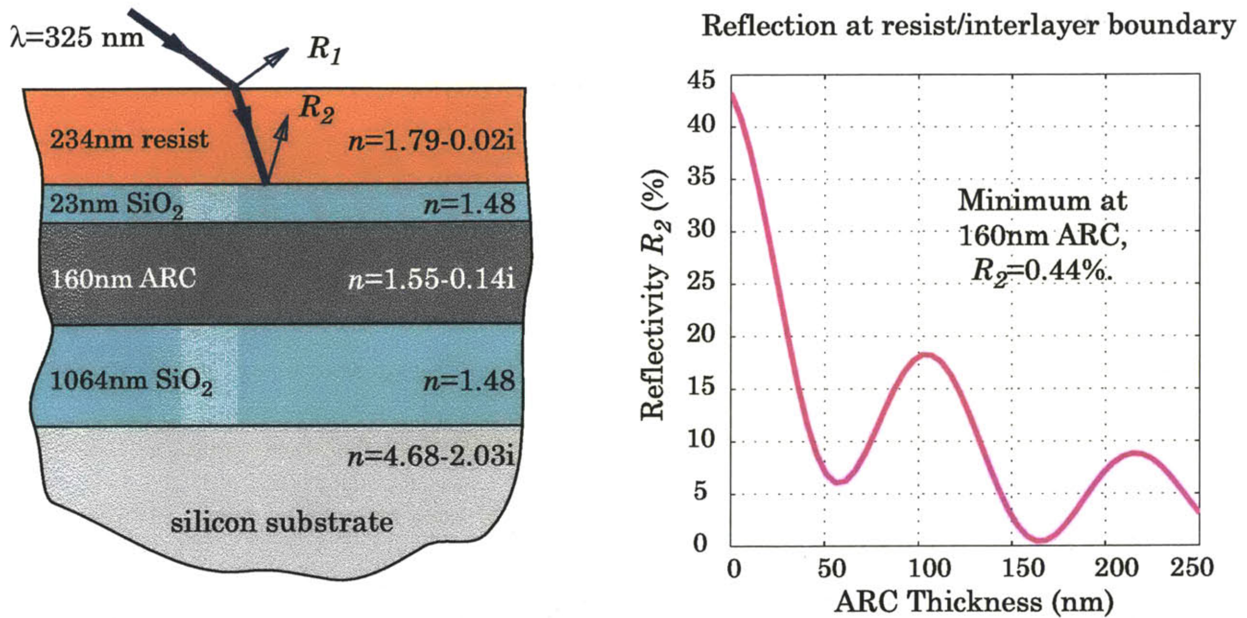


Figure 7.4 Left: schematic of the thin-film stack used for fabricating the distributed-feedback structures described in Section 7.4. Complex refractive index values are shown for each layer. Right: reflectivity at the resist/interlayer boundary (R_2) plotted as a function of the ARC thickness.

This tri-level resist process forms the basis for most of the processes used in the applications described later in this chapter. As an example, the process steps for fabricating the distributed-feedback (DFB) gratings described in Section 7.4 are detailed here. A silicon substrate thermally oxidized to form a 1064nm-thick layer of SiO_2 was prepared with an anti-reflection coating, a thin interlayer of SiO_2, and a positive photoresist ††††, as shown in Figure 7.4. The ARC thickness of 160 nm was chosen to minimize reflectivity at the resist/interlayer boundary, and the Lloyds-mirror interferometer was used to expose lines on a 350nm pitch (Figure 7.5a). This pattern is transferred to the silica interlayer using reactive-ion etching (RIE) [23]. High etch-selectivity between the photoresist and the silica obtained using CHF_3 as an etch gas for this step. Next, the pattern was etched by (RIE) through the ARC layer using an oxygen plasma. The ARC and resist, having a

†††† BARLi I-line ARC from AZ Electronic Materials. Interlayer deposited by e-beam evaporation. PFI-88 A3 positive photoresist from Sumitomo (Sumika) Chemical Inc..

fundamentally similar polymeric structure, etch at approximately the same rate in oxygen plasma, and the photoresist pattern is largely destroyed during this step. The role of the silica interlayer now becomes clear, acting as an etch mask for the ARC as the resist is consumed (Figure 7.5b). Only small triangular ridges of photoresist remain, but the interlayer is clearly intact. The ARC pattern is then used as a mask for etching into the thick oxide layer on the substrate, again using a CHF_3 plasma as the etchant, shown in Figure 7.5c. Finally, the remaining ARC is removed using the RCA1 cleaning process [24]. The final device, a 350nm-period grating etched 50 deep in SiO_2 is shown in Figure 7.5d.

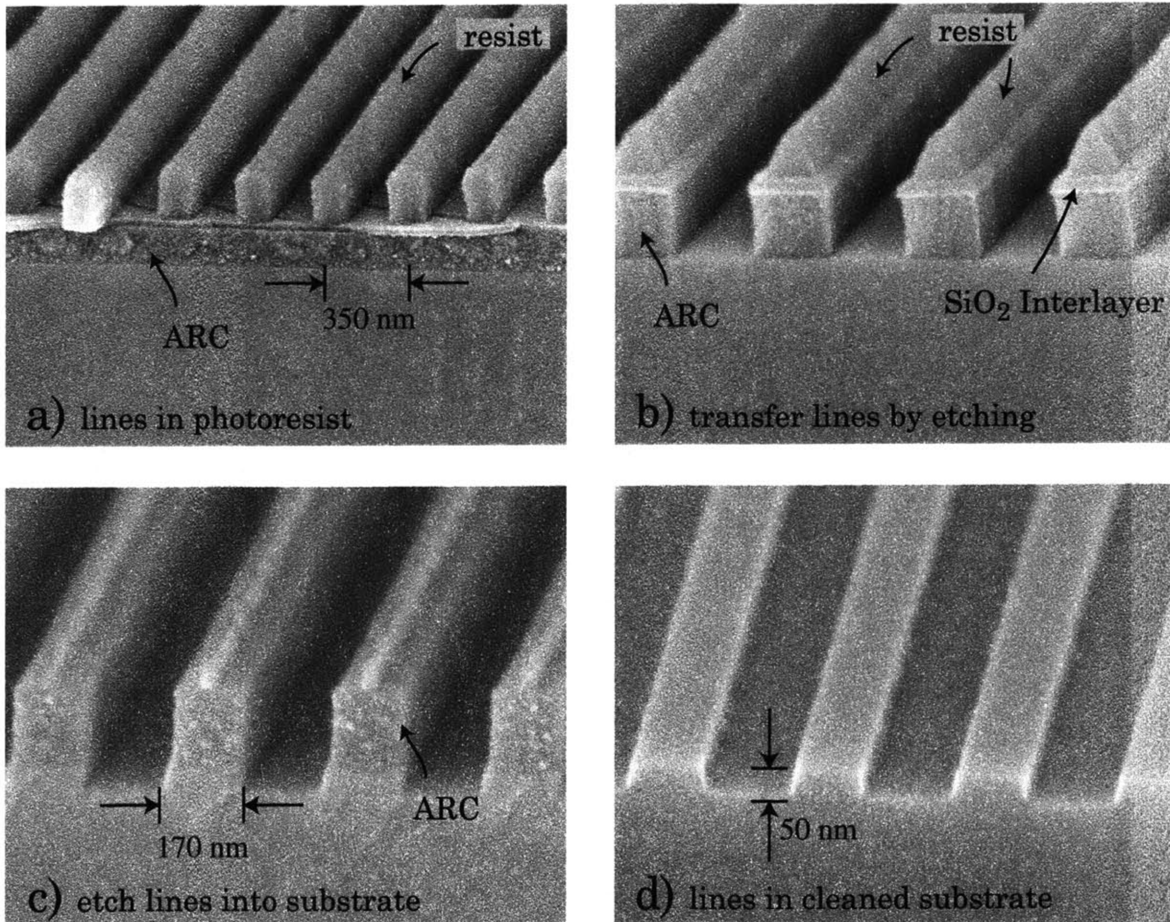


Figure 7.5 A series of micrographs depicting the fabrication of a DFB grating device in SiO_2. a) Photoresist lines over a 23nm silica interlayer over a 160nm anti-reflection coating on a SiO_2 substrate. b) After etching through the interlayer and ARC, the photoresist lines are severely degraded. C) Pattern is etched into the substrate using the ARC as a mask. Final device etched 50nm into the SiO_2 substrate.

Although illustrated only for a single device, the process remains largely unchanged at various feature sizes on various substrates. Details such as film thicknesses and etch chemistries may change, but the basic steps will stay the same. The patterned magnetic-media devices described in Section 7.2 include an additional ion-beam etching step described in detail in Chapter 8.

7.1.4 *Ultrathin Resist*

For extremely fine lines, it becomes necessary to use a very thin photoresist layer to prevent mechanical collapse of resist features caused by surface tension during development [25,26,27]. For example, it is predicted that the critical aspect-ratio for collapse (CARC) will be between 1 and 2 for dense sub-100 nm lines in certain resists [28], where aspect-ratio is defined as the ratio of the line height to its width. Examples of the mechanical failure of high-aspect-ratio resist structures are shown in Figure 7.6, where the resist lines can either (a) physically break, or (b) de-adhere from the substrate. Note that the aspect-ratio (~3) is higher than would be observed for dense features at the same linewidth (60nm lines on a 120 nm pitch) because their spacing is greater.

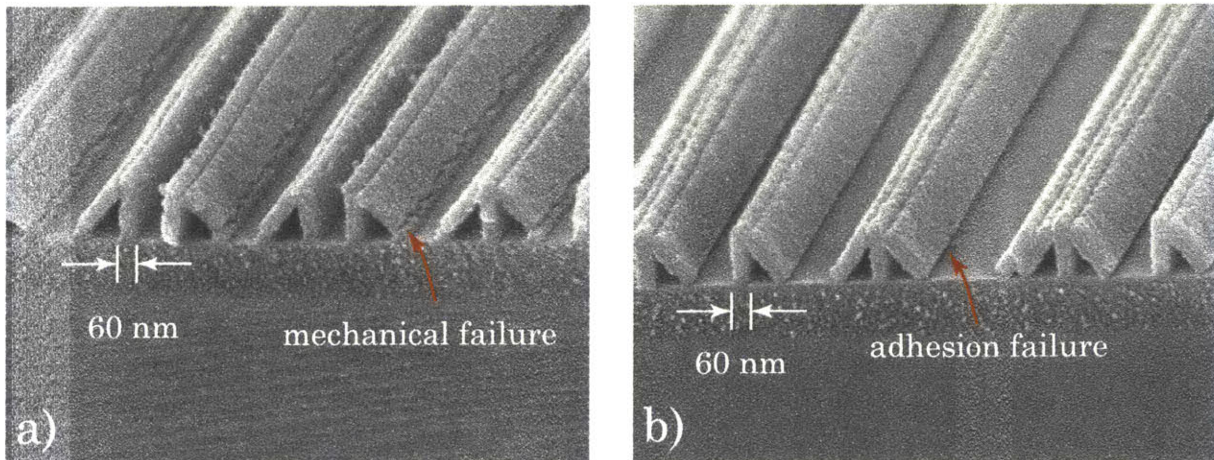


Figure 7.6 60nm lines on a 200 nm pitch depicting collapse caused by surface tension during the photoresist development. a) capillary forces are large enough to break the resist structures. b) reduced adhesion of resist on SiO_2 interlayer causes the entire line to de-adhere instead of breaking.

It has also been shown that an extremely thin film of resist, less than 1/4 the periodicity of the vertical standing-wave (Equation 7.1), can tolerate intensity variations in the layer despite the presence of a strong standing wave, eliminating the need for an anti-reflection coating [29,30]. Thus, if a reduced resist thickness is required to prevent collapse, the use of an anti-reflection coating may be unnecessary. This can also be attractive when it is desirable to minimize the number of process steps. Lines made using IL in ~40nm thick photoresist layers without an ARC are shown in Figure 7.7.

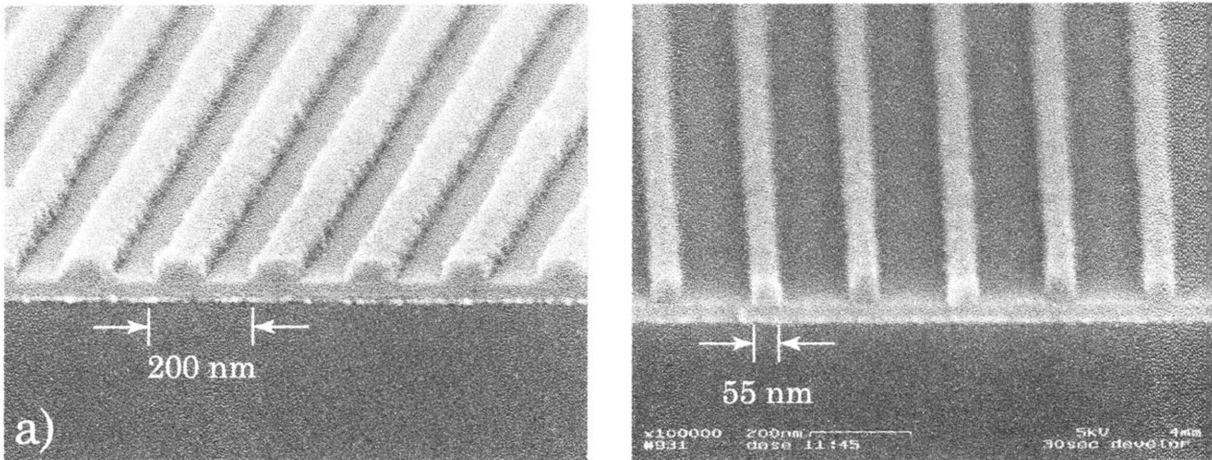


Figure 7.7 40nm-thick photoresist lines with different duty-cycle exposed over highly-reflecting substrate with no anti-reflection coating. The substrate is a silicon wafer coated with 15 nm of tungsten and 30nm of SiO_2.

7.1.5 Sensitometry

It was mentioned in Chapter 1 that the amplitude and periodicity are the only tunable parameters of a sinusoidal intensity pattern. In practice, the periodicity is set by the geometry of the exposure tool, and the amplitude is set by the available source power. Assuming the source power I_0 is constant, exposure time is the only parameter which can be easily varied. The incident dose *ID*, defined as the total energy of the exposure, is the product of the incident power per unit area I_0 and the exposure time *t*.

$$ID = I_0 \cdot t \tag{7.3}$$

Implicit in this relation is that resists obey the so-called Burns-Roscoe reciprocity relation: the resist responds equally to high-intensity over short times and low-intensity over long times [22]. A common reciprocity failure occurs when using high-intensity pulses, which can actually photo-ablate the resist [31,32]. This is usually observed with DUV illumination where the photon energy can be greater than many bond energies within the resist [33]. Most resist materials, however, exhibit reciprocity over a wide intensity range, and it is assumed throughout this work.

A convincing argument can be made that a lithography tool is only as good as the resist it exposes [34]. Thus, it is useful to understand the response of photoresist as a function of spatial-frequency since image properties such as contrast vary with spatial-frequency in most lithography tools. Interference lithography is an excellent tool for characterizing resist response at various feature sizes because spatial-frequency is the only variable that changes. The intensity pattern is the same high-contrast sinusoid for all periodicities.

The comparison of IL exposures at different spatial-periods requires a more rigorous definition of dose. One cannot simply assume a constant dose for exposures of different periods, even under identical exposure conditions. Consider

that the incident angle of the exposing light θ and the reflectivity of the wafer will both affect the percentage of incident power that acts to expose the resist. Figure 7.8 illustrates how the power density on the wafer decreases at oblique angles of incidence because of the increased area illuminated by the beam.

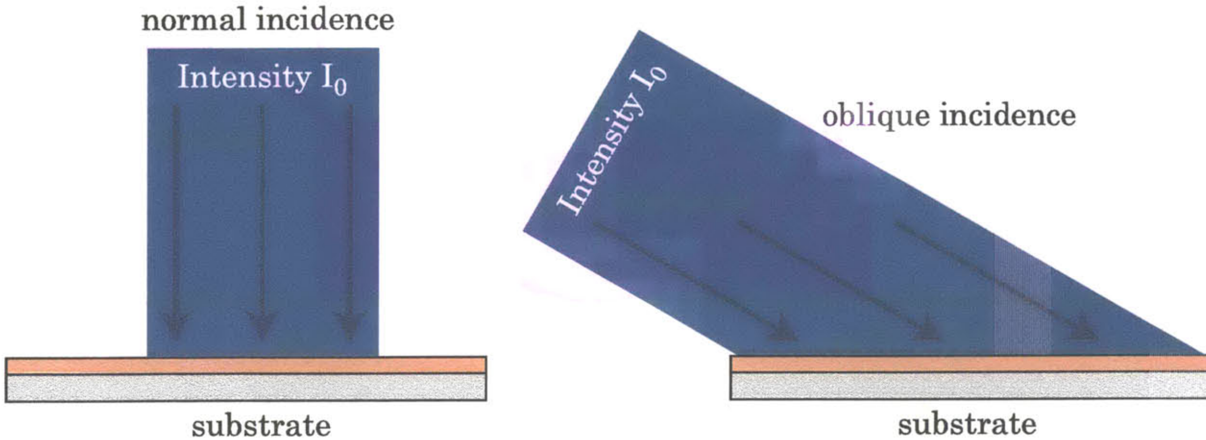


Figure 7.8 For a beam of fixed intensity, the power density on a substrate will be hightest at normal incidence. The substrate area intercepted by the beam will be larger at oblique incidence angles such that the power density decreases in proportion to the cosine of the incident angle.

Additionally, power reflected off the top surface of the photoresist (R_1) will reduce the effective intensity, while reflectivity at the lower surface (R_2) will increase the effective intensity. All of these effects are dependent on the angle of incidence, and therefore the periodicity of the exposed features. Refer back to Figure 7.4 for an illustration of R_1 and R_2.

We introduce the concept of the "equivalent-dose", *ED*, as a normalized value accounting for these effects, thereby facilitating a direct comparison of exposures at different periodicities and on different substrates.

$$ED = ID\left[(1 - R_1)(1 + R_2)\cos(\theta)\right] \qquad (7.4)$$

Reflectivity values R_1 and R_2 can be calculated for an arbitrary thin film stack using the same algorithm used to optimize the anti-reflection coating . The use of an ARC generally restricts R_2 to <1%, such that the difference between the incident- and

equivalent-dose is dominated by the top-surface reflections R_1 and the angle-of-incidence θ.

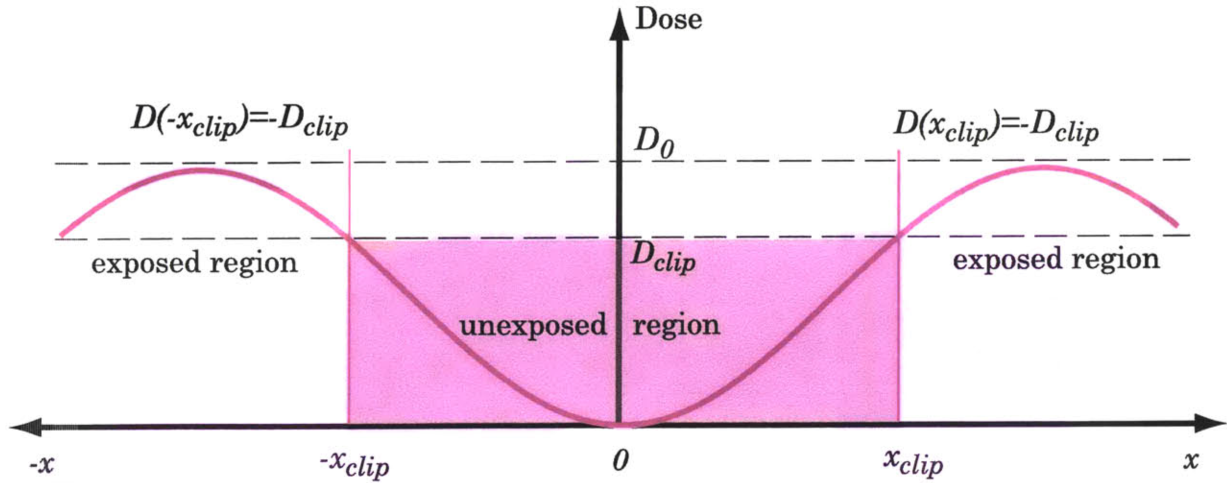


Figure 7.9 Schematic of the binary model for photoresist exposure. Regions where the dose is less than the threshold value D_{clip} remain unexposed, shown as the shaded region. Regions where the dose is above the threshold value are fully exposed.

The simplest and most common model for photoresists is the ideal binary response: above a certain threshold D_{clip} the resist is fully exposed, while below that threshold the resist is fully unexposed. This is shown schematically in Figure 7.9. Consider a sinusoidal spatial variation of the exposure dose (such as the one obtained using IL) where D_0 is the amplitude and P is the spatial-period.

$$D(x) = D_0\left[\sin\left(\frac{2\pi x}{P}\right) + 1\right] \tag{7.5}$$

Assume a positive photoresist, such that photoresist is undeveloped in regions where the dose is less than the threshold value ($D<D_{clip}$). The position x_{clip} is defined simply as the location where the intensity D is equal to the clipping level D_{clip}.

$$D(x_{clip}) = D_{clip} \tag{7.6}$$

Comparing the exposure of two sinusoidal patterns with periods P_1 and P_2, the dose at the positions where they clip must be equal, $D_1(x_1)=D_2(x_2)=D_{clip}$.

$$D_{clip} = D_1\left[\sin\left(\frac{2\pi x_1}{P_1}\right)+1\right] = D_2\left[\sin\left(\frac{2\pi x_2}{P_2}\right)+1\right] \qquad (7.7)$$

If the exposure energy is the same ($D_1=D_2$), the linewidth will be different but the *duty-cycle* of the exposed gratings must be the same.

$$\frac{x_1}{P_1} = \frac{x_2}{P_2} \qquad (7.8)$$

By using the equivalent-dose rather than the incident dose, the condition of equal energy is ensured and we can conclude that exposures made at different periodicities for the same equivalent-dose should have the same duty-cycle. Observing the duty-cycle as a function of equivalent-dose is therefore a useful way to characterize the response of photoresist to different spatial-frequencies.

