Multilevel Interference Lithography–Fabricating Sub-wavelength Periodic Nanostructures

by

Chih-Hao Chang

M.S., Mechanical Engineering
Massachusetts Institute of Technology, 2004

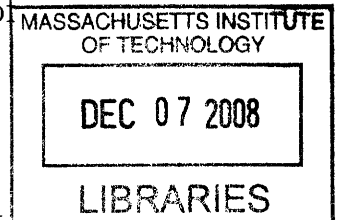
B.S., Mechanical Engineering
Georgia Institute of Technology, 2002

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEC 07 2008
LIBRARIES

September 2008

© Massachusetts Institute of Technology 2008. All rights reserved.

Author ....
Department of Mechanical Engineering
August 21, 2008

Certified by.....................................
Mark L Schattenburg
Senior Research Scientist
MIT Kavli Institute for Astrophysics and Space Research
Thesis Supervisor

Accepted by ....
Lallit Anand
Chairman, Department Committee on Graduate Students

ARCHIVES

Multilevel Interference Lithography–Fabricating Sub-wavelength Periodic Nanostructures

by

Chih-Hao Chang

Submitted to the Department of Mechanical Engineering
on August 21, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Periodic nanostructures have many exciting applications, including high-energy spectroscopy, patterned magnetic media, photonic crystals, and templates for self-assembly. Interference lithography (IL) is an attractive method for fabricating such structures, as it offers several advantages including large exposure area and high spatial-phase coherence. However, the spatial resolution of IL is limited, and the smallest attainable period is roughly half the wavelength of the light used. To overcome this wavelength-limited resolution, we have developed a multilevel interference lithography process that is capable of fabricating sub-wavelength periodic nanostructures over large areas.

In this process, multiple grating levels with different phase-offsets are overlaid and spatial-phase aligned to a common reference grating. Each grating level is pattern-transferred into a single hard mask layer, resulting in spatial-frequency multiplication. To ensure high grating overlay accuracy, each grating level is aligned to the reference grating with various interferometric techniques. In addition, an image-reversal process with plasma etch trimming was developed to control the linewidth of each grating level to nanometer-repeatability. Extensive optical simulations using rigorous coupled-wave analysis were used to examine the intensity distribution of exposures over multilayer periodic structures.

The immediate goal of this work is to extend the wavelength-limited resolution of interference lithography with high precision metrology and well-controlled fabrication processes. Using this multilevel process, we have successfully fabricated 50 nm-period gratings using light with 351.1 nm wavelength. This process presents a general scheme for overlaying periodic nanostructures, and can be used to fabricate more complex 2D and 3D geometries.

Thesis Supervisor: Mark L Schattenburg
Title: Senior Research Scientist
MIT Kavli Institute for Astrophysics and Space Research

Acknowledgments

I want to thank my advisor Dr. Mark Schattenburg for his guidance. Mark is an extraordinary mentor and an excellent role model. His bold visions frequently accomplish what others firmly believe are impossible. The breadth of his knowledge across diverse disciplines is truly amazing. Working with Mark makes six years of graduate research a pleasure.

I want to thank Professor Karl Berggren, Carol Livermore, and George Barbastathis for serving on my committee. Karl has provided invaluable insights and opened new possibilities for the future of this work. I want to thank Carol for her inputs and dedication. I am very touched and grateful. George has taught me all my fundamental knowledge in optics. His lectures, which at times include brilliant juxtaposition of Fourier Optics and Kobe Bryant wearing Michael Jordan's jersey, are absolutely fascinating. I also had the pleasure of answering to George in four of my seven Doctoral Qualifying exams. I want to thank Professor Hank Smith as well, for providing an excellent research environment in Nanostructures Laboratory (NSL) where I spent endless hours.

I want to thank the staff and students of Space Nanotechnology Laboratory for their support and company. Dr. Ralf Heilmann has played an instrumental role in almost all of my work. We have long exciting discussions ranging from x-ray spectroscopy to how to survive 72 hours in a synchrotron facility. Bob Flemming has provided excellent technical support, and trained me tirelessly and patiently on the interference lithography setup, helping me make a seamless transition to the lab. Soon-to-be Dr. Minseung Ahn and Dr. Yong Zhao have been perfect companions the last few years. Working in the lab with them feels like friends chatting in a bar. I also want to thank all the other students, Sarah the bright young physicist from Harvard, Dr. Juan Montoya the marathon runner, Dr. Carl Chen the muscular ramen eater, Dr. Chulmin Joo the future professor, Dr. Paul Konkola the windsurfer, and many others.

I especially want to thank my officemate Dr. Mireille Akilian. We walked in the

door within days of each other on a hot summer day in late August 2002, and six years later we are walking out together. We literally took every step together– going to our first class, giving our first conference presentations, studying for the Qualifying exams, hunting for our first jobs, and eventually defending our dissertations. I can not imagine going through this trying process without her sitting behind me.

I also want to thank all the staff and students of NSL. Jim Daley and Mark Mondol have provided unequal technical support, and I can not count how many favors I owe them. I have had countless fruitful discussions with the present and past students of NSL. I thank especially Dr. Mike Walsh, Dr. Euclid Moon, Dr. Feng Zhang, Dr. Will Arora, Tom O'Reilly, Joel Yang, and Amil Patel.

I thank all my friends at MIT. They added color to my life. I want to thank especially Dr Poe-Jou Chen and Dr Bae-Ian Wu. I have learned tremendously from their wisdom. I also want to thank my basketball buddies, with whom the slogan has changed over the years from "we will beat Lowell next year" to "we will beat BU next year."

I want to thank my family for their love. Since the very beginning my parents gave everything to emphasize on our education. Their sacrifice and support are unconditional, even at a point in time when I was ranked consistently at the bottom of my class. My big sister is the compass of my life, bluntly pointing out my sometimes erroneous ways and guiding me when I am lost. My little sister is the teddy bear of my life, offering sweet innocent comforts that I can warmly embrace.

Lastly I want to thank Peggy, the love of my life. It has been the hardest and most chaotic of times. But she aligned all aspects of my life to their destined positions, so that I may concentrate on this work without having to worry about anything else. She came, stood by me, and never left.

Contents

1 Introduction **23**

1.1 Periodic Nanostructures 25

1.2 Interference Lithography 31

1.3 The MIT Nanoruler 33

1.4 Spatial Resolution of Interference Lithography 36

1.5 Thesis Structure 38

2 Multilevel Interference Lithography **41**

2.1 Multilevel Interference Lithography Process 42

2.2 Theory of Spatial-frequency Multiplication 45

2.3 Summary 51

3 Grating Level Alignment **53**

3.1 Linear Phase Distortions 53

3.2 Optical Configuration 55

3.3 Reference Grating Stability 60

3.4 Grating Alignment Procedure 64

3.5 Summary 65

4 Design of Optical Multilayer **67**

4.1 Design of Reference Grating 68

4.1.1 Reference Grating in Silicon 69

4.1.2 Reference Grating in Silicon Nitride 72

4.2 Design of Multilayer Resist Stack over Homogeneous Layers 79
4.3 Design of Multilayer Resist Stack over Periodic Layers 86
4.3.1 Resist Stack Simulation for the Second Grating Level Exposure 87
4.3.2 Resist Stack Simulation for Third and Fourth Grating Level Exposure 92
4.3.3 Exposure Symmetry 94
4.4 Summary 103

5 Dose and Contrast Simulation 105
5.1 Dose and Contrast Modeling 106
5.1.1 Binary Resist Model 106
5.1.2 Lithography with Nanoruler: Optical Considerations 110
5.1.3 Lithography Using Nanoruler: Mechanical Considerations. . . 121
5.1.4 Dose, Contrast and Phase Calculation 127
5.2 Dose and Contrast Simulation 132
5.3 Summary 142

6 Fabrication Process for Linewidth Control 145
6.1 High Duty-cycle Grating Patterning 146
6.1.1 Negative Resist with SAFIER 147
6.1.2 Lift-off Process 149
6.1.3 Shadow Etch Process 150
6.1.4 Image-reversal Process 151
6.2 Line-width Control with Fabrication Processes 152
6.2.1 Reducing Duty-cycle with PE Trimming 153
6.2.2 Repeatability of PE Trimming 156
6.2.3 Linewidth Control of High Duty-cycle Patterns 159
6.3 Limitations of the Linewidth Control Process 161
6.4 Conclusion 164

7 Results and Discussion **167**

7.1 Phase Overlay of Two Grating Levels 168

7.2 Two-fold Spatial-frequency Multiplication 172

7.3 Higher-factor Spatial-frequency Multiplication 182

7.4 Future Work 186

7.5 Summary 189

8 Conclusion **191**

A Fabrication Process for Multilevel Interference Lithography **195**

A.1 Fabrication Process for Reference Grating 195

A.2 Fabrication Process of Single Grating Level 197

B Rigorous Coupled-Wave Analysis **201**

B.1 Optical Fields in Homogenous Layers 202

B.2 Optical Fields in Periodic Layers 203

B.3 Multilevel Matrix Methods 207

C Design of a Double-pass Acousto-optic Modulator **211**

C.1 Acousto-optic Effect 211

C.2 Double-pass Configuration 216

C.3 Spectral Leakage 219

List of Figures

1-1 A 1D structure that is periodic in the x-direction with period Λ. . . . 25

1-2 Applications for periodic nanostructures include (a) 2D patterned magnetic media [59], (b) 1D template for self-assembly [19], and 3D photonic crystals [58]. 26

1-3 A blazed x-ray diffraction grating (a) anisotropically etched in silicon, and (b) replicated in polymer. 27

1-4 A free-standing critical angle transmission grating for x-ray diffraction. The (a) top-view, (b) bottom-view, and (c) cross-sectional SEM images. 28

1-5 The concept of interference lithography. 31

1-6 Interference of (a) spherical waves results in hyperbolic grating phase and (b) plane waves results in grating with linear phase and non-linear phase distortion. 33

1-7 The concept of the Nanoruler, in which a substrate is scanned under a grating pattern. 34

1-8 The beam overlap scheme of the Nanoruler. 34

1-9 A 300 mm substrate covered with 400 nm-period grating patterned in the Nanoruler. 35

2-1 The alignment scheme for (a) multilevel optical lithography is similar to (b) multilevel IL. 42

2-2 A grating interferometer used to align the exposure by measuring the phase of the reference grating. The incident beams are aligned in a Littrow mount. 43

2-3 Fabrication process for the reference grating using (a) a bilevel resist stack. (b) The grating is exposed, and (c) pattern transfered into silicon nitride. 44
2-4 Fabrication process for spatial-frequency multiplication. Starting with (a) a bilevel resist stack on a substrate patterned with a reference grating, (b) the first grating level is exposed after aligning to the reference grating, and (c) pattern transfered into silicon nitride. (d) After spin-coating ARC and photoresist, (e) the second grating level is exposed with a π phase-offset. (f) The grating is pattern-transfered, resulting in a nitride grating with period of $\Lambda/2$. 44
2-5 Second grating level exposed with (a) perfect alignment and linewidth control. The cases of (b) overlay alignment error of d_{error} and (b) linewidth variation of Δw. 46
2-6 Normalized profile of (a) a periodic square grating and (b) two overlapped square gratings. 47
2-7 Simulated normalized fundamental spatial-frequency coefficient as functions of (a) phase overlay errors and (b) linewidth variation. 50

3-1 The second grating exposure with (a) period error and (b) angle error. 54
3-2 The heterodyne phase detection configuration, also called the 'reading' mode, as implemented in the MIT Nanoruler. 56
3-3 The homodyne phase detection and grating exposure configuration, also called the "writing" mode, as implemented in the MIT Nanoruler. 58
3-4 An unknown constant phase offset between the phase of the physical reference grating and sinusoidal exposure pattern. 59
3-5 The phase map of the reference grating in nm (a) before and (b) after the processing for a single grating level. Note possible particle-induced phase distortion due to vacuum chuck. 61
3-6 A particle is caught between the substrate and the vacuum chuck, inducing in-plane grating phase distortion. 62

3-7 The phase map of the reference grating distorted by a particle in nm (a) before and (b) after the processing for a single grating level. Note the particle-induced phase distortion has the opposite sign as the map shown in Figure 3-6(b). 63

4-1 The grating and incident beam geometry for the silicon reference grating. 69

4-2 The diffraction efficiency of the silicon reference grating as a function of etch silicon depth t for grating duty-cycles of 0.3, 0.4, and 0.5. . . . 70

4-3 The diffraction efficiency of the silicon reference grating as a function of nitride thickness t for grating duty-cycles of 0.1 and 0.2. 71

4-4 The grating and incident beam geometry for the silicon nitride reference grating. 73

4-5 The diffraction efficiency of the silicon nitride reference grating as a function of nitride thickness t for grating duty-cycles of 0.3, 0.4, and 0.5. 74

4-6 The grating and incident beam geometry for the 40 nm-thick nitride reference grating, with additional etch into silicon. 75

4-7 The diffraction efficiency of the 40 nm-thick nitride reference grating as a function of etch depth t into silicon for grating duty-cycles of 0.3, 0.4, and 0.5. 76

4-8 The diffraction efficiency of the 45 nm-thick nitride reference grating as a function of etch depth t into silicon for grating duty-cycles of 0.3, 0.4, and 0.5. 77

4-9 Cross-section SEM image of a fabricated silicon nitride reference grating. 78

4-10 Geometry of the homogenous multilevel stack. 81

4-11 Geometry of the homogenous multilevel stack with 220 nm resist, ARC thickness t, and 45 nm nitride on top of a silicon substrate. 82

4-12 Simulated reflection efficiency at the resist/ARC interface. 83

4-13 Simulated (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride layers. 85

4-14 Cross-section SEM image of an exposed grating in resist. 86

4-15 Geometry of the homogenous multilevel stack with 220 nm resist, ARC thickness t, and 45 nm nitride on top of a silicon substrate. 87

4-16 Simulated zeroth and first-order diffraction efficiencies at the resist/ARC interface. 88

4-17 Simulated zeroth and first-order diffraction efficiencies for a silicon nitride grating with duty-cycle of 0.875. 89

4-18 Simulated second exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers. 91

4-19 (a) Simulated zeroth and first-order diffraction efficiencies, and (b) stack geometry for the first through fourth interference lithography exposures. 92

4-20 Simulated third level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers. 95

4-21 Simulated fourth level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers. 96

4-22 Second to fourth grating level lithography for (a) order with all symmetric exposures and (b) order with some asymmetric exposures. . . 97

4-23 Simulated phase of the zeroth and first-order diffraction from reference grating consisting of 45 nm of nitride grating over silicon substrate. . 98

4-24 Simulated asymmetric second level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 100 nm ARC, and 45 nm nitride grating layers. 100

4-25 Lateral shifts of calculated resist centroid as a function of resist depth. 101

4-26 Simulated asymmetric second level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, and 45 nm nitride grating layers. 102

4-27 Lateral shifts of calculated resist centroid as a function of resist depth. 103

5-1 The binary photoresist model for exposure and development. Only the resist above the clearing dose is exposed. Linewidth can be controlled by increasing or decreasing the exposure dose. 107

5-2 The effects of contrast degradation for constant average dose. The yellow and light blue regions depict the linewidth of perfect and reduced contrast, respectively. 108

5-3 Duty-cycle with different interference fringe contrast as functions of overexposure in a binary resist model for, (a) positive and, (b) negative resist. 110

5-4 A Gaussian beam focused by a lens with focal length of f. The beam radius $w(z)$ and wavefront radius of curvature $R(z)$ both depend on propagation distance z. The wavefront is planar at the beam waist, shown here at $z = 0$. 111

5-5 The interference lithography setup in the Nanoruler. Two Gaussian beams interfere at the beam waists. 112

5-6 The fluorescent imaging system used to measure the intensity profiles of the beam on the wafer plane. 114

5-7 (a) The intensity maps of the left and right beams, and (b) the sum of the maps. 115

5-8 The calculated contrast map of the interference fringes. 116

5-9 The image contrast as the left beam is numerically displaced by Δx and Δy. 118

5-10 (a) The intensity maps of the numerically shifted left and right beams, and (b) the sum of the maps. 118

5-11 The calculated contrast map of the "numerically aligned" interference fringes. 119

5-12 (a) The intensity maps of the left and right beams for incident angles of 61.37^o, aligned for 200 nm-period grating exposure. 120

5-13 The phase-shifting interferometer used to measure the non-linear phase of the image grating. 121
5-14 The phase map for a 200 nm-period exposure, obtained using the PSI scheme. 122
5-15 The raster scan method for parallel writing. 123
5-16 The coordinate frame with y and x as the scan and step-over directions, respectively. The scan angle error is defined as θ_{scan}. 125
5-17 The coordinate frame with y and x as the scan and step-over directions, respectively. The period error is defined as $\delta\lambda$. 126
5-18 The desired input and output parameters of the simulation program. 128
5-19 The dose and contrast of a single scanned exposure for 200 nm-period grating. 130
5-20 The exposure dose and contrast profile of 200 nm period grating with a step-over of 0.45 mm. The dose and contrast variations are due to the non-Gaussian profile of the beams. 133
5-21 The exposure dose and contrast profile of a 200 nm period grating with a step-over of 0.45 mm. The scan angle and period measurement errors are 5 μrad and 0.005 nm, respectively. The dose and contrast variations are due to the non-Gaussian profile of the beams. 134
5-22 The dose and contrast of a single scanned exposure for a 200 nm-period grating with 10 % relative background intensity. 136
5-23 The dose and contrast profile for a 200 nm-period grating with 10 % relative background intensity. 137
5-24 The dose and contrast profile for a 200 nm-period grating with 100 μm depth variation. 139
5-25 The dose and contrast profile for a 200 nm-period grating with 5 μrad of scan angle error, 0.005 nm of period measurement error, 10 % relative intensity Gaussian background exposure, and 100 μm of depth variation. 140
5-26 The experimental contrast as a function of clearing dose, plotted on top of the binary model with a contrast of 0.68. 141

6-1 The SAFIER process to reduce the (a) initial spacing of a trench by, (b) coating the SAFIER polymer and (c) going through a thermal cycle. The reduced trench spacing is shown in (d). 147

6-2 The narrowing of e-beam resist trench spacing from 70 to 41 nm after three cycles at 120° C. 148

6-3 The lift-off process to reverse the tone of (a) the developed positive resist by (b) evaporating a metal, and then (c) dissolve the resist. . . 149

6-4 The shadow etch process starts with (a) a low duty-cycle profile in positive resist, and then (b) deposited with metal at an angle. (c) The shadowed narrow trench is then transfered into the substrate, and (d) the resist and metal are striped. 150

6-5 The image-reversal process starts with (a) a low duty-cycle grating profile in resist, which is then transfered into ARC. (b) A silicon-containing polymer is spin coated to planarize the ARC, which is (c) exposed by an etch back step. (d) The ARC is then etched away, and (e) the high duty-cycle pattern is transfered into the substrate. 151

6-6 The plasma trimming process starts with (a) a grating profile in positive resist. (b) The pattern is transfered into ARC. Using a high pressure oxygen PE, (c) the grating linewidth is reduced. 154

6-7 Cross-section SEM of (a) ARC grating with 45 nm linewidth. (b) The linewidth is reduced to 24 nm after 90 s PE trimming. 155

6-8 Linewidth of a 200 nm period ARC grating vs PE etch time. 157

6-9 Etched width of the ARC grating for three different trials. The difference in etch rate can be observed. 158

6-10 Process flow diagram for fabricating high duty-cycle grating profile with nm-repeatability. 160

6-11 The linewidth is being (a)-(c) PE trimmed, and (d) image-reversed. . 161

6-12 Process flow diagram for fabricating high duty-cycle grating profile with nm-repeatability. 162

6-13 (a) The linewidth of the grating is (b) increased by the image-reversal process. 164

7-1 (a) Top-view SEM image measuring the overlay accuracy of the first and second grating levels, denoted as "A" and "B," respectively. (b) Illustration of the cross-sectional profile of the multilayer stack. . . . 168

7-2 A 2D array of top-view SEM images spaced 2 mm apart to measure the phase overlay of two grating levels over large area. 170

7-3 Phase map of overlay error in nanometers for two 200 nm-period grating levels. No systematic distortion is evident, however there is a local particle induced phase distortion near position (16,2). 171

7-4 Top-view micrograph of two overlaid 200 nm-period grating levels to achieve a two-fold spatial-frequency multiplication, resulting in a grating with 100 nm period. 173

7-5 A top-view micrograph is used to extract (a) a single averaged linescan, and (b) the power spectral density. 175

7-6 Relative power map for (a) the two-fold multiplied and (b) fundamental spatial frequency. The effects of the particle can be readily detected. . 176

7-7 Power ratio map of the fundamental frequency and two-fold multiplied components. 177

7-8 Power ratio map of the fundamental frequency and two-fold multiplied components with depth variations removed. 178

7-9 The 100 mm substrate is immersed in water and illuminated with white light to show diffraction from the reference period region. The device region has subwavelength grating and does not diffract. 180

7-10 Zoomed-in pictures of the substrate with white light illuminating on device regions with (a) no phase distortion, and (b) local phase distortion induced by particle contamination. In the pictures the gamma of blue is increased to highlight the distortion. 181

7-11 Top-view micrograph overlaying three 200 nm-period grating levels to achieve a "fence-type" grating structure. 183

7-12 (a) Top-view micrograph overlaying four 200 nm-period grating levels to achieve a four-fold spatial-frequency multiplication, resulting in grating with 50 nm period. (b) The 2D spectrum of the 50 nm-period grating. 185

7-13 The multiple grating vectors and exposure patterns for (a) a rectangular pattern and (b) hexagonal pattern. 187

7-14 The 2D intensity profiles for three grating exposures, with the third grating having phase-offset of (a) 0, (b) $\pi/2$, and (c) π. 188

7-15 The simulated exposure contrast as a function of the phase-offset of the third grating exposure. 189

B-1 Geometry of the multilevel homogenous and periodic stack. 202

B-2 Permittivity profile in the periodic layer. 204

B-3 Permittivity profile for multiple grating levels in the periodic layer. . 207

C-1 Acousto-optic interaction in a crystal medium diffracts and frequency-shifts an incident beam. 213

C-2 Reference axis for (a) longitudinal wave, and (b) shear wave. The optical axis is in the $+z$-direction. 214

C-3 Optical setup of the double-pass AOM in shear mode. 217

C-4 Measured system efficiency of the double-pass AOM in shear mode as a function of input acoustic frequency. 219

C-5 Experimental setup to measure the magnitude of spectral leakage components. 220

C-6 Semi-logarithmic plot of measured intensity spectra with nominal acoustic frequency input of 105 MHz and Δf of 20 kHz for (a) a double-pass AOM, and (b) with the beam path after the AOMs blocked. 222

C-7 Relative intensity of the double-pass signal and single-pass leakage components over the FWHM band of the double-pass AOM. The signal intensity is scaled by 10^{-3}. 224

List of Tables

1.1 Major advantages and disadvantages of different fabricating methods for periodic nanostructures. 29

1.2 Roadmap of spatial resolution due for various factors. 37

4.1 Experimentally measured etch rates in nanometer per min for O_2, CF_4, and CHF_3 RIE. 80

Chapter 1

Introduction

Over the past few decades, the field of nanotechnology has been generating enormous interest from all areas of research. By building structures over the length scales of 10-100 times the size of individual atoms, nanoengineered materials behave significantly different from their respective bulk properties. Such ability opens new fields and enables exciting applications such as quantum computers, microscale medical analytical systems, and direct manipulation of photons. It is with such future promises and implications that scientists and engineers alike converge on the field of nanotechnology.

Historically, the field of nanotechnology was first introduced by Nobel Laureate Professor Richard P. Feynman in his famous lecture 'There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics [28]' on December 29th 1959 at the annual meeting of the American Physical Society at the California Institute of Technology. In this lecture Feynman considered the manipulation of individual atoms, and metaphorically called to 'write the entire 24 volumes of the Encyclopedia Brittanica on the head of a pin.' Some of the possible applications that Feynman envisioned, such as computer circuits with minimum feature sizes in the order 10s of nanometers, have already become a reality, while other predictions, such as micro-robots as remote surgerical tools, are actively being pursued.

Nanotechnology has the potential for revolutionary impact, however, many obstacles need to be overcome. One of the main challenges remains how to effectively make devices, or even machines, at the nanometer level. To tackle this problem two

general approaches are taken. The first is the 'bottom-up' approach, proposed as the solution by molecular chemists, biochemical engineers, and material scientists. In this approach individual nanoscale components or 'building blocks' are assembled to build a complex structure [68]-[69]. The assembly is based on intermolecular dynamics and other guiding mechanisms. This approach starts at the nanoscale and move towards the macroscale.

The second method is the 'top-down' approach, which is favored by applied physicists, electrical, and mechanical engineers. In this approach the nanostructures are constructed by a pattern-generating lithographic system, without any intermolecular interaction. Within this approach, there are different methods to generate the pattern, such as photons in optical lithography and electrons in electron-beam (e-beam) lithography. This approach starts at the macroscale and moves towards the nanoscale.

The main advantage of the this approach, often referred to as nanolithography, is the mature infrastructure available from the semiconductor industry. For decades microelectronics have been manufactured with optical projection lithography, in which a master mask pattern fabricated with e-beam lithography is replicated. Within the semiconductor industry, progress towards smaller electronics is driven by Moore's Law [29], which is described in a paper written by Gordon E. Moore in 1965. Interestingly, Moore's Law started out as a prediction that the number of transistor in a microelectronic device will double every two years, but ended up being a strict guideline the industry must follow. Currently, feature sizes as small as 45 nm are routinely being manufactured.

While these two general methods differ in the paths taken, they both solve a common problem. In fact they are quite complimentary, as the 'bottom-up' approach fails to self-assemble over long distances while the 'top-down' approach struggles at length scales below 10s of nanometers. Many envision that a powerful nanotechnology enabler will rely on a combination of these two methods. In this thesis, I will deal primary with 'top-down' approaches using advanced optical lithography.

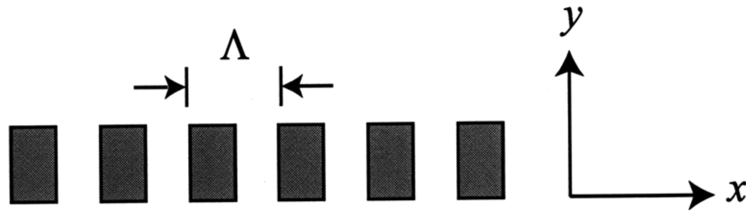


Figure 1-1: A 1D structure that is periodic in the x-direction with period Λ.

1.1 Periodic Nanostructures

Nanostructures are simply structures with geometric length scales of 1-100 nm. The focus is on the adjective 'nano,' since such structures have similar length scales to the forces they interact with. As a result both material and geometric properties need to be considered, and often nanostructures give rise to effective properties that are dramatically different, and often times unnatural in bulk material.

Periodic nanostructures are a class of nanostructures that have discrete translational symmetry. It describes a geometry which the material properties repeat in space over a period Λ, and for a one-dimensional (1D) case satisfies the relationship

$$f(x) = f(x + n\Lambda), \tag{1.1}$$

where $f(x)$ is a material property of interest and n is an integer. An example of a structure that is periodic in the x-direction is shown in Figure 1-1. The interesting property of such a structure is that it can be expressed as a sum of Fourier series given by

$$f(x) = \sum_i F_i \exp(jG_x ix), \tag{1.2}$$

where F_i is the amplitude of the $i - th$ Fourier component term, and $G_x = \frac{2\pi}{\Lambda}$ is also known as the reciprocal lattice vector. Therefore a periodic structure contains only spatial-frequencies defined by Λ and its harmonics. Such a property give rises to interesting phenomenon, for example a periodic potential in a crystal lattice gives

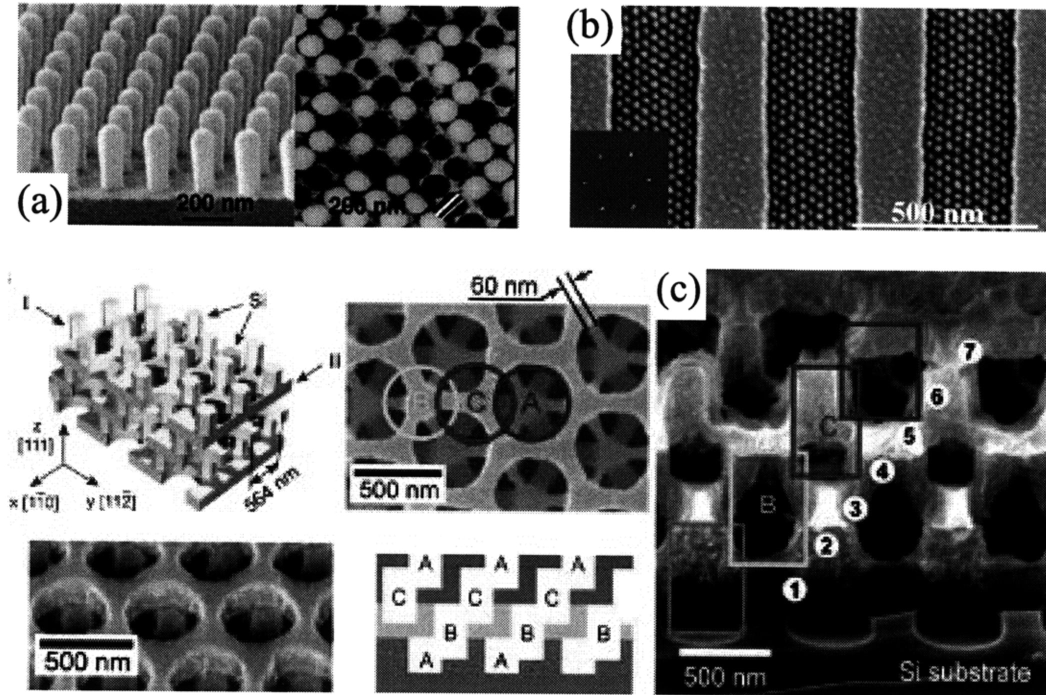


Figure 1-2: Applications for periodic nanostructures include (a) 2D patterned magnetic media [59], (b) 1D template for self-assembly [19], and 3D photonic crystals [58].

rise to energy bandgaps for electrons [40]. Eq. 1.2 describes the case for 1D periodic structures, but can be generalized to describe structures periodic in 3D.