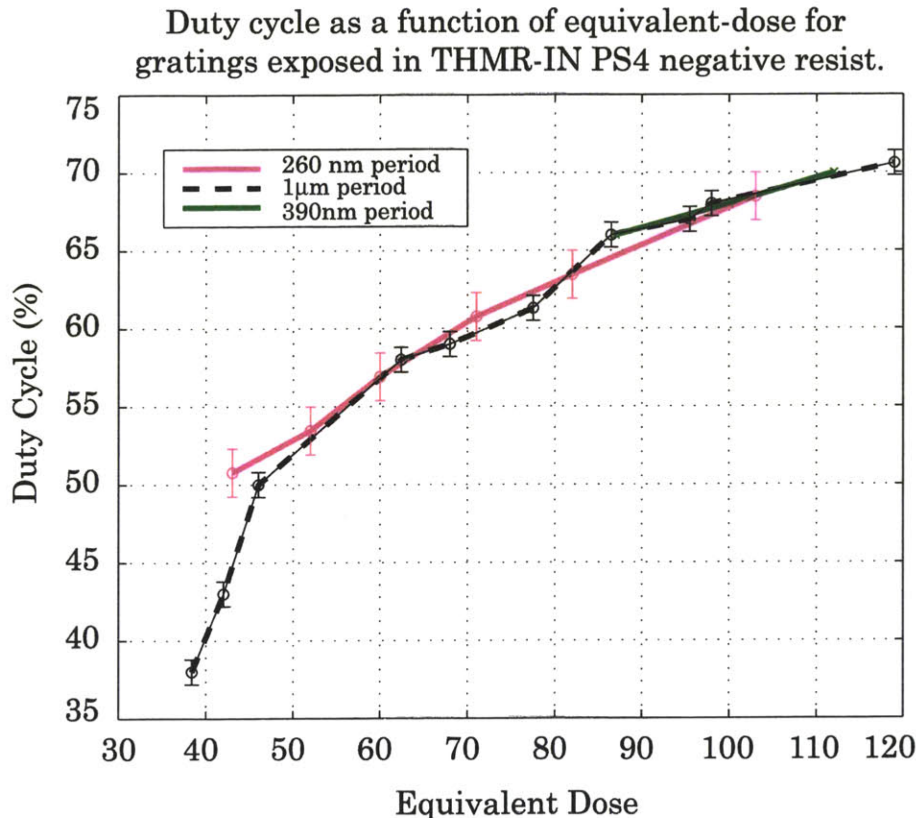


Figure 7.10 Duty-cycle as a function of equivalent-dose for 260nm and 1000nm-period gratings. Additional data points for a 390nm-period grating are also shown. Exposures were perfomed using the Lloyds mirror interferometer using THMR-iN PS4 negative photoresist. Note the invariance of the function to grating period except in the region DC<50%.

A plot of the duty-cycle as a function of equivalent-dose is shown in Figure 7.10 using data for 3 different grating periods exposed in negative resist. In the high-duty-cycle (high-dose) region, results are as expected, exhibiting an excellent agreement for all of the data. In the low-dose regime, however, there is a significant difference between the 260nm-period and the 1000nm-period gratings. The 260nm-period grating is unable resolve features with less than a 50% duty-cycle, while the 1000nm-period grating is able to resolve features down to 38% duty-cycle. That is to say, the dose which fully exposes the 1000nm-period grating at 38% duty-cycle is insufficient to fully expose the 260nm-period grating.

A similar disparity at low exposure-doses can also be seen in exposures using positive photoresist. Figure 7.11 shows gratings of three different periods. P_1=200nm , P_2=600nm, and P_3=1000nm, all exposed with the same equivalent-dose. Recall that the dependence of duty-cycle on dose for the positive-tone resist will be opposite that of the negative resist; increased dose produces narrower lines.

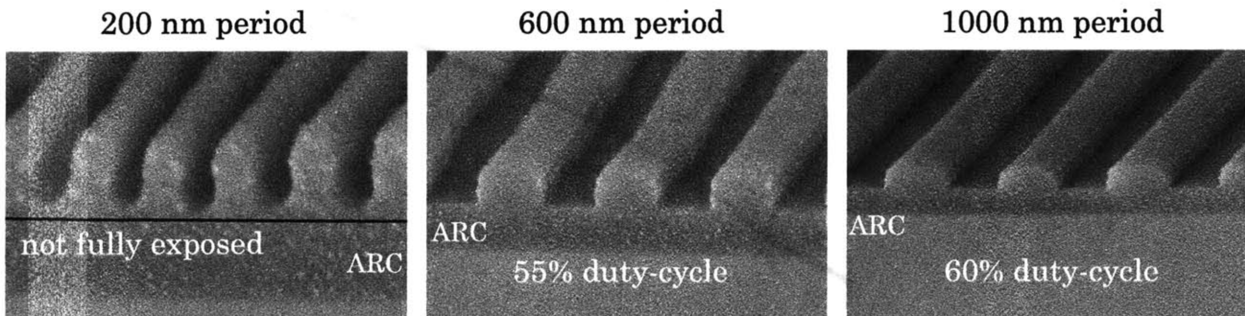


Figure 7.11 Three different periodicity gratings exposed in positive resist at the same equivalent dose. The black line across the 200nm-period grating marks the interface between the resist and ARC. Note that the duty cycle is ~50%, but the resist is not fully exposed. The larger-period samples, however, are fully exposed, with the duty-cycle increasing with the period.

It is clear from these results that the spatial-frequency of the features influences the behavior of the resist. While the dose shown is sufficient to expose the 1000nm- and 600nm-period lines, it is insufficient to fully expose the 200nm-period lines and the duty-cycle of all three gratings is different. As was observed in the negative resist, the lower spatial-frequency features require less dose to fully

expose than the high-frequency features. Additionally, the exposure latitude of both resists is reduced for high-spatial-frequency images. It is impossible for the 200nm-period gratings to achieve a duty-cycle beyond 50%, although this is clearly not the case for the 1000nm-period gratings.

The 200nm-period and 1000nm-period gratings from Figure 7.11 are shown again in Figure 7.12, this time exposed with a 14% increase in the equivalent-dose. The increased dose is now sufficient to fully expose the 200nm-period grating, which maintains a 50% duty-cycle. The higher dose also achieves a 50% duty-cycle in the 1000nm-period grating, and the predicted invariance of the dose/duty-cycle function to spatial-frequency is again observed. The pattern follows that of the negative resist (Figure 7.10), where the ideal behavior of the resist is observed only at doses above the necessary value for a 50% duty-cycle in the highest spatial-frequency image.

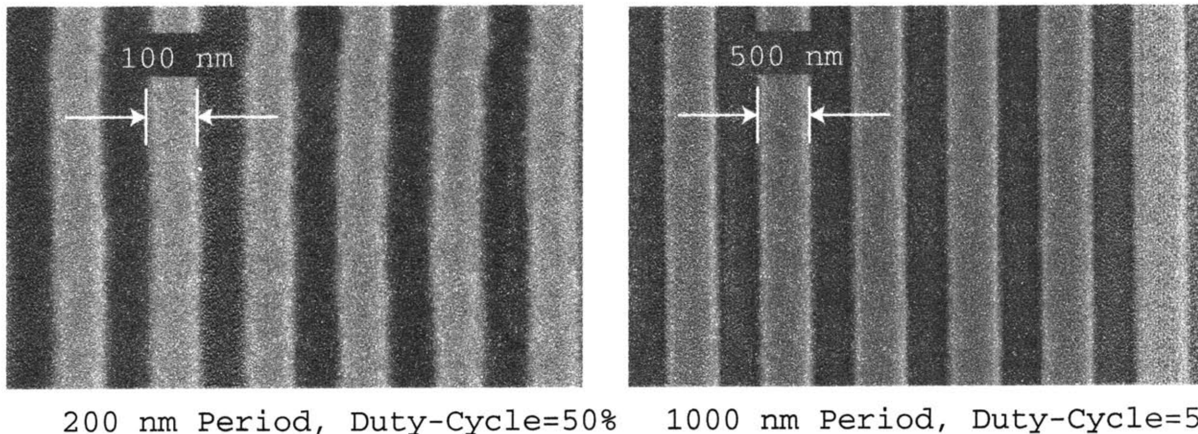


Figure 7.12 Plan-view scanning-electron-micrographs of gratings exposed at the same equivalent-dose, 14% higher than for the exposures shown in Figure 7.11. In this case, the 200nm-period grating and the 1000nm-period grating both have a 50% duty cycle, as would be expected from an ideal binary resist.

The results for both the positive and negative-tone resists indicate that the binary resist model is reasonably accurate for situations where the equivalent-dose is above threshold. The threshold itself, however, exhibits a spatial-frequency dependence such that the exposure of lower spatial-frequency images requires less dose than at higher spatial-frequencies. The wider range of duty-cycles indicates

that the resist action is closer to ideal in low-spatial-frequency images. A likely explanantion for this effect is that the diffusion of photoproducts in the resist layer occurs on a length scale which is comparable to the linewidth in high-spatial-frequency exposures. For example, a diffusion of photoproducts from the exposed to unexposed regions in a 200nm-period gratings would manifest as a lower exposure dose, causing the features to be underexposed as seen in Figure 7.11. The combination of the significantly steeper spatial-gradients driving diffusion with the reduced size of the features creates a situation where high-spatial-frequency images are far more likely than low-spatial-frequency images to visibly show the effects of diffusion.

In practical terms, operation near the 50% duty cycle point demands more stringent process control than is necessary in the overexposed regime. Small variations in dose can cause the incomplete exposure of fine-period structures, and the slope of the duty-cycle/dose curve for coarse-period structures is much steeper than at higher doses.

In the overexposed regime, the dependence of the duty-cycle (or linewidth) on equivalent-dose is seen to be approximately linear. The *DC(ED)* curve of positive resist is shown in Figure 7.13 for a 200nm-period grating exposed using the Lloyd's-mirror interferometer. Process latitude is observed to be quite high, with the minimum duty-cycle occurring at a dose almost 2.5 times that of the maximum duty-cycle. A linear fit to the data is seen to fall within the measurement error, with a slope of ~1nm per 10 seconds of exposure. As the exposure time can be controlled to better than 1 second, nanometer-level linewidth control should be readily achieved, although this outstrips the available metrology and is therefore difficult to confirm.

The invariance of the duty-cycle to spatial-frequency allows the characterizations seen in Figure 7.10 and Figure 7.13 to be applied at other grating periods by inverting Equation 7.4 to solve for the incident dose *ID*.

$$ID = \frac{ED}{\left[(1 - R_1)(1 + R_2)\cos(\theta)\right]} \qquad (7.9)$$

Thus, a resist can be well characterized by the observed duty-cycle as a function of equivalent dose on a single substrate at a single period. For an arbitrary grating period on an arbitrary substrate, the new reflectivity values R_1 and R_2 can be calculated and used with the new interference-angle to scale the known value of equivalent-dose at the desired duty-cycle.

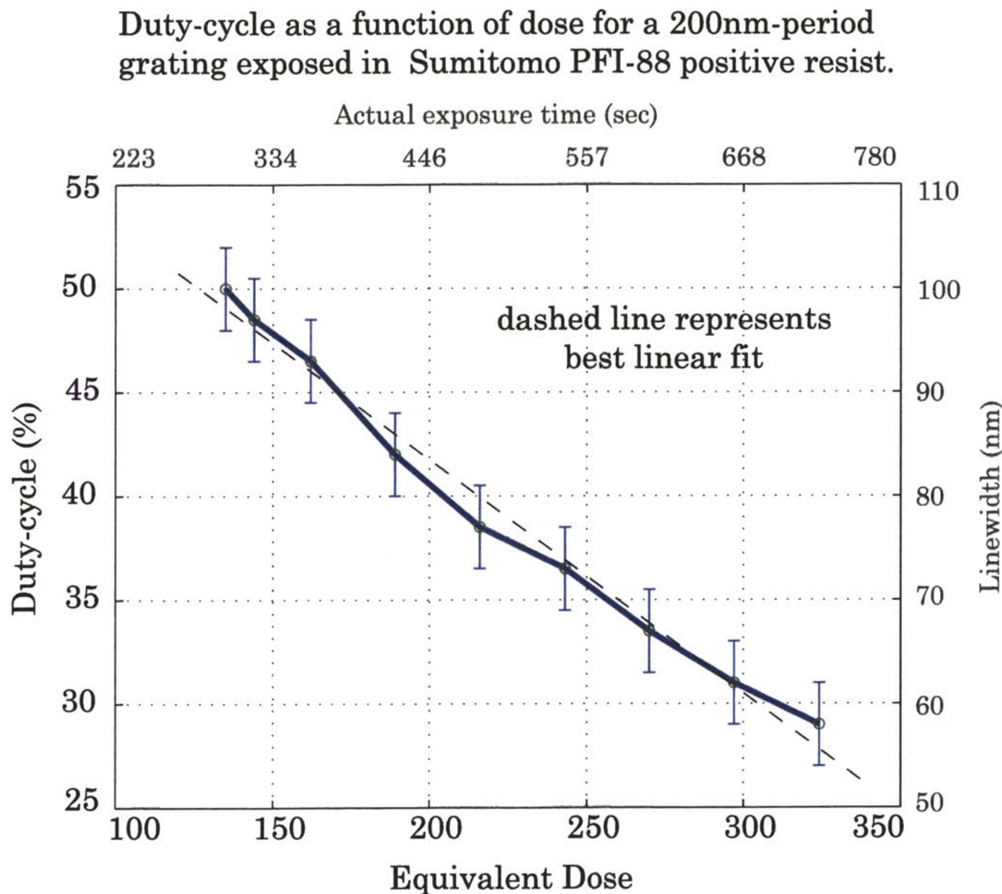


Figure 7.13 Duty-cycle as a function of equivalent dose for a 200nm-period grating in positive resist. Actual exposure times and linewidth are also shown. A linear-fit to the data is seen to fall within the measurement error. Measurement error of +-4nm is due to the pixelation of the scanning-electron-microscope images.

7.2 Patterned Magnetic Media

7.2.1 Introduction

In recent years, there has been a rapidly growing interest in the fabrication and study of nanometer-scale magnetic elements. The data-storage industry in particular stands to benefit from advances in magnetic nanostructures. The use of lithographically-patterned magnetic particles has been proposed as a paradigm for the next-generation of data-storage devices [35,36,37,38]. Current hard-disk technology utilizes a continuous magnetic thin-film composed of small (10-20nm) crystalline grains. Each of these grains has a preferred magnetization-direction which is randomly oriented, demanding that a large number of grains be used in each bit to meet signal-to-noise (SNR) requirements. Furthermore, the boundary between bits is rough due to the finite grain size, as depicted in Figure 7.14.

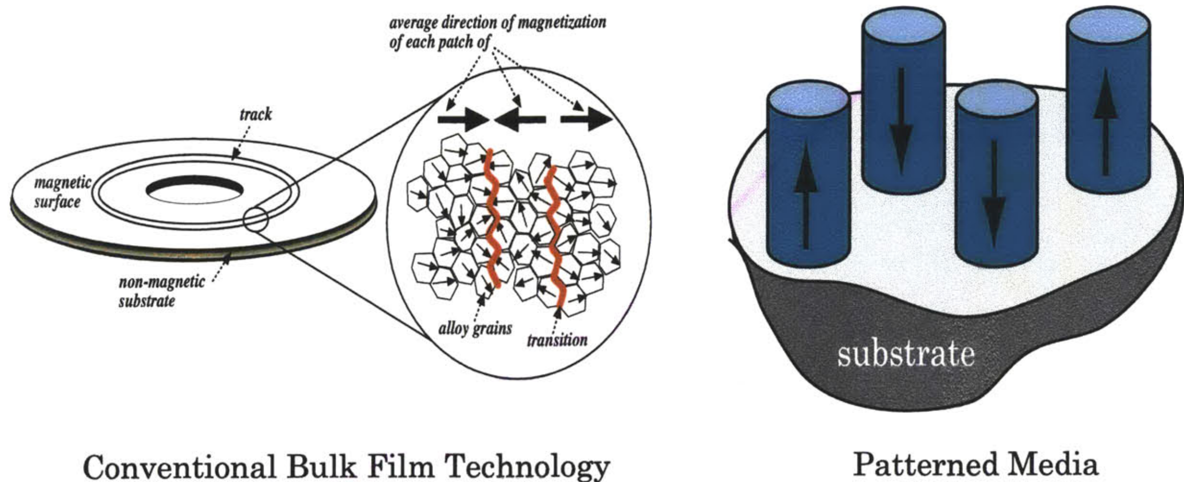


Figure 7.14 Bits of data in conventional hard disk (left) are composed of a number of individual nanocrystalline domains in a bulk magnetic film. Roughness of the bit edges due to the grain-structure reduce the signal-to-noise ratio when reading and writing these bits. The single-domain lithographically-defined bits in patterned magnetic-media (right) can have much smoother edges and require fewer grains per bit.

These issues can be ameliorated by reducing the size of the individual grains that compose the film. Unfortunately, a fundamental physical limitation prevents the grain size from becoming arbitrarily small. Known as the superparamagnetic limit, the magnetization becomes thermally unstable for particles less than ~10nm [39,40].

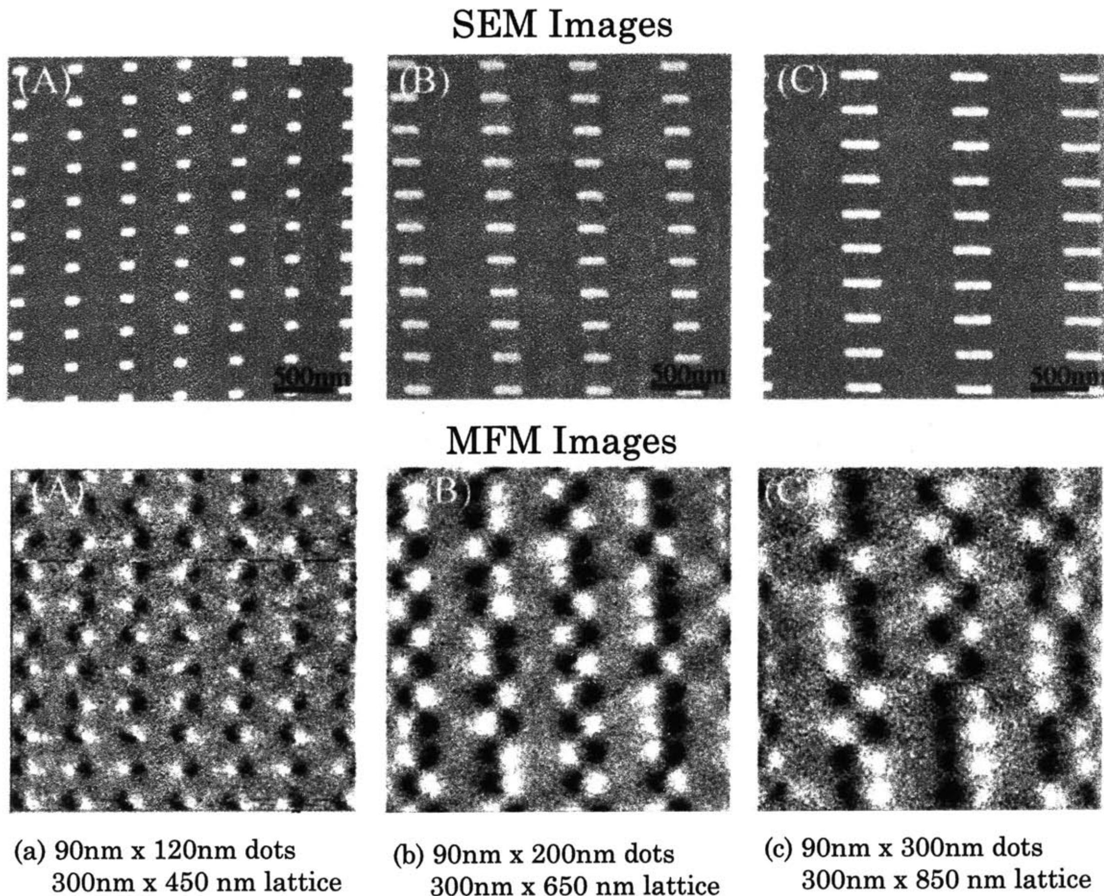


Figure 7.15 From [44]. Top: scanning-electron micrograph of magnetic dot arrays. Bottom: magnetic-force microscope (MFM) images of the magnetic field of the arrays at remanence. Dark spots represent fields pointing into the plane, white spots represent fields pointing out of plane. Array elements are seen to be single-domain dipoles.