Periodic nanostructures have many exciting applications. Figure 1-2(a) illustrates a structure periodic in 2D that can be used as patterned magnetic media [59] for storage memory. A 1D periodic template structure can be used to assist in self-assembly [19], as shown in Figure 1-2(b). Finally, a photonic crystal periodic in 3D [58] combines material and geometric properties to induce an energy bandgap for light [70] [71] [37], shown in Figure 1-2(c).

The primary motivation of our work is high energy spectroscopy [45] [62] [31] [39] for space instrumentation. For such an application a 1D periodic structure, known as a diffraction grating in optics, is used. In the past we fabricated high fidelity blazed diffraction gratings for high efficiency x-ray spectroscopy [10], shown in Figure 1-3. These gratings were designed for NASA's next generation x-ray telescope

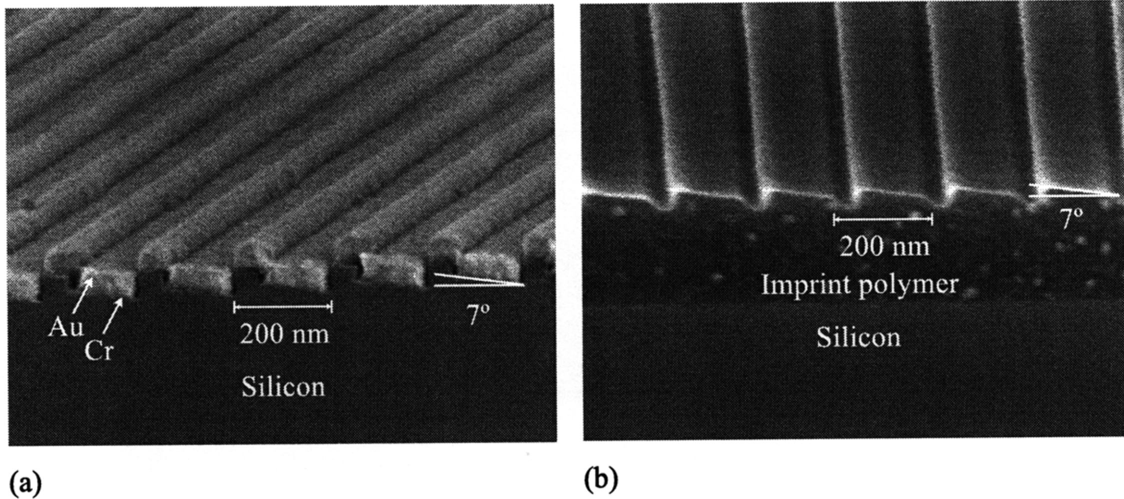


Figure 1-3: A blazed x-ray diffraction grating (a) anisotropically etched in silicon, and (b) replicated in polymer.

Constellation-X.

The 200 nm-period grating structure was patterned with optical lithography and anisotropically etched into silicon, resulting in a 7.5° blaze as shown in Figure 1-3(a). The grating is then coated with chromium and gold to increase x-ray reflectivity. The blazed grating can also be used as a master template, which can be replicated using a thermal nanoimprint process, yielding a polymer grating as shown in Figure 1-3(b).

Recently, our research group has also developed a new kind of blazed grating that consists of arrays of nanoscale mirrors. These gratings, named critical angle transmission (CAT) gratings [1] [2], are shown in Figure 1-4. The top-view and bottom-view micrographs are depicted in diagrams (a) and (b), demonstrating a linewidth variation of only 10 nm after etching through 4 μm silicon with room-temperature potassium hydroxide (KOH). The cross-sectional SEM image illustrating the high aspect-ratio profile is shown in diagram (c). These type of grating combines traditional advantages of tranmission gratings for ease of alignment, and blazing effects for a blazed reflection grating [33].

These x-ray diffraction gratings play an important role in our ability to probe into deep space and move forward in answering some of the most fundamental, sometimes

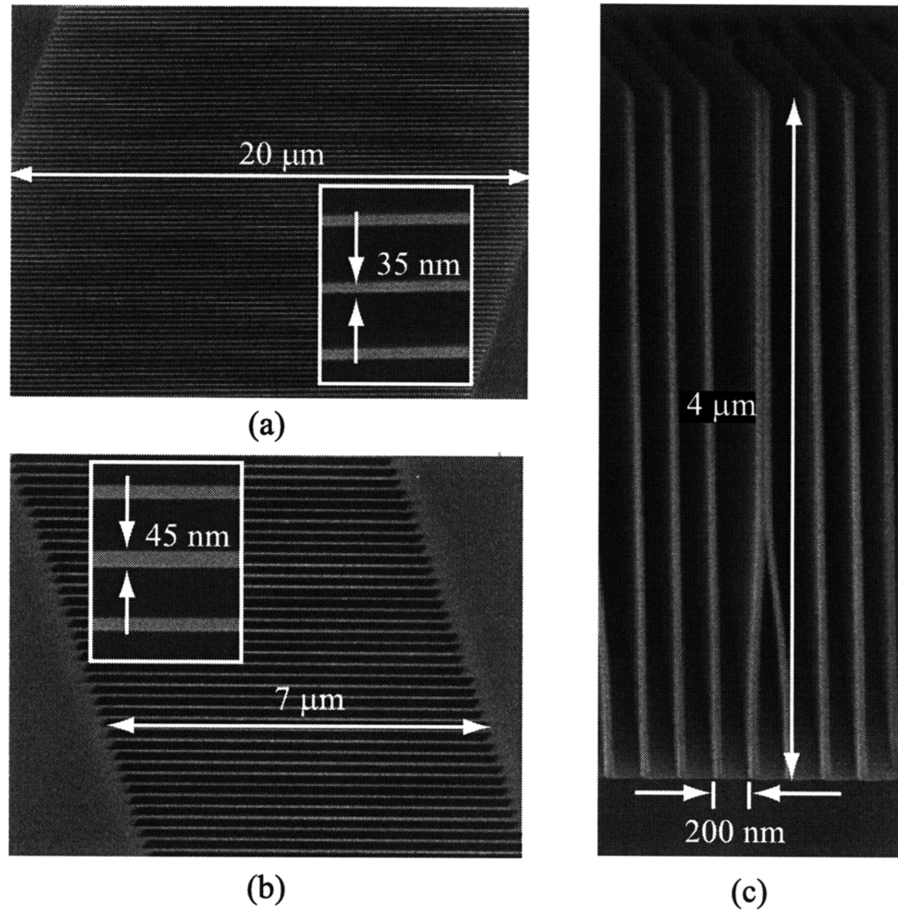


Figure 1-4: A free-standing critical angle transmission grating for x-ray diffraction. The (a) top-view, (b) bottom-view, and (c) cross-sectional SEM images.

even philosophical, questions about life. While these are all exciting and promising applications, the question of how to fabricate these wonderful structures in large quantities with high quality remains a challenge.

When describing the quality of periodic nanostructures there are three important characteristic parameters, including spatial resolution, spatial-phase coherence, and pattern area. First and foremost, spatial resolution, which probably receives the most attention among the research community, is a measure of how small the feature size can be. The importance of this parameter is embedded in the name 'nanostructure,' and is the driving force towards smaller and smaller devices.

The second parameter, which is arguably more important than spatial resolution

Table 1.1: Major advantages and disadvantages of different fabricating methods for periodic nanostructures.

	Probe-based	Self-assembly	Interference Lithography
Spatial resolution	period $\sim$ 30 nm	period $\sim$ 40 nm	diffraction limited
Spatial-phase coherence	low placement error	short distance defects	high
Pattern area	small $\sim \mu m^2/hr$	small $\sim \mu m^2/hr$	large $\sim m^2/hr$
Others	arbitrary pattern	no hardware limited geometry	none

in certain applications, is spatial-phase coherence. This parameter is a measure of how well structures at different locations relate to one another. The spatial frequency and phase are related by

$$\vec{f}(x,y) = \nabla\phi(x,y), \tag{1.3}$$

where $\vec{f}(x,y)$ is the spatial-frequency and $\phi(x,y)$ is the spatial-phase. In other words, spatial-phase coherence determines how periodic the structure is.

The third parameter, which is more critical to industry, is the pattern area. This parameter controls the technology cost in terms of time and money, and in the end will determine if the technology can be transfered from a laboratory setting to the real world. It thus defines the difference between being 'interesting' and being 'useful.' These three characteristic parameters can be used to evaluate the quality of a periodic nanostructure, and depending on the application different emphasis may be placed on different parameters.

There are many processes to fabricate periodic nanostructures, which can be divided into three major categories. Table 1.1 outlines the characteristics of these categories, which can be characterized as probe-based techniques, self-assembly systems, and interference lithography. The main advantages and disadvantages are discussed

in detail.

Probed-based techniques cover any lithography process in which a "probe" with a finite dimension is scanned over the substrate to write the pattern. Examples of such methods are e-beam, ion beam, atomic force microscope (AFM) [57], and scanning tunneling microscope (STM) [21]. The main advantage of this technique is spatial resolution. In all of these examples the probes can be as small as 10s of nm, and periods around $\sim$ 30 nm can be readily achieved. However, since these scanning probe techniques are serial in nature, the patterning is extremely slow. Furthermore, due to various time-varying conditions the placement errors of these techniques can be greater than its resolution, resulting in periodic structures with low spatial-phase coherence. Another advantage of the probe-based technique is its ability to write arbitrary patterns. However, the same reasoning makes it quite redundant to use these techniques for writing repeating patterns.

The second category is self-assembly, which is a bottom-up approach as mentioned previously. These chemical systems, such as block copolymer [19] and DNA [69], self-assemble into the most energy-favorable order due to intramolecular effects. Depending on the diameter of the particle, these systems can yield periods as small as $\sim$ 40 nm. However, due to defects the assembly can form undesirable grain boundaries, limiting spatial-phase coherence over long distances. One of the most appealing advantage of this method is that it requires no hardware at all, however, the geometry is dependent on the chemistry of the system, and can be quite limited.

The last technique, which will be explored exclusively in the context of this thesis, is interference lithography (IL). The main advantage of this method is that it can pattern a much larger area over a relatively shorter time than the aforementioned techniques. Due to the parallel exposure nature of the process, the pattern also exhibits much higher spatial-phase coherence. However, as with all optical techniques, the spatial resolution is diffraction-limited, and the smallest attainable period is governed by the wavelength of the light source.

IL holds the most promising prospects for fabricating large-area high-spatial-phase coherent periodic nanostructures. I will then examine the various issues in IL and

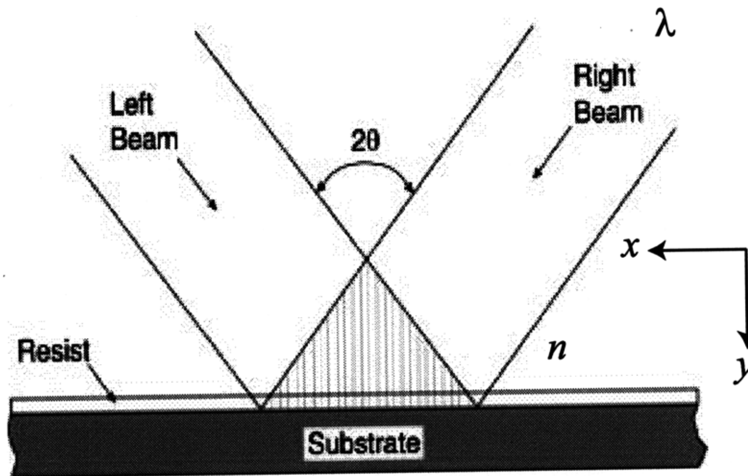


Figure 1-5: The concept of interference lithography.

techniques to push the limits of spatial resolution.

1.2 Interference Lithography

In this section IL will be presented in more detail, focusing on its limitations in spatial resolution, spatial-phase coherence, and patterning area. A typical IL setup is shown in Figure 1-5. In this method two mutually coherent plane waves interfere and produce a sinusoidal intensity pattern, which is then recorded by a photo-sensitive resist. The intensity pattern is then given by

$$I = |\vec{E}_L + \vec{E}_R|^2, \tag{1.4}$$

where $\vec{E}_L$ and $\vec{E}_R$ are the electric fields of the left and right incident beams, respectively. The fields can be expressed as

$$\begin{aligned} \vec{E}_L &= \hat{e}_L A_L \exp(-j\omega t - nkx\sin\theta + nky\cos\theta), \\ \vec{E}_R &= \hat{e}_R A_R \exp(-j\omega t + nkx\sin\theta + nky\cos\theta), \end{aligned} \tag{1.5}$$

where $\hat{e}_L$ and $\hat{e}_R$ are unit vectors denoting the polarization, and A_L and A_R are the amplitudes of the left and right beams, respectively, $\omega = 2\pi f$ is the optical frequency,

n is the index of the medium, and $k = \frac{2\pi}{\lambda}$ is the wave vector. The intensity distribution along the x-direction is then given by

$$I(x) = A_L^2 + A_R^2 + 2A_L A_R(\hat{e}_L \cdot \hat{e}_R)\cos(2nkx\sin\theta). \tag{1.6}$$

The period Λ of the intensity pattern is then given by

$$\Lambda = \frac{\lambda}{2n\sin\theta}. \tag{1.7}$$

From this equation, the period is fundamentally governed by the wavelength of the interfering beam and the smallest attainable period is $\lambda/2n$.

Another parameter of interest is the contrast Γ, or fringe visibility, given by

$$\Gamma = \frac{2A_L A_R(\hat{e}_L \cdot \hat{e}_R)}{A_L^2 + A_R^2}. \tag{1.8}$$

Examining only the polarization effects for now, this equation shows that to maximize contrast, the polarization of the two interfering beams has to be in the same direction. Thus in terms of optimizing contrast, IL is done strictly with TE-polarized beams. Such is the case for most of our work, and for the remainder of this thesis I will only consider only the interference for TE-polarization unless specified otherwise.

In these calculations the interference of two plane waves are examined, and the resulting sinusoidal intensity pattern has linear phase. This case is not practical from an experimental point of view, and in reality any non-planar wavefront error will result in non-linear phase distortion. This case is illustrated in Figure 1-6(b). An alternative is interfering two spherical waves instead, as depicted in Figure 1-6(a). However, for regions where the paraxial approximation is not valid, the grating will result in a hyperbolic phase [27]. Both of these phase distortions are undesirable and degrade spatial-phase coherence. However, the advantage of IL is clearly illustrated. By using large optics, a grating can be patterned in a single exposure as long as the intensity and contrast are sufficient. The patterning area of IL then is limited only by the size of the optics.

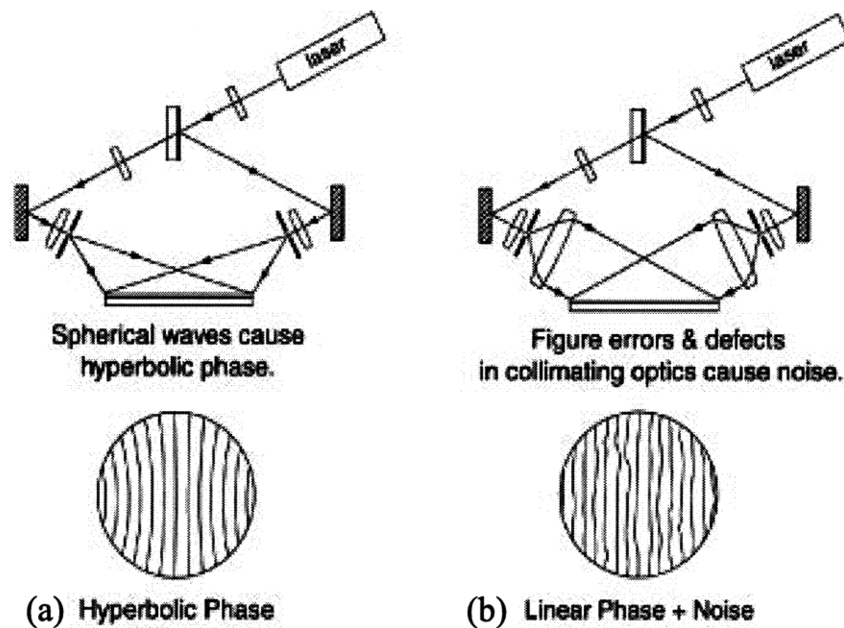


Figure 1-6: Interference of (a) spherical waves results in hyperbolic grating phase and (b) plane waves results in grating with linear phase and non-linear phase distortion.

1.3 The MIT Nanoruler

In an effort to fabricate even larger gratings with higher spatial-phase coherence, the MIT Nanoruler has been developed previously in our group [15] [44] [14] [43] [34]. The Nanoruler uses a scanning-beam interference lithography (SBIL) scheme, in which a substrate is scanned under a relatively small image grating pattern (~ 1 mm diameter) generated by interfering two Gaussian beams. By using a real-time heterodyne fringe-locking system [35], the image grating is phase-locked to the motion of the interferometer stage. The concept of the Nanoruler is illustrated in Figure 1-7. The Nanoruler overlaps individual scans so that the whole substrate can be patterned, as shown in Figure 1-8.

There are several advantages to the SBIL scheme. The first is that by scanning a relatively small image grating, any high spatial-frequency phase distortions can be spatially averaged. This is due to the fact that the resist dose is averaged over the entire exposure field. Also, the exposure process is a repetitive (or periodic) process,

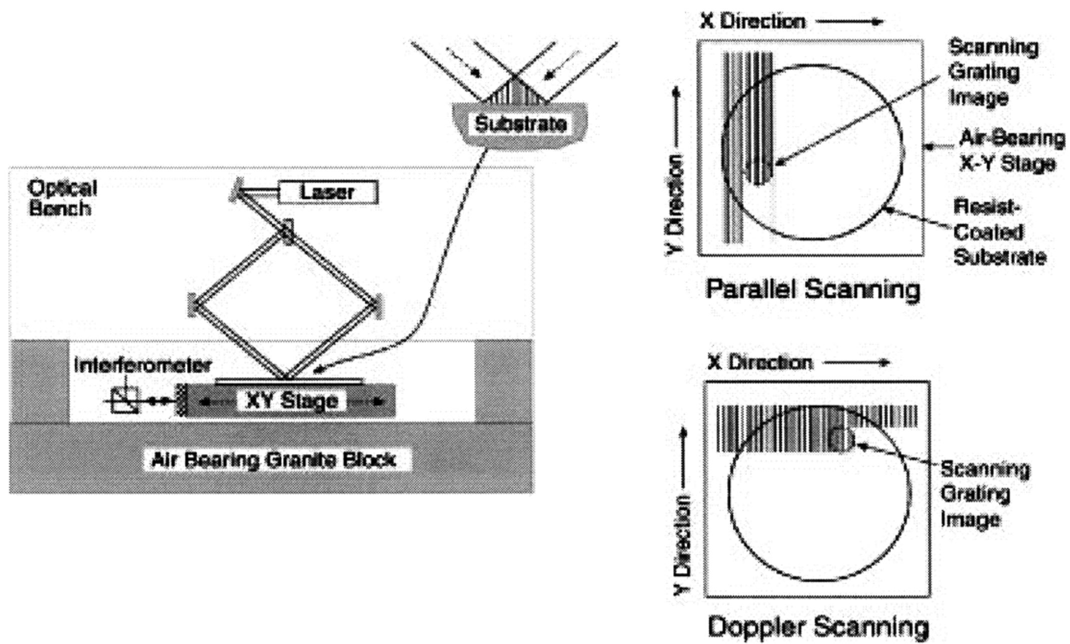


Figure 1-7: The concept of the Nanoruler, in which a substrate is scanned under a grating pattern.

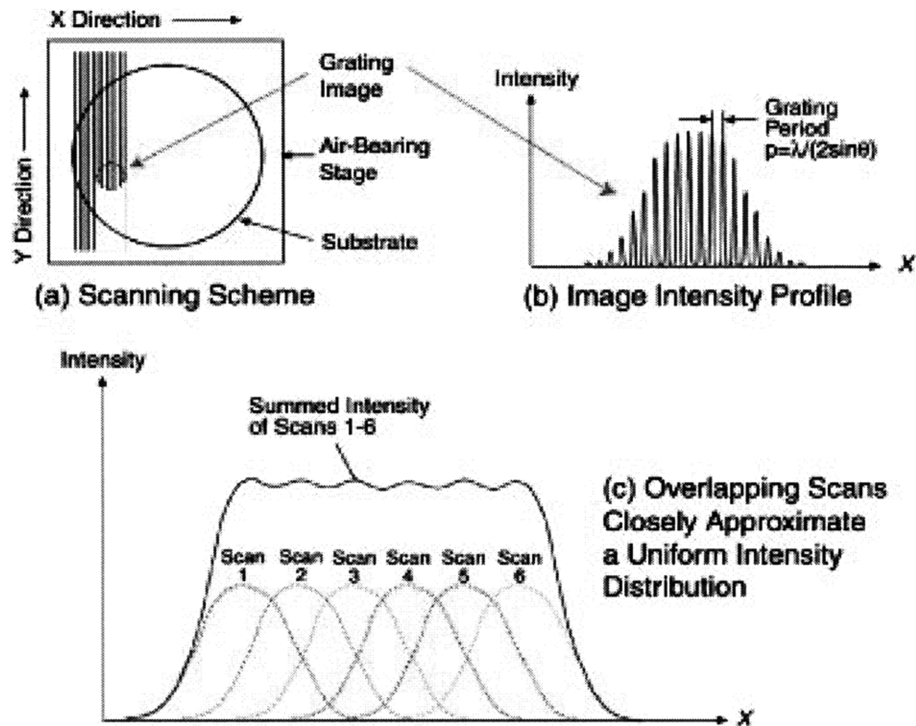


Figure 1-8: The beam overlap scheme of the Nanoruler.

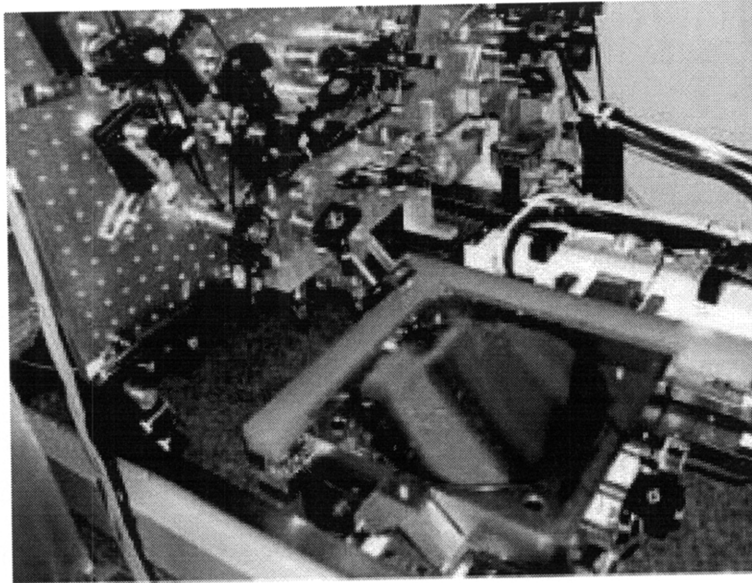


Figure 1-9: A 300 mm substrate covered with 400 nm-period grating patterned in the Nanoruler.

so the exposure conditions are the same over the whole sample. The dominant factor preventing the Nanoruler from patterning perfectly linear grating is the stability of the reference frame.

The second advantage of the Nanoruler is, quite obviously, the pattern area. As mentioned previously, traditional static IL setups are generally limited by the size of the optics. The Nanoruler suffers no such constraint, and the pattern area is instead limited by the traveling distances of the interferometer stage. The prototype system at MIT, the Nanoruler I, is capable of fabricating a grating over an entire 300 mm substrate, as shown in Figure 1-9. In this case the substrate was fully patterned with 400 nm-period grating by the Nanoruler in around 45 minutes. A next generation SBIL tool, the Nanoruler II, is currently located at Plymouth Grating Laboratory[1] and is capable of exposing gratings over a 900 mm by 500 mm area.

[1]http://www.plymouthgrating.com/

1.4 Spatial Resolution of Interference Lithography

The Nanoruler proves to be a viable solution for addressing the issues of spatial-phase coherence and pattern area. However, as with all optical lithography techniques in the far-field, the spatial resolution is diffraction-limited. As previously shown in Eq. 1.7, the smallest attainable period is roughly half of the light wavelength. The Nanoruler currently utilizes a 0.5 W argon-ion laser with $\lambda = 351.1$ nm, and in air can expose gratings with periods down to 200 nm. In many applications the performance of the periodic nanostructure increase as period is decreased. There are even some interesting phenomenon which can only be observed with sub-100 nm feature sizes, such as subwavelength optics [38]. The main goal outlined in this thesis is to extend the spatial resolution of IL in the neighborhood of 50 nm period.

To increase the spatial resolution of IL, the most fundamental methods are to use a shorter wavelength and higher immersion index. For example, a F_2 excimer laser with $\lambda = 157$ nm has been used to fabricate gratings with period as small as 44 nm [67] [6]. But the equipment cost is high, the pattern fidelity is low, pattern area is limited, and the grating has not been transfered into a hard mask. Other processes use extreme-ultraviolet IL to fabricate gratings with sub-50 nm period [63] [64] [65], but these methods require complex sources such as synchrotrons while having limited exposure area. Other non-optical fabrication processes have been demonstrated to multiply the spatial-frequency of patterned gratings by an even factor down to 50 nm period [74] [22], but the process is limited to 1-dimensional patterns and spatial-phase coherence over long range is still questionable.

Conceptually, the simplest method to increase the spatial resolution of the Nanoruler is to mount a 157 nm laser or an EUV source. However, due to various material and optical issues at such high photon energy, these sources are difficult to work with. For example, the most viable refractive material for 157 nm is CaF_2, which is birefringent and extremely expensive. Furthermore, excimer lasers are pulsed, which makes real-time phase metrology, a critical component of the Nanoruler, a daunting task. EUV IL is even more challenging, as EUV or x-ray sources that are coherent can currently

Table 1.2: Roadmap of spatial resolution due for various factors.

λ (nm)	NA	n	Spatial-frequency multiplication	period (nm)
351.1	0.9	1	1	195.1
351.1	0.9	1.45	1	134.6
351.1	0.9	1	4	48.8
193.0	0.9	1	1	107.2
193.0	0.9	1.54	1	69.6
193.0	0.9	1	2	53.6

only be achieve from synchrotrons. Furthermore, EUV optics typically consists of multilayer reflection components, and are extremely expensive.

The lack of commercially available shorter continuous-wave wavelength lasers puts stringent limitations on the spatial resolution of optical-based lithography. Instead of being on the endless and extremely expensive path of wavelength reduction, I have developed a multilevel interference lithography technique based on the concept of multiple exposures [8] to reduce the grating period. In this process multiple grating levels are overlaid and spatial-phase aligned to a common reference grating using interferometry. Each grating level is patterned with offset phase-shifts and etched into a single layer to achieve spatial-frequency multiplication.

The goal of the multilevel spatial-frequency multiplication process is outlined in Table 1.2. The spatial resolution of the Nanoruler is outlined in the case presented in the first row. For a wavelength of 351.1 nm, numerical aperture (NA) of 0.9 (corresponding to incident angle of 64.16°), no immersion, and no spatial-frequency multiplication process, the smallest period is around 195.1 nm. For an immersion index of 1.45, the period can be reduced to around 134.6 nm. If instead one uses a fabrication process that can multiply the spatial frequency of the grating by a factor of four, the period can be reduced to as small as 48.8 nm. This process then by-passes the issue of immersion, and the lithography can be done in air. For comparison purposes, immersion with $\lambda = 193$ nm yields a minimum period of around 70 nm.

Moreover, using 193 nm a spatial-frequency multiplication factor of only two is needed for a period around 50 nm.

Using such a spatial-frequency multiplication process, I have fabricated 50 nm-period gratings using 351.1 nm wavelength light. This period is smaller than the diffraction-limited resolution by a factor of four. This work is done without any significant hardware upgrades, and have many future implications for manufacturing multi-dimensional periodic nanostructures.

1.5 Thesis Structure

In this thesis I present the multilevel interference lithography process with the following structure. The process is outlined in Chapter 2 in the context of spatial-frequency multiplication. This application is expecially interesting as a novel method to fabricate sub-100 nm period gratings. Major challenges such as overlay accuracy and linewidth control of grating layers and their effects on spatial frequency are discussed in Chapter 3.

Chapter 4 covers the design of multilayer homogeneous and periodic resist stacks for various cases involved in this process. Specifically I analyze the balance of diffraction orders for a reference grating, the reflection efficiencies for various lithography steps, and the effects of asymmetric exposure. The design of the resist stack is especially critical, since the exposure conditions can be significantly different. The design simulations are based on rigorous coupled-wave analysis and a binary resist model.

Chapter 5 explores image contrast issues during lithography and their effects on the linewidth of the residual resist grating. Various contrast degradation sources related to beam intensity profiles, wavefront distortions, and other mechanical factors are examined. These factors are analyzed analytically to develop a numerical model which simulates exposure contrast and dose during lithography. The simulated models are compared with experimental results to test the validity of the model and sensitivity of the exposure conditions.

Chapter 6 furthers the discussion on linewidth control, focusing on improving re-

peatability based on fabrication processes. An image-reversal process, in conjunction with plasma etching, was developed to fabricate high duty-cycle grating profiles with nm-repeatability. Multiple metrology iterations are critical for high controllability, and will be discussed in detail.

The experimental and testing results are summarized in Chapter 7. I will present the samples fabricated using multilevel IL. I have successfully demonstrated spatial-frequency multiplication factors up to four. Various metrology methods will be used to evaluate the success of the multilevel IL process. I will also explore other applications for multilevel IL.

Chapter 2

Multilevel Interference Lithography

In this chapter I will present the multilevel interference lithography process, focusing mainly in the context of spatial-frequency multiplication. In this process, multiple grating levels with different phase-offsets are overlaid by aligning to a common reference grating. Each grating level is then pattern-transfered into a single hard mask layer, effectively reducing the grating period. I attempt to use multiple grating levels to extend the diffraction-limited resolution.

There are many challenges that need to be overcome in order for this process to work. The three most important are alignment accuracy of the grating levels, optical design of multilayer homogenous and periodic stacks, and linewidth control of the grating pattern. I will introduce these issues, as well as present some potential remedies.

The goal of this work, simply put, is to use this multilevel interference lithography process to fabricate 50 nm-period gratings using 351.1 nm wavelength light. While this goal is quite specific, spatial-frequency multiplication is only one application of this process. This work actually presents a general method to fabricate multiple levels of periodic nanostructures with arbitrary spatial phase and linewidth.

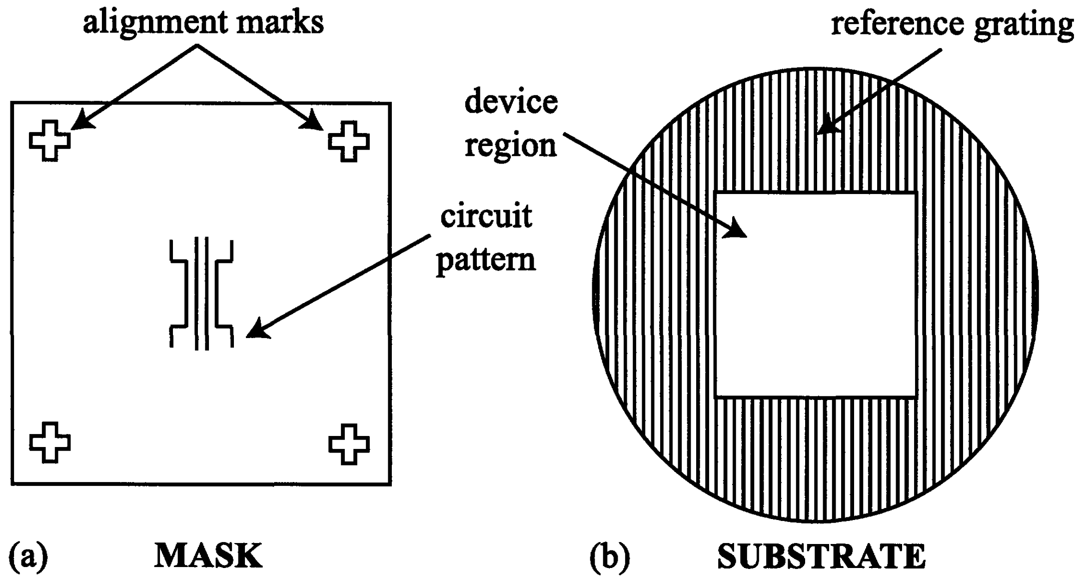


Figure 2-1: The alignment scheme for (a) multilevel optical lithography is similar to (b) multilevel IL.

2.1 Multilevel Interference Lithography Process

The fundamental principle of multilevel interference lithography (IL) is establishing a fiducial grid which then serves as a reference frame for each level of lithography [61]. It is conceptually simple, and is directly analogous to multilevel optical lithography. The parallel between the two lithographic techniques are highlighted in Figure 2-1.

Figure 2-1(a) depicts the alignment scheme that is typically used in optical lithography, in which alignment marks are used to align multiple exposure masks to one another. Multilevel IL, on the other hand, utilizes a reference grating on the substrate as the alignment mark, as shown in Figure 2-1(b). Using this method, every grating level exposure in the device region is aligned to the reference grating prior to lithography. While the area of the reference grating is comparable to the device area in this illustration, this condition is not necessary. In principle only three small patches with mm^2 area is sufficient to establish the reference frame of the grating. Gratings are nearly perfect for the role of being the alignment mark, as they are basically a ruler with a nanometer length scale. For the case of multilevel IL it is quite convienient

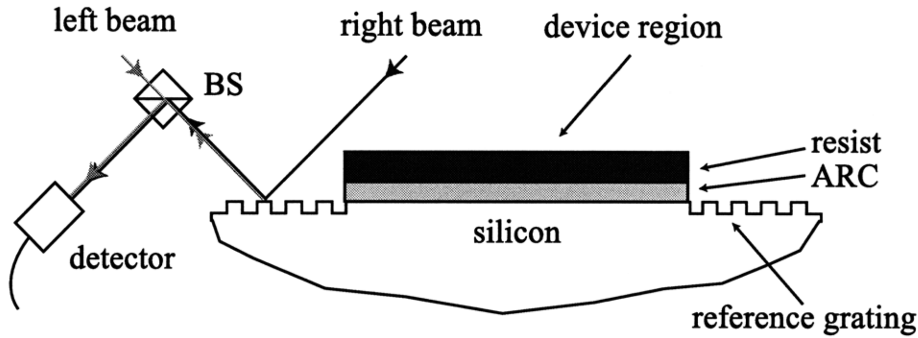


Figure 2-2: A grating interferometer used to align the exposure by measuring the phase of the reference grating. The incident beams are aligned in a Littrow mount.

and interesting, as the pattern we are exposing is a pattern we can align to.

A grating interferometer can then be used to measure the phase of the reference grating and align the exposure beams. The setup is illustrated in Figure 2-2. The substrate depicts a cross-section profile of the reference grating and device region during alignment. The incident beams are aligned in Littrow configuration, where for each beam the first-order diffraction is back-diffracted at the incident angle. Therefore the zeroth reflected order of the right arm and the back-diffracted first order of the left arm co-propagate and interfere. Using a beamsplitter (BS) cube, the beams can be picked-off, and the interference intensity signal is measured by a photodetector.

The interference signal measures the relative phase-offset between the zeroth and first-order diffraction. By scanning the substrate, the principle of phase-shifting interferometry can be used to determine the relative phase of the reference grating [49]. It is essential to map the phase of the reference grating before every exposure to study possible distortions caused by fabrication processes. Both homodyne and heterodyne phase detection schemes are used, the former by setting the same incident beam frequencies, while the latter a frequency offset is introduced. Both of these methods are discussed further later in this chapter.

The first step to the multilevel IL process is the patterning of the reference grating. The fabrication process for the reference grating is illustrated in Figure 2-3. A bilevel resist stack comprised of antireflection coating (ARC) and resist is used, as shown in

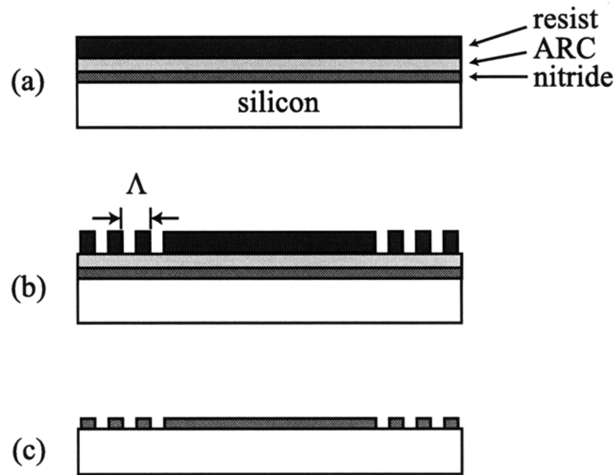


Figure 2-3: Fabrication process for the reference grating using (a) a bilevel resist stack. (b) The grating is exposed, and (c) pattern transfered into silicon nitride.

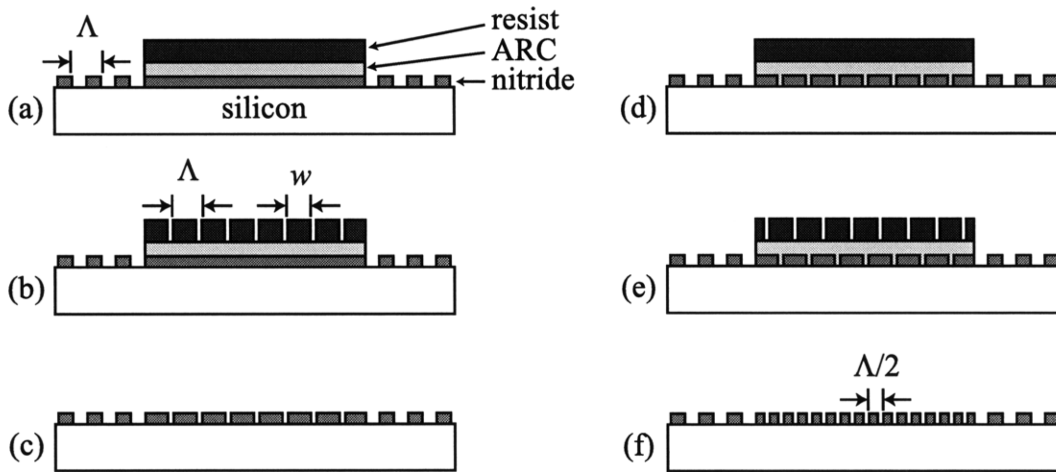


Figure 2-4: Fabrication process for spatial-frequency multiplication. Starting with (a) a bilevel resist stack on a substrate patterned with a reference grating, (b) the first grating level is exposed after aligning to the reference grating, and (c) pattern transfered into silicon nitride. (d) After spincoating ARC and photoresist, (e) the second grating level is exposed with a π phase-offset. (f) The grating is pattern-transfered, resulting in a nitride grating with period of $\Lambda/2$.

Figure 2-3(a). The grating pattern with period $\Lambda = 200$ nm is then exposed at the outer region of the substrate with no alignment using IL, as shown in Figure 2-3(b). The grating is then transfered into the nitride layer using reactive ion etching (RIE), and the residual polymer is stripped using an RCA clean, as shown in Figure 2-3(c). The etched grating is then established as the reference grating. The design of the resist stack, including material and thickness modeling, will be presented in Chapter 3. The details are intentionally omitted in this figure, and the process is presented for illustration only.

The fabrication process for the multilevel IL process is presented in Figure 2-4. The substrate, patterned with reference grating in the outer region, is spincoated with ARC and resist. The polymer in the reference grating region is removed by oxygen plasma, as shown in Figure 2-4(a). After aligning to the reference grating, the first grating level is exposed and developed with period $\Lambda = 200$ nm and linewidth w. The pattern is etched with RIE into the ARC and nitride layers, as shown in Figure 2-4(b)-(c). The process is then repeated to expose the second grating level with a π phase-offset, as shown in Figure 2-4(d)-(e). After transferring the second grating pattern into the nitride layer, the final grating has a resulting period $\Lambda/2 = 100$ nm, as shown in Figure 2-4(f).

This multilevel process presents a conceptually simple technique for spatial-frequency multiplication. Note that this process can be further repeated for higher multiplication factors, depending on the duty-cycle w/Λ of each grating level. For example, duty-cycles of 0.75 and 0.875 are needed for two-fold and four-fold factors, respectively. An appealing advantage of this approach is that since this is solely a fabrication process, finer period gratings can be fabricated without needing to change the light source, optical components, or resist material.

2.2 Theory of Spatial-frequency Multiplication

In this section I examine the geometric theory of spatial-frequency multiplication and the challenges involved. The concept of multilevel IL is intuitive, and the process

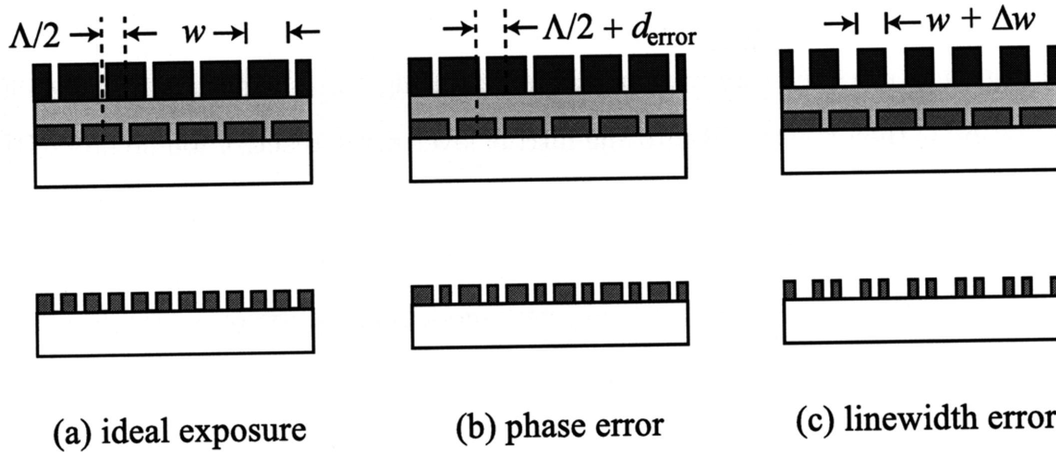


Figure 2-5: Second grating level exposed with (a) perfect alignment and linewidth control. The cases of (b) overlay alignment error of d_{error} and (b) linewidth variation of Δw.

appears relatively simple in the cartoon illustrations as presented in the previous section. However, there are many optical, mechanical, and fabrication challenges that need to be addressed for the process to succeed. The major challenges such as overlay accuracy, linewidth control, and optical layer design are explored further.

The first major challenge is the overlay accuracy between grating levels. In the multilevel IL process the grating exposures are aligned to the reference grating. Any alignment errors will result in phase-offsets and overlay error of the grating levels in the device region. The effects of overlay accuracy is illustrated in Figure 2-5. In this figure the case of a second grating level exposure with perfect alignment of $\Lambda/2$ phase offset is shown in diagram (a). If the exposures of either the first or the second grating level have alignment errors, the resulting grating structure then will have an overlay error of d_{error}, as shown in diagram (b).

The second major challenge is the linewidth control of each grating level. In the spatial-frequency multiplication scheme the linewidth of every grating level has to be consistent in order to successfully increase the spatial frequency. For other multilevel IL applications having precise linewidth control also improves the versatility of the process. In the context of spatial-frequency multiplication, linewidth control

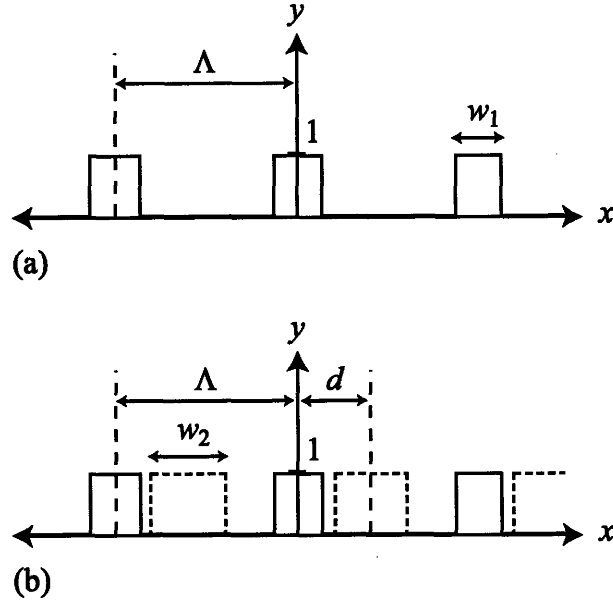


Figure 2-6: Normalized profile of (a) a periodic square grating and (b) two overlapped square gratings.

is necessary. The case of linewidth variation between two grating levels is shown in Figure 2-5(c).

Figure 2-5 depicts pictorially the consequences of overlay and linewidth errors. These cases can be examined more comprehensively by examining the Fourier series. Consider a periodic rectangular profile with period Λ and linewidth w_1, as shown in Figure 2-6(a). The profile can be written as

$$y(x) = \mathrm{rect}(\frac{x}{w_1}) * \mathrm{comb}(\frac{x}{\Lambda}). \tag{2.1}$$

I will examine the amplitude of the Fourier coefficients, given by

$$F_n = \frac{1}{\Lambda}\int_{-w_1}^{w_1} e^{(-j2\pi\frac{n}{\Lambda}x)}\mathrm{d}x. \tag{2.2}$$

The amplitude of the fundamental frequency, $\frac{1}{\Lambda}$, is then

$$F_1 = \frac{1}{\pi}\sin(\pi\frac{w_1}{\Lambda}). \tag{2.3}$$

The goal then is to suppress or even eliminate this fundamental frequency by adding a second profile, as illustrated in Figure 2-6(b). In the figure the general case is

considered, where the second profile is phase offset by d and has a linewidth of w_2. The combined profile in spatial coordinates is then

$$y(x) = \text{rect}(\frac{x}{w_1}) * \text{comb}(\frac{x}{\Lambda}) + \text{rect}(\frac{x}{w_2}) * \text{comb}(\frac{(x-d)}{\Lambda}). \tag{2.4}$$

The Fourier coefficient of the fundamental frequency is then

$$F_1 = \frac{1}{\pi}[\sin(\pi\frac{w_1}{\Lambda}) + \sin(\pi\frac{w_2}{\Lambda})e^{-j2\pi\frac{d}{\Lambda}}]. \tag{2.5}$$

From this equation one can see that for $d = \Lambda/2$ and $w_2 = w_1$, the two terms cancel and F_1 goes to zero. This is the desired case and the fundamental principle of spatial-frequency multiplication.

Note that this multilevel process exploits the fact that even though the period of a single rectangular grating level might be limited by wavelength, the structure profile nevertheless has higher-order spatial-frequency harmonics. The task at hand then is to minimize the fundamental frequency, while maximizing the higher-order spatial-frequencies. If instead the grating profile is purely sinusoidal, which is the intensity profile of the IL exposure, the multilevel process would not be possible. The non-linear, binary response of the resist is then a critical step in this process, as it transfers a sinusoidal intensity profile to a square grating profile.

Let us examine the previous two major challenges using Eq. 2.5. First I assume perfect linewidth control, $\Delta w = 0$, and a overlay error of d_{error}. The Fourier coefficient of the fundamental frequency is then

$$F_{1,overlay} = \frac{1}{\pi}\sin(\pi\frac{w_2}{\Lambda})[1 - e^{-j2\pi\frac{d_{error}}{\Lambda}}]. \tag{2.6}$$

I can then plot the normalized Fourier coefficient, $F_{1,overlay}/F_1$, for a range of overlay error, as shown in Figure 2-7(a). Note the overlay error is normalized as a function of the period. The normalized coefficient represents the effects of the second exposure, and a value of unity means the fundamental frequency does not change at all. As expected, the fundamental frequency is completely suppressed with no overlay error. In reality this result is unlikely, so I examine the overlay tolerances. To suppress the fundamental frequency by a factor of 10, overlay accuracy of better than $\sim 0.07\Lambda$,

or 14 nm, is needed. The challenges of overlay accuracy is examined later in this chapter.

The next step is to examine the effect of linewidth variation between two grating levels. Assuming no overlay error, $d_{error} = 0$, and $w_2 = w_1 + \Delta w$, the Fourier coefficient of the combined profile is

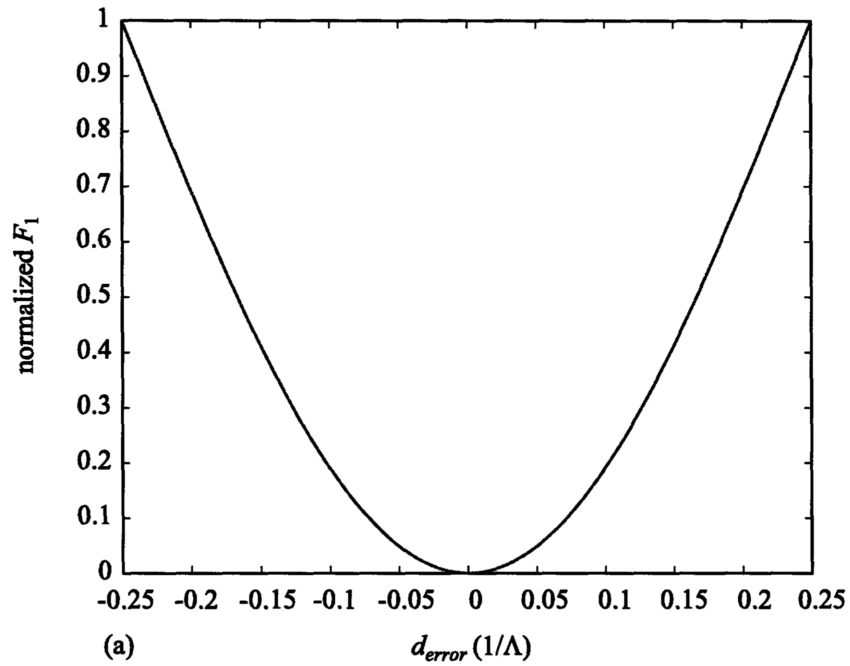
$$F_{1,linewidth} = \frac{1}{\pi}[\sin(\pi\frac{w_1}{\Lambda}) - \sin(\pi\frac{w_2}{\Lambda})]. \tag{2.7}$$

Setting $w_1 = 0.25\Lambda$, the normalized $F_{1,linewidth}$ for a range of linewidth variation Δw is plotted in Figure 2-7(b). As expected, the fundamental frequency is fully suppressed when $\Delta w = 0$. Using the factor of 10 as a guideline, the linewidth variation has to be smaller than $\sim 0.03\Lambda = 6$ nm. In terms of relative linewidth, the variation tolerance is 12% of w_1. Chapter 4 and 5 will examine the various issues influencing linewidth control.

While I examine the scenarios of overlay and linewidth errors separately, the combined effect on the suppression of the fundamental frequency is more stringent. These calculations give some general indicator of the requirements on overlay accuracy and linewidth control. While I use the term 'spatial-frequency multiplication' rather loosely in our discussion, it is not technically true unless perfect alignment can be achieved. The process is then focused on suppressing the fundamental spatial-frequency as much as possible.

There are also other factors, such as height variation and sidewall slopes, which are left out of these calculations. While mathematically these factors can influence the spatial-frequency coefficients, they are less important from a fabrication point of view. This is because the pattern will be use as a hard mask for further processing, and any discrepancies related to height can be mitigated.

The third major challenge is the effects of lithography on multilevel homogeneous and periodic stacks. Mathematically the second grating profile can be "placed" without any difficulties and influences from the first grating profile. However experimentally the effect of the first grating posses significant questions. How does the underlying grating layer effect the lithography of the second, third, or fourth expo-



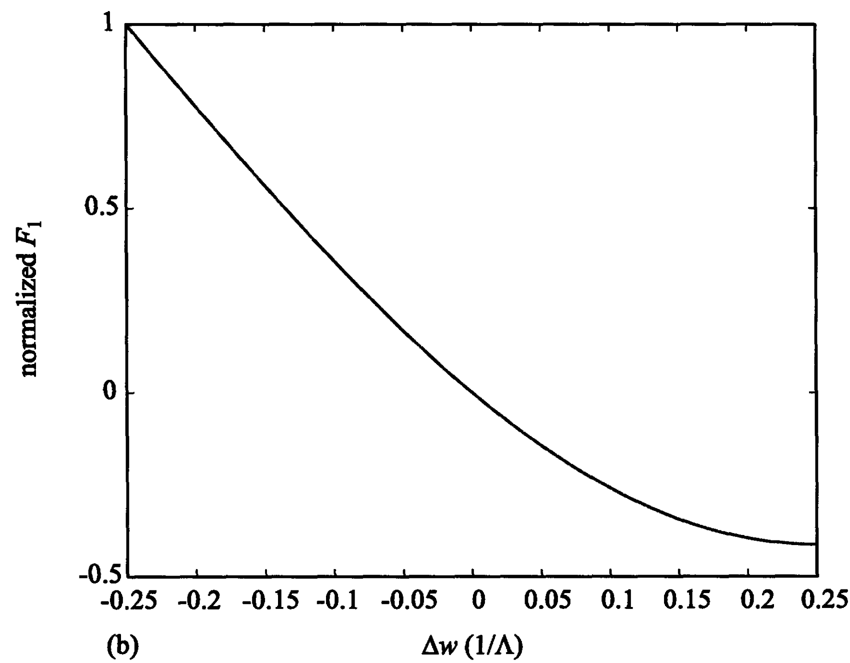


Figure 2-7: Simulated normalized fundamental spatial-frequency coefficient as functions of (a) phase overlay errors and (b) linewidth variation.

sures? To answer this question I have to examine the contrast, intensity, and phase of every grating level exposure. Chapter 3 will examine this topic extensively.

While the multilevel IL process has been introduced in the context of spatial-frequency multiplication, it has many other applications. The process I am proposing is a general technique to overlay multiple levels of periodic structures. Much like using optical lithography to overlay multiple masks in a semiconductor device, I want to be able to fabricate complex periodic nanostructures by using multiple levels of interference lithography.

2.3 Summary

In this chapter I introduced the multilevel IL process, focusing in the context of spatial-frequency multiplication. The most important feature of this process is patterning a reference grating in the outer region of the substrate and establishing it as a fiducial grid. Every grating level is then aligned to the reference grating and exposed with a relative phase shift. Using this process one can extend the diffraction-limited resolution by using high precision metrology and well-controlled fabrication techniques.

I examined the principle of spatial-frequency multiplication from the perspective of two separate exposures. Two critical challenges of this process are phase overlay and linewidth variation as they significantly influence the spatial frequency. I determined that to suppress the fundamental spatial-frequency by a factor of 10, overlay accuracy of 14 nm and linewidth variation of 12% are required. These calculations serve as general guidelines for the multilevel process.

Another major challenge I identified is lithography over multilayer periodic structures. The amplitude and phase of optical fields in these structures must be examined thoroughly. Analyzing these issues is essential to implement the multilevel IL process.

Chapter 3

Grating Level Alignment

In this chapter we discuss the challenges regarding the alignment of multiple grating levels. This is critical for our process, as grating misalignment will induce systematic phase overlay error. We will analyze linear phase distortion due to misalignment, and set some general guideline for alignment tolerances.

We also introduce various optical configurations that are used in the multilevel interference lithography process. There are optical subsystems specifically designed for metrology, allowing for high precision grating level alignment. We also use these metrology tools to examine the time and processing stability of the reference grating, which is essential in our multiexposure process.

3.1 Linear Phase Distortions

In this section we analyze the effects of grating level misalignment and the systematic phase overlay distortions it can cause. Specifically we examine grating period mismatch and angle alignment error between two grating levels. These two cases are critical, because they represent solely linear phase distortion, as we will show.

We examine two cases of grating alignment errors, based on geometry, as illustrated in Figure 3-1. In both of these cases the first grating pattern, with period Λ and oriented in the y-direction, is represented by the black dotted lines. For illustration purposes, these lines represents the peaks of a periodic pattern.