The use of patterned-media, shown schematically in Figure 7.14, allows the bit size to be reduced beyond the realm possible using bulk films. Bits are physically separated from one another, and bit boundaries can be well-defined. Precision of the bit boundaries is now limited by lithography rather than the grain

structure of the film, increasing the SNR. More importantly, it has been shown that nanometer-scale particles can act as a single magnetic domain [41,42,43]. Figure 7.15 shows arrays of magnetic particles of three different sizes, and in each case all of the particles are acting as single domains [44]. For single-domain magnetic elements, the superparamagnetic limit applies to the size of the *entire particle* rather than the size of the constituent grains, thus enabling smaller absolute bit sizes. Estimates place the minimum size of a thermally stable bit at ~10nm per side (1000 nm^3) [39,40]. The potential data density for bits of this size is >1000 Gb/in^2, more than an order-of-magnitude increase over the potential of current technology.

Although random distributions of nanoscale magnetic particles have received considerable study [45,46,47], only within the past 15 years or so have fabrication and computational capabilities becomes sufficiently advanced to allow the study of controlled arrays of magnetic particles. Fundamental to the design of data-storage devices is a thorough understanding of the hysteresis (switching behavior) of the individual elements. Each bit must reliably switch when desired, yet be immune to switching at other times. In collaboration with the research group of Prof. Caroline Ross in the Department of Materials Science and Engineering, the switching behavior of magnetic nanoparticles was experimentally measured and compared to evolving theoretical models.

7.2.2 Fabrication patterned magnetic media using IL

Interference-lithography is an ideal method for patterning nanoscale magnetic elements, as particle arrays large enough to be measured in bulk can be easily fabricated ‡‡‡‡. Furthermore, as was discussed in Chapter 1, the size of the elements that can be fabricated using IL is smaller than that of most other

‡‡‡‡ A large number of particles is necessary to obtain a measurable magnetic moment using standard tools such as the vibrating-sample magnetometer (VSM), alternating-gradient magnetometer (AGM) and superconducting quantum-interference detector (SQUID).

lithographic techniques. Serial processes such as electron-beam lithography are capable of competitive feature sizes, but are generally too slow [48]. A wide range of particles sizes, aspect-ratios and densities are available by tuning the periodicity, linewidth, and rotation-angle for grid exposures using IL.

Both the tri-layer resist process described in Section 7.1.3 and the ultrathin resist process (Section 7.1.4) were used in the fabrication of samples for this study. The majority of the samples were prepared by depositing a magnetic thin-film over the entire substrate. Etch-masks of ~15-20nm tungsten (W) were deposited over the unpatterned magnetic films prior to spin-casting the resist stack. Titanium (Ti) was also used as an etch-mask for some samples. Patterns exposed using IL were transferred through the resist-stack using reactive-ion etching (RIE) as described in Section 7.1.3. This pattern was subsequently etched into the metallic hard-mask, also using RIE. Both W and Ti are etched in a fluorine chemistry similar to that used to etch silicon [10]. Although it would be advantageous to use an etch chemistry such as CHF_3 that is highly selective to the ARC, the presence of hydrogen acts to catalyze the deposition of W from its volatile compound WF_6. Not surprisingly, the observed etch rate of W in CHF_3 is approximately zero.

Remains of the resist stack are removed, and the patterned W or Ti is used as a mask for etching the magnetic film. Magnetic materials generally cannot be etched using RIE, making them considerably more difficult to pattern than many semiconductor materials. A novel ion-beam etch process developed as part of this project is detailed in Chapter 8. Using this process, sub-100nm features can be patterned in a variety of magnetic materials, including cobalt, nickel, and permalloy (~80% Ni, ~20% Fe by weight) with excellent dimensional control.

7.2.3 Switching behavior of patterned magnetic media

The rigorous definition of a single-domain magnetic particle is one in which the magnetization is parallel throughout the element. Although this state can only

be achieved for the impractical case of perfectly ellipsoidal structures, it provides an excellent point of reference from which to characterize the behavior of other magnetic elements with shape anisotropy that can be approximated as ellipsoidal. A well-known model for the switching of these ideal elements, the so-called Stoner-Wolfarth model [49], describes the magnetization rotation as coherent. That is, the spin states of all the atoms within the element remain parallel at all times. Under this model, the magnetization aligns either parallel or antiparallel to the long axis of the ellipse in the absence of an external field, thus having the ability to store 1 bit of data. A schematic of this switching behavior is shown in Figure 7.16.

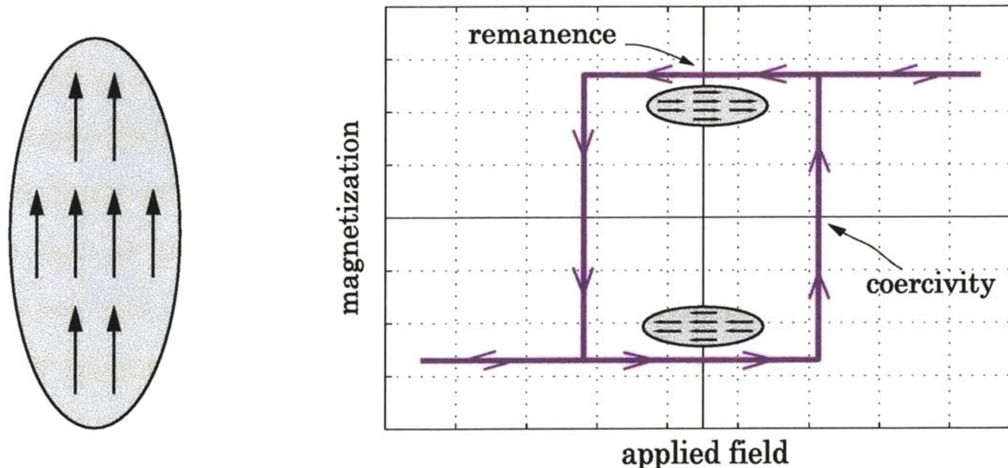


Figure 7.16 Uniform magnetization in a single-domain ellipsoidal particle (left). The Hysteresis loop of a single-domain particle is square (right). Magnetization does not change in the presence of an applied field until a certain point, where it switches entirely to the opposite state. The field at which this occurs is known as the coercivity. The magnetization of the particle in the absence of an applied field is called the remanence.

Magnetic elements fabricated experimentally (rather than theoretically) will never be perfectly ellipsoidal. Further, magnetocrystalline anisotropy due to the grain structure of the films may prevent uniform magnetization. Therefore, the term single-domain is generally used to refer to the particles with one domain that show only small deviations from uniform magnetization.

7.2.4 Hysterisis loops

A large number of patterned-media samples were prepared with numerous combinations of aspect-ratio and feature size. Only a few of those samples, in conjunction with their measured hysteresis loops, will be described here. The reader is referred to the literature for thorough coverage of the work done by the author and collaborators [50,51,52].

Figure 7.17 shows arrays particles with three different aspect-ratios, (a) ~1.5, (b) ~2.5, and (c) ~3.5 [53]. Hysteresis loops measured for each sample are shown to the right of the micrographs. All elements were etched from a multilayer stack of (5nm Cr)/(5nm Co)/(3nm Cu)/(2 nm Co) deposited by sputtering on a silicon wafer. The chromium layer is used as to promote adhesion of the cobalt. It was expected that the two cobalt layers would switch individually, an effect that should be seen in the hysteresis loops. However, only a hint of this is seen in Figure 7.17(b), where a slight shoulder is visible indicating the switching of the top Co layer. It is likely that the unprotected top layer of 2nm-thick Co has been largely oxidized at some point during the process such that the switching of only the lower 5nm-thick layer is observed.

The coercivity along the major axis for samples (a), (b) and (c) is 200 Oe, 360 Oe, and 435 Oe, respectively. The coercivity is seen to increase with the aspect-ratio and a decrease in the volume of the particle. However, the coercivities are lower than the theoretical values predicted for a single-domain ellipsoid, which are 280 Oe, 680 Oe, and 980 Oe, respectively [54]. Switching fields lower than the value predicted by the Stoner-Wolfarth model indicate that the particle switching occurs incoherently.

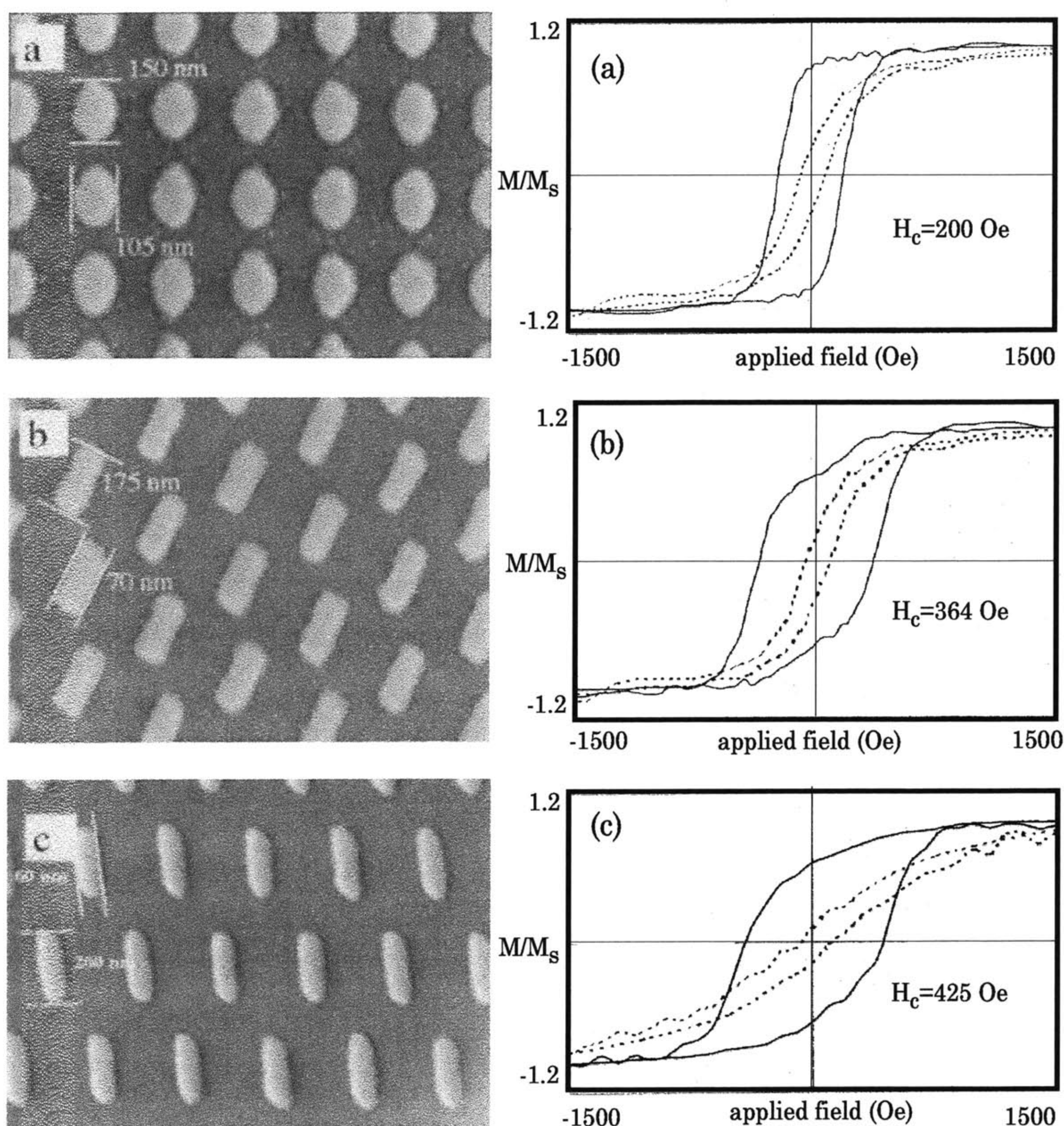


Figure 7.17 SEM micrographs (left) and measured hysteresis loops (right) for three different arrays. Data plotted by the solid lines are parallel to the long direction of the dots, dotted lines are measured perpendicular to the long axis. [53]

The sheared shape of the hysteresis loops for the arrays are qualitatively different than the square loop expected from of any single particle. Micromagnetic simulations have been used to determine the cause of this difference [55]. Although shear of the hysteresis loop can be characteristic of strong magnetostatic

interactions between particles, the interparticle spacing is large enough that arrays shown in Figure 7.17 are only weakly interacting. The shear must therefore be caused by a large distribution in the switching fields of the individual particles. Recall that the measured signal is the aggregate magnetization of a large number of particles (~10^9). Thus, although each particle individually may have an ideal square loop, the sum hysteresis of the array will not be square unless each particle switches at exactly the same applied field. Simulations indicate that the standard-deviation of the switching-field for the particles in Figure 7.17(a) is as high as 150 Oe, out of an average coercivity is 200 Oe. The likely causes of this large spread in switching fields are small irregularities in the shape of the particles, and a distribution of the magnetocrystalline anisotropy due to the random grain structure of the cobalt films.

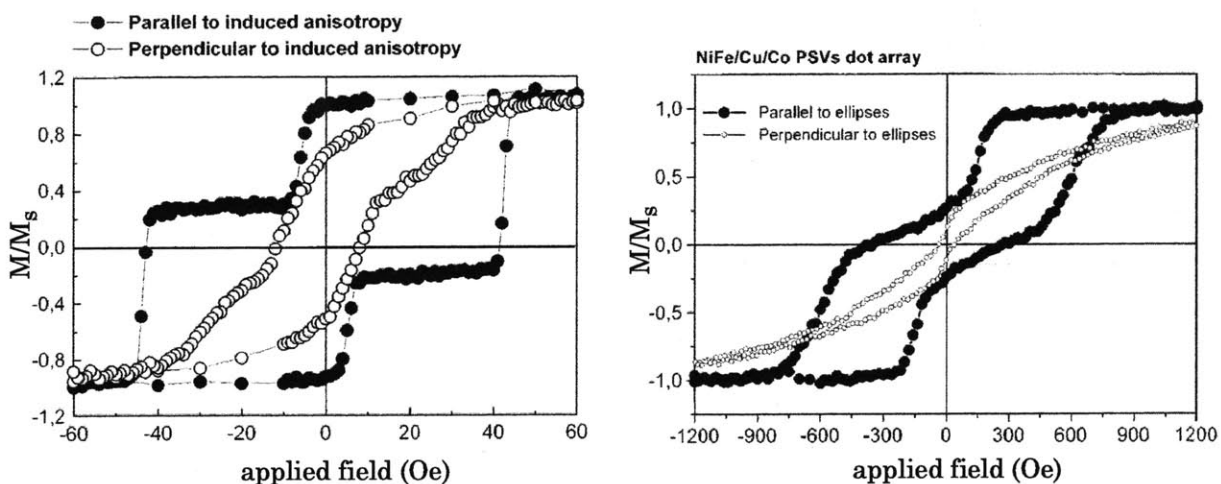


Figure 7.18 Hysteresis loops for unpatterned (left) and patterned (right) films of (6 nm NiFe)/(6 nm Cu)/(4nm Co)/(4nm Cu. Anisotroopy induced during deposition in the unpatterned film is dominated by the shape anisotropy in the patterned film, which is perpendicular to the easy-axis of the film. [56]

Measurements of a similar thin-film stack, (6 nm NiFe)/(6 nm Cu)/(4nm Co)/(4nm Cu), indicate that the inclusion of a thin copper capping layer is sufficient to allow for the patterning at nanoscale dimensions while maintaining distinct

switching of the two layers [56]. Figure 7.18 shows the hysteresis loops for: (a) the unpatterned bulk films and (b), the film patterned with an array of 80nm x 140 nm dots. Distinct two-layer switching is observed in both cases. The patterned NiFe layer reverses at ~150 Oe, while the Co layer reverses at ~600 Oe. These switching fields are significantly higher than those of the bulk film, which are 5 Oe and 45 Oe respectively.

Anisotropy in the unpatterned film, evidenced by the difference between the black and white curves in Figure 7.18(a), was deliberately induced through the use of an applied magnetic field during the deposition. However, shape anisotropy in the patterned films was dominant to the induced film anisotropy. The long-axis of the patterned ellipses, set perpendicular to the easy-axis of the film, overwhelmed the film anisotropy in determining the preferred axis of the features.

Examining the hysteresis of the patterned films, the NiFe layer is seen to reverse direction *before* reversal of the applied field. This is due to the strong magnetostatic coupling of the two layers. The field from the Co layer as seen by the NiFe layer is opposite to the direction of the field in the Co layer itself, and is sufficient to reverse the direction of the NiFe layer despite the parallel external field. The anti-parallel magnetization of the two layers is observed by the low combined magnetization of the pair at remanence. This is a desirable property in a memory device because the state of a bit held in the hard layer (Co) can be read by flipping the soft layer (NiFe) without changing the bit itself.

7.3 Templated self-assembly

7.3.1 Introduction

Self-assembled nanostructures tend to exhibit short-range order at best, but are often completely random. A combination of top-down and bottom-up techniques called templated self-assembly (TSA) is one method to induce long-range order in a self-assembled process by using lithographic templates. Templated self-assembly is a broader outgrowth of work performed here at MIT and elsewhere in the 1970's and 1980's on the orientation of microstructure in thin films using templated substrates, know as graphoepitaxy [57,58,59]. This technique was also used to demonstrate the orientation of liquid-crystal molecules [60,61]. Recent interest in TSA has included the ordering of block-copolymers. [62,63,64], and binary mixtures of incompatible polymers [65], as well as the crystallization of organic semiconductors [66]. Results from Cheng *et al.* are shown in Figure 7.19

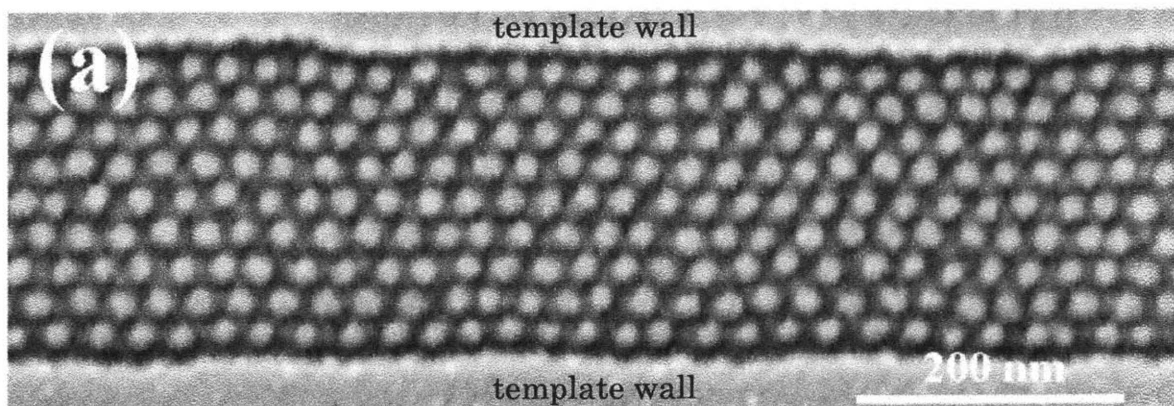


Figure 7.19 An ordered array of ~30 nm dots formed by annealing a diblock copolymer inside a shallow silicon trench. The particles show a hexagonally close-packed structure with a spacing of ~56nm [64].

7.3.2 Directed assembly of SiGe quantum-dots

The de-wetting of a thin-film into an array of particles, either during deposition or annealing, is called agglomeration. Also known as quantum-dots, a variety of devices exploiting these particles have been proposed, including quantum computers and single-photon lasers [67,68]. However, poor control over the location of the particles is an important factor inhibiting the development of these devices.

Silicon-germanium alloys ($Si_{1-x}Ge_x$) grown on silicon will form quantum-dots under conditions of high film-strain, corresponding to a high germanium content. Formation of these islands is a way for the film to relieve strain caused by a lattice mismatch between the SiGe alloy and the silicon substrate. It has been determined that the size distribution and spatial ordering of the quantum-dots depends on the diffusion of atoms across the substrate, and that the distribution of island sizes can be controlled by the growth conditions [69,70,71]. Recently, it has been demonstrated that substrate topography can also influence the development of the islands [72,73,74]. The precise mechanisms for this, however, are not well understood. In collaboration with researchers at MIT and IBM, attempts have been made to understand these processes through real-time monitoring of the nucleation of Ge islands on substrates patterned using interference lithography [75].

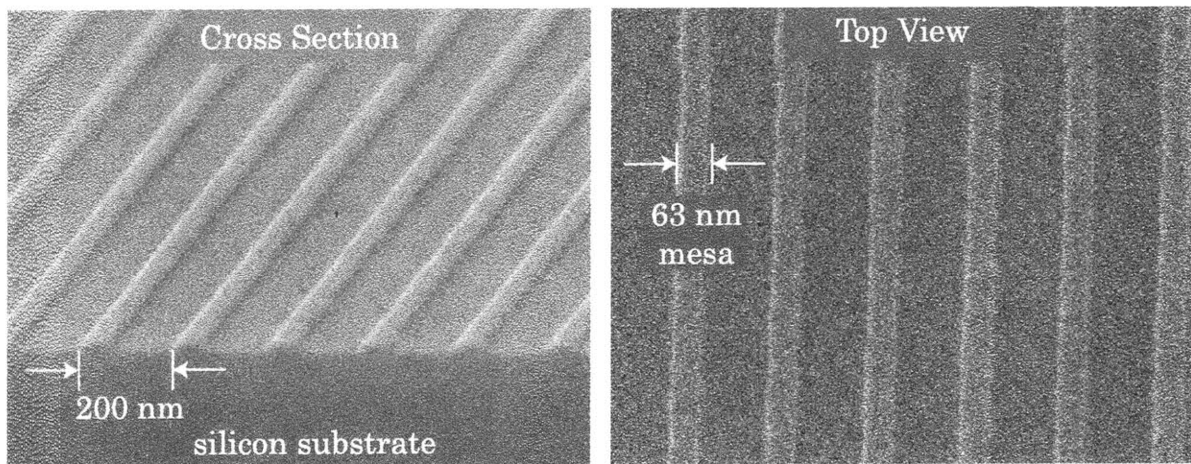


Figure 7.20 Scanning-electron microscope (SEM) images of the deposition template in cross-section (left) and top-view (right). The substrate is a silicon wafer with ~65 nm lines etched 20 nm deep on a 200nm pitch.