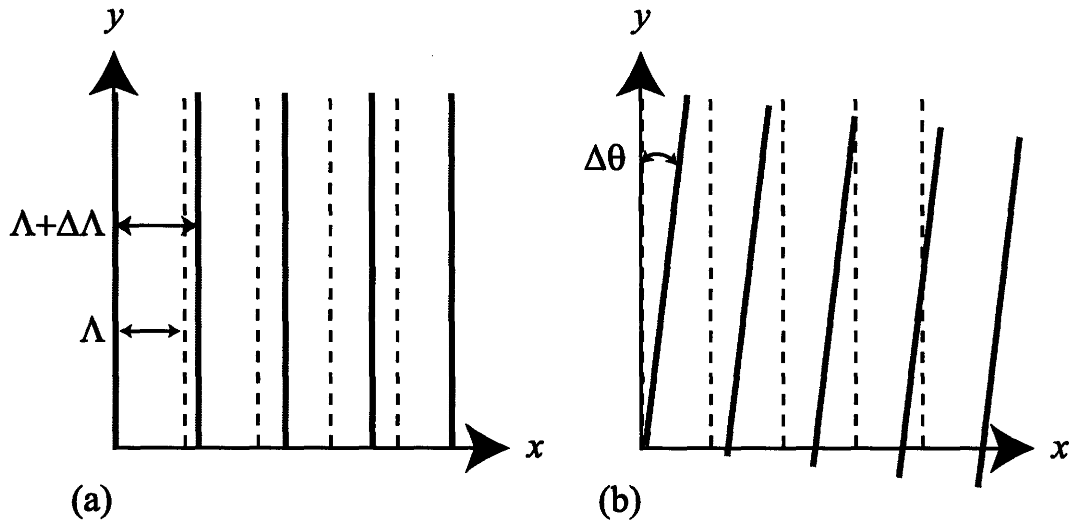


Figure 3-1: The second grating exposure with (a) period error and (b) angle error.

Figure 3-1(a) illustrates the case when the second grating exposure has slightly different period of $\Lambda + \Delta\Lambda$. The two grating profiles are assumed to have perfect phase overlay at $x = 0$. From this simple diagram we can see that even if the grating is perfectly overlapped at one point, there will be phase errors at other locations if the periods are not the same. The phase offset between the two profiles can then be written as

$$\phi_{\Delta\Lambda}(x) = 2\pi\left(\frac{1}{\Lambda} - \frac{1}{\Lambda + \Delta\Lambda}\right)x. \tag{3.1}$$

From this equation we can determine that the phase overlay error due to period mismatch is linear in the x-direction. We can rearrange to approximate the relative period error

$$\frac{\Delta\Lambda}{\Lambda}(x) = \frac{x_{\Delta\Lambda}}{x}, \tag{3.2}$$

where $x_{\Delta\Lambda}$ is the overlay error with unit of length. This equation can be used as a guideline on the tolerance of period mismatch between grating levels. For example, to achieve overlay accuracy of better than $0.1\Lambda = 20$ nm over a distance of $x = 10$ mm, the period error has to be less than 2 part per million (ppm). This means that if the period of the first grating is 200 nm, the largest period the second grating can have is 200.0004 nm.

The next step is to examine the accuracy tolerance on the angular mismatch between two grating layers, as illustrated in Figure 3-1(b). In this case the angle of the second grating level is misaligned. The phase difference can be written as

$$\phi_{\Delta\theta}(x,y) = \frac{2\pi}{\Lambda}x(\cos\Delta\theta - 1) + \frac{2\pi}{\Lambda}y\sin\Delta\theta. \tag{3.3}$$

Since $\Delta\theta \ll 1$, the first term can be eliminated. From this equation we can determine that the phase overlay error due to angle misalignment is linear in the y-direction. We can then simplify and express this equation as angle alignment error

$$\Delta\theta(y) = \frac{x_{\Delta\theta}}{y}, \tag{3.4}$$

where $x_{\Delta\theta}$ is the overlay error with unit of length. For the same case previously described, $x_{\Delta\theta} = 20$ nm over a distance of $y = 10$ mm, the angle alignment error has to be less than 2 μrad. These calculations demonstrates the strict period and angle tolerances that are required for the multilevel IL process. We can also combine these two effects and write

$$\phi_{total}(x,y) = \frac{2\pi}{\Lambda}(\frac{\Delta\Lambda}{\Lambda}x + \Delta\theta y) + \phi_o, \tag{3.5}$$

where ϕ_o is the constant phase offset. From this equation we can see that period and angle errors contribute to linear phase errors parallel and perpendicular to the grating direction, respectively. Eq. 3.5 describes the phase overlay map, and will be an important tool to examine overlay accuracy. The map of phase overlay errors then determine the suppression of the fundamental spatial frequency.

3.2 Optical Configuration

In this section we present the lithography setup and alignment scheme for grating exposure. Specifically, we will highlight the optical configuration used for heterodyne, homodyne phase measurement, and grating exposure. These setups play integral parts in the multilevel interference process, and are all used in a single grating level exposure for alignment purposes.

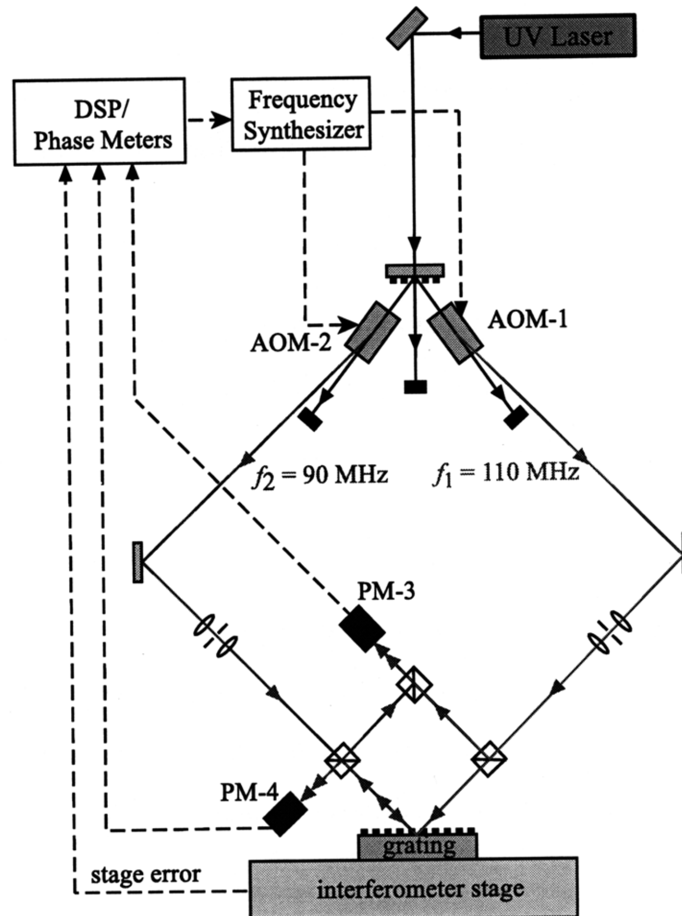


Figure 3-2: The heterodyne phase detection configuration, also called the 'reading' mode, as implemented in the MIT Nanoruler.

We will first introduce the optical configuration for heterodyne phase detection. The main purpose of this configuration is to measure the spatial phase of the reference grating. Heterodyne phase detection is desirable because it is not sensitive to intensity fluctuations due to noise. By mapping the phase of the reference grating before the exposure of every grating level, any in-plane distortion induced by fabrication processes can be detected. This procedure is essential to ensure that the fiducial grid used as the reference frame does not change. The "heterodyne phase detection configuration" is also called the "reading mode", as it is used to "read" the phase of a grating. We will use the two terms interchangeably.

The reading mode as implemented in the MIT Nanoruler is shown in Figure 3-2. A

UV laser beam with wavelength 351.1 nm is split into right and left interference arms. The arms are frequency shifted by 90 and 110 MHz, respectively, by two acousto-optic modulators (AOM) before being spatial-filtered by two 4F systems [30]. A portion of each beam is then picked off by beamsplitters before reaching the grating, and they interfere and are recorded by phase meter 3 (PM-3) as a 20 MHz reference heterodyne signal. The remaining beams are then incident on the grating, and phase meter 4 (PM-4) then measures the interference between the reflected order of the right arm and the back-diffracted order of the left arm.

The phase meters use a zero-crossing technique to determine the temporal phase of a heterodyne interference signal [77]. PM-1 measures the phase difference between the right and left arms, $\phi_R - \phi_L$, plus the phase error induced by the motion of the metrology block, ϕ_m, where all the beamsplitters are mounted. The phase measurement of PM-3 can be written as

$$\phi_3 = \phi_R - \phi_L + \phi_m. \tag{3.6}$$

PM-4 measures the same parameters, as well as the grating phase, ϕ_G. The phase of the grating is included in the phase of the back-diffracted first order of the left arm, and not in the reflected zeroth order of the right arm. The phase measurement of PM-4 is then

$$\phi_4 = \phi_R - \phi_L + \phi_m + \phi_G. \tag{3.7}$$

The grating phase can then be determined by taking the difference $\phi_4 - \phi_3$. Note that the grating phase measurement is relative, as this configuration does not establish an absolute reference frame. However, by scanning the substrate we can measure the grating phase as a function of position, effectively generating a phase map of the reference grating [43]. This is a useful technique to measure the possible distortion of the reference grating.

Next, we present the optical configuration for both homodyne phase detection and grating exposure, as shown in Figure 3-3. In this configuration the two arms are both frequency shifted by 100 MHz, and a third beam, the heterodyne arm, is frequency shifted by 120 MHz by the AOMs. The right and left arm are each combined with the

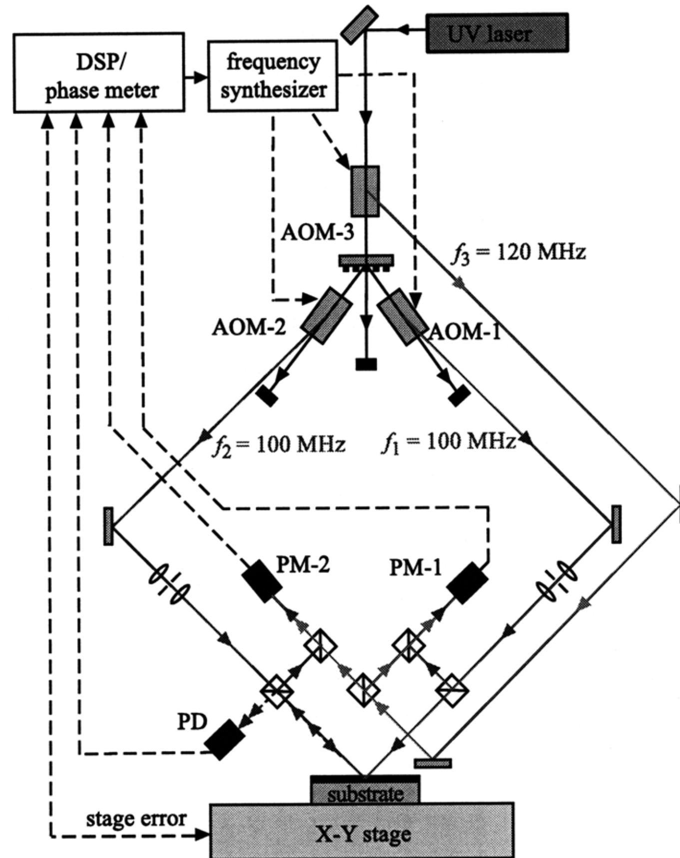


Figure 3-3: The homodyne phase detection and grating exposure configuration, also called the "writing" mode, as implemented in the MIT Nanoruler.

heterodyne arm, and the 20 MHz heterodyne phase signals are measured by phase meter 1 (PM-1) and phase meter 2 (PM-2), respectively. This is the heterodyne fringe-locking system in which the motion of the stage is fed-back into the controller to shift the phase of the left arm by using AOM-2 [35]. In this configuration the fringes of the interference pattern are phase-locked to the stage interferometer so that the intensity pattern is stationary relative to the resist. This configuration is also called the 'writing mode,' which we will use interchangeably.

A homodyne phase detection capability is also added in this configuration, as shown in Figure 3-3. The reflected zeroth order and the back-diffracted first order of the right and left arms, respectively, are picked off by a beamsplitter and measured by a photodetector (PD). Since the frequency of the interfering arms are the same,

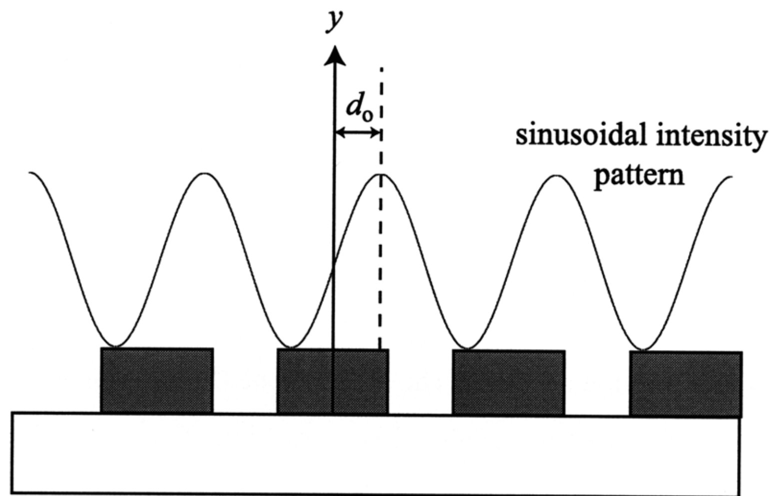


Figure 3-4: An unknown constant phase offset between the phase of the physical reference grating and sinusoidal exposure pattern.

this is a homodyne detection scheme. Note that this method also only determines the relative phase of the reference grating.

This homodyne grating interferometer plays a critical role for the alignment of each grating level exposure. We can now change the phase of the sinusoidal exposure, by changing the phase of one arm, and have intensity measurement on the phase. However, in this configuration the phase measurement does not give us any information on the relative phase offset between the reference grating and the sinusoidal intensity pattern that results from the two interfering arms. This problem is illustrated in Figure 3-4. The figure depicts a constant phase offset $\phi_o = \frac{2\pi}{\Lambda} d_o$ between the phase of the reference grating and the sinusoidal intensity pattern. This result is due to the fact that the phase of the diffracted order depends on the grating material and geometry.

Therefore, using this configuration, we can only align each grating level to the reference grating at an unknown phase offset ϕ_o. This is the primary reason why we can not use the reference grating as the first grating level. Instead, the first grating level is exposed with ϕ_o phase offset to the reference grating, and the second grating level is exposed at $\phi_o + \pi$. This process ensures that the two grating levels

are phase offset by π, even if the relative phase to the reference grating is unknown. An important assumption that can not be overlooked is that the reference grating is stable during the processing of the first grating level.

3.3 Reference Grating Stability

As discussed in the previous section, the stability of the reference grating is critical to the alignment of the grating levels. The definition of stability in this case is in the optical sense, and simply put the reference grating needs to give the same phase reading for each grating exposure. The phase measurement is dependent on the geometry and material of the reference grating, and the stability of these two factors is essential.

The geometry of the reference grating is defined in the initial fabrication step, as shown in Figure 2-3, and must remain undisturbed. Therefore this constraint is taken into consideration when designing the fabrication processes for each grating level. For example, during all RIE etch processes in the device region the reference grating is protected by a metal mask. As a result the reference grating is only in contact with an oxygen plasma etching (PE), RCA and Piranha clean for short periods of time. None of these processes should significantly alter the geometry of the reference grating.

These processes, however, might change the material properties of the reference grating, resulting in changes in refractive index. We specifically examine the oxidation of silicon from these processes. The temperature used in the RCA and Piranha clean are $\sim 80°$, significantly less than what is required to accelerate diffusion of oxide into silicon. We then experimentally tested the effects of both PE and RIE with oxygen by etching an area of a reference grating, while the other areas are masked, for 10 minutes. Using heterodyne phase detection, there was no detectable phase discontinuity across the interface, demonstrating that any possible distortion is less than the noise floor, which is around 1 nm. The oxygen plasma therefore does not oxidize the silicon enough to induce phase distortions above the measurable limit of around 1 nm.

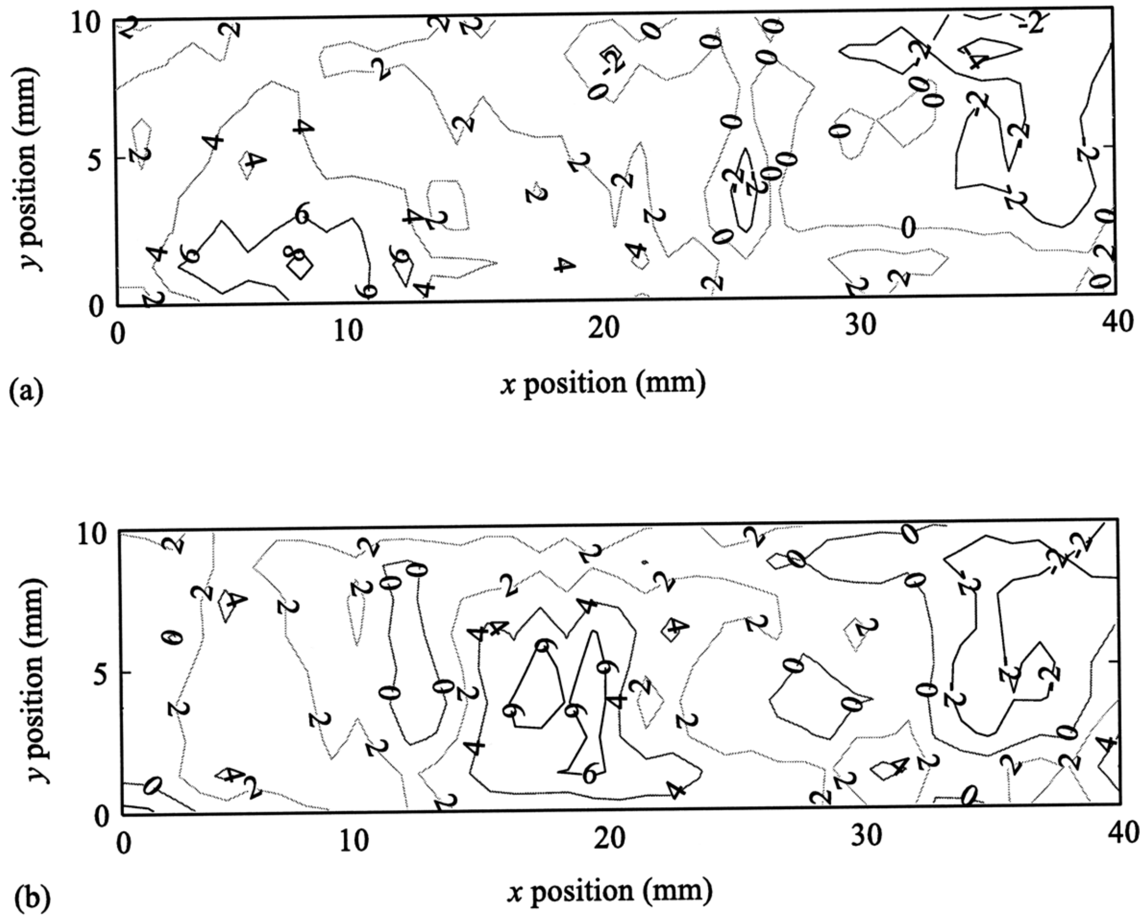


Figure 3-5: The phase map of the reference grating in nm (a) before and (b) after the processing for a single grating level. Note possible particle-induced phase distortion due to vacuum chuck.

To further ensure that the reference grating is stable, we examined the phase after all the processing steps in fabricating a grating level, which are described in Chapter 6. Initially, the phase of the reference grating is mapped using heterodyne phase detection, as shown in Figure 3-5(a) with units of nm. After numerous fabrication processes to transfer the first grating level into the device region, the reference grating phase is mapped again, as shown in Figure 3-5.

Both of the phase maps were measured over an area of 40 mm × 10 mm, centered at the same location within 1-2 mm. Note that while there are some phase differences, the magnitude is relatively small. They are of the same order as the measurement

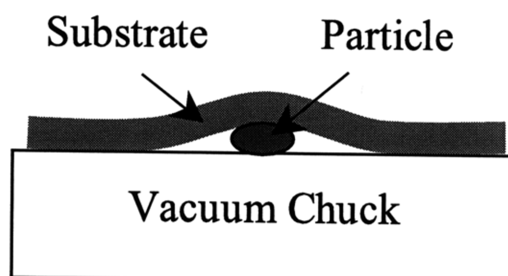


Figure 3-6: A particle is caught between the substrate and the vacuum chuck, inducing in-plane grating phase distortion.

stationary repeatability (sample not removed from vacuum chuck) of our metrology tool, which is around 3 nm (3σ). There appears to be no systematic distortions, as no phase contours are evident. It is important to note that these two phase maps are taken days apart with many fabrication process steps in between.

However, there may be some characteristics of particle-induced distortion in Figure 3-5(b). During the mounting of the substrate, particles might be present between the substrate and the vacuum chuck, as shown in Figure 3-6. The particle bows the substrate upwards, and the grating on top is stretched in opposite directions. The grating region on the left and right sides of the particle is then expected to have negative and positive phase distortions, respectively. This may be the distortion seen in (x = 15, y = 5). This in-plane grating phase distortion is then coupled to the out-of-plane distortion induced by the particle.

The particle can also be caught under the substrate during the exposure of the reference grating. The mechanism then works in reverse. The particle stretches the substrate and resist outward during lithography, and when the vacuum is released the substrate changes back to its original shape, distorting the grating. The signs of the phase distortion is then opposite to the case of particle-induced distortion while reading the phase.

This case is shown for another reference grating, as depicted in Figure 3-7(a). In this experiment the reference grating was distorted by a particle during lithography.

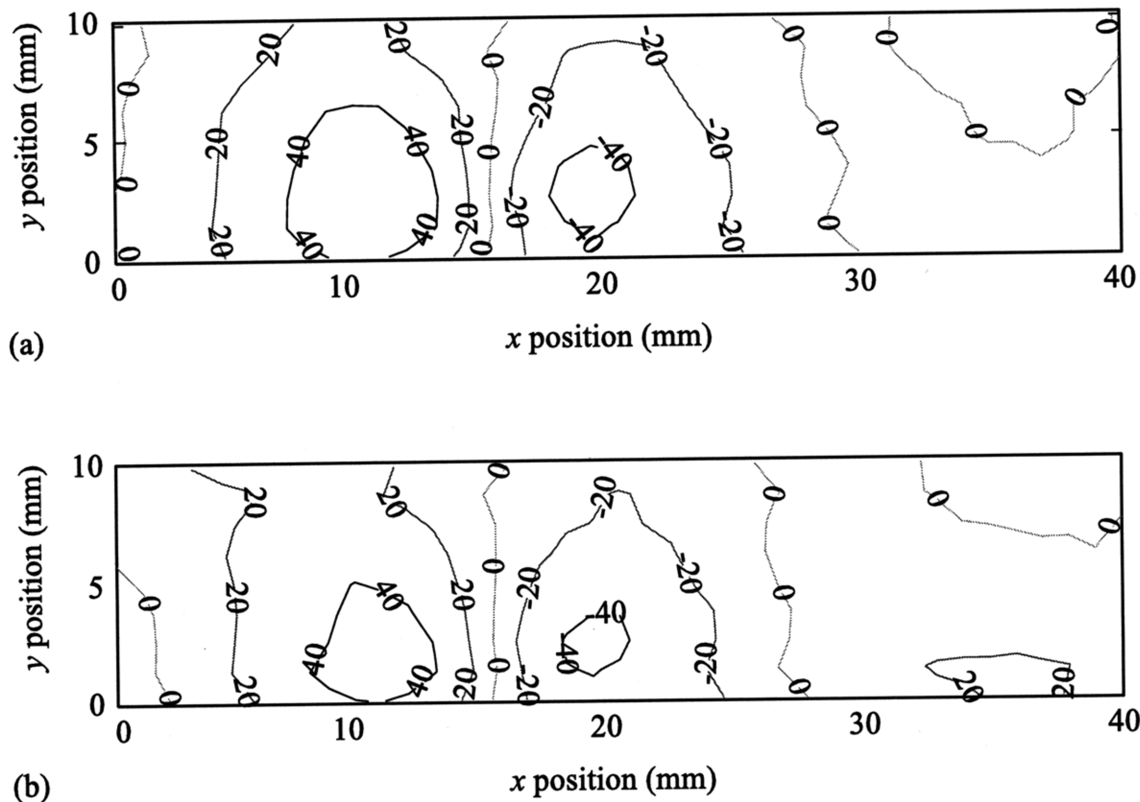


Figure 3-7: The phase map of the reference grating distorted by a particle in nm (a) before and (b) after the processing for a single grating level. Note the particle-induced phase distortion has the opposite sign as the map shown in Figure 3-6(b).

The particle in this case was significantly larger, resulting in much higher distortion. This distortion is etched into the reference grating, and therefore cannot be removed. The phase map taken after the processing of the first grating level in the grating device region, shown in Figure 3-7(b), demonstrate this point. However, this result is favorable, since no other phase distortion can be observed.

These heterodyne reading map experiments prove that the phase of the reference grating is stable with time and fabrication processes for the grating levels. The reference grating can then be used as a stable and reliable fiducial grid for the multilevel IL process. However, particles can introduce local phase distortions during both lithography and metrology. This can become significantly problematic if it occurs in the grating device region. A vacuum pin chuck is utilized to mitigate this issue, but does

not eliminate it altogether. The particle issue and hand cleaning turn out to be the most critical step in this process.

3.4 Grating Alignment Procedure

The various optical configurations and subsystems described in the previous section give us the required tools to overlay multiple grating levels. We can then develop a set of procedures to align each grating level at a relative phase shift to the reference grating. In others word the exact procedure we follow in Figure 2-4(b) will be outlined. Much of this work is done in collaboration with Yong Zhao [76], who has done excellent work in temperature control to improve the repeatability of the Nanoruler.

The first step of the alignment procedure is to map the phase of the reference grating using heterodyne phase detection. To purpose of this step is to examine the reference frame and determine if there are in-plane distortions related to fabrication processes. In order for the multilevel IL technique to be successful, the reference grating has to be distortion-free throughout the process. It would be pointless to align to a fiducial grid that is varying its shape.

After ensuring the reference grating is free from distortion, the next step is to use homodyne phase detection to measure the period and angle of the reference grating relative to the exposure arms. This is done by scanning the reference grating by a fixed distance and counting the number of oscillations in the intensity signal [14] [15]. The period and angle of the reference grating can be determined in this fashion by scanning perpendicular and parallel to the grating lines, respectively. With knowledge of these two parameters, in principle any linear phase overlay error can be eliminated.

The final step is to determine the relative phase of the reference grating using homodyne phase detection. The grating levels for spatial-frequency doubling are then exposed at, for example, the maximum and minimum values from the homodyne intensity signal (i.e. π phase difference). Using these procedures each grating level can be exposed with the desired phase shift. We have demonstrated overlay accuracy of -1 ± 2.8 nm for two 574 nm period gratings over 35×25 mm^2 areas [75]. We want

to extend the capability of this process by overlaying four 200 nm-period gratings. As we will present in Chapter 7, we have achieved this goal with similar phase overlay results over an area of 16 mm $\times$ 30 mm.

3.5 Summary

In this chapter we examined the issues regarding the alignment of grating levels. This is critical for the multilevel IL process, and misalignment of a single grating level can introduce significant phase errors. Our approach uses various metrology steps, including heterodyne and homodyne interferometry, to align each exposure to the reference grating.

We analytically examined systematic linear phase distortions that might be present. We determined that they are induced by period mismatches and angle misalignments between grating levels. We calculated that to have less than phase overlay error of 20 nm over a distance of $\pm x = 10$ mm, period mismatch of less than 2 ppm and grating alignment accuracy of better than 2 μm are required. These calculations serve as general guidelines for the alignment requirements of our multilevel process.

We also examined the geometric and material stability of the reference grating. Various reference grating phase maps were measured using heterodyne interferometry before and after fabrication processes over the length of several days. The results show that the reference grating can function as a reliable fiducial grid for each grating level.

Chapter 4

Design of Optical Multilayer

The design of layer stacks for the multilevel interference lithography is critical to the success of the process. The layer design must satisfy important factors such as optical performance, pattern transfer fidelity, and process compatibility. The design process can be focused onto two regions, namely the reference grating in the outer region and the device grating in the center region of the sample. This chapter discusses the layer design and modeling of these two regions.

The main design goal of the reference grating is to provide a high quality alignment signal for multilevel overlaid exposures. Since the reference grating serves as a fiducial grid, the grating profile fidelity must be preserved and the fabrication process must not distort the grating over time. Other issues such as process transfer and compatibility to subsequent process steps are also considered.

The device grating region is the center area in which multiple grating levels are overlaid and etched. The main question to address is the stability of the exposure conditions. In the proposed multilevel process the resist may be exposed over layered periodic structures, and rigorous modeling is required to ensure pattern fidelity. Simulations involving layered homogeneous and periodic structures are presented, with the main design goal of minimizing the formation of standing waves in the resist layer. In the final part of this chapter exposure symmetry and its effect on grating phase is examined.

4.1 Design of Reference Grating

The reference grating is critical to the success of the multilevel interference process, as it provides the frame of reference for each grating level exposure. As previously outlined the reference grating is mounted in the two-beam Littrow configuration, and the diffracted zeroth-order of the right arm co-propagates with the first-order of the left arm. The two beams interfere and provides an intensity signal, which can be described as,

$$I(t) = |A_0|^2 + |A_1|^2 + 2A_0A_1\cos(\theta_1(t) - \theta_0(t)), \tag{4.1}$$

where A_0 and A_1 are the amplitudes, and $\theta_0(t)$ and $\theta_1(t)$ are the phases of the diffracted zeroth and first order beams, respectively. Note that the incident right and left beams are assumed to have equal intensity, which is the case during resist exposure. This assumption is important in the grating alignment, and will be discussed further in Chapter 5.