Using interference lithography followed by reactive-ion etching (RIE), 200nm-period grating were etched ~20 nm deep into silicon substrates. These templates were then mounted in a transmission-electron microscope (TEM) modified to allow for chemical-vapor deposition (CVD) of silicon and germanium during imaging [76,77,78]. Using this tool, evolution of the island formation can be studied in real-time. Comparisons were made between templates of varying dimensions, with and without a high-temperature annealing step (900°C, 5 minutes) prior to the deposition. The anneal acts to smooth the nearly vertical sidewall of the etched structure - this is depicted in the cross-sections at the right of Figure 7.21.

Early experiments used 200nm-period gratings with 65nm-wide mesas and 130nm-wide valleys, shown in Figure 7.20. On the un-annealed (straight sidewall) templates, Ge dots first formed on the valley floor, appearing later on the mesa tops after additional deposition. This is shown in Figure 7.21(a)§§§§. Samples annealed prior to deposition, with gently sloping sidewalls, showed the opposite behavior. Fully annealed samples, seen in Figure 7.21(b), showed preferential island formation on top of the mesas, with dot formation occasionally occurring on the valley floor only much later in the process. One can conclude from this that the slope of the sidewall has enough influence over the island formation to switch the preferential nucleation sites from the valley to the mesa.

Three mechanisms can be identified which influence the location of preferential island formation. First, germanium atoms with high surface-mobility will tend to diffuse towards the mesa tops due to the local reduction of strain caused by greater compliance of the thin mesa line [79,80]. The increased Ge film thickness that results from this diffusion will cause nucleation to occur on the mesas before it occurs in the valleys.

§§§§ Note that the images, recorded using reflection-electron microscopy, are highly foreshortened in the horizontal direction. The dark-field images are formed from scattering of grazing incidence electrons off of topographical irregularities such as the SiGe islands.

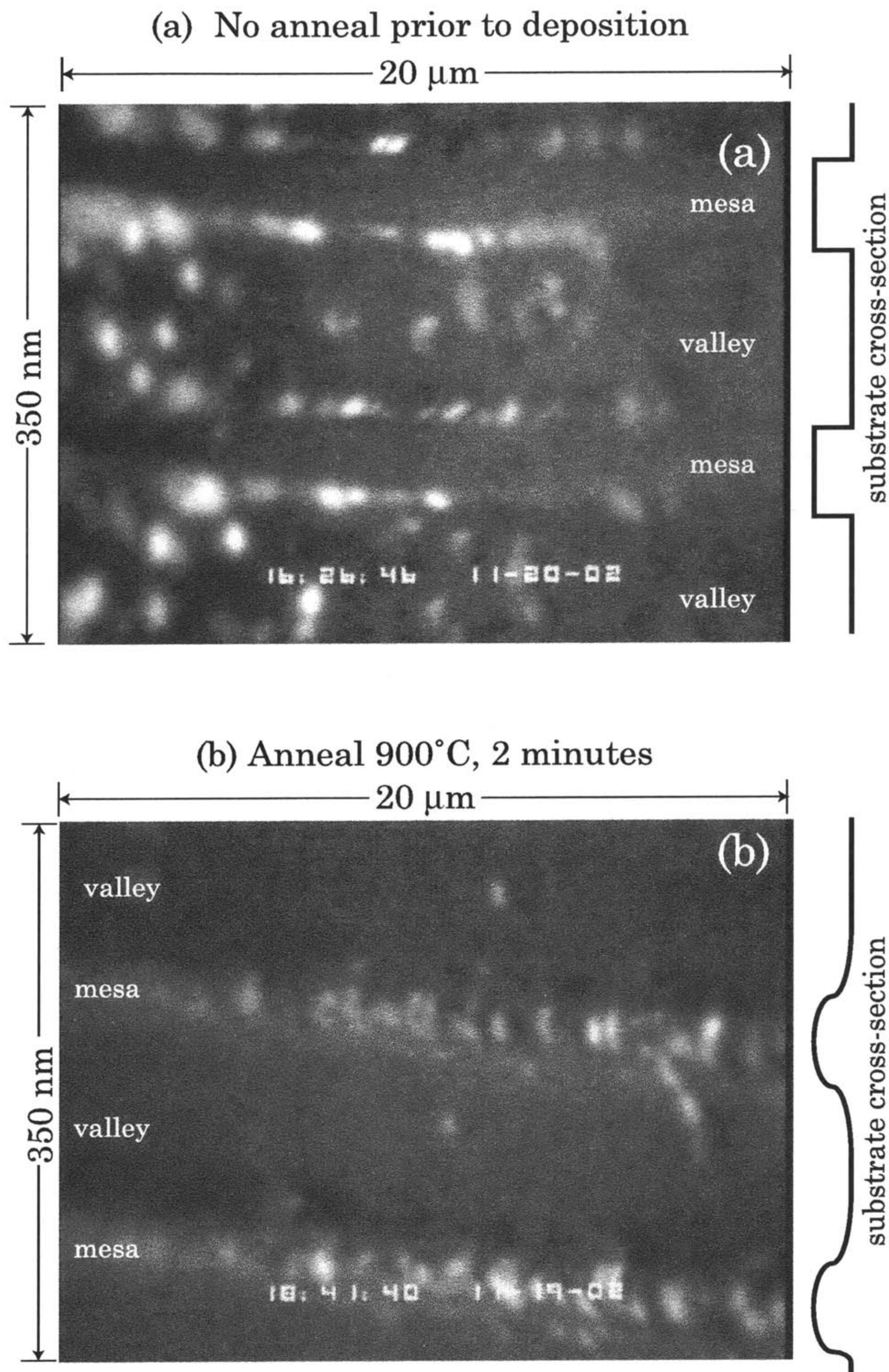


Figure 7.21 (a) Quantum dots seen as white areas form on the valley floor rather than the mesa tops in the un-annealed sample. The cross section (at right) shows vertical mesa sidewalls. (b) Annealed samples show the islands forming preferentially on the tops of the mesas. The cross section of the annealed samples shows smoothly sloped sidewalls.

In competition with this are two mechanisms that oppose the formation of dots at the top of the narrow mesas. The lattice mismatch between silicon and germanium creates a large amount of strain in the deposited film, and the quantum-dots form as a way to relieve this strain. However, strain at the top of the mesa will be lower than the strain in the valley, therefore the "driving force" towards the formation of the islands should be reduced on the mesa tops compared to the valleys. Also, the width of the mesa compared to the valley means that more of the mesa area is closer to the edge (sidewall) of the structure. The sidewalls appear to act as "sinks" for the germanium, which may disrupt the formation of the islands. This effect was observed on these structures by ceasing the flux of germanium and monitoring the evolution of the quantum-dots over time. On all of the samples, the islands shrank over time and eventually disappeared, a behavior which is consistent with diffusion of the Ge to the mesa sidewalls. In contrast, quantum-dots grown on a flat substrate continue to cluster into a few large islands even after the deposition has ended [78].

For all the samples, both un-annealed and fully annealed, the relative effects of the three mechanisms should be equal due to the identical template geometry. It is clear, however, that the reduced sidewall slope of the annealed samples enables the preferential formation of the islands on the mesa tops. This can be explained if we consider that the surface diffusion of the Ge atoms is impeded by steep mesa sidewalls on the un-annealed samples. In contrast, the tendency for mobile Ge to cluster on the mesas clearly outweighs the opposing factors on the annealed templates. If the germanium were still able to diffuse freely over the steep sidewalls, we would expect the same result for both cases. Thus, we can conclude that the steep sidewalls act as a diffusion barrier such that the mesas and valleys act independently of one another. The valleys, which are further from the sidewalls and do not have the benefit of the topological strain relief of the mesas, become the preferred location for island formation.

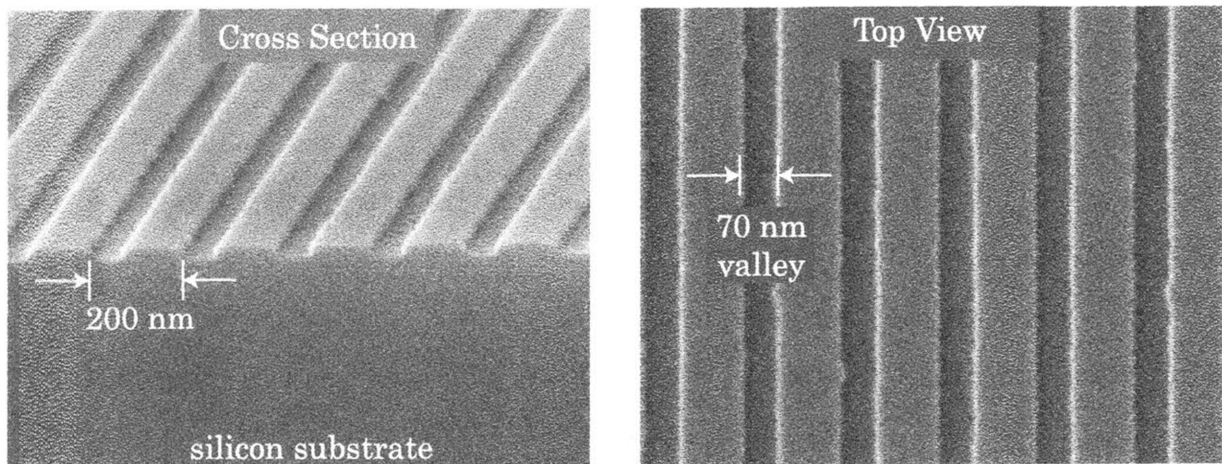


Figure 7.22 Substrates prepared which have the mesa and valley widths inverted from those in Figure 7.20. 130nm-wide lines etched 20 nm deep into silicon substrates on a 200nm pitch are seen in cross-section in the left-hand micrograph and top-down in the right-hand micrograph.

The geometry of the narrow-mesa template was such that the preferential location for dot formation would switch from mesa to valley depending on the sidewall slope. If the template were reversed, with a 130nm-wide mesa and a 70 nm-wide valley as shown in Figure 7.22, then the preferred location of nucleation should always be the mesa. Interestingly, tests with these samples showed that the islands formed preferentially in the valleys even when annealed before the Ge deposition. Reasons for this are still unclear. To clarify this surprising result, an additional set of samples was produced to maintain the same valley width (130nm) and depth (20nm) as in the first experiment, but increase the mesa width. These templates, shown in Figure 7.23, have periodicities of (a), 260nm at a 50% duty cycle, and (b), 390nm at a 66% duty cycle. Samples with significantly larger dimensions at spatial-periods of 1μm and 2μm were also produced (not shown). Experiments testing the island formation of these samples are ongoing.

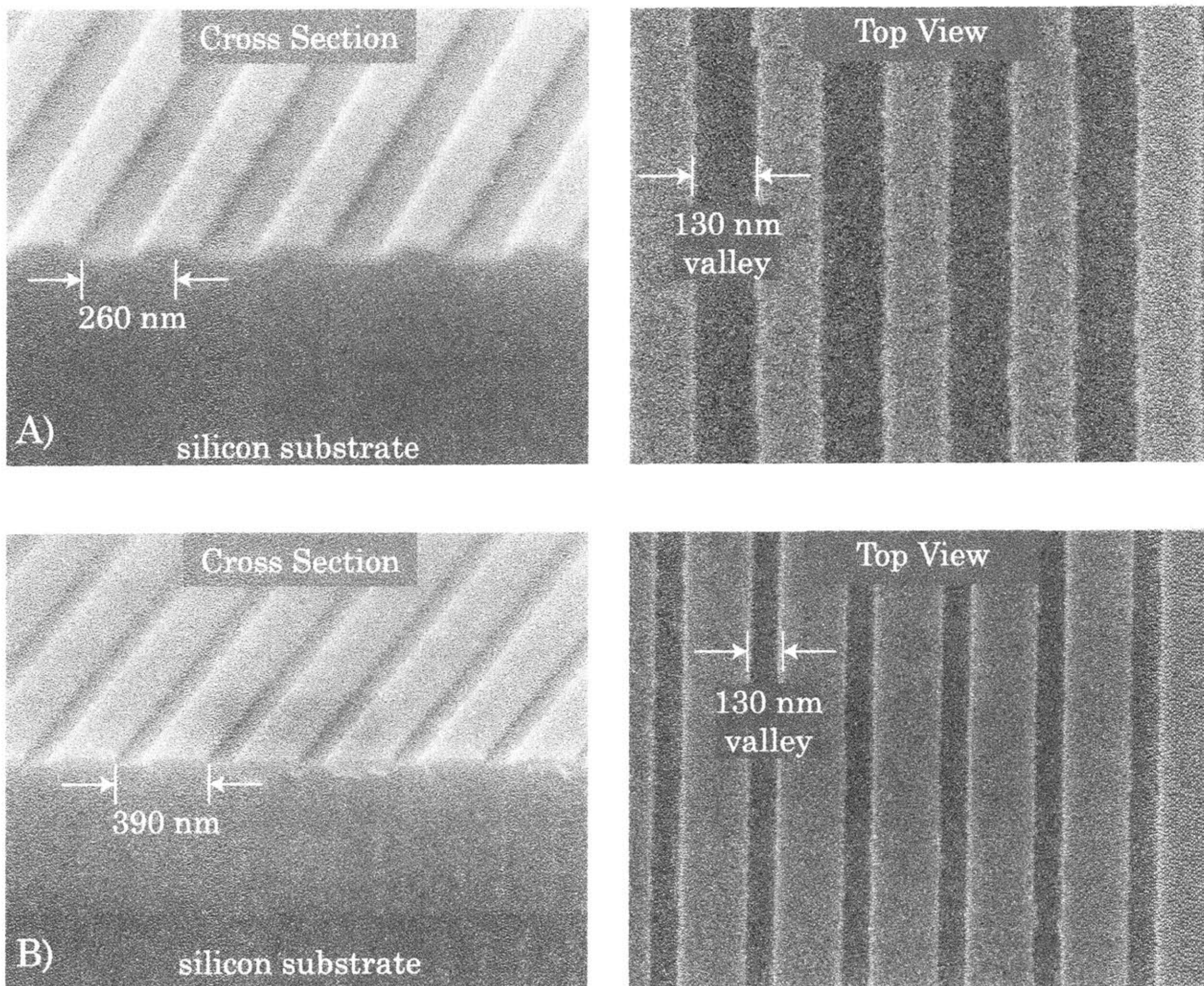


Figure 7.23 Additional substrates prepared with the same width of the valley floor (130 nm) as those shown in Figure 7.20, but with different mesa widths. (a) The top images show samples with equal mesa and valley width. (b) The lower images show mesas twice as wide as the valley. It is expected that the substrates with wider mesas will exhibit quantum-dot formation preferentially on top of the mesas for all annealing conditions.

7.4 Semiconductor Nanocrystal Lasers

7.4.1 Introduction

Another excellent example of the synthesis of top-down and bottom-up nanotechnology is the development of a semiconductor-nanocrystal laser in collaboration with the group of Prof. Moungi Bawendi at MIT [81]. Colloidal semiconductor nanocrystals, also known as quantum-dots (QD), consist of self-assembled inorganic semiconductor crystals, typically less than 10nm in diameter, surrounded by a layer of organic ligands [82]. The electronic and optical properties of these nanocrystals are tunable through the size of the particle, while the organic shell allows them to be chemically integrated into a variety of materials. Using the quantum-dots as a gain medium in conjunction with a distributed-feedback (DFB) structure patterned using interference lithography and RIE, a highly tunable and robust laser system can be developed.

Crystals of diameter between 1nm and 6nm operate in the so-called strong confinement regime, allowing greater than 1 eV tuning of the emission wavelength. In this regime, size has been reduced to the point that the spacing of energy levels in the QD is significantly greater than the available thermal energy, leading to properties which are robustly insensitive to temperature [83]. Low lasing thresholds are also predicted for nanocrystal gain media [84]. Additionally, color selectivity of the gain profile can be easily controlled by changing the size or material of the nanocrystals.

It has been shown that a film of CdSe nanocrystals can exhibit amplified stimulated emission of wavelengths between 550-650nm using femtosecond pulses at 400nm-wavelength for optical pumping [85]. A spin-coatable titania composite can be synthesized with a high volume-fraction of nanocrystals, either using CdSe or CdSe(Zn) [86,87]. Thus, device fabrication can be as simple as spin-coating the nanocrystal gain-medium over a DFB structure. A variety of different DFB

gratings with periodicities between 300nm and 400nm were fabricated using the process described in Section 7.1.3. A representative example is shown in Figure 7.24. A 330nm-period grating etched in silica is shown with and without the composite nanocrystal film.

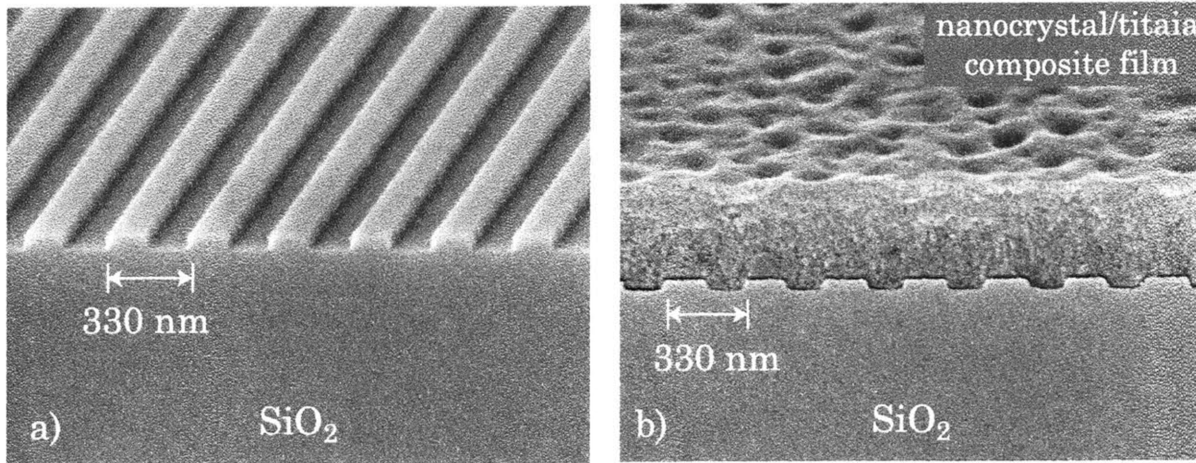


Figure 7.24 Scanning-electron micrographs of the DFB structure. a) 330nm-period grating etched 50nm-deep into a SiO_2 substrate. b) Same grating with the titania/nanocrystal composite cast on top.

The refractive index of the film will vary between n=~1.65 and n=~1.82 depending on the volume fraction of nanocrystals, which can be as high as 12%. This is important as the Bragg condition of the DFB structure, seen below, includes the refractive index of the film [88].

$$P = \frac{m\lambda_0}{2n_{eff}}, \tag{7.10}$$

where the period of the DFB is P, m is the order of diffraction, λ_0 is the emission wavelength and n_{eff} is the effective index of the composite nanocrystal film. Thus, either the periodicity of the DFB structure or the film index - or both- can be varied to achieve a certain emission wavelength.

As an example of this flexibility, four different lasers are shown in Figure 7.26. Solid curves show the normalized emission intensity as a function of photon energy above the lasing threshold, while the dashed curves are the

photoluminescence spectra below the lasing threshold. Each curve represents a nanocrystal of different diameter cast in a film whose index has been tuned such that all the lasers operate *using the same period DFB structure*, in this case P=350 nm. Note that the width of the emission below threshold, ~30 nm full-width half-max (FWHM) sharpens to ~1.5nm FWHM above threshold. This is an excellent demonstration of the simplicity of tuning emission wavelength with crystal size, and also of the utility of refractive-index tuning via the volume-fraction of crystals in the film. The laser spectrum of a similar device made using CdSe(Zn) nanocrystals and operating at room temperature is shown in Figure 7.25. The linewidth of this device does not change over the range from 80°K to room-temperature, realizing the predicted temperature insensitivity of a nanocrystal-based laser.