In Eq. 4.1, both heterodyne and homodyne phase detection schemes can be used [35], depending on the frequencies of the incident beams. In the heterodyne scheme, the incident beams have a frequency offset Δf, and the intensity results as a sinusoidal time signal with frequency Δf. The change of this frequency offset is then used to determine the relative phase of the two beams. In the homodyne scheme the two incident beams have the same frequency, and phase-shifting interferometry is used to determine the phase.

In either of these two cases, high fringe contrast in the interference signal is essential to obtain an accurate phase measurement. The fringe contrast Γ_{align}, or visibility, of the alignment signal is given by

$$\Gamma_{align} = \frac{2A_0A_1}{|A_0|^2 + |A_1|^2}. \tag{4.2}$$

In order to obtain near perfect contrast $\Gamma_{align} \sim 1$, A_0 and A_1, which depend solely on the diffraction properties of the reference grating, should have similar magnitude. It is thus important to design the reference grating with the optimal material and geometry so that the zeroth and first-order diffraction efficiencies are balanced.

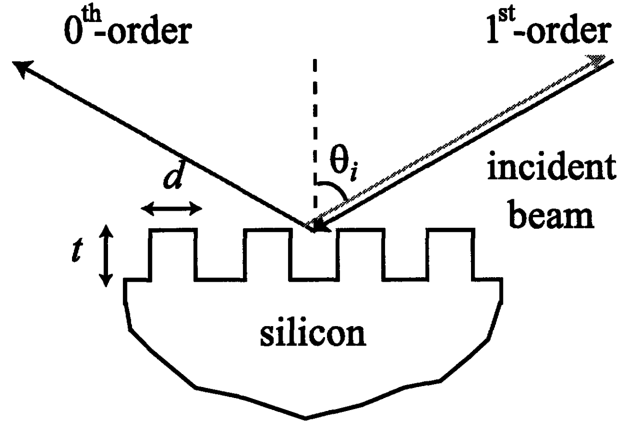


Figure 4-1: The grating and incident beam geometry for the silicon reference grating.

In the following sections, rigorous coupled-wave analysis (RCWA) is used to numerically solve Maxwell's equations to calculate the field in layer periodic structure. The solutions to the wave equation are Bloch waves in periodic medium and plane waves in homogeneous medium [72] [37]. A transfer matrix can then be used to match the boundary conditions of the modes to solve for the transmitted and reflected fields. A detailed introduction of RCWA can be found in Appendix B.

4.1.1 Reference Grating in Silicon

The first material to examine, quite logically, is silicon, since that is the bulk material of the substrate. The silicon reference grating is then a periodic modulation of air and silicon. The refractive indices of air and silicon at $\lambda = 351$ nm are $n = 1$ and $n = 5.47 - j * 2.99$, respectively.

The grating and incident beam geometry is depicted in Figure 4-1. Parameters d and t are the linewidth and thickness of the reference grating. The incident beam has incident angle θ_i and wavelength λ, and the grating period is Λ. The angle of the $m-^{th}$ diffracted order is given by the grating equation,

$$\sin(\theta_m) = \frac{m\lambda}{\Lambda} - \sin(\theta_i). \tag{4.3}$$

Using $\lambda = 351.1$ nm and $\Lambda = 200$ nm, the incident angle necessary to satisfy the Littrow condition is calculated to be 61.34°. In this configuration, the zeroth-order

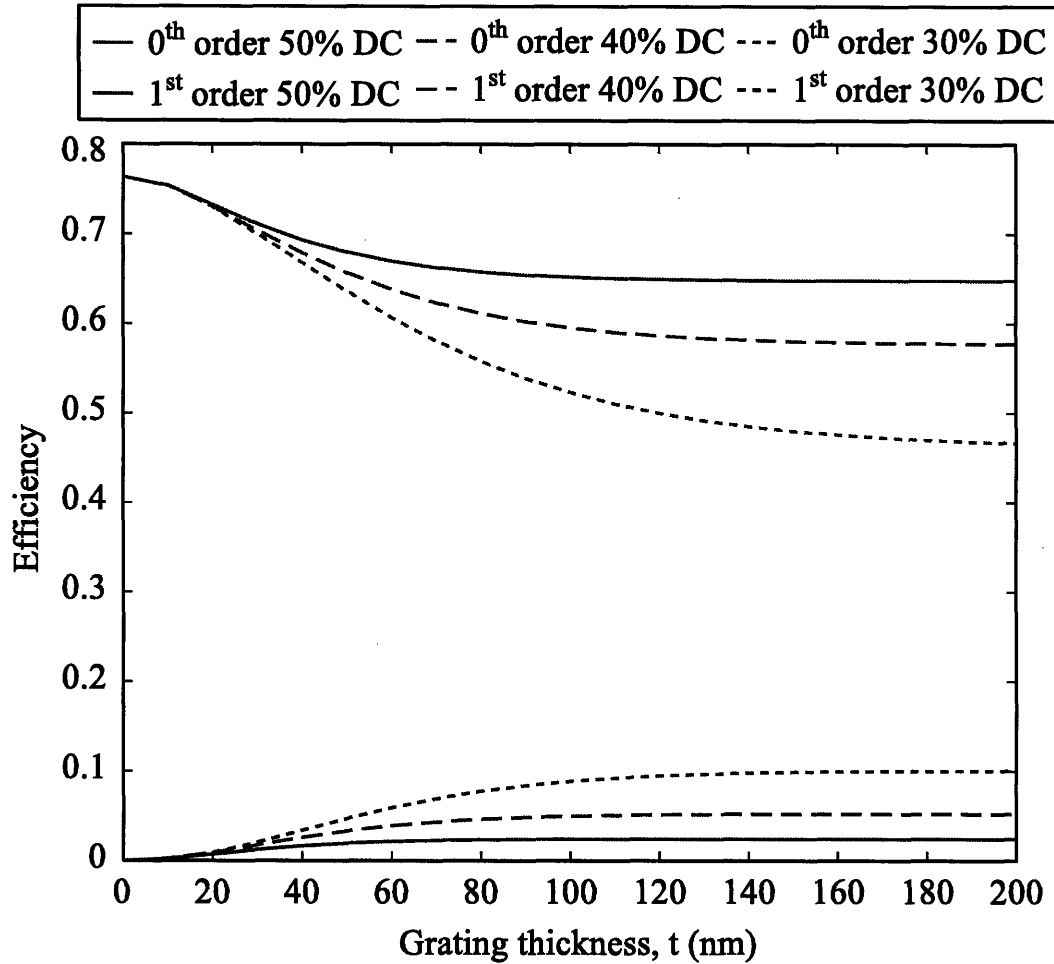


Figure 4-2: The diffraction efficiency of the silicon reference grating as a function of etch silicon depth t for grating duty-cycles of 0.3, 0.4, and 0.5.

is specularly reflected and the first-order is back-diffracted at the incident angle, as shown in Figure 4-1. Note that Eq. 4.3 only defines the angles of the diffracted orders, through geometric arguments. The efficiencies of the diffracted orders depend on material and geometric properties of the grating, and must be calculated using more rigorous models.

For the diffraction configuration shown, the diffraction efficiencies of the orders can be simulated for various values of d and t using RCWA. The simulated efficiencies, defined as the percentage of incident intensity, are shown in Figure 4-2. The zeroth and first-order diffraction efficiencies for three different duty-cycles (DC $= d/\Lambda$) are

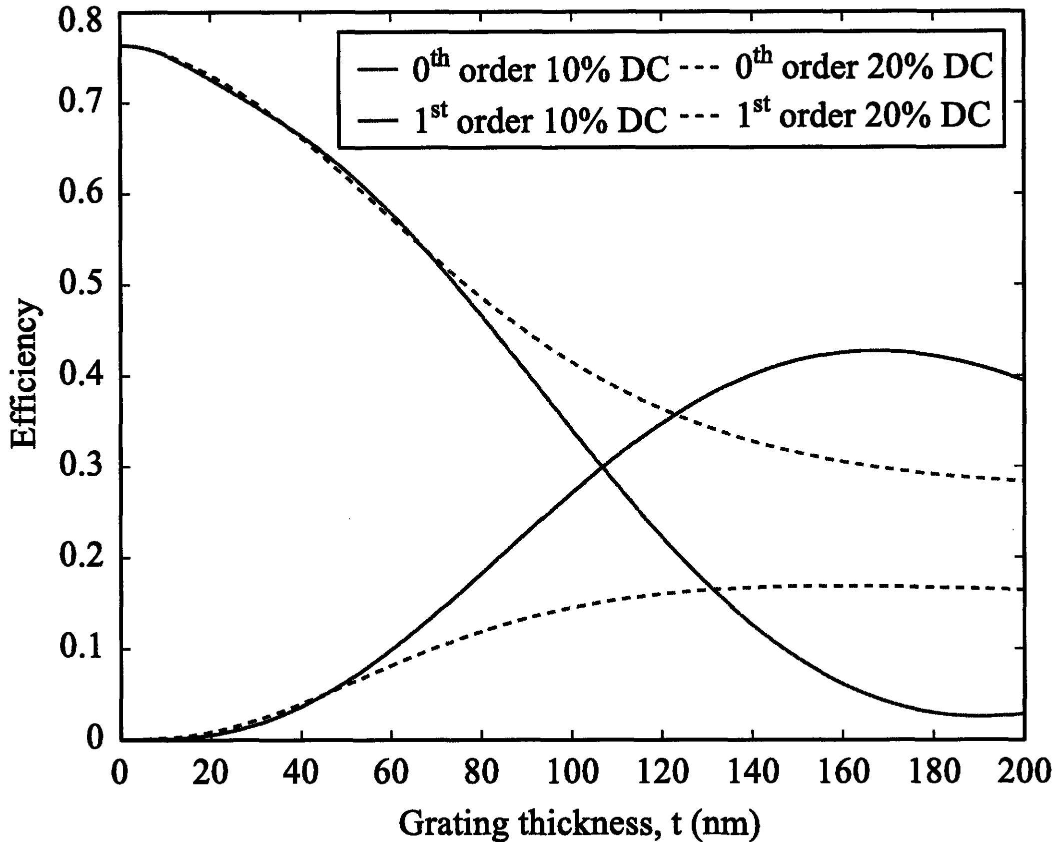


Figure 4-3: The diffraction efficiency of the silicon reference grating as a function of nitride thickness t for grating duty-cycles of 0.1 and 0.2.

simulated as functions of t. In the case of $t = 0$, which represents a homogeneous air/silicon interface, the efficiency curves converge to the case of specular reflection.

From Figure 4-2, it can be seen that for any value of d and t simulated the efficiencies can not be balanced for duty-cycles of 0.3, 0.4, and 0.5. Even in the best case scenario of 0.3 DC and $t = 200$ nm, the ratio of the zeroth-to-first order efficiencies is around 5. At this level the alignment signal contrast Γ_{align} is around 0.75. Furthermore, the diffractive characteristics of the reference grating starts to saturate at around thickness of 100 nm. This may be due to the strong absorption ($k = 2.99$) from silicon at this wavelength, and light does not actually reach the bottom of gratings.

These values of d and t are simulated because they fall within the boundary of

being relatively easy to fabricate. For smaller values of d and larger values of t, the profile fidelity after etching might be poor. To keep the fabrication process simple, etching depth of less than $\sim$ 50 nm is desirable. This is especially critical when the feature sizes of the grating is around $\sim$ 100 nm. For deeper etching depth trilayer processes would have to be used [60], and would significantly increase process steps.

For completeness the efficiencies for duty-cycles of 0.1 and 0.2 are simulated, and plotted in Figure 4-3. Note for a duty-cycle of 0.1 and grating thickness of around 100 nm, the diffraction efficiencies can in fact be balance. However, this case requires a grating linewidth d of 20 nm, while having an etch depth of 100 nm. This geometry has a relative high aspect ratio of 5, and the sample will require non-liquid drying to ensure mechanical stability. The tall structures also present more challenges for subsequent processes which include numerous liquid planarization and plasma etching steps.

From these simulations a silicon reference grating was determined to be unsuitable. Instead of trying to modify the geometry to extremes and generate a major fabrication challenge, a new material is considered. From the saturation effect shown in Figure 4-2, the absorption of silicon is concluded to be too high.

4.1.2 Reference Grating in Silicon Nitride

Silicon nitride is a common material often used as an etch mask in microfabrication processes. It can be easily deposited using chemical vapor deposition (CVD), readily etched using reactive ion etching (RIE), and it is resistant to the anisotropic etchant potassium hydroxide (KOH), which is often use in MEMS processes. All these fabrication advantages make silicon nitride an attractive material.

The optical properties of silicon nitride depends on the relative content of silicon. Using low-pressure chemical vapor deposition (LPCVD) to deposit a low-stress nitride, the complex refractive index is measured to be 2.47-j*0.195 using a spectroscopic ellipsometer. The refractive index of nitride appears favorable, as the real part has a significant index contrast to air and the imaginary part is not too high.

The geometry of the silicon nitride reference grating and the illumination condition

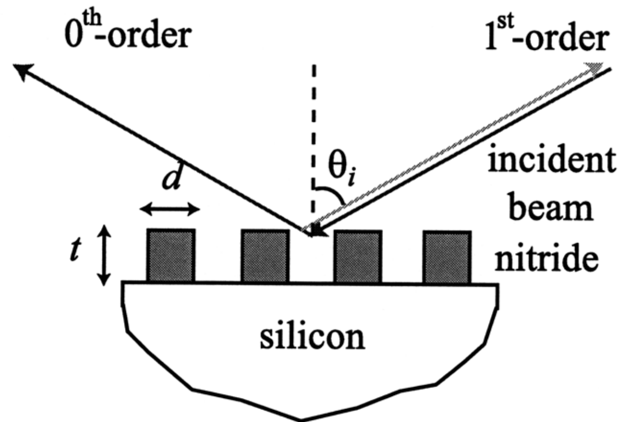


Figure 4-4: The grating and incident beam geometry for the silicon nitride reference grating.

are depicted in Figure 4-4. The simulation parameters are the same as described in the previous section, with the only exception that d and t now describe the linewidth and thickness of the silicon nitride reference grating.

The diffraction efficiencies of the silicon nitride reference grating can then be simulated as functions of t for duty-cycles of 0.3, 0.4, and 0.5, as shown in Figure 4-5. In the figure the efficiencies of the zeroth-order drops to similar magnitude as the first-order starting at a nitride thickness of around 50 nm. The efficiency curves demonstrate that by using a silicon nitride reference grating, the zeroth and first-order diffraction efficiencies can be balanced. This result is encouraging, as it indicates that a fabrication friendly material can be used.

In fact, various combination of d and t can be chosen to balance the two diffracted orders. For example, for a duty-cycle of 0.3, around 62 nm of silicon nitride is needed to balance the orders, while for a duty-cycle of 0.5, 53 nm is needed. The control of linewidth depends on lithography and pattern transfer processes, which will be discuss in details in Chapters 5 and 6. In general, from a process point of view it is difficult to fabricate grating patterns with 0.5 duty-cycle. In a typical fabrication process involving only lithography and pattern transfer, a duty-cycle between 0.3 and 0.4 is more realistic. The duty-cycle can be controlled by additional fabrication steps, but ideally should be avoided. These issues need to be considered to develop a robust

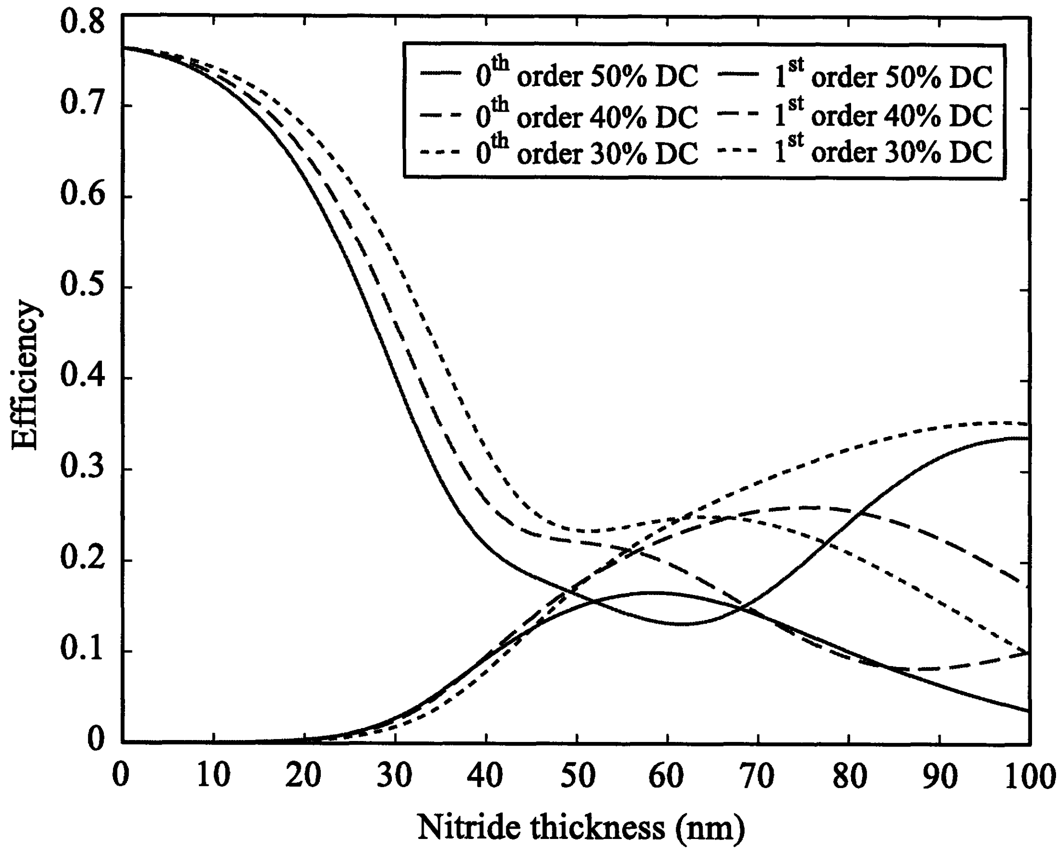


Figure 4-5: The diffraction efficiency of the silicon nitride reference grating as a function of nitride thickness t for grating duty-cycles of 0.3, 0.4, and 0.5.

fabrication process for the nitride reference grating.

In addition to the linewidth d, the thickness t also provides some constraints. As previously mentioned, hard mask materials that are thicker than 50 nm can be difficult to etch. So while Figure 4-5 suggests that nitride thicknesses from 50 to 70 nm might provide optimal efficiency balance, thicknesses around 40 nm are considered. The main rational for this thickness constraint is that even though thicker nitride could be used in the reference grating region, it will pose great pattern-transfer challenges in the device grating region, where feature sizes as small as 25 nm are expected.

Under this thickness constraint, the zeroth-to-first order ratio can be as high as three, depending on the duty-cycle. To further balance the two orders, additional etching into the underlying silicon substrate was examined. This is a feasible ap-

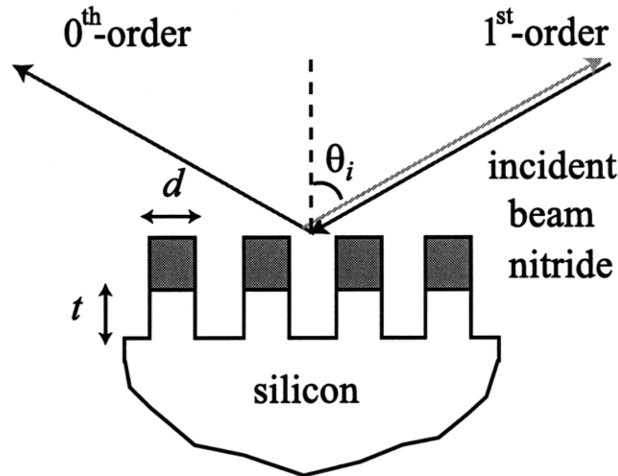


Figure 4-6: The grating and incident beam geometry for the 40 nm-thick nitride reference grating, with additional etch into silicon.

proach, since it does not limit the feature size in the device grating region. This case is illustrated in Figure 4-6, where t is now defined as the etch depth into silicon. The thickness of the silicon nitride is set as a constant, limited to around 40 nm. In this geometry the reference grating essentially consists of a silicon nitride grating of a constant thickness over a silicon grating of variable thickness.

The diffraction efficiencies are then simulated as functions of etch depth into silicon for grating duty-cycles of 0.3, 0.4, and 0.5, as plotted in Figure 4-7. The simulation results show that the zeroth-order can be further suppressed by etching into the silicon substrate, but the increase in first-order diffraction efficiency is limited. Furthermore, saturation of the efficiencies can be observed, similar to the previously described case of a bare silicon grating. This is again due to the high absorption of silicon at this wavelength. However, for a relatively shallow etch of around 30 nm, the zeroth-order efficiency can be reduced by half for all duty-cycles, which will result in a significant improvement in the phase signal contrast.

Another important issue is the thickness control for the silicon nitride film. Since the deposition is CVD, which has a relatively high deposition rate, the thickness of the film is difficult to control and a thickness error of up to 5 nm is unavoidable. Furthermore, in the fabrication process up to 15 silicon substrates are placed in a

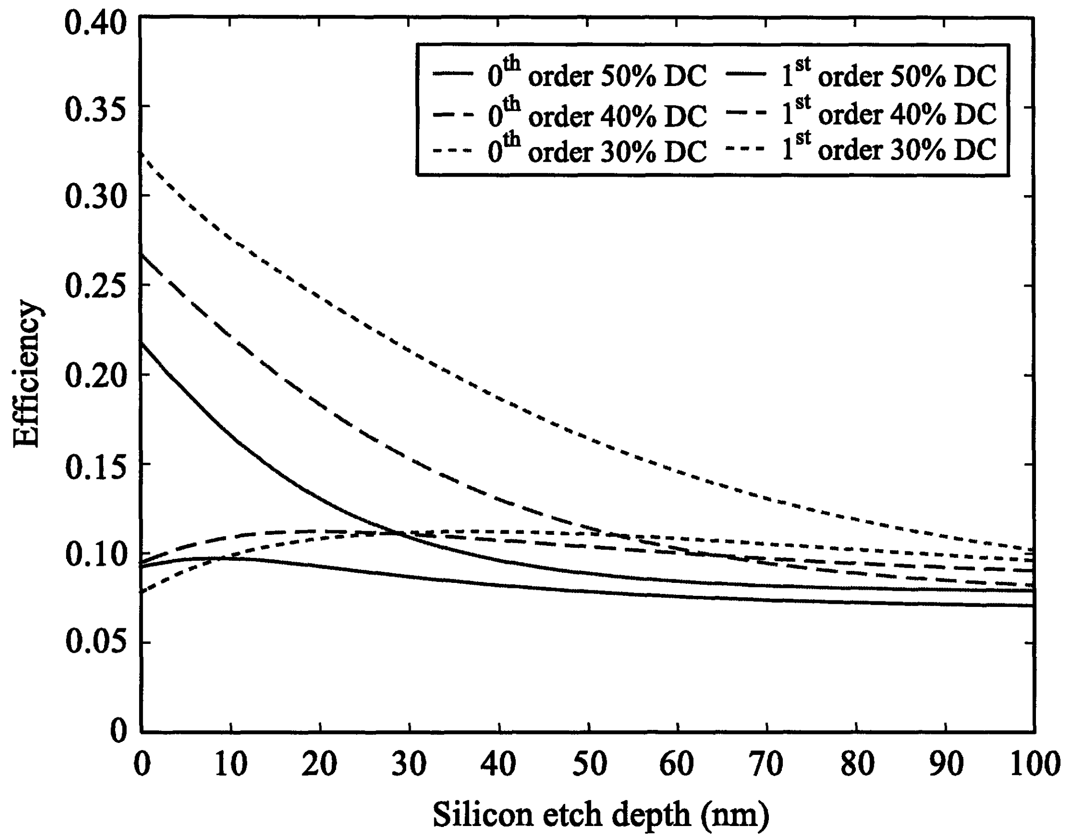


Figure 4-7: The diffraction efficiency of the 40 nm-thick nitride reference grating as a function of etch depth t into silicon for grating duty-cycles of 0.3, 0.4, and 0.5.

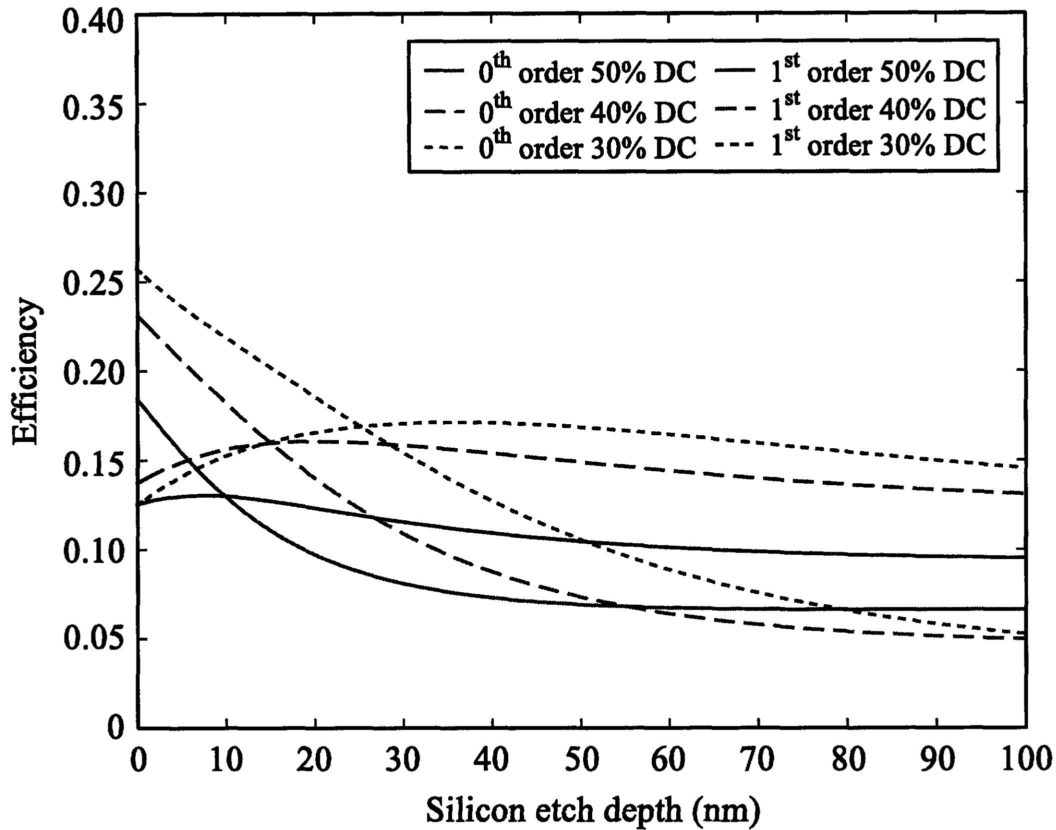


Figure 4-8: The diffraction efficiency of the 45 nm-thick nitride reference grating as a function of etch depth t into silicon for grating duty-cycles of 0.3, 0.4, and 0.5.

vertical stack and coated in a single run. As a result thickness variations of up to 10-20% exist between the top-most and the bottom-most substrate. Therefore it is important to simulate and examine the diffraction efficiencies for a range of silicon nitride thickness.

To test the sensitivity of nitride thickness variation, the same configuration shown in Figure 4-6 is simulated with a nitride thickness of 45 nm. The resulting efficiency curves are illustrated in Figure 4-8. From these results the diffracted orders can be readily balanced by etching 10-30 nm for all grating duty-cycles. The diffraction performance of the 45 nm thick nitride is significantly better than 40 nm, even though there is only a thickness difference of 5 nm.

Based on these simulations, the target of the reference grating is designed to be 45

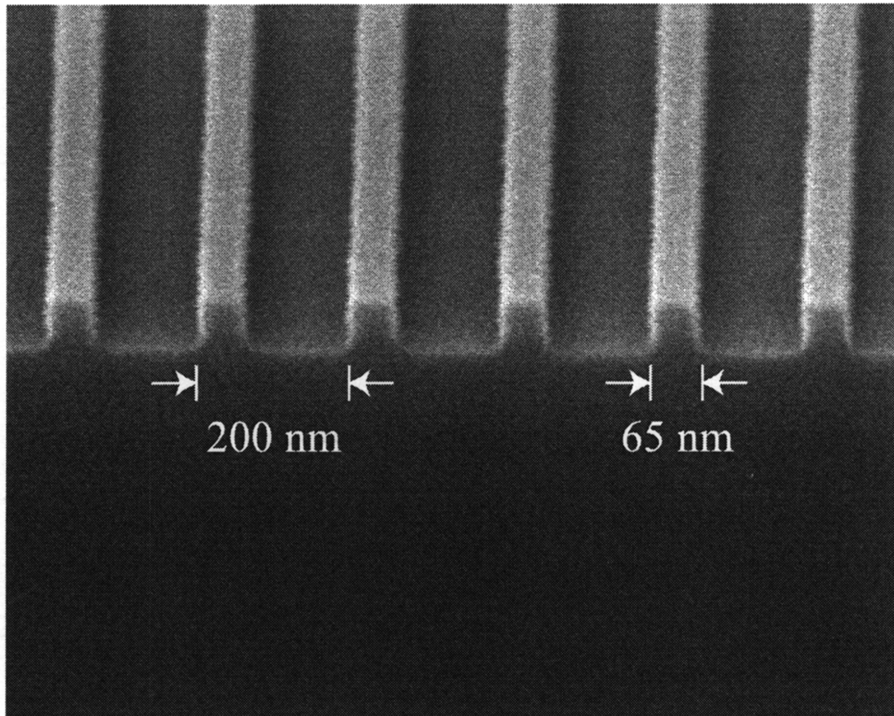


Figure 4-9: Cross-section SEM image of a fabricated silicon nitride reference grating.

nm thick nitride and etched around 30 nm deep into silicon. The zeroth and first-order diffraction efficiencies are comparable under these conditions, which will result in a high contrast phase measurement signal in the grating interferometer. A cross-section scanning electron microscopy (SEM) image of a fabricated nitride reference grating is shown in Figure 4-9. Measuring from the image, the duty-cycle is $\sim$0.33, the nitride thickness is $\sim$45 nm, and the etched depth into silicon is $\sim$26 nm. Experimental testing for this reference grating shows the zeroth and first-order efficiencies vary by less than 30%, resulting in a contrast of around 0.99.

There are several factors that are not taken into account in the simulation, and may result in the final efficiencies being unbalanced. The grating profile is assumed to be perfectly vertical, as the refractive index profile used is periodic square waves. In reality the walls tend to be slightly sloped due to pattern transfer. There are also some line-edge roughness (LER), which is not taken into account. Furthermore the exact grating duty-cycle, nitride thickness, and silicon etch depth are difficult

to control. Based on all these fabrication induced errors, the resulting zeroth and first-order may still be unbalanced. However, as long as the efficiencies are of the same magnitude the contrast will be high enough.

4.2 Design of Multilayer Resist Stack over Homogeneous Layers

This section describes the design of the resist stack that consists of multiple layers of homogenous films. The main design objectives are optimizing exposure contrast in the resist layer and ensuring high pattern-transfer fidelity. In the multilevel interference lithography process, there are two cases when such a stack is used, namely the exposure of the reference grating and the first grating level in the device region. The layer stacks must be properly designed to ensure high grating profile fidelity.

The typical resist stack consists of resist, antireflection coating (ARC), and a hard mask on top of silicon. From the previous section, the hard mask is determined to be ~45 nm of silicon nitride. Since the nitride film is typically deposited over the entire substrate, the process would be simpler if the same thickness can be used in the device grating area. The goal then is to design a layer stack on a silicon substrate with around ~45 nm of nitride

The first important factor to consider in designing the resist stack is related to fabrication issues. The grating exposed in resist must be pattern-transfered using RIE into ARC, nitride, and silicon with high pattern fidelity. Depending on the gas used, the masking material must have adequate thickness to ensure proper pattern transfer. Table 4.2 tabulates experimentally measured etch rates of resist (Sumitomo PFi-88A2), ARC (Brewer Science i-CON-16), silicon nitride, and silicon for etch gases of O_2, CF_4, and CHF_3.

The O_2 RIE etch process uses RF power of 100 W, pressure of 7 mTorr, and gas flow of 19 standard cubic centimeter per minute (sccm). The CF_4 RIE etch process uses RF power of 200 W, pressure of 10 mTorr, and gas flow of 15 standard cubic

Table 4.1: Experimentally measured etch rates in nanometer per min for O_2, CF_4, and CHF_3 RIE.

	O_2	CF_4	CHF_3
resist (PFi88-A2)	60	40	15
ARC (i-CON-16)	75	N/A	N/A
silicon nitride	0	30	10
silicon	0	20	0

centimeter per minute (sccm). The CHF_3 RIE etch process uses RF power of 200 W, pressure of 10 mTorr, and gas flow of 15 standard cubic centimeter per minute (sccm).

In fabricating the device and reference gratings, the grating pattern needs to be transfered to the nitride and silicon level, respectively. The pattern-transfer is performed using an O_2 RIE to etch into ARC, and then CF_4 to etch into nitride and silicon. As a general guideline the resist thickness needs to be more than twice the ARC thickness to ensure good pattern-transfer. Also, both the resist and ARC thicknesses need to be much thicker than the desired etch depth through nitride and silicon. These are rules of thumb that should be considered when designing the homogenous layer stack.

Another important factor in the layer stack design is its optical performance. A typical homogenous multilayer stack is shown in Figure 4-10. The refractive index and thickness of the $i-th$ layer is represented by n_i and d_i. The figure depicts a single exposure arm with incident angle θ_i, and the red and blue arrows represent transmitted forward and reflected backward propagating plane waves in each layer. The angles of the transmitted and reflected beams inside each stack are given by

$$\sin(\theta_i) = \left(\frac{n_0}{n_i}\right) \sin(\theta_0). \tag{4.4}$$

Since the multilevel stack is homogenous, only a transmitted and reflected mode are allowed.

In interference lithography, the top most stack is the resist layer. The resist records

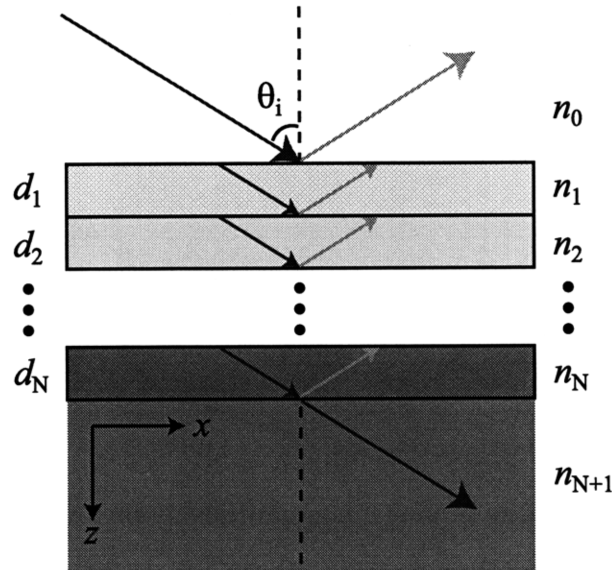


Figure 4-10: Geometry of the homogenous multilevel stack.

the interference pattern along the x-direction between the transmitted beams (only one arm shown in Figure 4-10 in the first stack, depicted in red. The primary concern is then the reflected beam at the bottom of the resist interface, depicted in blue. The reflected beam will interfere with the transmitted beam to a create standing wave in the z-direction, resulting in wavy resist sidewalls after exposure. This is a direct result of having four-beam instead of two-beam interference.

To reduce the reflected wave intensity, an ARC layer is used under the resist. The ARC layer is absorbing and has a real index similar to the resist index. The resist/ARC is then optically matched, yielding little reflection, while the ARC absorption can reduce the reflection from the underlying layers. Therefore, by varying the thickness of the ARC layer, the reflection intensity at the resist/ARC interface can be minimized. Optimizing the ARC thickness is thus the primary focus in the design of the multilevel resist stack with homogenous layers.

To design for the optimal ARC thickness, a transfer-matrix method [73] can be used. This method is widely used to calculate multilevel dielectric stack with homogenous layers [60]. Implementing in Matlab, the RCWA simulation software described in Appendix B was simplified from a $(2M \times 2M)$ to (2×2) matrices to simulate the

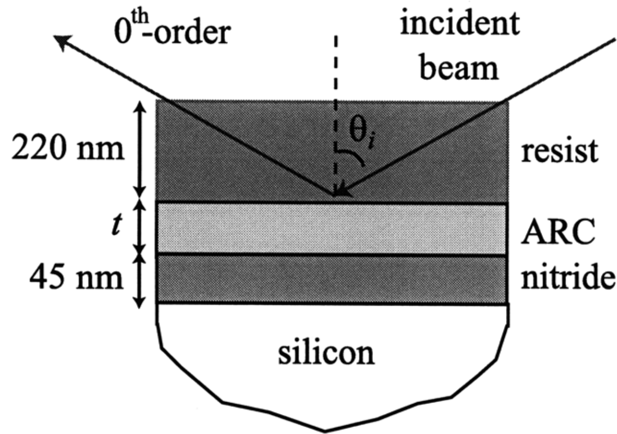


Figure 4-11: Geometry of the homogenous multilevel stack with 220 nm resist, ARC thickness t, and 45 nm nitride on top of a silicon substrate.

case for multilevel homogenous stacks. All of the following simulations were generated using this method.

The layer stack geometry to simulate is shown in Figure 4-11. The multilayer stack consists of 220 nm resist, ARC thickness t, and 45 nm nitride on top of a silicon substrate. Note the 45 nm of nitride is needed for zeroth-first order balance, as described in the previous section. The thickness of resist is chosen to be 220 nm because the thick polymer mask can be used effectively for pattern-transfer. There is an upper limit to the resist thickness. Since resist linewidth of 60-80 nm is expected after exposure, having very thick resist will result in mechanical instability. Therefore the resist grating should have a maximum height-to-linewidth aspect ratio of around 4.

The simulated reflection efficiency at the resist/ARC interface is plotted as a function of ARC thickness in Figure 4-12. The efficiency is defined as the intensity of the reflected order relative to the intensity of the incident beam. Often times a relative efficiency is used, which is defined as the ratio of the reflected and transmitted orders in the resist. The efficiency curve has two local minima, at thicknesses of 13 and 120 nm. However, the efficiency at t = 13 nm is 3%, significantly larger than at t = 120 nm, where only 0.5% of the incident beam is reflected. A thickness of 13

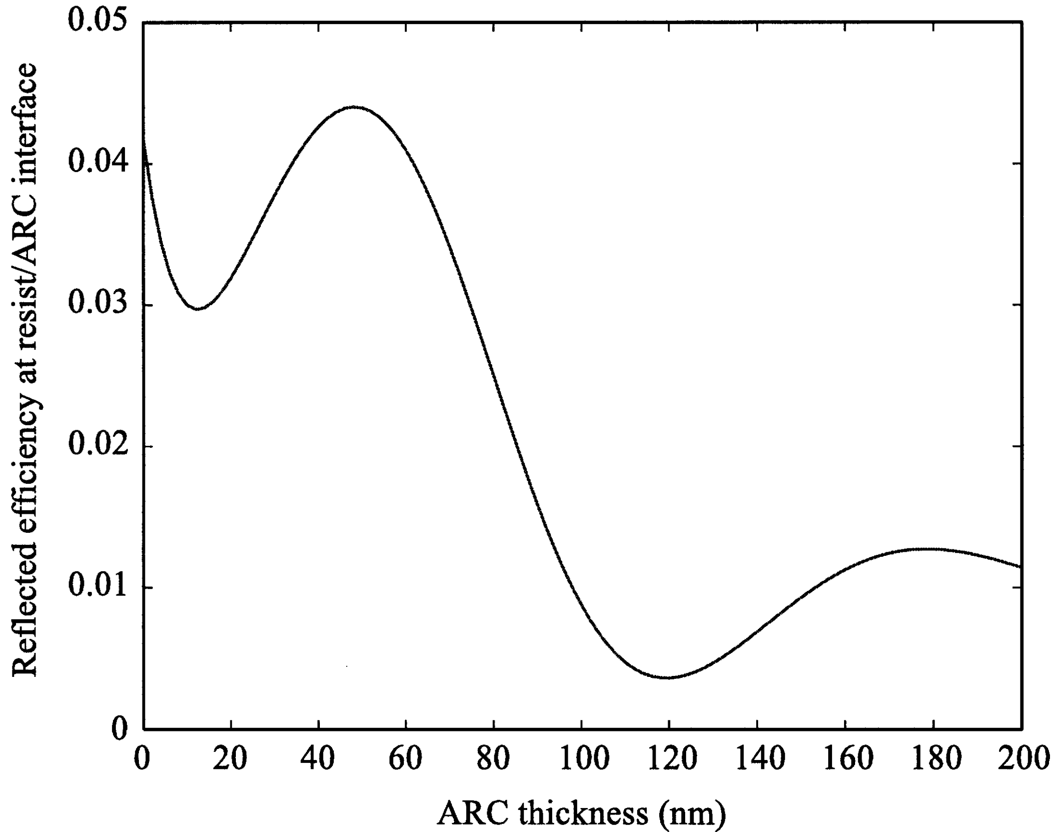


Figure 4-12: Simulated reflection efficiency at the resist/ARC interface.

nm is also extremely difficult to spincoat, and the large slopes surrounding the point makes fabrication tolerances extremely tight. The minimum at $t = 120$ nm is much more desirable, as it offers lower reflection efficiency, larger thickness tolerance, and also allows easier pattern transfer.

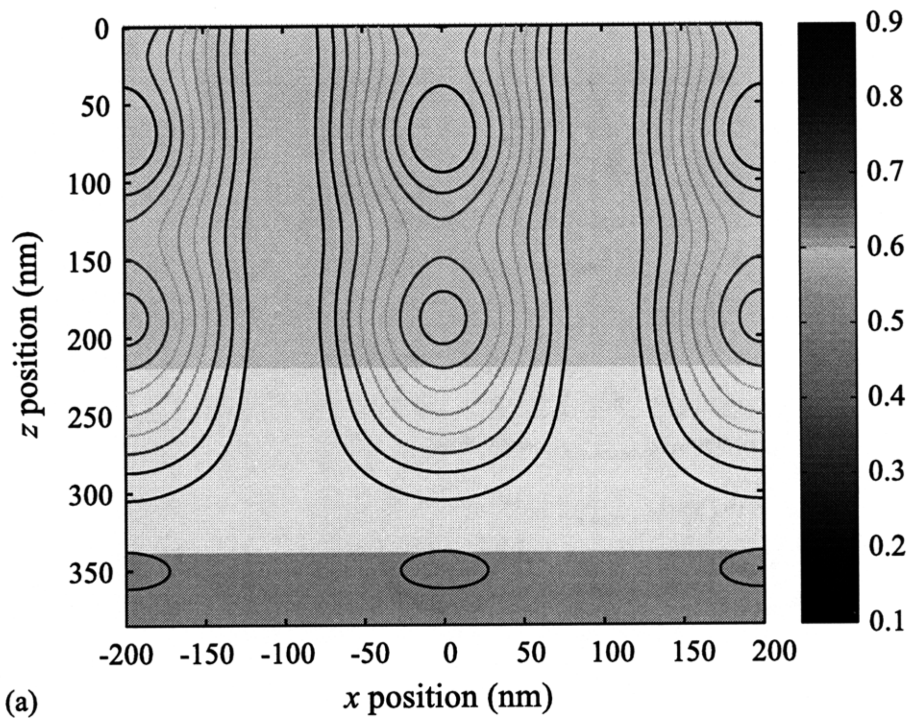
From the simulation results, the designed stack layer consists of 220 nm resist, 120 nm ARC, and 45 nm nitride. The intensity pattern for two-beam interference can then be calculated, using parameters such as the amplitude and phase of each mode. The simulated intensity map in the multilevel layer stack is illustrated in Figure 4-13(a). The figure depicts a cross-section normalized intensity profile of the multilevel stack during exposure for two periods. The sample surface, or the air/resist interface, is given at $z = 0$ nm. The resist, ARC, and nitride layers are shaded in white, yellow, and gray, respectively.

The intensity contours provides additional insight into the exposure process during lithography. The simulation shows that reflection from the resist/ARC has not been completely eliminated, as standing waves in the z-direction still exist. This is a direct result of having 0.5% reflection efficiency by using an ARC thickness of 120 nm. However it has been suppressed and the remaining intensity variation is not a factor during exposure. Also note that the optical fields are being absorbed quite readily as they enter the ARC layer.

The intensity profile can also be used to predict the profile of the residual resist after exposure and development. I use a simple binary model, which will be discussed further in Chapter 5, to simulate the resist profile. In the binary resist model any resist above a defined clipping dose is exposed, while those under the threshold is unexposed. Note that such binary response of the resist is the case when the resist material contrast is infinite.

In this model I use a positive resist, in which the exposed region is dissolved, while the unexposed regions remain. Setting the clipping dose at a relative intensity of 0.2 with exposure time of 1 s, the resulting resist profile is depicted in Figure 4-13(b). As seen in the figure, the wavy intensity contours translate directly into non-vertical resist sidewalls. However, based on the model one can conclude that the design parameters are feasible.

An exposed resist grating cross-section profile is imaged with the SEM and shown in Figure 4-14. Note the slight waviness in the resist sidewalls are consistent with the simulated resist profile shown in Figure 4-13(b). The thick resist then serves as an effective mask to pattern-transfer the grating into ARC for the case of the device grating, and into nitride and silicon in the case of the reference grating.



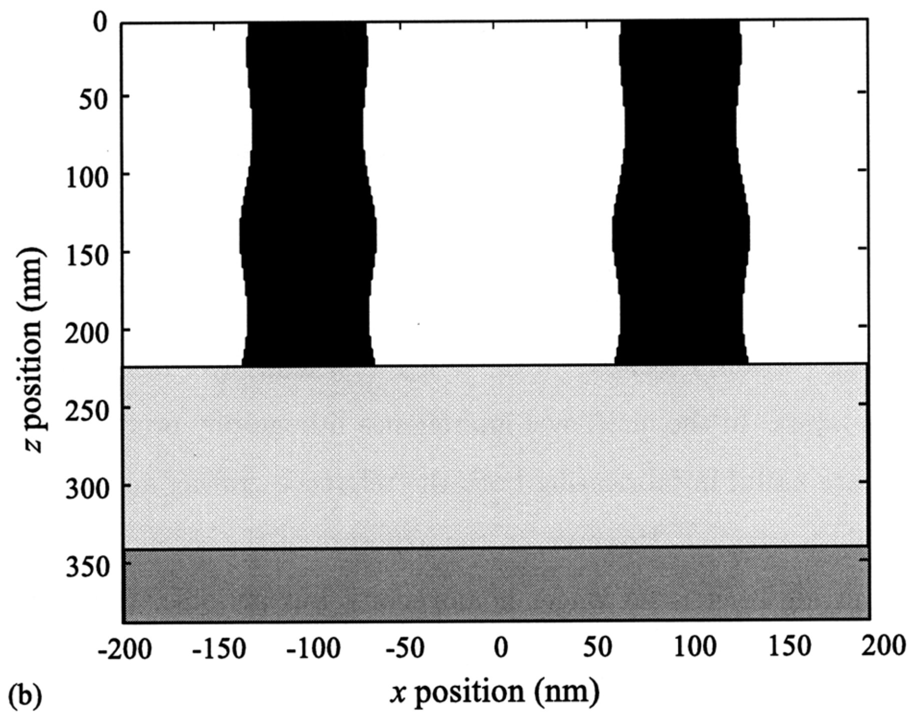


Figure 4-13: Simulated (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride layers.

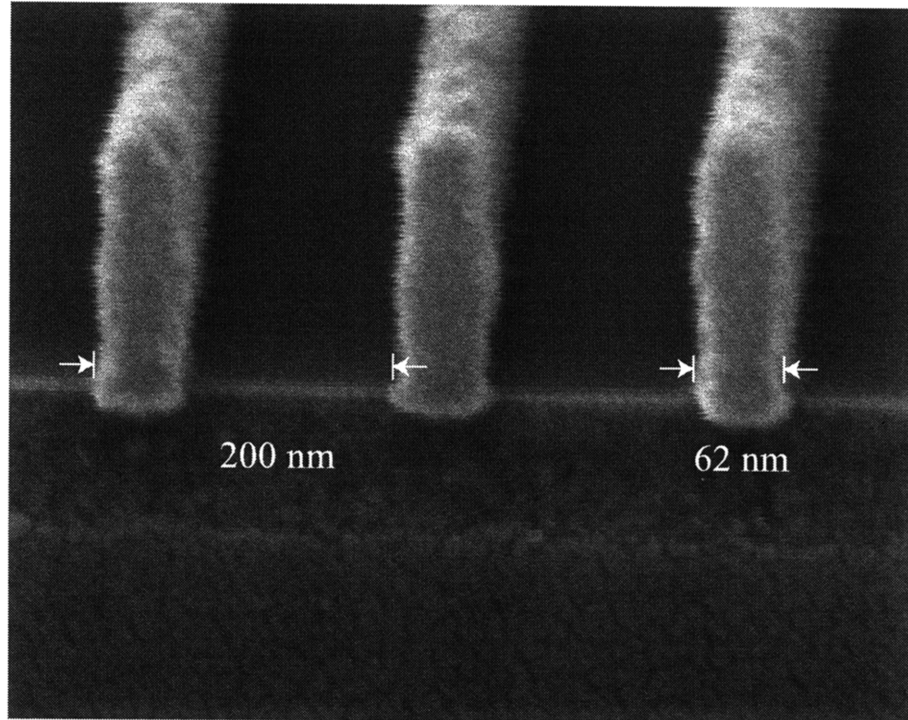


Figure 4-14: Cross-section SEM image of an exposed grating in resist.

4.3 Design of Multilayer Resist Stack over Periodic Layers

In the previous section, I focused on the design of a multilayer resist stack limited to homogenous layers. In the multilevel interference lithography process, such homogenous stacks are useful in fabricating both the reference grating and the first grating level in the device region. However, in the exposure of the second grating level, the underlying nitride layer is no longer homogenous, but periodic. The model for designing the multilayer stack with homogenous layers is thus inadequate, and RCWA needs to be used.

In the multilevel interference lithography process, one of the main challenges is to examine the optical effects of a multilayer stack with both homogenous and periodic layers. This is critical, since in the resist exposure beyond the first grating level, there will be an underlying grating layer. The primary concern is the reflection at

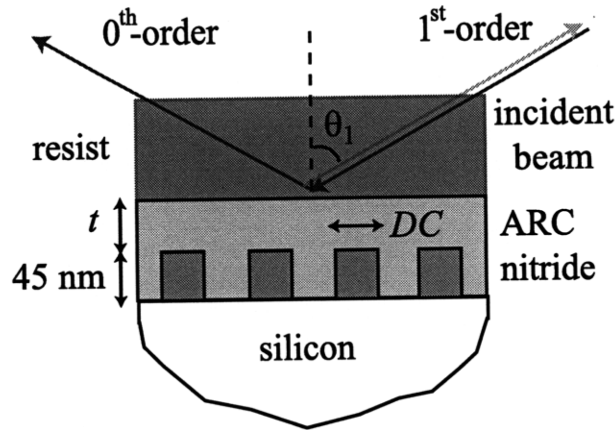


Figure 4-15: Geometry of the homogenous multilevel stack with 220 nm resist, ARC thickness t, and 45 nm nitride on top of a silicon substrate.

the resist/ARC interface, which now includes additional propagating orders due to the periodic layer. I then need to be able to model the optical effects of the changing material and geometric grating layer properties for each lithography step.

4.3.1 Resist Stack Simulation for the Second Grating Level Exposure

To start the analysis, I consider the multilayer geometry shown in Figure 4-15. This layer geometry closely resembles the case for the exposure of the second grating level. The stack consists of 220 nm of resist, ARC thickness t, and 45 nm of ARC/nitride grating layer on top of the silicon substrate. The duty-cycle is defined as the relative linewidth of the nitride to the period. The incident beam satisfies the Littrow condition and has angle of θ_1 in the resist layer.

As illustrated in the figure, there are additional propagating modes due to the underlying grating layer. The area of particular interest is the resist/ARC interface. Previously for the case of homogenous layers there are only reflected orders, but now there is also a back-diffracted first order. The effect of this order on exposure, similar to that of the reflected order, is the addition of a standing wave in the z-direction. In the worst case where both the zeroth and first orders are not minimized, the exposure

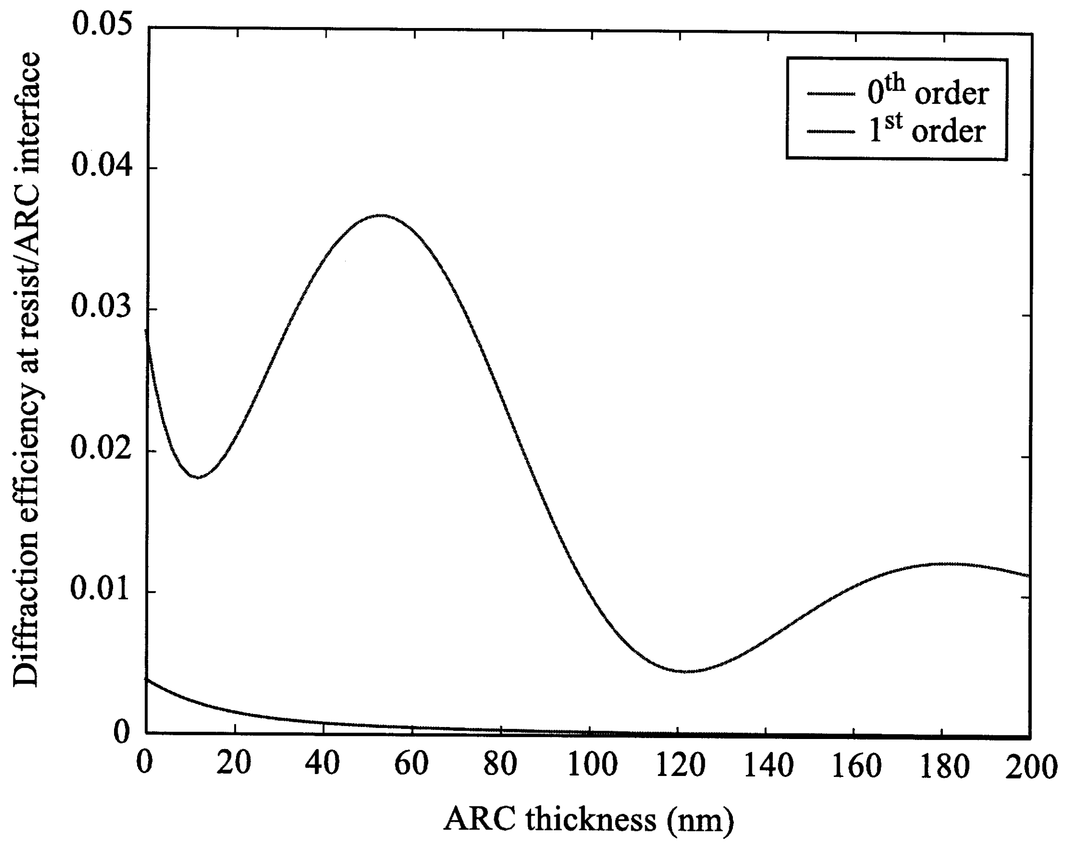


Figure 4-16: Simulated zeroth and first-order diffraction efficiencies at the resist/ARC interface.

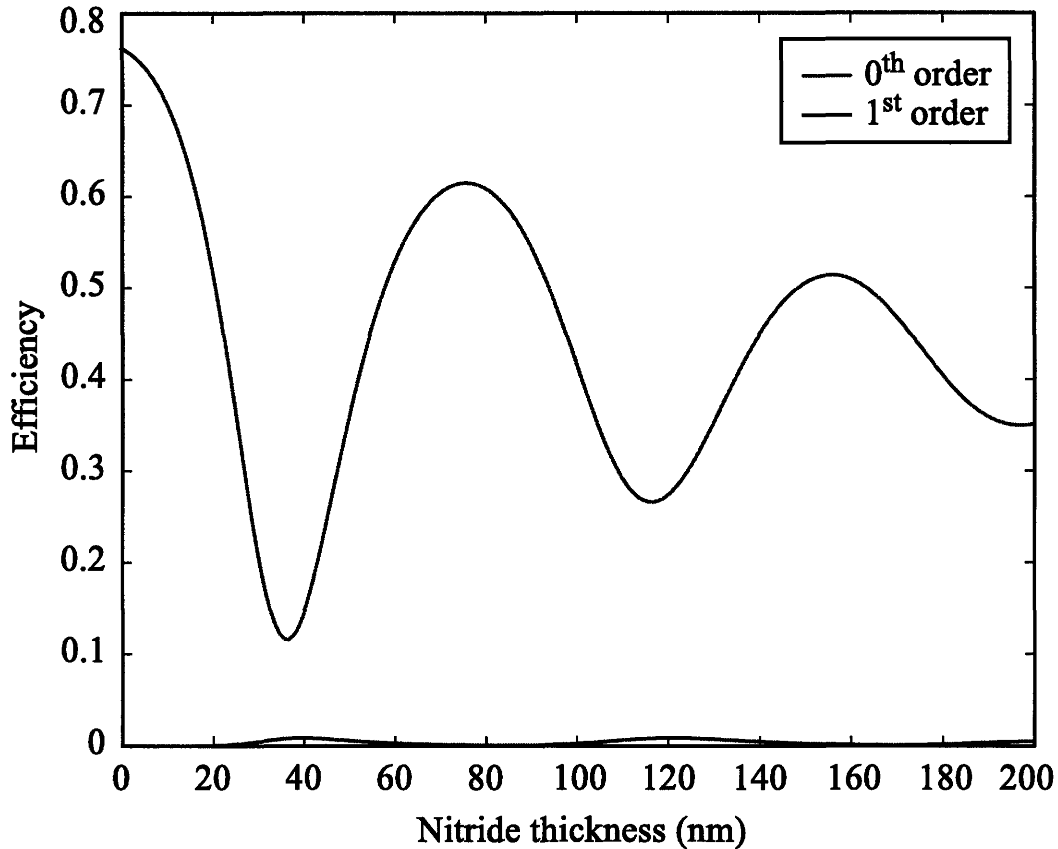


Figure 4-17: Simulated zeroth and first-order diffraction efficiencies for a silicon nitride grating with duty-cycle of 0.875.

pattern will be the result of six-beam interference. In the exposure of a multilayer stack with a periodic layer, an optimal ARC thickness to reduce both the zeroth and first-order diffraction is critical.