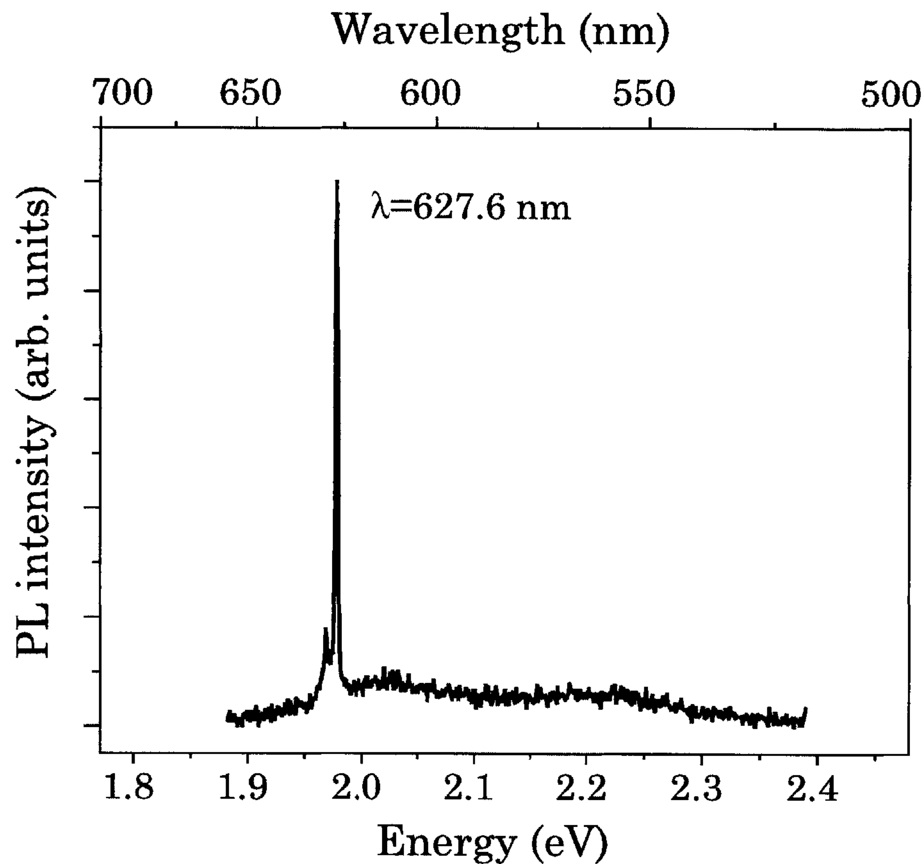


Figure 7.25 Room temperature emission spectrum a CdSe(Zn) nanocrystal DFB laser. The narrow bandwidth of the emission, similar to that of the devices operating at 80°K seen in Figure 7.26, is evidence of the temperature insensitivity of the device.

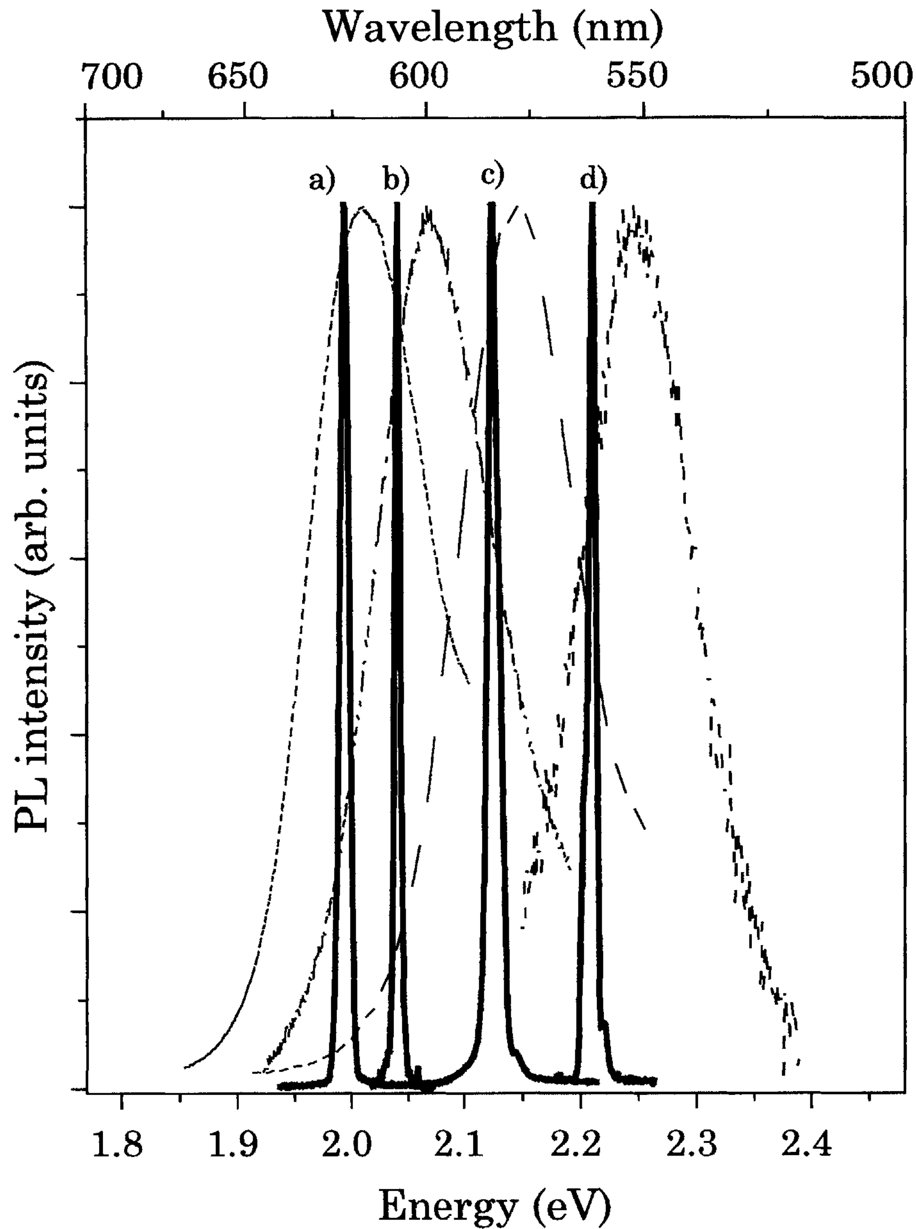


Figure 7.26 Photoluminescence emission spectra for nanocrystals of different sizes below threshold (dashed lines) compared with the emission spectra above the laser threshold (solid lines) for DFB devices operating at 80°K. Nanocrystal composition and wavelength of the laser emission is a) CdSe 621nm, b) CdSe 607nm, c) CdSe 583nm, d) CdSe(Zn) 560nm.

7.5 References

[1] S.R.J. Brueck, S.H. Zaidi, X. Chen, Z. Zhang, "Interferometric Lithography - from periodic arrays to arbitrary patterns". *Microelectronic Eng.* p.145 (1998)

[2] X. Chen, S.H. Zaidi, S.R.J. Brueck, D. Devine, "Interferometric lithography of sub-micrometer sparse hole arrays for field emission displays". *J. Vac. Sci. Technol. B* **14**(5) p.3339 (1996)

[3] A. Fernandez, J.Y. Decker, S.M. Herman, D.W. Phillion, D.W. Sweeney, M.D. Perry, "Methods for fabricating arrays of holes using interference lithography". *J. Vac. Sci. Technol. B* **16**(6) p.2439 (1997)

[4] J. Klein-Weile, P. Simon, "Fabrication of periodic nanostructures by phase-controlled multiple-beam interference". *App. Phys. Lett.* **83**(23) p.4707 (2003)

[5] V. Berger, O. Gauthier-Lafaye, E. Costard, "Photonic band-gaps and holography". *J. Appl. Phys.* **82** p.60 (1997)

[6] T.A. Brunner, "Optimization of optical properties of resist processes". *Proc. SPIE* **1466** p.297 (1991)

[7] T. Brewer, R. Carlson, J. Arnold, "The Reduction of the Standing-Wave Effect in Positive Photoresists", *J. Appl. Photographic Engineers* **7**(6) p.184 (1981)

[8] Y. Lin, S. Jones, G. Fuller, "Use of antireflective coating in bilayer resist process". *J. Vac. Sci. Technol. B* **1**(4) p.1215 (1983)

[9] R.R. Dammel, R.A. Norwood, "Modeling of Bottom Anti-Reflection Layers: Sensitivity to Optical Constants". *Proc SPIE* **2724** p.754 (1996)

[10] M.E. Walsh, "Nanostructuring magnetic thin films using interference lithography". M.S. Thesis, Massachusetts Institute of Technology (2000)

[11] D.R. Lide, ed. *CRC Handbook of Chemistry and Physic, 73rd edition.* CRC Press, Ann Arbor MI, (1993)

[12] E. Palik, ed., *Handbook of Optical Constants of Solids*, Academic Press, New York, *Volume 1* (1985), *Volume 2* (1991), *Volume 3* (1998)

[13] Measurements done by the author at the MIT Institute for Soldier Nanotechnology using an M-2000 spectroscopic ellipsometer from the J.A. Woollam Company.

[14] M.L. Schattenburg, R.J. Aucoin, R.C. Fleming, "Optically matched tri-level resist process for nanostructure fabrication". *J. Vac. Sci. Technol. B* **13**(6) p.3182 (1995)

[15] M. Farhoud, J. Fererra, A.J. Lochtefeld, T.E. Murphy, M.L. Schattenburg, J. Carter, C.A. Ross, H.I Smith, "Fabrication of 200nm-period nanomagnet arrays using interference lithography and a negative resist". *J. Vac. Sci. Technol. B* **17**(6) p.3182 (1999)

[16] Two primary i-line photoresists were used for this work: a) PFI-34 and PFI-88 positive resists from the Sumitomo/Sumika corporation, and b) THMR-iN PS4 negative resist from Tokyo Ohka Kogyo (TOK).

[17] H. Ito, C.G. Willson, *Polym. Sci. Eng.* **23** p.1012 (1982)

[18] J. Frechet, H. Ito, C.G. Willson, *Proc. Microcircuit Eng.* **82** p.260 (1982)

[19] H.Ito, "Chemical Amplification Resists: History and development within IBM". *IBM J. Res. Dev.* **41**(1) p.69 (1997)

[20] S.A. MacDonald *et al.*, "Airborne Chemical Contamination of a Chemically Amplified Resist", *Proc. SPIE* **1466** p.2 (1991)

[21] W.D. Hinsberg *et al.,* "Quantification of Airborne Chemical Contamination of Chemically Amplified Resists Using Radiochemical Analysis" *Proc. SPIE* **1672** p.24 (1992)

[22] W. Moreau, *Semiconductor Lithography: Principles, Practices and Materials,* Plenum Press, New York, (1988)

[23] J. Coburn, "Plasma etching and Reactive-Ion Etching". *American Vacuum Society Monograph Series*, Ed. N. R. Whetton, Amer. Inst. Phys., New York, (1983)

[24] W. Kern, D.A. Pluotinen, "Cleaning Solutions Based on Hydrogen Peroxide for Use in the Semiconductor Industry". *RCA Review*, **31** p.137 (1970)

[25] T. Tanaka, M. Morigmi, N. Atoda, "Mechanism of resist pattern collapse during development process". *Jpn. J. Appl. Phys.*, Part 1 **32**, p.6059 (1993)

[26] L. Ocala, D. Tennant, G. Timp, A. Novembre, "Lithogrpahy for sub-60nm resist nanostructures". *J. Vac. Sci. Technol. B* **17**(6) p.3164 (1999)

[27] J. Simons *et al.*, "Image Collapse Issues in Photoresist". *Proc SPIE* **4345** p.19 (2001)

[28] H.B. Cao, P.F. Nealey, W.D. Domke, "Comparison of resist collapse properties for deep-ultraviolet and 193nm resist platforms". *J. Vac. Sci. Technol. B* **18**(6) p.3303 (2000)

[29] N.N Efremow, N.P. Economou, K. Bezijian, S.S. Dana, H.I Smith, "A simple technique for modifying the profile of resist exposed by holographic lithography". *J. Vac. Sci. Technol.* **19**(4) p.1234 (1981)

[30] M.E. Walsh, Y.Hao, C.A. Ross, H.I. Smith, "Optimization of a lithographic and ion-beam etching process for nanostructuring magneto-resistive thin-film stacks". *J. Vac. Sci. Technol. B* **18**(6) p.3539 (2000)

[31] R. Srinivasan, "Kinetics of the ablative photo-decomposition of organic polymers in the far-ultraviolet (193) nm." *J. Vac. Sci. Technol. B*, **4** p.923 (1983)

[32] S. Rice, K. Jain, "Direct high-resolution excimer laser etching". *App. Phys. A* **33** p.195 (1984)

[33] V. Srinivasan, M.A. Smrtic, S.V. Babu, "Excimer laser etching of polymers". *App. Phys. Lett.* **43** p.1076 (1986)

[34] R. Dammel, "Photoresists for microlithography, or the Red Queens race". *J. Microlith. Microsys. Microtech.* **1**(3) p.270 (2002)

[35] S.E. Lambert, I.L. Saunders, A.M. Patalch, M.T. Krounbi, S.R. Hetzler, "Beyond discrete tracks: Other aspects of patterned media". *J. Appl. Phys.* **69** p.4724 (1991)

[36] J.F. Smyth, S. Schultz, D.R. Fredkin, D.P. Kern, S.A. Rishton, H. Schmid, M. Cali, T.R. Koehler, "Hysteresis in lithographic arrays of permalloy particles: experiment and theory (invited)" *J. Appl. Phys.* **69** p.5262 (1991)

[37] B.A. Everitt, A.V. Pohm, R.S. Beech, J.M. Daughton, "Pseudo-spin-valve MRAM cells with sub-micrometer critical dimensions". *IEEE Trans. Magn.* **34** p.1060 (1998)

[38] R.L. White, R.H.M. New, R.F.W. Pease, "Patterned Media: A Viable Route to 50GB/in^2 and Up for Magnetic Recording?". *IEEE Trans. Magn.* **33** p.990 (1997)

[39] P. Lu, S.H. Charap, "Thermal instability at 10 Gbits/in^2 magnetic recording". *IEEE Trans. Magn.* **30** p.4230 (1994)

[40] S.H. Charap, P. Lu, Y. He, "Thermal Stability of Recorded Information at High Densities", *IEEE Trans. Magn.* **33** p.978 (1997)

[41] G.A. Gibson, J.F. Smyth, S. Schultz, D.P. Kern, "Observation of the switching fields of individual Permalloy particles in nanolithographic arrays via magnetic-force microscopy". *IEEE Trans. Magn.* **27** p.5187 (1991)

[42] R.M.H. New, R.F.W. Pease, R.L. White, "Physical and magnetic properties of submicron lithographically defined patterned magnetic islands". *J. Vac. Sci. Technol. B* **13**(3) p.1089 (1995)

[43] S.Y. Chou, M.S. Wei, P.R. Krauss, P.B. Fischer, "Single-domain magnetic pillar array of 35nm diameter and 65Gbits/in^2 density for ultrahigh density quantum magnetic storage". *J. Appl. Phys.* **76**(10) p.6673 (1994)

[44] Y.Hao, F.J. Castaño, C.A. Ross, B. Vögeli, M.E. Walsh, H.I. Smith, "Magnetization reversal in lithographically patterned sub-200-nm Co particle arrays". *J. Appl. Phys.* **91**(10) p.7989 (2002)

[45] F.E. Luborsky, *J. Appl. Phys.* **32** p.171 (1961)

[46] G.T.A. Huysmans, J.C. Lodder, J. Waki, "Magnetization curling in perpendicular iron particle arrays (alumite media)". *J. Appl. Phys.* **64** p.2016 (1988)

[47] T. Fujiwara, *IEEE Trans. Magn.* **23** p.3125 (1987)

[48] H.I. Smith, *Submicron and Nanometer Structures Technology*. NanoStructures Press, Sudbury MA (1994)

[49] E.C. Stoner, F.R.S. Wolfarth, E.P. Wolfarth, "A mechanism of magnetic hysteresis in heterogeneous alloys". *Philos. Trans. R. Soc. London* **240** p.599 (1948)

[50] C.A. Ross, H.I. Smith, T.Savas, M. Schattenburg, M. Farhoud, M. Hwang, M. Walsh, M.C. Abraham, R.J. Ram, "Fabrication of patterned media for high density magnetic storage". *J. Vac. Sci. Technol. B* **17**(6) p.3168 (1999)

[51] C.A. Ross, S. Haratani, F.J. Castaño, Y. Hao, B. Vögeli, M. Farhoud, M. Walsh, H.I. Smith, "Magnetic Behavior of lithographically patterned particle arrays". *J. Appl. Phys.* **91**(10) p.6848 (2002)

[52] Y.Hao, "Magnetic Properties of Lithographically Patterned Thin Film Magnetic Elements for Magnetic Random Access Memory Applications" Ph.D. Thesis, Massachusetts Institute of Technology (2003)

[53] Y.Hao, M. Walsh, M. Farhoud, C.A. Ross, H.I. Smith, J.Q. Wang, L. Malkinski, "In-plane anisotropy in Arrays of Magnetic Ellipses". *IEEE Trans. Magn.* **36**(5) p.2996 (2000)

[54] J.A. Osborn, "Demagnetization factors of the general ellipsoid". *Phys. Rev.* **67** p.351 (1945)

[55] M. Hwang, M. Farhoud, Y. Hao, M. Walsh, T.A. Savas, H.I. Smith, C.A. Ross, "Major Hysteresis Loop Modeling of Two-Dimensional Arrays of Single Domain Particles". *IEEE Trans. Magn.* **36**(5) p.3173 (2000)

[56] F.J. Castaño, Y.Hao, S. Haratani, C.A. Ross, B. Vögeli, M. Walsh, H.I. Smith, "Magnetic Switching in 100nm Patterned Pseudo Spin Valves". *IEEE Trans. Magn.* **37**(4) p.2073 (2001)

[57] D.C Flanders, H.I Smith, "Surface relief gratings of 3200Å-period - Fabrication and influence of thin-film growth." *J. Vac. Sci. Technol.* **15** p.1001 (1978)

[58] H.I. Smith, M.W. Geis, C.V. Thompson, H.A. Atwater, "Silicon-on-insulator by graphoepitaxy and zone-melting recrystalization of patterned films", *J. Cryst.Growth.* **63** p.527 (1983)

[59] T. Yonehara, H.I. Smith, C.V. Thopson, J.E. Palmer, "Graphoepitaxy of Ge on SiO_2 by Solid-State Surface-Energy-Driven Grain Growth". *Appl. Phys. Lett.* **45** p.631 (1984)

[60] D.C. Flanders, D.C. Shaver, H.I. Smith, "Alignment of Liquid Crystals using sub-micrometer-periodicity gratings". *App. Phys. Lett.* **32** p.597 (1978)

[61] H.V. Känel, J.D. Lister, J. Melngailis, H.I. Smith, "Alignment of Nematic Butoxybenzilidene Octylaniline by Surface Relief Gratings" *Phys Rev. A* **24** p.2713 (1981)

[62] R.D. Peters, X.M. Yang, Q. Wang, J.J. de Pablo P.F. Nealey, "Combining advanced lithography technology and self-assembly of thin-film diblock copolymers to produce templates for nanofabrication". *J. Vac. Sci. Technol. B* **18** p.3530 (2000)

[63] X.M. Yang, R.D. Peters, P.F. Nealey, H.H. Solak, F. Cerrina, "Guided Sel-Assembly of symmetric diblock copolymer films on chemically nanopatterned surfaces" *Macromolecules*, **33** p.9575 (2000)

[64] J.Y. Cheng, C.A. Ross, E.L. Thomas, H.I. Smith, G.J. Vansco, "Templated Self-Assembly of Block Copolymers: Effect of Substrate Topography" *Adv. Mater.* **15**(9) p.1599 (2003)

[65] M. Boltau, S. Walheim, J. Mlynek, G. Kraush, U. Steiner, "Surface-induced structure formation of polymer blends on patterned substrates". *Nature* **391** p.877 (1998)

[66] D.J. Mascaro, "Directed growth of organic semiconducting thin-films for electronic and optoelectronic devices" Ph.D. Thesis, Massachusetts Institute of Technology, (2004)

[67] A. Immomoglu, D.D Awschalom, G. Burkhard, D.P. DiVincenzo, D. Loss, M. Sherwin, A. Small, "Quantum information processing using quantum dots and cavity QED, ". *Phys. Rev. Lett.* **83** p.4204 (1999)

[68] K. Eberl, M.O. Lippinski, Y.M. Manz, W. Winter, N.Y. Jin-Phillip, G.O. Schmidt, "Self Assembling Quantum dots for optoelectronic devices on Si and GaAs". *Physica E* **9** p.164 (2001)

[69] J. Drucker, "Coherent islands and microstructural evolution". *Phys Rev. B* **48** p.18203 (1993)

[70] F.M. Ross, F.K. LeGoues, J. Tersoff, R.M. Tromp, M. Reuter, "In situ transmission electron microscopy observations of the formation of self-assembled Ge islands on Si". *Microsc. Res. Technique* **42** p.281 (1998)

[71] F.M. Ross, "Growth Processes and phase transformations studied by in situ transmission electron microscopy". *IBM J. Res.* **44** p.489 (2000)

[72] G.Jin, J.L. Liu, K.L. Wang. "Regimented placement of self-assembled Ge dots on selectively grown Si mesas". *Appl. Phys. Lett.* **76** p.3591 (2000)

[73] T. Schwartz-Selinger, Y.L Foo, D.G. Cahill, J.E. Greene, "surface mass-transport and island nucleation during growth of Ge on laser textured surfaces". *Phys Rev. B* **65** p.125317-1 (2002)

[74] J.L. Grey, R.Hull, J.A. Floro, "Control of surface morphology through variation of growth rate in SiGe/Si(100) epitaxial films: Nucleation of 'quantum fortresses'". *Appl. Phys. Lett.* **81** p.2445 (2002)

[75] F.M. Ross, M. Kammler, M.E. Walsh, M.C. Reuter "In situ reflection elctron microscopy of Ge island formation on mesa structures". *Microscopy and Microanalysis,* **10** p.104 (2004)

[76] F. LeGoues, M.C. Reuter, J. Tersoff, M. Hammar, R.M. Tromp, "Cyclic growth of strain-relaxed islands", *Phys. Rev. Lett.* **73** p.300 (1994)

[77] M. Hammar, F. LeGoues, J. Tersoff, M.C. Reuter, R.M. Tromp, "In situ ultrahigh vacuum transmission electron microscopy studies of heteroexpetaxial growth. I. Si(001)/Ge". *Surf. Sci.* **349** p.129 (1995)

[78] F.M. Ross, J. Tersoff, R.M. Tromp, "Coarsening of self-assembled Ge quantum dots on Si(100), *Phys. Rev Lett.* **80** p.984 (1998)

[79] T.I. Kamins, R.S. Williams, "Lithographic positioning of self-assembled Ge islands on Si(001)". *Appl. Phys. Lett.* **71** p.1201 (1997)

[80] G. Jin *et al.*, "Controlled arrangement of self-organized Ge islands on patterned Si(001)substrates". *Appl. Phys. Lett.* **75** p.2752 (1999)

[81] H.J. Eisler, V.C. Sundar, M.G. Bawendi, M.E. Walsh, H.I. Smith, V. Klimov, "A color-selective semiconductor nanocrystal laser". *App. Phys. Lett.* **80** (24) p.4614 (2002)

[82] C.B. Murray, C.R. Kagan, M.G. Bawendi, *Annu. Rev. Mat. Sci.*. **30** p.545 (2000)

[83] M.Asada, Y.Miyamoto, Y.Suematsu, "Gain and threshold of three-dimensional quantum-box lasers", *IEEE J. Quantum Electron.* **QE-22** p.1915 (1986)

[84] Y.Arakawa, H.Sakaki, "Multidimensional quantum-well laser and temperature dependence of its threshold current". *Appl. Phys. Lett.* **40** p.939 (1982)

[85] V.I. Klimov, A.A. Mikhailovsky, S. Xu, A. Malko, J.A. Hollingsworth, C.A. Leatherdale, H.J. Eisler, M.G. Bawendi, "Optical gain and stimulated emission in nanocrystal quantum dots". *Science* **290** p.314 (2000)

[86] C.B. Murray, D.J. Norris, M.G. Bawendi, " Synthesis and Characterization of Nearly Monodisperse CdE(E = S, Se, Te) Semiconductor Nanocrystallites". *J. Am. Chem. Soc.* **115** p.8706 (1993)

[87] B.O. Dabbousi, J. Rodriguez-Viejo, F.V. Mikulec, J.R. Heine, H. Mattoussi, R.Ober, K.F. Jensen, M.G. Bawendi, "(CdSe)ZnS Core-Shell Quantum Dots: Synthesis and Characterization of a Size Series of Highly Luminescent Nanocrystallites". *J. Phys. Chem. B* **101** p.9463 (1997)

[88] H. Kogelnik, C.V. Shank, "Coupled-wave theory of distributed feedback lasers". *App. Phys. Lett.* **43** p.2328 (1972)

Chapter 8
Etching Magnetic Thin Films

8.1 Introduction

8.1.1 Reactive-ion etching

In Chapter 7, it was stated that the reactive-ion etching (RIE) of magnetic materials is extremely difficult. Although some effort has been devoted towards reactive etch processes for selected magnetic materials, results are mixed [1,2,3,4]. RIE processes are based upon a chemical reaction between the plasma species of the etchant gas and the substrate, leading to substrate atoms being bound into a volatile compound [5]. A quick, albeit simplistic, indication of whether or not a volatile by-product is possible can be found by examining the boiling-point of compounds of the etchant and substrate [6]. Low-boiling points indicate volatility and hence etching, while high boiling points indicate either that an etch will only occur at high temperatures, or not at all. Table 8.1 lists a few of the simple compounds formed by common etch gasses with cobalt and nickel, and their

respective boiling points. Compounds formed by the more easily etched materials tungsten (W) and silicon (Si) are included for comparison. It is clear from these data that the magnetic materials are not amenable to reactive-ion etching.

Table 8.1

Material	***Etchant***	***Compound***	***Boiling Point °C***
Co	Fluorine	CoF_2	1400
Co	Chlorine	$CoCl_2$	1049
Ni	Chlorine	$NiCl_2$	Subl. 973
W	Fluorine	WF_6	17.5
Si	Fluorine	SiF_4	-86

In the context of patterned-media devices, such as those described in Chapter 7, one must also consider the variety of materials which may need to be etched. In addition to magnetic elements such as cobalt and nickel, alloys such as NiFe, and CoCrPt are also of great interest. Multilayer stacks consisting of very thin (<10 nm) layers can include a variety of magnetic and non-magnetic materials. Examples include Co and NiFe, conductors such as Cu and Ta, and potentially even insulators such as SiO_2 or Al_2O_3. Even assuming an RIE process were available for etching one magnetic material, e.g. NiFe, it is unlikely to etch other layers in the stack. Further, an etch chemistry appropriate for one material may actually damage others in the stack.

The lack of available RIE processes, combined with the broad range of potential materials in a multilayer stack, necessitates that an alternate etch method be used. Wet chemical etching does not afford the necessary process control for deep sub-micron resolution. Although arguably a step backwards technologically, broad ion-beam etching (also known as ion-milling) can provide

both the necessary resolution and the ability to etch *any* material. However, there are some limitations to this technique that must be understood in order to use it successfully.

8.2 Ion-beam etching

8.2.1 Etch Selectivity

Ion-beam etching, also known as ion-milling, is a non-reactive etch process by which material is removed through physical sputtering [7,8]. Sputtering refers to the removal of atoms from the target surface under bombardment by energetic ions [9,10]. If the energy of the incident ions is greater than a certain threshold, ~10 eV [11,12,16], sufficient energy can be coupled to atoms in the surface layers of the target that they will be ejected. The energy efficiency of the process is quite low, with only highly-energized atoms near the surface being ejected. Those atoms deeper in the target will have their energy lost as heat. One estimate indicates that about 98% of the incident ion energy is lost to heating and substrate damage [13]. This phenomenon was first observed by Grove in 1852 [14], although not until more than 100 years later was it first deliberately applied for etching [15]. The theory of sputtering has been well researched over approximately the past 50 years, although the primary application of sputtering has not been etching, but film deposition [16,17,18].

In the context of etching, a couple of problems arise due to the non-reactive mechanism of the etch. First, there is low etch-selectivity between the mask and target. Etch selectivity is defined as the ratio of the etch rate of the target material to the mask material.

$$selectivity = \frac{etch\ rate\ of\ target}{etch\ rate\ of\ mask} \qquad (8.1)$$

The etch rate S of a material, given in Equation 8.3, is proportional to the ion current-density J_{ion}, the sputtering yield Y, and the ratio of atomic mass M_t to density ρ. Sputtering yield is defined as the number of atoms ejected from the target material per incident ion. The sputtering yield is generally on the order of unity, such that etch rates for different materials rarely differ by more than about a factor of 2 [19,20,21,22].

$$Y = \frac{ejected\ atoms}{incident\ ion} \qquad (8.2)$$

$$S = \left(1.04 \times 10^{-8}\right)\frac{J_{ion} M_t Y}{\rho} \qquad (8.3)$$

The etch rates of a selection of relevant materials under standard conditions***** are shown in Table 8.2 [23].

Table 8.2

Material	***Etch Rate Å/min***
Co	550
Ni	660
Fe	530
Cu	1100
Pt	880
Cr	580
Ti	380
W	380
SiO_2	400

Ideally, we want infinite mask selectivity - a mask which is not affected at all during the etch. Reactive-ion etching can enable high selectivity, often of the order

* 500 eV normally-incident argon ions with a current density of 1.0 mA/cm^2

of 10:1 to 100:1. A modification of ion-milling, called reactive-ion-beam etching (RIBE), has been demonstrated in which the sputtering occurs in an atmosphere of reactive gas such as oxygen or chlorine [24,25]. In the case of oxygen, the reaction does not act to increase the etch rate of the target, but to decrease the etch rate of the mask. Oxides of the mask materials such as aluminum can have much lower sputtering rates than the original metal. Also, oxygen is often preferentially sputtered out of the mask, but can be replaced from the atmosphere. Under these conditions, selectivity can increase dramatically. Although this is an attractive technique for certain material systems, the properties of magnetic materials can be altered or destroyed by oxidation, making RIBE an unacceptable option for these materials. In general, the minor differences in etch-rate between materials leads to low etch selectivity.

8.2.2 Faceting

Compounding the difficulty caused by poor selectivity is a phenomenon known as faceting [26,27,28]. Faceting occurs when square corners of a feature become angled and etch at a greater rate than flat regions. Figure 8.1 shows this evolution schematically, and the resultant change in the profile of the etched feature. Due to increased exposed surface-area at exterior corners of a feature, atoms which would otherwise be too deep within the target are permitted to escape. For this reason, sharp exterior corners etch at a greater rate than flat sections, thus beginning the facet. Once initiated, faceting continues because sputtering yield (and etch rate) are strong functions of the incident-angle of bombarding ions [29].

Facets develop at the angle corresponding to the maximum etch-rate and propagate through the mask at this increased rate. Therefore, the actual etch rate of a faceted mask can be much higher than the expected etch rate for normally incident ions. In this context, it is useful to mark the distinction between sputtering yield and etch rate. The sputtering yield determines how many atoms

are dislodged from the target per incident ion. At oblique angles, more energy is transferred to the atoms closer to the surface, and yield increases.

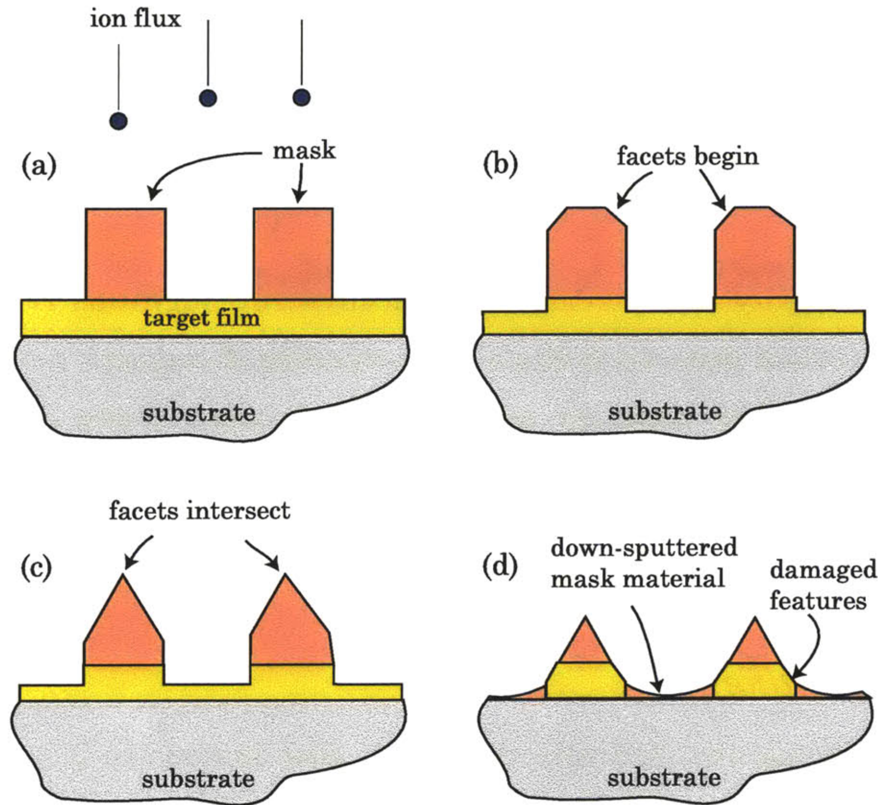


Figure 8.1 Evolution of faceting effects on the mask during ion milling

Etch rate, however, is proportional to both sputtering yield and ion flux. Thus, although the sputtering yield on the facet is increased, the ion flux decreases with the cosine of the angle. The decrease in ion flux is analogous to the decrease of optical power density at oblique angles of incidence, pictured in Figure 7.8.

$$S(\theta) \propto \frac{(J_{ion}\cos(\theta))M_tY(\theta)}{\rho} \qquad (8.4)$$

In conjunction with the redeposition of sputtered target material (discussed in Section 8.2.3), the evolution of the surface contours can be quite complex. This has been studied extensively, and a number of models have been presented [26,28, 30,31]. Etch rates peak at angles between 45° and 60°, and can be a factor of two or more greater than at normal incidence. Even more damaging, the "phase velocity"

of the intersection of the facet with the target, or the intersection of two facets, can cause far more rapid mask erosion than one would expect from just the increased etch-rate of the facet itself. Mask-wear and linewidth reduction are highly accelerated while the etch-rate of the substrate is unchanged.

Although it would be intuitive to increase the mask height to compensate for the increased wear due to faceting, this can be problematic for two reasons. One is redeposition of substrate material on the mask; this will be addressed shortly. Secondly, material sputtered off the mask facets has a high probability of being ejected down towards the substrate, where it inhibits etching of the target material. A tall mask can actually reduce the already low selectivity by depositing unwanted material on the target film.

8.2.3 Redeposition

Material ejected from the surface of the target by normally-incident ions will have an angular distribution as measured from the substrate normal which follows a cosine law [32]. The majority of the material is ejected normal to the surface, with very little ejected parallel to the substrate. The sputtered efflux from off-normal ion bombardment will have a preferential direction, e.g. the down-sputtered mask material described in the previous section. Sputtered material is generally atomic, with energy ranging from a few to a few tens of eV. Although this energy is too low to itself induce sputtering, the efflux is very efficient at coating anything it encounters. This is known as redeposition, and is especially problematic when it occurs on the sidewalls of the mask, illustrated in Figure 8.2.

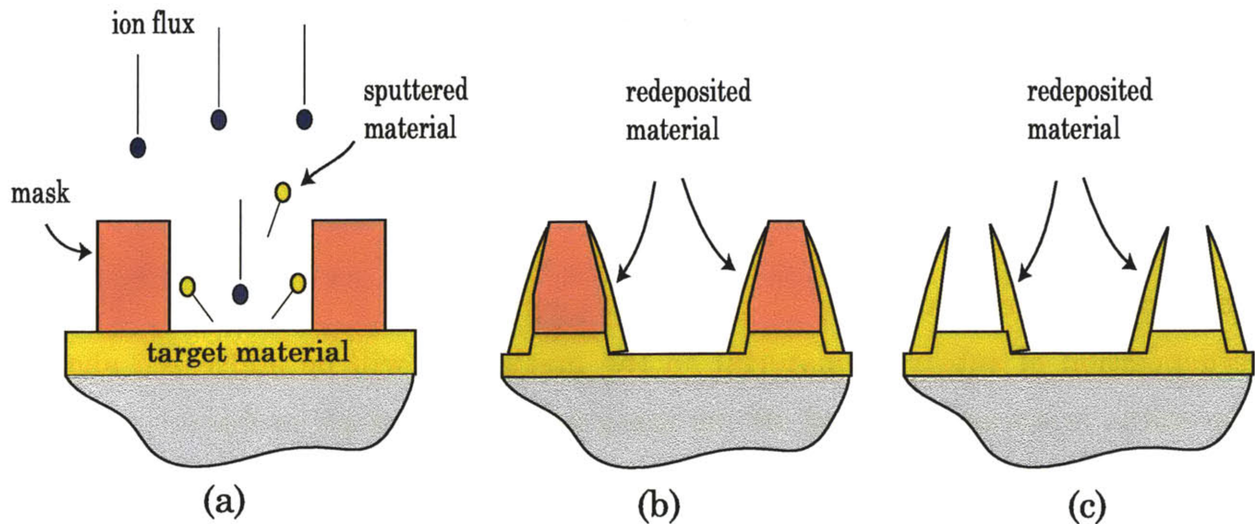


Figure 8.2 Redeposition process. (a) Some sputtered material cannot escape and is deposited on the mask. (b) Redeposit after ecthing. (c) redeposit remains after mask is stripped.

Any etch-mask of non-zero thickness necessarily blocks the escape of some sputtered substrate material. The taller the mask, the more material will be blocked. However, mask height alone is not the appropriate quantity to consider. The solid angle occluded by the mask, determined by the ratio of the mask height to the width of the exposed target, is a better way of gauging the severity of the redeposition problem for small feature sizes. Thus, to reduce redeposition, mask height should be small in comparison to the features. The formation of "wings" or "crowns", seen in Figure 8.2(c), can cause a multitude of problems including loss of linewidth control and detrimental surface topography [33,34]. They can also radically alter the properties of magnetic structures, which are very sensitive to shape irregularities. Multilayers relying on isolation of one layer from another can be ruined by "short-circuits" created by the redeposit. SEM micrographs of dramatic redeposition are seen in Figure 8.3.

8.3 Mask Design for Ion Milling

8.3.1 Introduction

A conflict in mask design arises if the problems of faceting and redeposition are to be addressed simultaneously. A tall mask is desirable to overcome the increased mask erosion from faceting, yet a thin mask is desirable to overcome redeposition. Strategies that have been developed to overcome redeposition come with increased faceting and the resultant narrowing of features. One such technique finds an optimal angle for the incident ions at which the redeposited material is etched away at the same rate it is deposited [35]. An alternate implementation of the same idea begins with a mask that has tapered sidewalls rather than vertical sidewalls. In either case, faceting is enhanced and narrowing of the etched features is observed. Any successful ion milling process must take into account both faceting and redeposition, often trading-off one against the other.

8.3.2 Thick polymeric masks

In the majority of published work pertaining to ion-milling, photoresist was used as the mask material with some success. It should be noted, however, that the scale of most devices was many microns. Problems with faceting and redeposition only become worse for the sub-100nm feature sizes of interest to patterned magnetic media. Early experiments performed by the author for patterning magnetic media with nanometer-scale features utilized the simplest and most familiar masking scheme: the tri-layer resist described in Chapter 7. In retrospect, this process was destined for failure, although it seemed reasonable at the time. Results are presented as a real-world example of the many challenges associated with the ion-milling of nanometer-scale features.

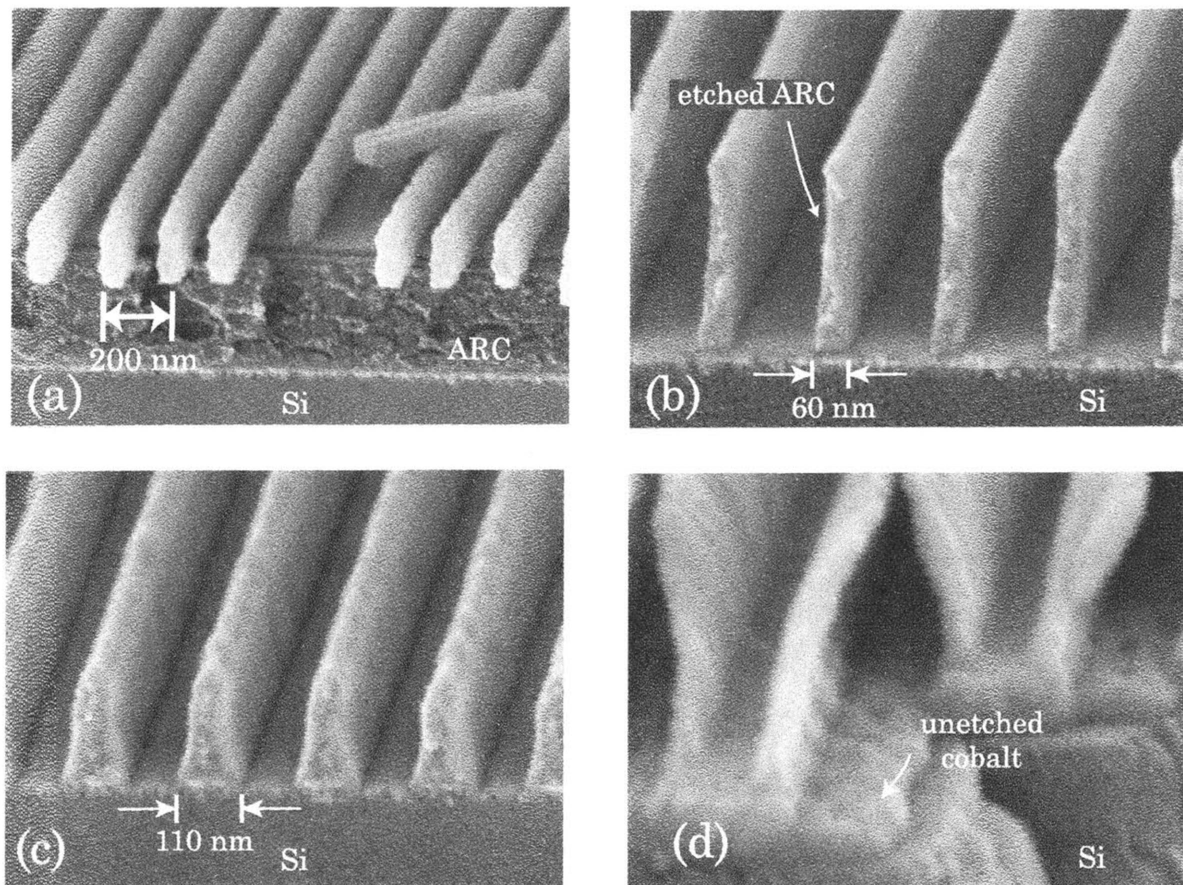


Figure 8.3 SEM micrographs of the process steps for ion-milling a 200nm-period grating into cobalt using an ARC mask. (a) Lines in photoresist produced by IL. (b) ARC mask before ion-milling. (c) Facted mask after ion-milling is twice as wide as it began. (d) Redeposit after ARC strip. Note the incomplete etch of the underlying cobalt.