To simulate the diffraction efficiency, the multilayer RCWA outlined in Appendix B is used. The zeroth and first-order efficiencies at the resist/ARC interface are plotted as functions of ARC thickness, shown in Figure 4-16. The grating layer duty-cycle is chosen to be 0.875, which is the case for four-fold spatial-frequency multiplication.

The simulation results show the reflected zeroth order demonstrates slightly reduced efficiency, but with similar behavior as the homogenous layer case. The reduction in efficiency is expected, as some intensity has been shifted to the diffracted first

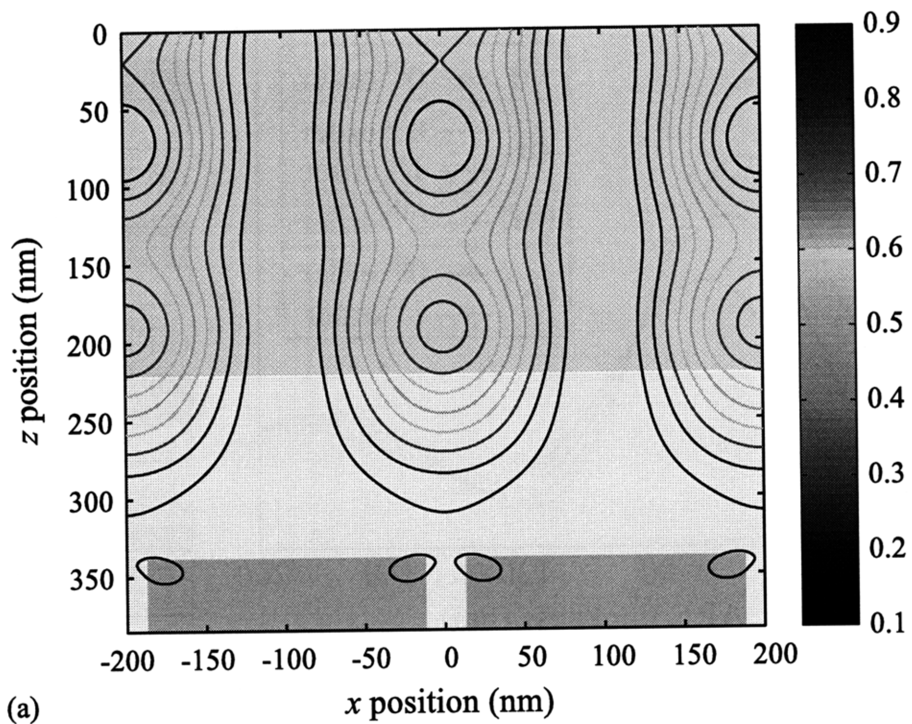
order. The overall efficiency of the first order, however, is significantly lower than that of the zeroth order. This result is due to the choice of duty-cycle for the nitride grating level. The first-order efficiency also decreases rapidly with ARC thickness.

In the previous section I have demonstrated that the zeroth and first order beams from a nitride reference grating can be balanced for duty-cycles of 0.3, 0.4, and 0.5, shown in Figure 4-5. For a nitride reference grating with a duty-cycle of 0.875, there is a dramatic difference in diffraction efficiencies, as shown in Figure 4-17. Thus I can conclude that a nitride grating with such high duty-cycle at this incident angle functions primarily as a reflector, which explains the low first-order diffraction efficiency in Figure 4-16. Therefore having a duty-cycle of 0.875 for the nitride grating, which is necessary in the process for four-fold multiplication, helps reduce the first-order diffraction at the resist/ARC interface.

Examing the efficiency curves in Figure 4-16, an optimal ARC thickness of ~ 120 nm is chosen to minimize both the zeroth and first order. I can then simulate the intensity contour and residual resist profile for two-beam interference lithography, shown in Figure 4-18. In the figure the layer stack consisting of resist, ARC, and nitride grating layers are illustrated in the background.

For this simulated case, the spatial phase of the nitride grating is illustrated, so that the residual resist will be located in the middle of each nitride ridge. Additional fabrication processes, which will be presented in Chapter 6, can then be used to transfer the two grating levels into a single layer. The resulting grating is then spatial-frequency doubled and has a period of 100 nm with duty-cycle of 0.75.

It is important to note that the simulated contours are extremely similar to those plotted for the homogenous layers, shown in Figure 4-13. This is due the first-order in the grating stack being heavily suppressed, while the reflected zeroth order has similar magnitude. These simulation results are encouraging, and arguablely the most desirable scenario in designing the layer stacks. I have shown that by choosing optimal duty-cycle in the grating layer and thicknesses of ARC and nitride, the intensity pattern in the resist layer can be nearly identical for the first and second grating level exposures.



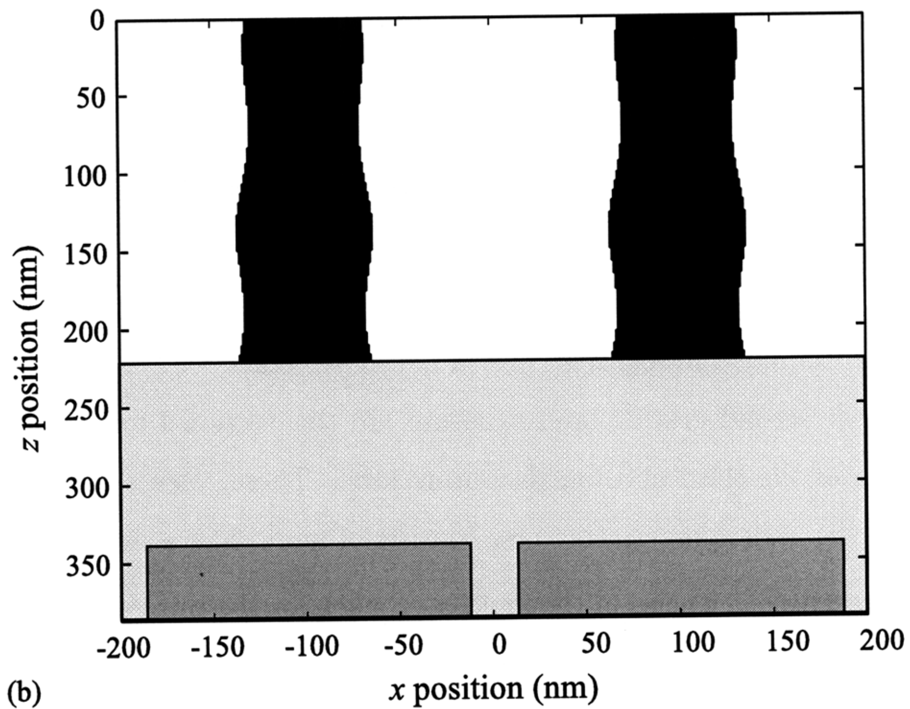


Figure 4-18: Simulated second exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers.

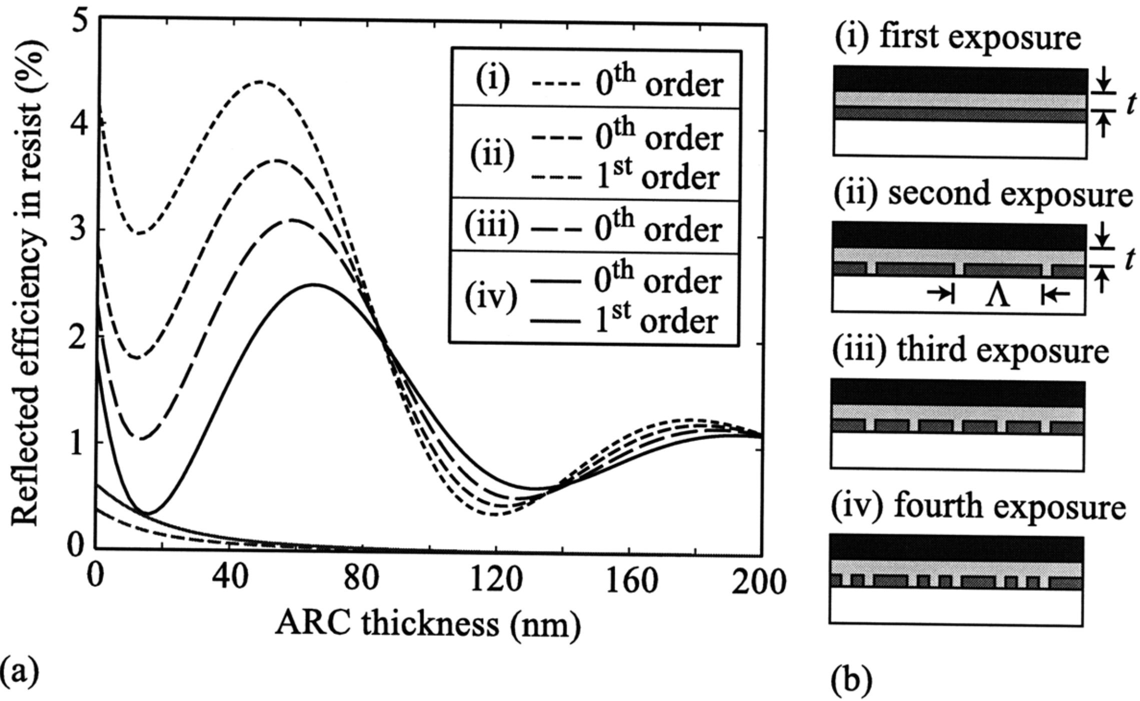


Figure 4-19: (a) Simulated zeroth and first-order diffraction efficiencies, and (b) stack geometry for the first through fourth interference lithography exposures.

4.3.2 Resist Stack Simulation for Third and Fourth Grating Level Exposure

Previously, I examined lithography over a multilayer stack with a periodic layer for the case of the second grating level exposure. In this section I focus on simulating the exposure of the third and fourth grating level. These cases are relatively more complex, as the underlying grating now consists of more than a single grating.

The diffractive efficiencies of the multilayer stack for the third and fourth grating level exposure are simulated and plotted in Figure 4-19(a). The efficiency curves for the first and second exposure are also plotted on the same figure for comparison. For better illustration, the multilayer resist stacks associated with each simulation condition are depicted in Figure 4-19(b). For each case the stack consists of 220 nm of resist, ARC thickness t, and 45 nm of nitride grating/homogeneous layer over silicon substrate. Note that the only difference between the exposure steps is in the

nitride layer while all other parameters remain identical.

For the third grating level exposure, the underlying nitride grating layer has a spatial-frequency doubled period of 100 nm. This geometry is illustrated in Figure 4-19(b) as diagram (iii). For light with $\lambda = 351.1$ nm, the underlying grating is subwavelength and therefore there is no diffraction. The zeroth-order reflection is then the only propagating order, and its efficiency is plotted and denoted as (iii) in Figure 4-19(a). The periodic nitride layer acts as a partially reflecting dielectric mirror.

For the fourth grating level exposure, the underlying nitride layer consists of three 200 nm-period gratings, each with a 50 nm lateral phase-shift from one another. The layer stack geometry is illustrated in Figure 4-19(b) as diagram (iv). This fence-like grating can be thought of as a superposition of a 200 nm-period grating with a 100 nm-period grating. Therefore, as Figure 4-19(a) shows, the diffraction in the resist stack in the fourth exposure has both zeroth and first order. However, similar to the case of second grating level exposure, the efficiency of the first-order diffraction is considerably lower than that of the zeroth-order.

The comparison of the four grating layer exposure steps shown in Figure 4-19 is insightful. One of the most interesting observations is the relative efficiencies for the four different lithography steps. Disregarding the first order, the diffraction efficiency of the zeroth order decreases from the first to the fourth level exposure. This decrease is due to the decrease in volume of the higher index nitride, resulting in a lower effective index in the grating layer. The multilayer stack thus has less index contrast at the interface, and results in weaker reflection.

The efficiency curves generally have similar trends, with slight offsets for the location of optimum ARC thickness. Note while minimizing the reflection is desirable, it is extremely difficult to control layer thicknesses to better than ± 5 nm during spincoating. However, ARC thickness error of ~ 10 nm is unlikely to have catastrophic effects, as the reflected orders still have relatively low efficiencies. To support this argument, while the optimal ARC thickness for both third and fourth level exposures is ~ 130 nm, a thickness of 120 nm is used to simulate the exposure pattern.

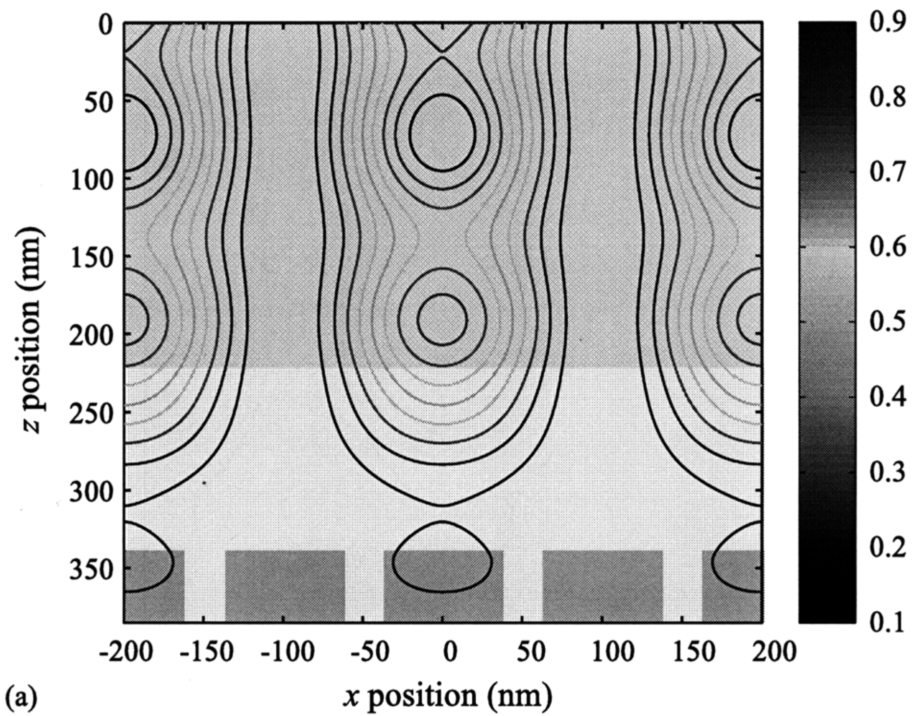
The interference intensity patterns of the third and fourth exposures are shown in Figure 4-20(a) and Figure 4-21(a), respectively. The corresponding simulated residual resist profiles are depicted in Figure 4-20(b) and Figure 4-21(b). The intensity contours can be compared to those simulated earlier for the first and second grating level exposures, shown in Figure 4-13 and Figure 4-18. Note that while slight differences can be observed in the intensity contours of the four exposures, the simulated residual resist profiles are nearly identical.

In this section I have examined the lithography conditions for the third and fourth grating level exposure steps. Using RCWA, a multilayer stack can be designed so that the intensity pattern in the resist layer is not effected by the underlying layers. Such a design goal is critical to ensure process stability for each level of lithography. The simulated residual resist profiles demonstrate that the designed multilayer stacks satisfy this goal.

4.3.3 Exposure Symmetry

The relative grating phase to the two incident beams has important implications during lithography. This is the case because unlike a homogenous layer, a periodic layer does not have continuous translational symmetry. Therefore there are scenarios where the relative grating phase to the left and right interfering beams are different, resulting in asymmetric exposure. When symmetry is satisfied, the grating profile should have even symmetry about the normal direction which bisects the incident beams. In the simulated cases shown in previous sections, the exposure steps were chosen so that the exposure condition was always symmetric.

The concept of exposure symmetry can be readily visualized, as shown in Figure 4-22. The figure illustrates two different exposure orders from the second to the fourth grating level. The exposure order described and simulated in the previous section is shown in Figure 4-22(a), where every exposure is symmetric. A different exposure order is shown in Figure 4-22(b), in which the second grating level is exposed at a $\pi/2$ phase-offset to the first level. The third and fourth grating levels are then exposed at $\pi/2$ additional phase-offsets. In this exposure order the second and third exposures



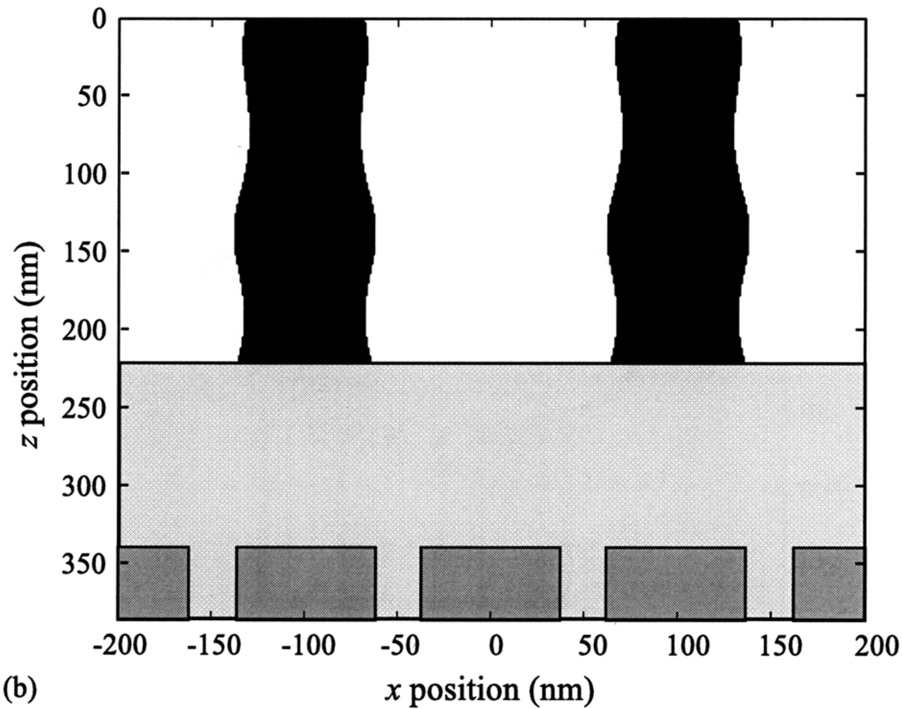
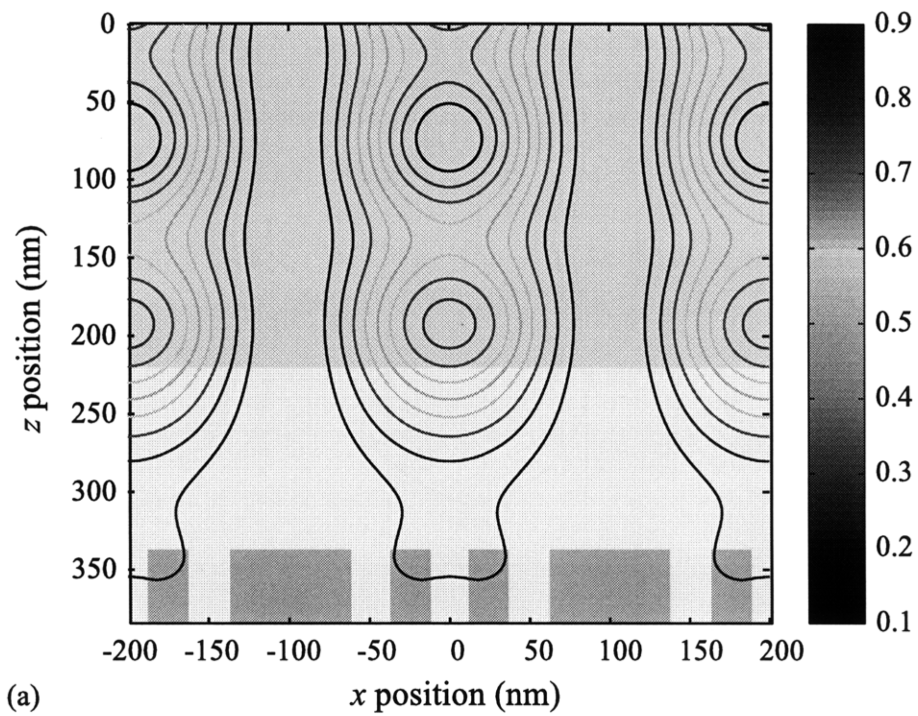


Figure 4-20: Simulated third level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers.



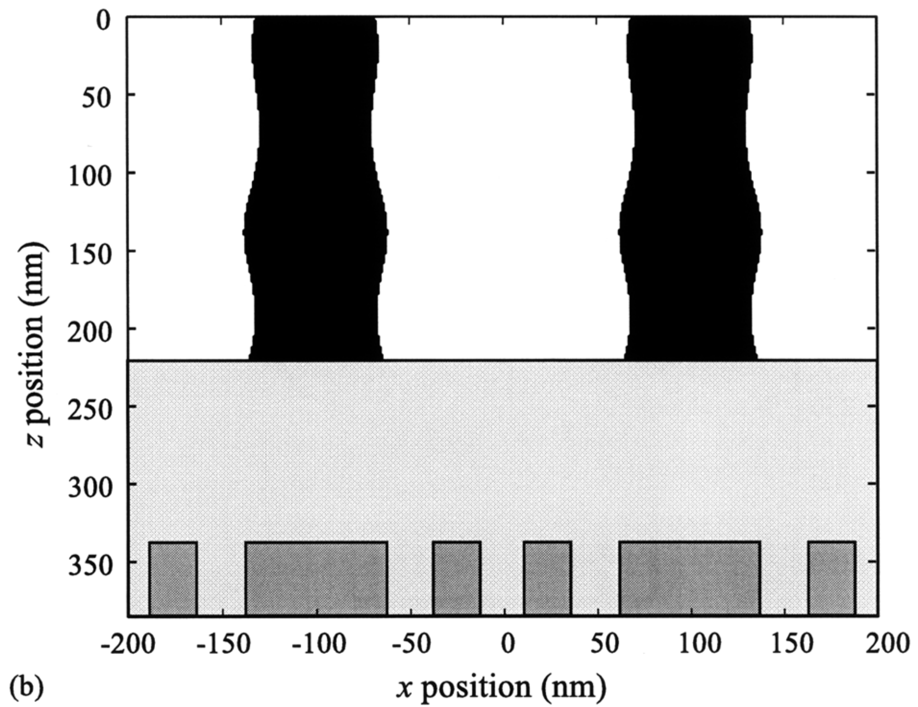


Figure 4-21: Simulated fourth level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 120 nm ARC, and 45 nm nitride grating layers.

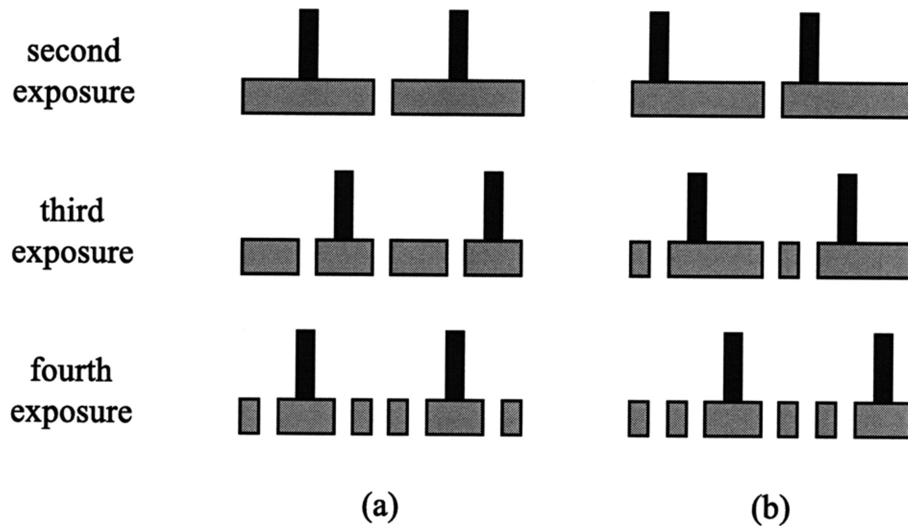


Figure 4-22: Second to fourth grating level lithography for (a) order with all symmetric exposures and (b) order with some asymmetric exposures.

are asymmetric.

One very important point to highlight is that even though a symmetric exposure order can be chosen for four-fold spatial-frequency multiplication, it might not be the case for other applications of multilevel interference lithography. In fact for processes with multiplication factors of 2^N, where N is an integer, there is always an order in which the exposure can be symmetric. However, for other factors, such as 3 and 5, some exposure steps will have to be asymmetric. This is also true for other applications where grating levels might be designed with arbitrary phase-offsets. Therefore it is critical to examine the effects of asymmetric exposures.

Asymmetric exposures can potentially be problematic. While the diffraction intensity efficiencies are not sensitive to the grating phase, the phase of the diffracted order are effected. For example, the phase of the zeroth and first-order diffraction from the reference grating is plotted in Figure 4-23. The phase plotted is the relative phase calculated at the air/grating interface. The results show that the phase of the zeroth order is independent of the grating phase, while the phase of the first order changes as a linear function of the grating phase. These dependencies are quite intuitive, as the average index of the grating is constant, resulting in no phase-shifts for the

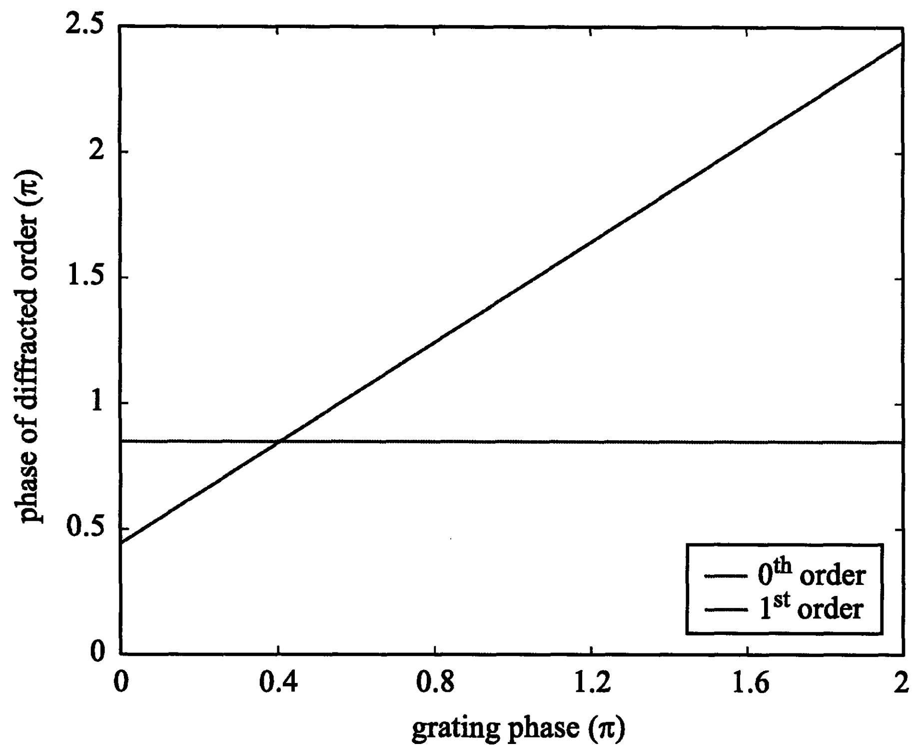


Figure 4-23: Simulated phase of the zeroth and first-order diffraction from reference grating consisting of 45 nm of nitride grating over silicon substrate.

reflected order. Note that these results are the fundamental bases for phase-shifting interferometry, which are discussed in more detail in Chapter 3.