The resist-stack and process steps are shown in Figure 8.3. Photoresist lines of 200nm-period atop the SiO_2 interlayer, ARC and 20nm of Co are seen in Figure 8.3(a)†††††. Note that the sidewalls of the resist show evidence of a minor standing wave. The resist is entirely removed during the ARC etch, and the SiO_2/ARC combination, shown in Figure 8.3(b), is used as the mask for ion-milling. The ARC mask, however, is not well suited for the task. Facet formation and redeposition of

††††† Resist: 200nm-thick PFI-88 (Sumitomo Chemical Inc.). Interlayer: 20nm evaporated SiO2. ARC: 220nm ARC-XL (Brewer Science). Cobalt: 20nm-thick e-beam evaporated.

target material lead to significantly broadened triangular mask profiles. Beginning at ~60nm, the mask width almost doubles during the etch to ~110nm, seen in Figure 8.3(c). Additionally, down-sputtering of mask material off the facets dramatically slows the etch rate of the cobalt target. In Figure 8.3(d) the etch of the cobalt layer is seen to be incomplete. The etch time, 4 minutes using 500eV Ar+-ions at 0.5mA/cm^2, is significantly longer than would be expected from the etch rate data in Table 8.2, and about twice as long as successful etches described in the following sections. The redeposit is clearly apparent after the ARC has been stripped. The large wing-like structures protruding from the grating lines, likely a mixture of ARC and cobalt, are not easily removed.

It is clear that, although the tri-layer stack is well-tuned for the lithography and RIE steps, ARC proves to be a dismal mask for ion-milling. The failure can be attributed to the high aspect-ratio of the mask which created a situation of heavy redeposit of both mask and target material.

8.3.3 Influence of atomic weight on sputtering yields

To solve the dual problems of faceting and redeposition, we adopt a strategy in which the use of a very thin etch mask is facilitated by an appropriate choice of sputtering gas. A very thin mask is desirable to reduce redeposition, but generally cannot be realized due to poor selectivity and faceting. In reactive-ion etching (RIE), the choice of etch gas greatly affects the selectivity of the etch, and is chosen with regard to both the mask material and the substrate material. In ion milling, a non-reactive process, the sputtering ions are generally argon. It will be shown that when the range of choices for ion-milling is broadened to include all of the noble gasses, the selectivity of the etch can be enhanced by choosing the right gas for a given combination of mask and substrate material [36].

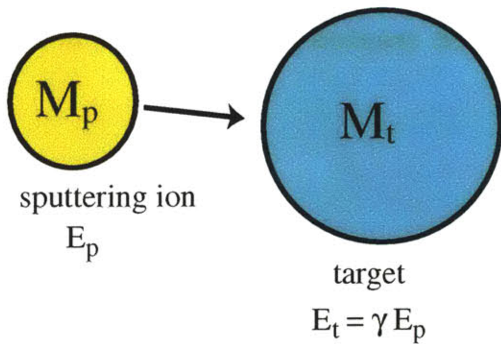
A simple, qualitative model assumes the sputtering yield is proportional to the efficiency of energy transfer in an elastic collision between two atoms. The

energy-transfer parameter γ is given below, where M_p is the mass of the sputtering ion and M_t is the mass of the target atom.

$$Y \propto \gamma \tag{8.5}$$

$$\gamma = \frac{4 \cdot M_t \cdot M_p}{(M_t + M_p)^2} \tag{8.6}$$

As seen in Figure 8.4, the energy-transfer is maximized when M_t=M_p, and decreases for M_t>M_p and M_t<M_p. This suggests that for a given material M_t, we can either increase or decrease the efficiency of the energy transfer through choice of the sputtering ion M_p.



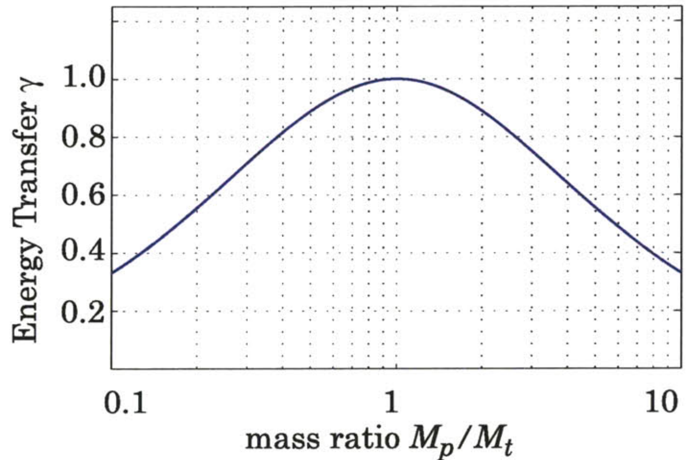


Figure 8.4 Sputtering yield is assumed to be proportional to the energy transfer parameter γ in an elastic collision model. The plot shows γ as a function of the mass ratio of the bombarding ion (M_p) to the target (M_t), which peaks when the two are equal.

In the context of ion milling, etch selectivity can potentially be increased if the sputtering gas is allowed to be something other than argon. Based on our simple model, the ion should be chosen to have a large mass differential with respect to the mask, and a small mass differential with respect to the substrate. That is, the energy transfer between the sputtering gas and the mask should be minimized, while the efficiency for sputtering the substrate should be maximized.

8.3.4 Ion milling using Ne^+ and thin tungsten masks

To implement this strategy we developed a process in which a very thin (15 nm) layer of tungsten (W) is used as a hardmask. Redeposition is effectively eliminated due to the minimized solid-angle occluded by the ultrathin mask. The use of such a thin mask is enabled by a number of factors. First, the sputtering rate of W is among the lowest of all metals. Perhaps more important, W has a very low dependence of sputtering yield on the incident angle [32,29], limiting the development of facets.

The target material is assumed to be cobalt, although multilayer stacks of Co/Cu/Co were also tested. The large difference in the atomic weights, (M_W=184 amu, M_{Co}=60amu), provides an opportunity to exploit the sputtering ion to increase selectivity. A light element such as neon (M_{Ne}=20amu) should have poor energy transfer to the heavy W, coupling energy more efficiently into the more closely matched Co atoms. Figure 8.5 plots the Co/W selectivity for Ne^+, Ar^+, Kr^+, and Xe^+ based tabulated sputtering rate data [23]. As predicted, lighter sputtering ions increase selectivity for this mask/target pair. Normally incident Ne^+ ions at 500eV etch Co 2.3 times more rapidly than they do W. This represents an increase of about 50% over the selectivity when etching with argon of 1.5. It is interesting to note that the relative etch-rate of tungsten increases as its mass becomes more closely matched to that of the sputtering ion. For krypton and xenon ions, the selectivity falls below 1, meaning that tungsten actually etches *faster* than cobalt.

Although the extreme case of sputtering with helium is not included on the plot, it is a good example of the magnitude of the effects possible by simply switching the sputtering gas. The Co/W etch selectivity increases to 14 when sputtering with helium, an order of magnitude greater than when sputtering with argon. Because of the low ionization cross-section of helium and very low sputtering rates, ion-milling with helium is not considered a practical alternative.

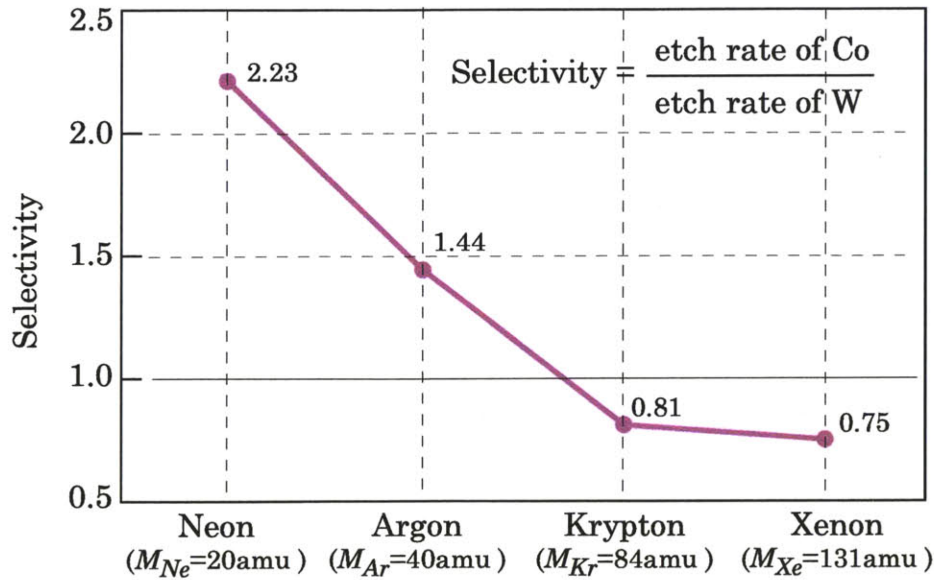


Figure 8.5 Variation in etch selectivity with sputtering gas for a cobalt target with a W mask. Etch rates are for 500 eV normally incident ions at 1mA/cm2 current density [23].

Etch experiments on a 20nm-thick Co layer were conducted to compare the effects of using Ne^+ ions over the more commonly used Ar^+ ions. Patterns of lines and dots were generated using IL in an ultrathin resist (Section 7.1.4) to minimize extra process steps. The complete film stack consisted of 50 nm photoresist, 30 nm SiO_x, 15 nm W, and 15nm Co on a silicon substrate‡‡‡‡‡. The resist pattern was transferred into the oxide layer by RIE with CHF_3, then etched into the tungsten by RIE in CF_4 and O_2. The remaining resist and oxide were removed from the sample, leaving the mask pattern in W atop the Co film. SEM micrographs of these steps are shown in Figure 8.6.

Ion-milling was performed using a Kaufman-type ion source [23]. Transfer of the W patterns into the Co layer was compared for etches using Ar^+ and Ne^+ ions.

‡‡‡‡‡ The resist was Sumitomo PFI-88A2 positive resist, thinned 1:3 in methyl-amyl-ketone. The SiO_2, W and Co films were deposited by e-beam evaporation.

Beam parameters were 1000 eV ion energy, 0.4 mA/cm^2 current-density for the neon etch, and 500 eV, 0.5 mA/cm^2 for the argon etch.

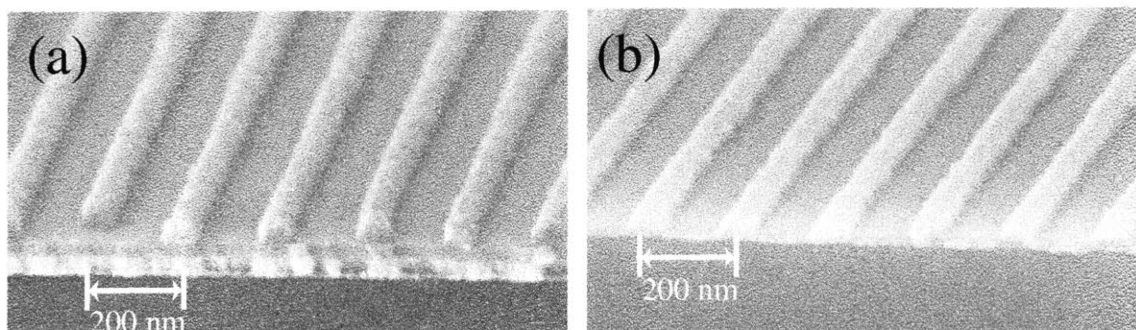


Figure 8.6 (a) 200nm-period lines in a 50nm-thick layer of photoresist before RIE etching. (b) Grating pattern etched in W after removal of residual resist and SiO_2.

Although it would have been preferable to use 500eV for both cases, limitations of our ion-source necessitated that the ion energy be increased to 1000eV to achieve a suitable current density during the neon etch. It is assumed that conducting the neon etch using 1000 eV instead of 500 eV ions increases the etch-rate of both materials without changing selectivity. For the energy range below 500 eV, the dependence of etch-rate on ion-energy can be non-linear, although this energy range was not investigated in these experiments. Simulations of this regime are presented in Section 8.3.6.

A comparison of the resulting linewidths after Ar+ and Ne+ ion etching is shown in Figure 8.7. Etch-time was normalized such that each sample was etched until the 15nm Co layer was etched through. The mask shown in Figure 8.6(b) is shown in plan-view in Figure 8.7(a), and has a linewidth of 75nm. The linewidth after Ne+ milling is unchanged, shown in (b), indicating that little mask wear has occurred. The consequences of lower selectivity under Ar+ sputtering are clear in (c), where the linewidth, reduced by 40%, is now 45 nm. Note that these results are in stark contrast to the prior results shown in Figure 8.3 for Ar^+ etching with an ARC mask, where the lines actually broadened.

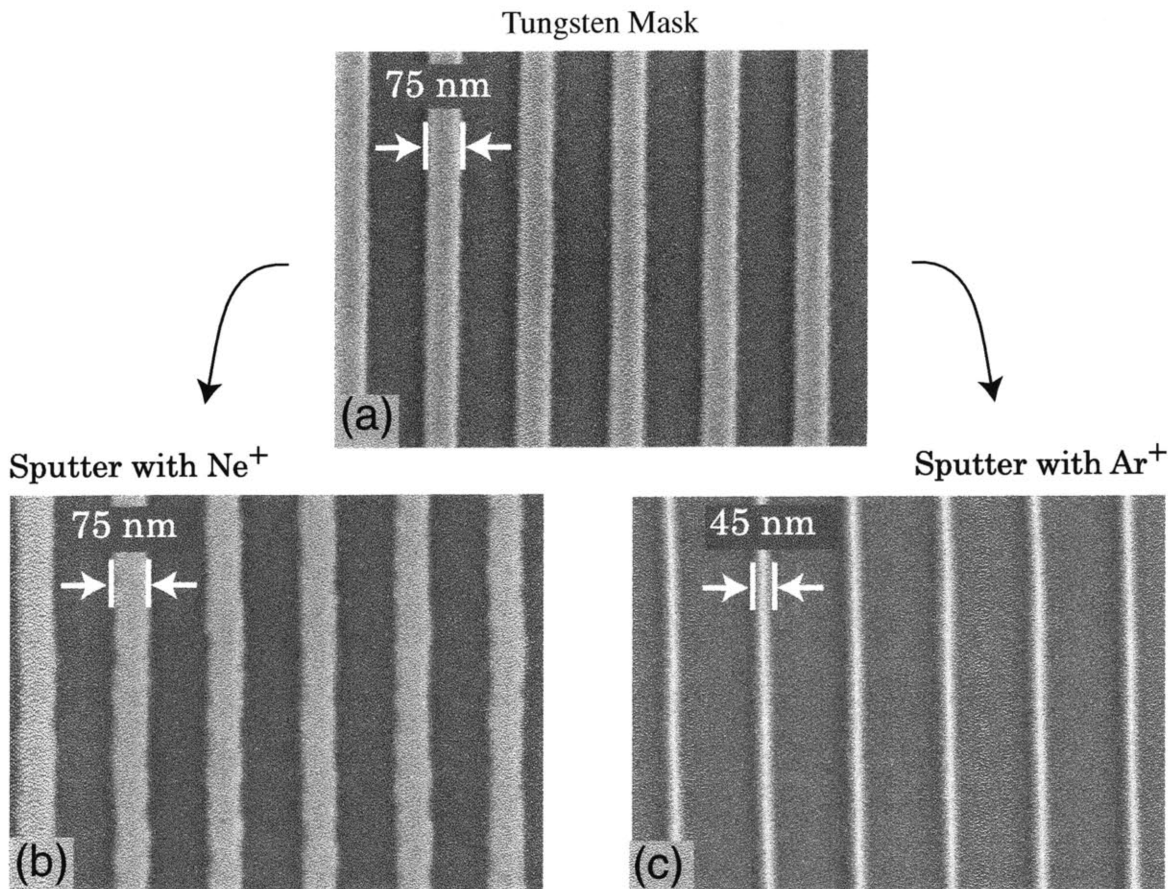


Figure 8.7 A comparison of ion-beam etching with Ar^+ and Ne^+ ions. (a) The W mask before ion-milling (same as Figure 8.6) is shown in plan-view. (b) After etching the Co layer with Ne^+ ions, the linewidth of 75nm is unchanged from the original mask. (c) Etching with Ar^+ ions severely facets the mask leading to rapid erosion and and a 40% reduction in linewith.

The results of Cheng [37] using this process for a similar investigation provide further evidence of the advantage of Ne+ sputtering over Ar+ sputtering for this materials system. In this case, the initial patterning was performed using a self-assembled block-coploymer, which produced an array of circular spots ~35nm in diameter. The saturation magnetization of a 10nm-thick Co film for increasing etch times under Ar+ and Ne+-ion bombardment is shown in Figure 8.8. Saturation magnetization is proportional to the amount of cobalt remaining in the film, and is normalized to the unetched film at 100%. In both cases, the magnetization during

the initial portion of the etch, and hence the volume of Co, is reduced linearly as the unmasked portions of the pattern are removed.

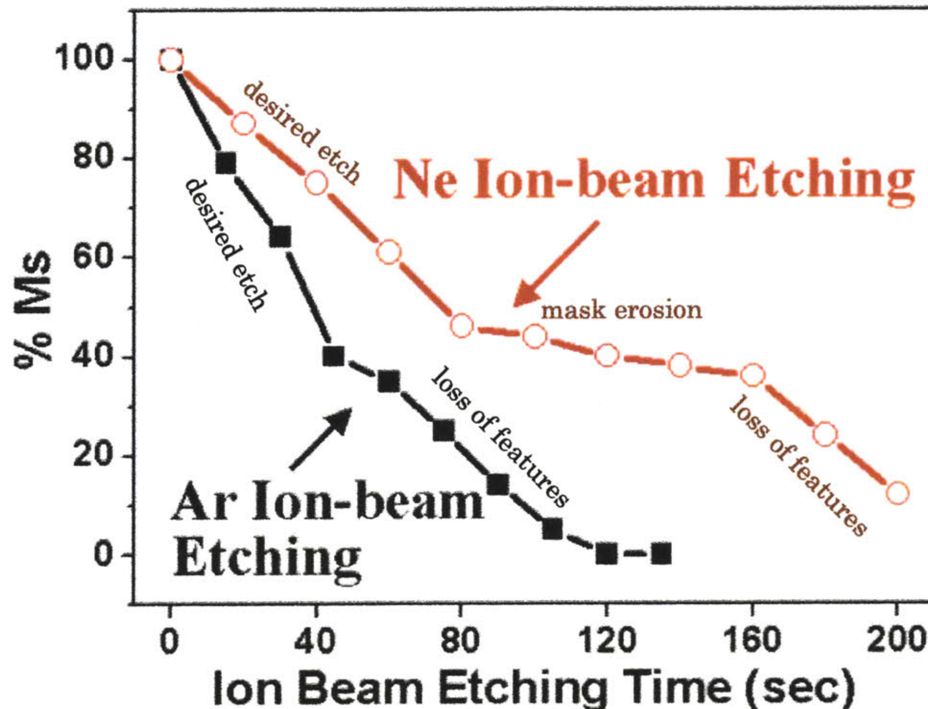


Figure 8.8 The evolution of magnetization of 35nm-diameter Co dots for ion-milling using Ne^+ and Ar^+ ions and a W mask. The Ar+-etched samples show rapid mask wear as evidenced by the minimal plateau of magnetization, which is proportional to Co volume. The wide plateau of magnetization during the Ne+ etch occurs because the size of the features remains approximately the same while the msak prevents continued etching. Once the mask is worn away, magnetization continues to decrease.

When the unmasked portions of the film are completely etched, the magnetization should remain constant if the mask is successfully preventing the further erosion of material. In the Ar^+ etch, shown in black, there is only a slight shoulder in the curve before the magnetization continues to decrease. The Ne^+ etch, however, shows a broad plateau over which the W mask remains intact and approximately the same size. Additionally, note that the magnetization at this plateau is higher for the Ne^+ etch than for the Ar^+ etch. Two conclusions can be drawn from this data. First, the size of the Co particles is smaller for the Ar+ etch, most likely due to faceting of the etch mask. That is, the mask diameter has already

been reduced by the time the Co layer is fully etched. Second, the W mask is almost completely eroded during the early stages of the Ar^+ etch, as evidenced by the continued reduction in magnetization after the etching of the unmasked areas is complete. Mask integrity is seen to be much improved under Ne^+-ion etching, which exhibits a broad magnetization plateau before the mask is compromised and the desired features begin to etch. It is clear that the process window is greatly improved for the Ne+ etch, leading to more uniform and repeatable features.

8.3.5 Ion-milling using Xe^+ and photoresist masks

Consider that most polymer materials, including photoresists, are predominantly composed of carbon and hydrogen§§§§§ such that energy transfer between the bombarding ion and the photoresist will occur due to an ion collision with either H or C. The weight of both these atoms is very low, M_H=1amu and M_C=12amu. Based on the simple model presented in Equations 8.5 and 8.6, it would be expected that sputtering with a heavy ion such as xenon (M_{Xe}=131amu) would provide high selectivity for a comparatively heavy target like Co (M_{Co}=59amu). We demonstrate that thin photoresist can be used alone as an ion-milling mask, without the addition of a tungsten hardmask, if the etch gas is switched from argon to xenon. Although the height of the resist mask makes it more prone to the problems of redeposition, and therefore less desirable than the thinner W mask, it provides a second clear demonstration that an informed choice of the sputtering ion can have a large effect on the quality of the etch.

Figure 8.9(a) shows the resist mask after reactive-ion etching through a 30 nm thick oxide layer on top pf the Co, but before ion milling. After even a partial etch of the Co layer with Ar^+-ions, seen in Figure 8.9(b), the mask has faceted to the point of being practically useless. In contrast, the same sample etched using Xe^+-

§§§§§ Essentially all i-line photoresists, including the Sumitomo PFI-88 positive resist used for these experiments, use a class of phenolic-formaldehyde resins known as novolac.

ions, shown in Figure 8.9(c), can be completely etched with excellent mask integrity. Although not shown, results using the resist mask with a Ne^+ sputtering ion (M_{Ne}=20amu) are characteristically similar to the Ar^+ results, as would be expected due to more efficient momentum transfer to the mask materials. This emphasizes that the best choice of etch gas is very specific to the mask and the material being etched; a given gas will not provide the best results in every case. Sputtering with a light gas such as neon, which was clearly advantageous when using a tungsten mask, is a poor choice when using a resist mask.

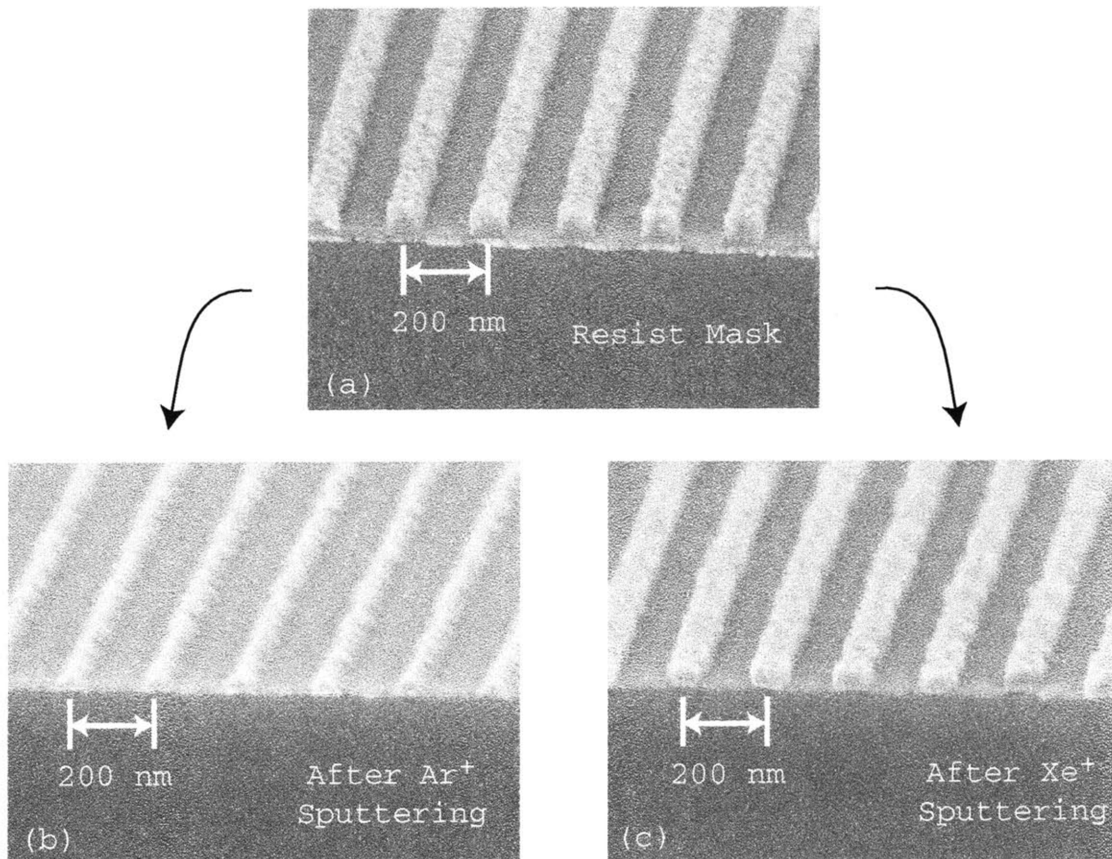


Figure 8.9 (a) Lines in photoresist before ion-milling. (b) After etching with Ar^+ ions, the resist mask is essentially destroyed without successfully etching the underlying film. (c) After etching with Xe^+, the resist mask is well preserved, exhibiting little evidence of faceting.

8.3.6 A theory of low-energy sputtering

The most widely referenced set of sputtering-yield data, due to Wehner *et al.*, actually dates back to the 1960's. In the context of film deposition, they studied the effect of atomic weight of the sputtering-ion on the yield of elemental materials. The etch rate data shown in Table 8.2 was calculated from these yields using the formula given in Equation 8.3. Unfortunately, the scarcity of high-quality vacuum equipment is responsible for some characteristic inaccuracies in Wehner's results, as well as others from that era. The accuracy of data from these early studies has been linked to the reactivity of the target with oxygen, and measured yield values can deviate from more recent UHV results by ~50% [38]. Additionally, measured yield values tend to be limited to a few standard ion-energies, such as 500 or 1000eV. Given these limitations in the available data, a theoretical model to assist in the optimization of mask and ion materials, as well as the sputtering energy, would be extremely valuable.

The most accurate and widely accepted physical theory of sputtering is due to Sigmund [10,39]. Sigmund's model, based on the Boltzmann transport equation, describes a cascade of collisions of the bombarding ion with the target, which is assumed to be a homogenous and random. Atoms are sputtered from the surface of the target if they acquire enough energy from the secondary collisions to overcome their binding energy. At normal incidence, direct collisions with the incident-ion transfer momentum *into* the target, and therefore the momentum *away* from the target necessary for surface material to be ejected must be acquired through secondary or tertiary collisions. The analytical expression for sputtering yield based on Sigmund's theory is given as

$$Y(E) = \frac{3.56}{U}\left(\frac{Z_t Z_p}{\sqrt{Z_t^{2/3} + Z_p^{2/3}}}\right)\left(\frac{M_p}{M_t + M_p}\right)\alpha S_n(\varepsilon), \qquad (8.7)$$

where U is the binding energy (eV), Z and M are the atomic number and mass, respectively, of the projectile and target, α is a dimensionless parameter defined in

Equation 8.10, and $S_n(\varepsilon)$ is the elastic stopping as a function of the reduced energy ε [40]. The reduced energy ε is given as

$$\varepsilon = E_0\left[\frac{1}{32.5}\left(1+\frac{M_p}{M_t}\right)Z_t Z_p\sqrt{Z_t^{2/3}+Z_p^{2/3}}\right]^{-1}, \qquad (8.8)$$

where E_0 is the energy of the incident ions in keV. The nuclear stopping S_n as a function of reduced energy is due to Wilson [41].

$$S_n(\varepsilon) = \frac{0.5\,\ln(1+1.2288\varepsilon)}{\varepsilon + 0.173\sqrt{\varepsilon} + 0.008\varepsilon^{0.1504}} \qquad (8.9)$$

The dimensionless parameter α is derived empirically [42].

$$\alpha = \frac{0.15+0.05(M_t/M_p)}{1+0.05(M_t/M_p)^{1.6}} \qquad (8.10)$$

Using this model, one can study the use of different mask and ion combinations without concern about the quality of published data. Additionally, we can determine the selectivity of the etch as function of ion-energy, for which very little data is actually available. It should be noted, however, that this formulation does not include an angular dependence, and therefore cannot determine what materials will be more or less susceptible to faceting. A more sophisticated adaptation of Sigmunds theory should be able to account for angular effects.

Figure 8.10 plots the etch selectivity as a function of ion energy for the Co/W mask/target system over the energy range from 50eV to 1000eV. Three different sputtering ions are shown: neon, argon, and xenon. The agreement with the selectivity calculated from Wehner's data is excellent, although the values calculated from Table 2 are slightly lower. The selectivity for all case is seen to increase for lower ion energies, with the greatest improvement seen under Ne^+ bombardment. Based on the agreement of the theoretical model with tabulated etch rates, as well as observed results, we can conclude that it can be a useful tool for process design. The intuition given by the simple model in Equation 8.6 is seen to be consistent with the rigorous analytical model in Equation 8.7. In addition to

maximizing the difference in atomic weight between the mask and sputtering ion, the use of low-energy sputtering ions is also seen to improve selectivity.

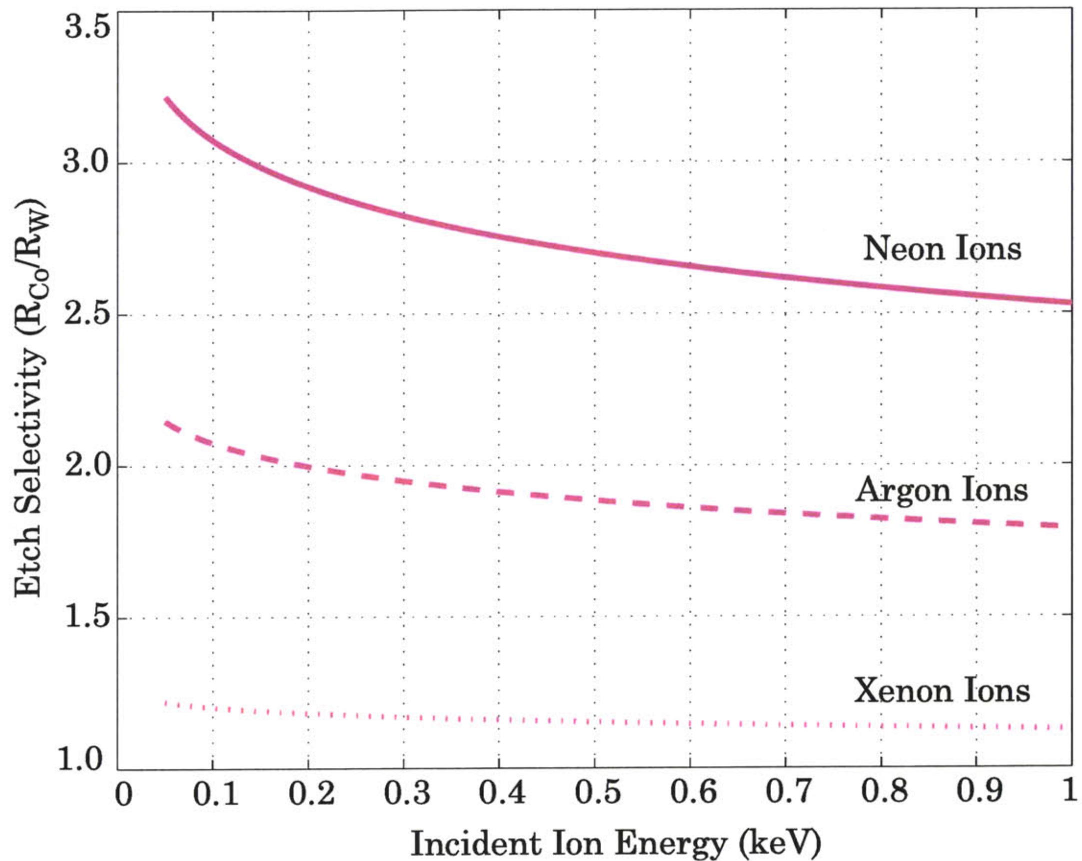


Figure 8.10 Selectivity of Co masked with W, etched using Ne^+,Ar^+ and Xe^+ ions calculated using Sigmund's sputtering theory. Better selectivity is achieved by using a lighter sputtering ions. Low ion-energy improves selectivity for all three cases, although the most improvement is seen for neon.

8.4 References

[1] M. J. Vasile, C.J. Mogab "Chemically assisted sputter etching of permalloy using CO or Cl_2 " *J. Vac. Sci. Technol. A* , **4** p.1841 (1986)

[2] K.B. Jung, H. Cho, K.P. Lee, J. Marburger, F. Sharifi, R.K. Singh, D. Kumar, K.H. Damen, S.J. Pearton, "Development of chemically assisted dry etching methods for magnetic device structures". *J. Vac. Sci. Technol. B* **17**(6) p.3186 (1999)

[3] M.S.P. Andriesse, T. Zijlstra, E. van der Drift, J.B.A. van Zon, "High speed anisotropic dry etching of CoNbZr for next generation magnetic recording" *J. Vac. Sci. Technol. B* **18**(6), (2000)

[4] K. Kinoshita, K. Yamada, H. Matsutera, "Reactive ion etching of Fe-Si-Al for thin-film head", *IEEE Trans. Magn.* **27**(6) p.4888 (1991)

[5] J. Coburn, "Plasma etching and Reactive-Ion etching". *American Vacuum Society Monograph Series*, Ed. N. R. Whetton, Amer. Inst. Phys., New York, (1983)

[6] D.R. Lide ed., CRC Handbook of Chemistry and Physics, 73rd edition. CRC Press, London, 1992

[7] J. Vossen, W. Kern, *Thin Film Processes II*. Academic Press, Boston, (1991)

[8] C.M. Mellar-Smith, "Ion-etching for pattern delineation", *J. Vac. Sci. Technol.* **13**(5), (1976)

[9] G. Carter, J.S. Colligon, *Ion Bombardment of Solids*. Elsevier, New York, (1968)

[10] P. Sigmund, "Theory of Sputtering I. Sputtering Yields of Amorphous and Polycrystalline Targets", *Phys. Rev.* **184**(2) p.383 (1969)

[11] R.V. Stuart, G.K. Wehner, "Sputtering Yields at Very Low Bombarding Energies". *J. Appl. Phys.* **33** p.2345 (1962)

[12] J. Bodhansky, J. Roth, H.L. Bay, "An analytical formula and important parameters of low-energy ion sputtering". *J. App. Phys.* **51** p.2861 (1980)

[13] H.I. Smith, "Ion Beam Etching", Proceedings Symposium on Etching for Pattern Definition, The Electrochemical Society Inc., 133 (1976)

[14] W.R. Grove, "" *Philos. Trans. R. Soc. Lond.* **142** p.87 (1852)

[15] R. Castaing, P. Laborie, *C. R. Acad. Sci.* (Paris) **238** p.1885 (1954)

[16] D.T. Hawkins, "Ion milling (ion beam etching), 1975-1978: A bibliography". *J. Vac. Sci. Technol.* **16**(4) p.1051 (1979)

[17] G.K. Wehner, G.S. Anderson, "The nature of physical sputtering", Handbook of Thin Film Technology, L.I. Maissel and R. Glang Eds. New York, McGraw-Hill, 1970 , Ch.3

[18] P. Sigmund, "Sputtering of single and multiple component materials", *J. Vac. Sci. Technol.* **17** p.396 (1980)

[19] G.K. Wehner et al., "Sputtering Yield Data in the 100-600 eV Energy Range". *General Mills Report 2309*, General Mills Electronic Division, Minneapolis Minn. (1962)

[20] J. Comas, C.B. Cooper, "Sputtering Yields of Several Semiconducting Compounds under Argon Ion Bombardment". *J. App. Phys.* **37** p.2820 (1966)

[21] N. Laegried, G.K. Wehner, "Sputtering Yields of Metals for Ar+ and Ne+ Ions with Energies from 50 to 600 eV". *J. App. Phys.* **32** p.365 (1961)

[22] D. Rosenberg, G.K. Wehner, "Sputtering Yields for Low Energy He+, Kr+ and Xe+ Ion Bombardment". *J. App. Phys.* **33** p.1842 (1962)

[23] H.R Kaufman and R.S. Robinson, *Operation of Broad Beam Sources*, Commonwealth Scientific Corporation, Alexandria VA, (1987)

[24] L.D. Bollinger, "Ion Beam Etching with Reactive Gases [sic]" *Solid State Tech.* (1983)

[25] M. Cantagrel and M. Marchal, "Argon Ion Etching in Reactive Gas," *J. Mater. Sci.*, **8** p.1711 (1973)

[26] H.I. Smith, R.C. Williamson, W.T. Brogan, "Ion-beam etching of reflective array filters". *Proc. Ultrasonics Symp., IEEE.* p.198 (1972)

[27] J.P. Ducommun, M. Cantagrel, and M. Marchal, "Development of a general surface contour by ion erosion. Theory and computer simulation." *J. Mater. Sci.* **9**, p.725, (1974)

[28] O. Auciello, "Ion interaction with solids: Surface texturing, some bulk effects, and their possible applications". *J. Vac. Sci. Technol.* **19** p.841 (1981)

[29] G.K. Wehner, "Influence of the Angle of Incidence of Sputtering Yields", *J. App. Phys.* **30** p.1762 (1959)

[30] R.E. Chapman, "Redeposition: a factor in ion-beam etching topography". *J. Mat. Sci.* **12** p.1125 (1977)

[31] A.R. Neureuther, C.Y. Liu, C.H. Ting, "Modeling ion milling". *J. Vac. Sci. Technol.* **16**(6) p.1767 (1979)

[32] G.K. Wehner, D.L. Rosenberg, "Angular Distribution of Sputtered Material". *J. Appl. Phys.* **31** p.177 (1960)

[33] F.Rousseaux, D. Decanini, F. Carcenac, E. Cambril, M. Ravat, C. Chappert, N. Bardou, B. Bartenlian, P. Veillet, *J. Vac. Sci. Technol. B* **13** p.2787 (1995)

[34] N. Bardou, B. Bartenlian, C. Chappert, R. Megy, P. Veillet, J.P. Renard, F.Rousseaux, M. Ravat, J.P. Jamat, P. Meyer. *J. Appl. Phys.* **79** p.5848 (1996)

[35] P.G. Glöersen, "Ion-beam etching". *J. Vac. Sci. Technol.* **12** p.28 (1975)

[36] M.E. Walsh, Y. Hao, C.A. Ross, H.I. Smith, "Optimization of a lithographic and ion-beam etching process for nanostructuring magneto-resistive thin-film stacks". *J. Vac. Sci. Technol. B* **18**(6) p.3539 (2000)

[37] J. Cheng, "Fabrication and Characterization of Nanostructures from Self-assembled Block Copolymers". Ph.D Thesis, Massachusetts Institute of Technology (2003)

[38] M.P. Sheah, "Pure element sputtering yields using 500-1000 eV argon ions". *Thin Solid Films* **81**(3) p.279 (1981)

[39] P. Sigmund, "Sputtering of single and multiple component materials", *J. Vac. Sci. Technol.* **17**(1) p.396 (1980)

[40] P.Zalm, "Energy dependence of the sputtering yield of silicon bombarded with neon, argon, krypton and xenon ions". *J. Appl. Phys.* **54**(5) p.2660 (1983)

[41] W.D. Wilson, L.G. Haggmark, J.P. Biersack, "Calculations of nuclear stopping, ranges, and straggling in the low-energy region" *Phys. Rev. B* **15**(3) p.2458 (1977)

[42] C. Garcia-Rosales, W. Eckstein, J. Roth, "Revised formulae for sputtering data". *J. Nucl. Mat.* **218** p.8 (1994)

Appendix A
Tabulation of Optical Constants

Accurate knowledge of the optical constants of substrate and thin-film materials is essential for the development of processes for interference lithography. A tabulation of the complex refractive index at a selection of relevant wavelengths, 157nm, 193nm, 325nm, and 351nm, is presented for materials commonly encountered in the MIT NanoStructures Lab. These data have been compiled from a variety of sources, which are indicated as a color code on the table shown on the following page. The sources and their corresponding color code are as follows.

Color	Source
(olive green)	D.R. Lide *ed., Handbook of Chemistry and Physics, 73rd Edition.* CRC Press, Cleveland , (1992)
(pink)	E. Palik, ed., *Handbook of Optical Constants of Solids,* Academic Press, New York, *Volume 1* (1985), *Volume 2* (1991), *Volume 3* (1998)
(light grey)	Data supplied by the manufacturer
(light blue)	Measured using spectroscopic elliposometry at the J.A. Woollam Co.
(light yellow)	Measured at MIT using a Woollam M-2000 spectroscopic ellipsometer

	157 nm 7.9 eV	**193 nm** 6.4 eV	**325 nm** 3.81 eV	**351 nm** 3.53 eV
Aluminum	0.075-1.70i	0.11-2.17i	0.33-3.95i	
Chromium		0.85-1.66i	1.12-2.95i	1.31-3.16i
Copper			1.34-1.81i	
Cobalt		0.94-1.53i	1.30-2.33i	1.35-2.56i
Gold	1.46-1.10i	1.34-1.10i	1.63-1.79i	
Molybdenum		0.79-2.36i	3.04-3.31i	
Nickel		1.01-1.14i	1.69-1.99i	
Palladium		0.73-1.29i		
Platinum		1.42-1.29i	1.53-2.37i	1.60-2.55i
Silicon		0.9-2.25i	4.68-2.03i	5.47-3.0i
Silver	0.94-0.75i	1.27-1.20i	0.56-0.53i	
Titanium		1.08-1.14i	1.68-2.25i	1.30-2.01i
Tantalum		1.55-1.75i	1.77-2.16i	1.85-2.30i
Tungsten		1.32-3.04i	2.99-2.56i	3.24-2.7i
SiO_2	1.68	1.56	1.48	1.48
Al_2O_3		1.79-0.01i	1.80	
CaF_2	1.565	1.50		
LiF	1.49	1.44	1.40	1.40
MgF_2	1.47	1.44(3)		
Ta_2O_5		1.99-1.23i	2.235	2.19
Si3N4	2.68-0.90i	2.60-0.27i	2.06-0.00i	2.04-0.0002i
GaAs		1.5-1.99i	3.50-1.90i	3.53-2.00i
DUO 193		1.8-0.27i	1.51	
XHRi		1.55-0.42i	1.52-0.20i	1.69-0.38i
BS WiDe				

	157 nm 7.9 eV	**193 nm** 6.4 eV	**325 nm** 3.81 eV	**351 nm** 3.53 eV
BARLi	1.498-0.256i	1.595-0.442i	1.55-0.14i	1.58-0.27i
ARC-XL			1.71-0.122i	1.68-0.14i
AR-19		1.79-0.40i		~1.57
Teflon AF			1.369	1.365
FOX HSQ	1.705	1.534	1.436	1.430
SU-8		1.47-0.59i	1.705-0.002	1.67
TARF+py0.25%		1.725-0.047i	1.572-0.021	1.577
TARFP6111		1.72-0.02i	1.56	1.549
PMMA		1.685	1.51	1.51
Shipley 1813			1.75-0.02i	
PFI-34			1.729-0.035i	1.70-0.02i
PFI-88	1.512-0.188i	1.395-0.745	1.79-0.02i	1.72-0.04i
OHKA PS-4	1.494-0.203i	1.395-0.747i	1.681-0.016	1.667-0.016