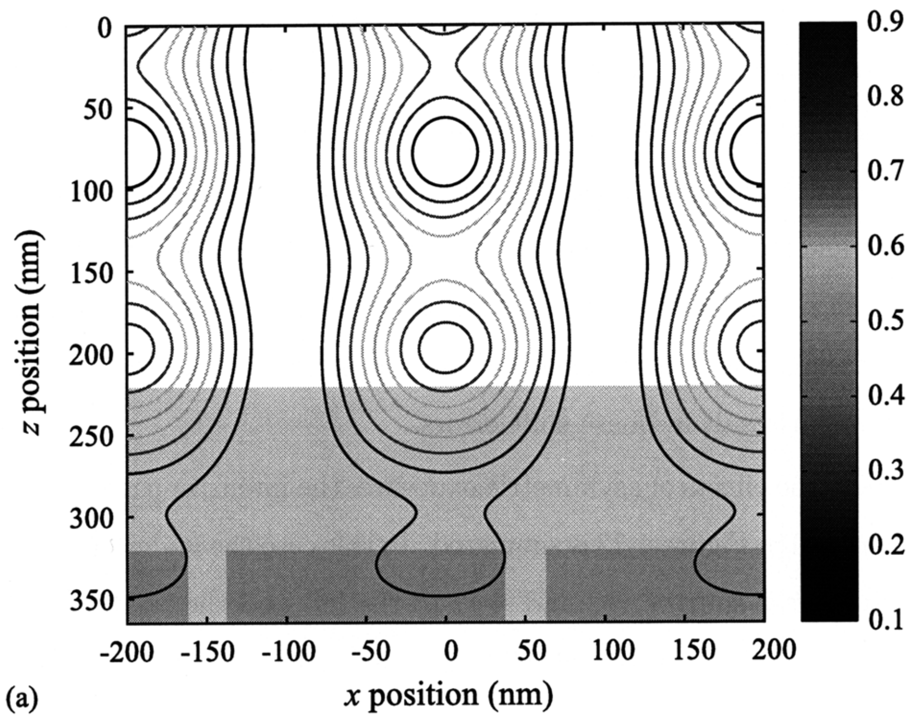
Based on this result, I can limit the focus on the reflected first-order diffraction at the resist/ARC interface. The main concern is that since the first-order diffraction from the left and right incident beams have different phase, the resulting interference intensity pattern may have a lateral phase-shift. As a consequence, even if the grating level are aligned with any desired spatial phase to previous grating pattern, the asymmetry might result in phase-shift errors.

To examine the effects of asymmetric exposure, the intensity pattern of the second exposure outlined in Figure 4-22 is simulated. In this case the underlying grating layer has a relative phase-shift of $\pi/2$ and $-\pi/2$ to the left and the right incident beams, respectively. The stack consists of 220 nm resist, 100 nm ARC, and 45 nm nitride grating layers. The simulated intensity contour and residual resist profile are shown in Figure 4-24.

From Figure 4-24(a), a slight asymmetry in the intensity contour can be detected. This is especially true in the nitride grating layer, in which the optical wave is guided towards the nitride, which has higher index. The asymmetric effect in the resist layer is significantly lower, due to the fact that the intensity of the diffracted orders has been greatly reduced by the ARC layer. As expected, the incident beams remain the more dominant factor in the intensity profile.

The resist profile, shown in Figure 4-24(b), can yield more interesting results. Upon visual inspection the profile seems to yield perfectly symmetric profile. However, by calculating the centroid of each individual linescan in the z-direction from $z = 0$ to $z = 220$ nm, quantitative information can be retrieved. The centroid shifts are then calculated with respect to its desired position (for example $x = 100$ nm), and then plotted as function of resist depth, shown in Figure ??.

The curve shows that the resist has non-uniform lateral phase-shifts that are dependent on depth. However, the magnitude of the shifts are extremely small, with peak-to-valley error of 2 nm. The average phase error is around -0.5 nm, and phase error of this magnitude is insignificant. The reason for the small asymmetric effect



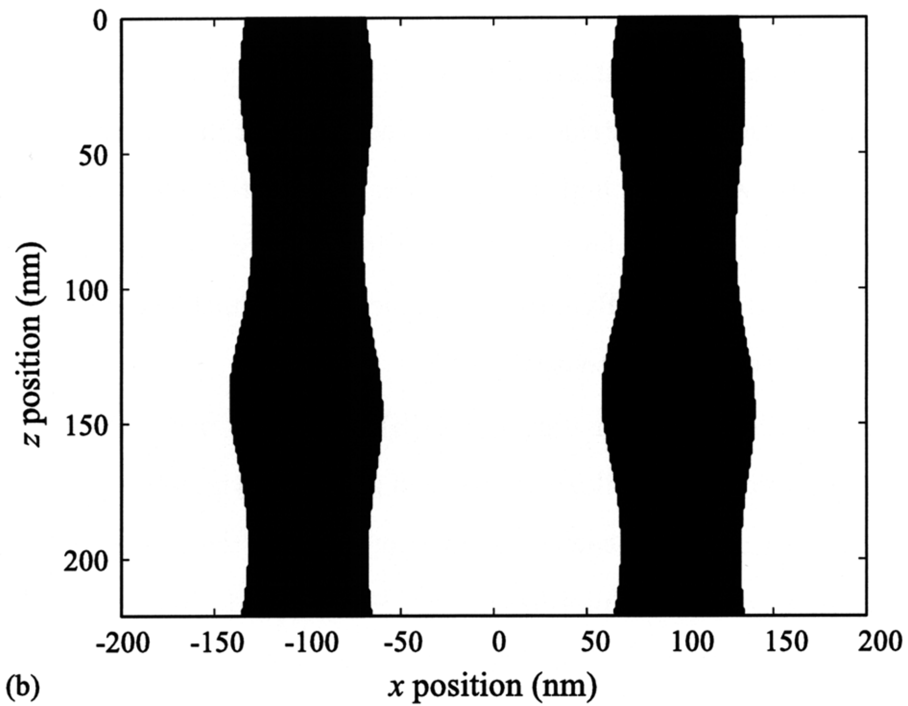


Figure 4-24: Simulated asymmetric second level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, 100 nm ARC, and 45 nm nitride grating layers.

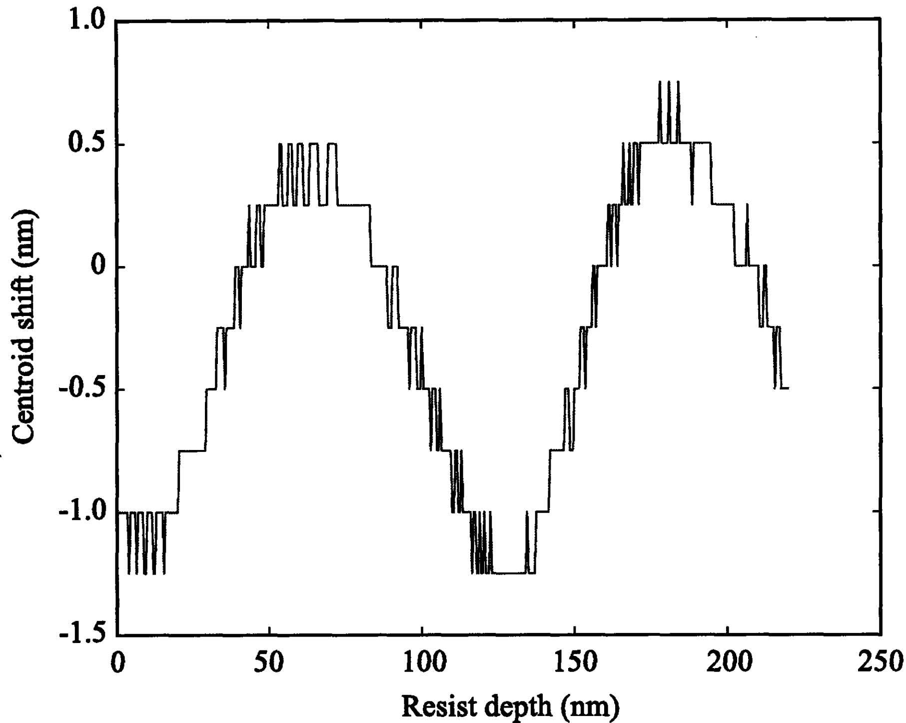
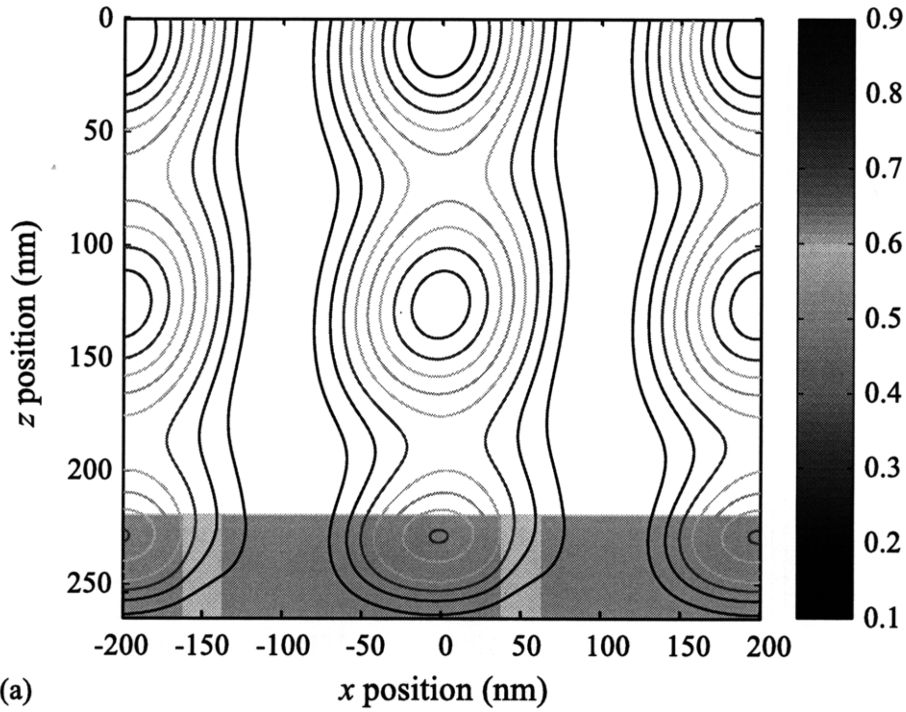


Figure 4-25: Lateral shifts of calculated resist centroid as a function of resist depth.

is partly due to the near optimal ARC thickness. Having a carefully designed ARC layer is even more critical for asymmetric exposures.

For comparison, I simulated the extreme case of an exposure with the same parameter except having no ARC layer. The intensity contours and residual resist profile are shown in Figure 4-26(a) and (b), respectively. The asymmetric effect in this case is much greater, and the resist profile is visually distorted. This is a direct result of the first-order diffraction having higher relative intensity. Using the same method, the centroid profile of the resist is plotted in Figure 4-27. From the figure a peak phase error of nearly -12 nm is observed, which is quite dramatic considering the desired linewidth of the resist grating is around 25 nm.

The two asymmetric exposures for the second grating level demonstrate the upper and lower limits of possible phase distortion. The two cases represent two extreme cases of the ARC thickness being near-optimal and having no ARC layer. These results again demonstrates the importance of designing and optimizing a multilayer



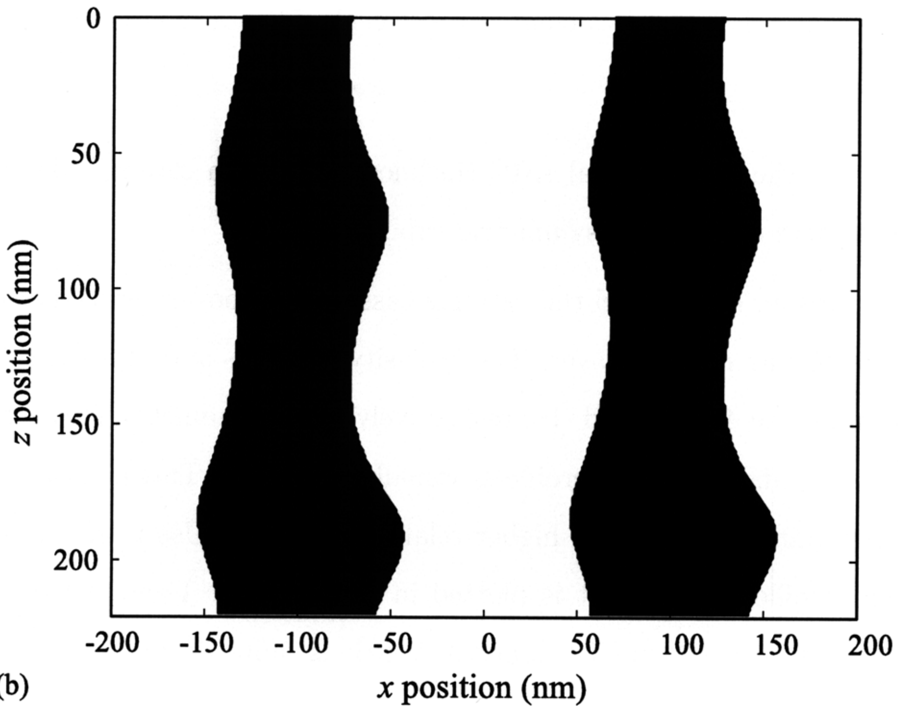


Figure 4-26: Simulated asymmetric second level exposure (a) interference intensity pattern, and (b) residual resist profile of the designed multilevel layer stack with 220 nm resist, and 45 nm nitride grating layers.

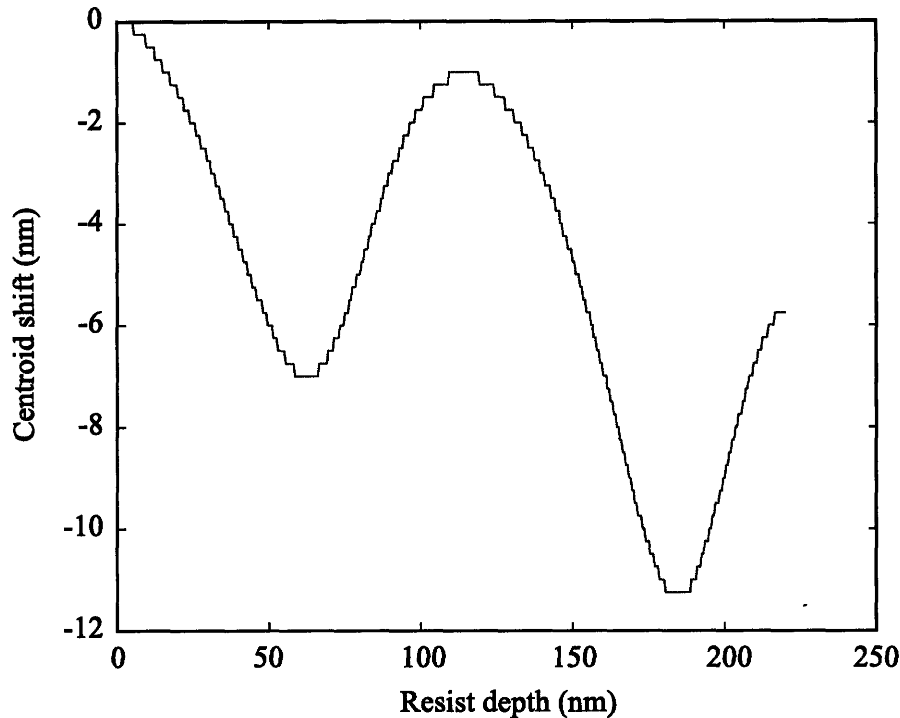


Figure 4-27: Lateral shifts of calculated resist centroid as a function of resist depth.

resist stack for each grating level exposure.

4.4 Summary

In this chapter I examined the design of optical multilayer stacks for various applications in the multilevel interference lithography process. RCWA and transfer matrix methods were implemented to simulate optical fields in these multilayer periodic nanostructures to achieve the desired design criteria. Critical issues such as the reference grating, resist stack for different grating level exposures, and asymmetric exposures were investigated.

I was able to design reference gratings with balanced zeroth and first-order diffraction efficiencies, allowing high signal contrast during phase measurement. The design of the multilayer resist stack for grating level exposures has the different requirement of minimizing back-diffracted efficiencies. By carefully designing the layer

stack, similar exposure conditions in the resist layer can be achieved for different exposure steps.

Lastly I examined the issue of asymmetric exposures, in which the grating has different phase-offsets to the left and right exposure beams. I found that the asymmetric effect is governed by the multilayer stack, and a carefully designed layer stack and minimize its effect. These studies are extremely critical and have exciting implications for future applications of the multilevel interference lithography process.

Chapter 5

Dose and Contrast Simulation

In the multilayer fabrication process described previously in Chapter 2, the linewidth control of each layer is essential. To achieve high order spatial-frequency multiplication, high duty-cycle is essential. Furthermore, the linewidth must be repeatable between layers. In the case of spatial-frequency multiplication, differences in linewidth between grating layers will result in residuals of the fundamental frequency. It is then critical that the linewidth of each grating layer be controlled to a small fraction of the period. In order the have a working multilayer process, linewidth control in the nm-range needs to be achieved.

There are a number of factors that contribute to the linewidth during the numerous fabrication processes. The linewidth is typically defined in the lithography steps, and then its variation in the subsequent steps are controlled. The important factors in the lithography step are exposure dose and contrast of the image grating, which will be described further in this chapter. Other processing variables such as post-exposure bake temperature and time, development temperature and time are kept constant to allow repeatability of the process. The pattern transfer processes, such as RIE, might also introduce linewidth variations and needs to be carefully monitored.

In order to have high precision linewidth control, a crucial concept to keep in mind is process latitude. Throughout the various fabrication steps there are process parameters that might change with time, or might be simplify uncontrollable. It is important that the process have enough process latitude so that it can tolerate minor

parameter fluctuations. In this chapter we examine the various exposure conditions that determines linewidth.

5.1 Dose and Contrast Modeling

Lithography is the most important step in determining the linewidth of the grating. During lithographic exposure, the photoresist, a photo-sensitive polymer material, absorbs photons and starts chemical reactions. Depending on the amount of energy received locally, the photoresist will either be exposed or unexposed. After developing, most of the time in a base solution, the exposed/unexposed region dissolves, depending on the tone of the resist. The linewidth of a periodic pattern is directly related to the ratio of the exposed/unexposed areas.

In interference lithography the energy per unit area received, or dose, for a sinusoidal intensity pattern in x-direction is given as,

$$D(x) = \int I(x,t)\ \mathrm{d}t. \tag{5.1}$$

As shown in Eq. 5.1, the amount of dose received at each x-position is dependent on the time integral of the intensity pattern. Note if the intensity pattern is not time variant, dose is linearly proportional to exposure time. In order to control linewidth, it is essential to model the exposure and development process.

The exposure and development of photoresist can be modeled with many methods. There are commercially available software, such as PROLITH^{TM} [41], that is capable of simulating three-dimensional developed resist patterns for a complex optical system. There are models that predict the profile of resist by simulating the chemical diffusion during developing, but due to the highly non-linear nature of certain resist chemistries, a simple binary model is sometimes sufficient to estimate the linewidth.

5.1.1 Binary Resist Model

The binary model is a simple technique to model linewidth, in which the resist is assumed to have infinite material contrast. In this model, as shown Figure 5-1,

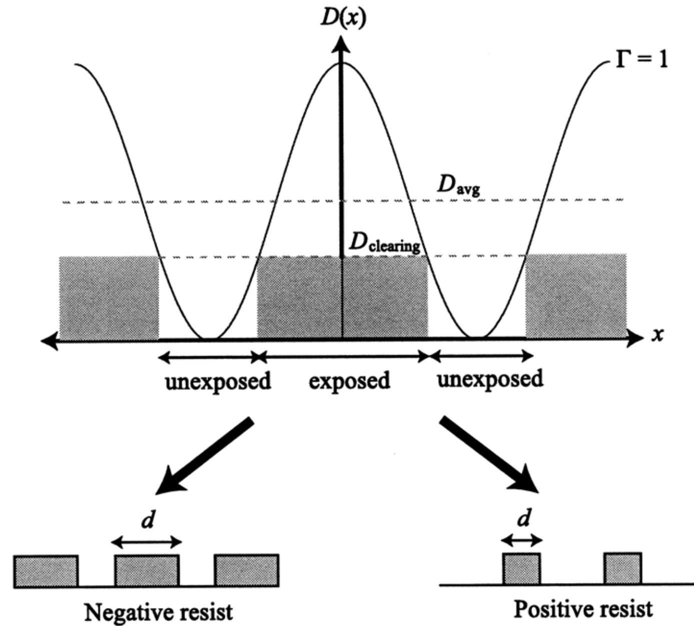


Figure 5-1: The binary photoresist model for exposure and development. Only the resist above the clearing dose is exposed. Linewidth can be controlled by increasing or decreasing the exposure dose.

photoresist regions receiving a dose above the clearing dose are fully exposed, while all the other regions are unexposed. Then for positive resist the exposed regions are fully removed by the developer, while for negative resist the unexposed regions are removed. The linewidth of the grating is then defined as d for both positive and negative resist.

In the binary model there are only two parameters, exposure dose and contrast, that determine the linewidth of the grating. The effect of dose is apparent, and it is the most common way to control linewidth. As depicted in Figure 5-1, by increasing the average dose D_{avg}, so that $D(x)$ scales vertically, more regions will have dose above the clearing dose. The exposed region thus increases as average dose is increased, and the opposite is true as average dose is decreased. Using both positive and negative resists, the linewidth of resist can thus be very efficiently controlled from 0 to approaching Λ by using $D_{avg} < D_{clearing}/2$ and $D_{avg} = \infty$, respectively. This method of controlling linewidth is favorable, as dose can be easily controlled by

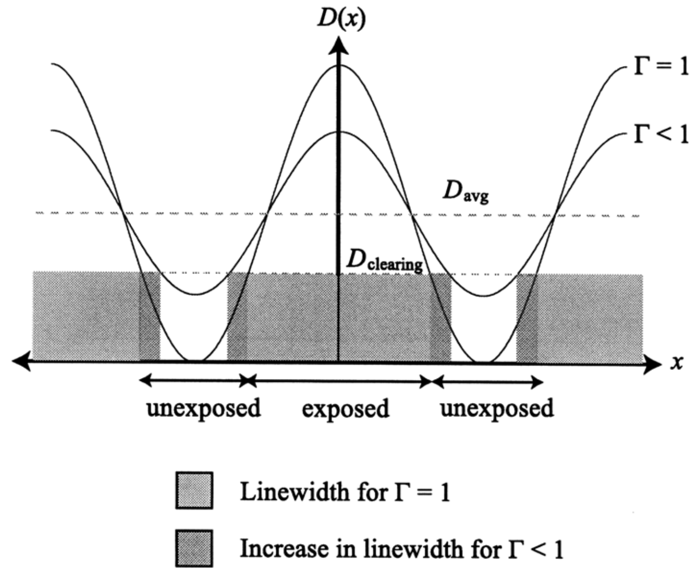


Figure 5-2: The effects of contrast degradation for constant average dose. The yellow and light blue regions depict the linewidth of perfect and reduced contrast, respectively.

exposure time, shown previously in Eq. 5.1

The other parameter is contrast, denoted as Γ, and it is the visibility of the fringes relative to the background intensity. It can also be described as the ratio of the AC to the DC intensity terms. Contrast of a sinusoidal intensity pattern is defined as,

$$\Gamma = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{5.2}$$

The lowest possible contrast is when $\Gamma = 0$, when the intensity pattern is purely DC. At the other extreme, a perfect contrast is achieved when $\Gamma = 1$, at which the amplitude of the AC term is the same as the DC term. With perfect contrast, such as the case shown in Figure 5-1, the intensity minimum is purely black and receives no exposure at all. This is the ideal case to control the linewidth of resist with dose, as it will take infinite dose for all the resist to be exposed. In reality during interference lithography the contrast is not perfect, and sufficient exposure over the clearing dose will fully expose the resist.

Since perfect contrast is unattainable during interference lithography, it is important to analyze the effects of contrast on linewidth. Figure 5-2 describes a scenario in

which the contrast reduces from $\Gamma = 1$ to $\Gamma < 1$ while the D_{avg} remains the same. It is evident from the illustration that the exposed region increases with the reduction of contrast, as shown in light blue. Thus it is established that the linewidth of resist is dependent on both the average exposure dose and contrast of the interference fringes.

In a binary model the linewidth can be written as a function of average dose and contrast. The dose profile during interference lithography, as depicted in Figure 5-1 and Figure 5-2, can be expressed as

$$D(x) = D_{avg}[1 + \Gamma \cos(\frac{2\pi}{\Lambda}x)]. \tag{5.3}$$

Letting $x_{clearing}$ be the x-position at which $D(x) = D_{clearing}$, Eq. 5.3 can be written as

$$D_{clearing} = D_{avg}[1 + \Gamma \cos(\frac{2\pi}{\Lambda}x_{clearing})], \tag{5.4}$$

where the resist in region $x \leq |x_{clearing}|$ will be exposed. The linewidth of the exposed region is then $2x_{clearing}$, and Eq. 5.4 can be rearranged to be

$$\frac{d}{\Lambda} = \frac{1}{\pi}\cos^{-1}[\frac{1}{\Gamma}(\frac{D_{clearing}}{D_{avg}} - 1)]. \tag{5.5}$$

For a periodic structure (or periodic time signals for temporal analogy), the linewidth to period ratio, $\frac{d}{\Lambda}$, is also known as duty-cycle. The duty-cycle will be used extensively throughout this chapter, as it is useful in being a scaled quantity. Another important term in is $\frac{D_{avg}}{D_{clearing}}$, which defines the amount of overexposure relative to the clearing dose.

Using Eq. 5.5, the duty-cycle of both positive and negative resist for different image contrast and dose can be predicted. In Figure 5-3 contrast from 0.7-1 and $\frac{D_{avg}}{D_{clearing}}$ from 0-10 are simulated. The duty-cycle decreases with overexposure for positive resist, and vice versa for negative resist, as expected. It is important to note the effects of fringe contrast even with constant dose, as duty-cycle changes dramatically for both positive and negative resist as contrast is varied.

Even though it is possible to simulate the duty-cycle for a given fringe contrast and exposure dose, it is difficult to accurately control linewidth. It is important to examine the process latitude of contrast and dose variation. This is especially

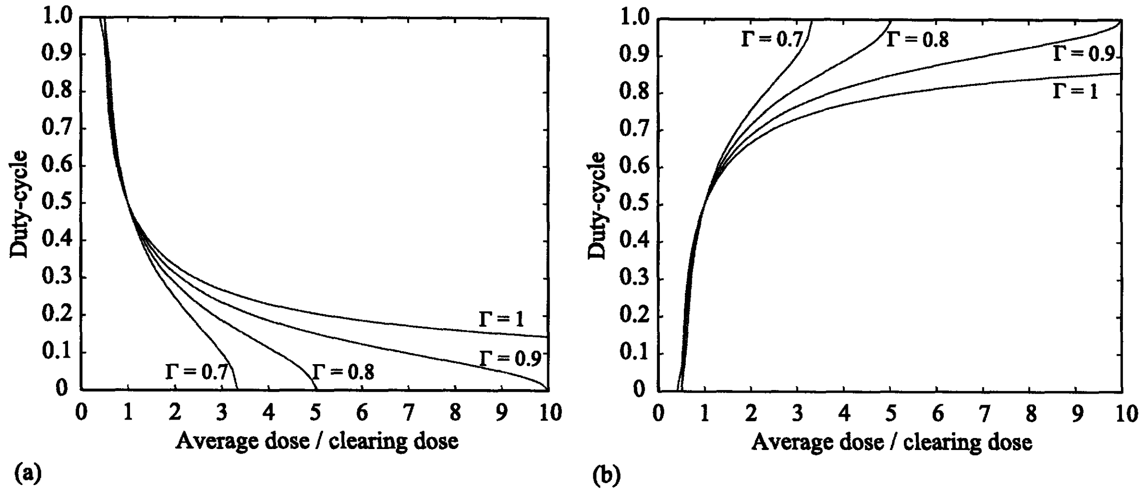


Figure 5-3: Duty-cycle with different interference fringe contrast as functions of overexposure in a binary resist model for, (a) positive and, (b) negative resist.

critical at duty-cycle near 50% or in the highly overexposed cases. These issues will be explored further in later sections. Using the binary resist model and Eq. 5.5, the response of resist during lithography can be predicted.

The binary resist model is an effective simulation tool in roughly predicting the duty-cycle of resist grating. However, it is important to remember the simplicity of the model, as the theory stems only from dimensional arguments. The binary model does not take into account depth related factors such as absorption and bleaching of resist. Therefore it will have difficulties simulating resist much thicker than the period being exposed. In those cases the resist grating will tend to have sloped walls. The model also ignores any time-dependent effects, such as chemical diffusion and development rates. Thus chemically amplified resists do not follow the model well due to their diffusion-based chemistry.

5.1.2 Lithography with Nanoruler: Optical Considerations

The exposure process of the Nanoruler is more complicated than that of traditional interference lithography. This is a direct result of having a dynamic scanning lithography system as oppose to one that is static. Previously in the binary model formula-

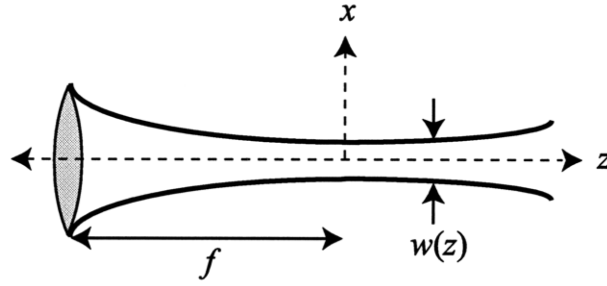


Figure 5-4: A Gaussian beam focused by a lens with focal length of f. The beam radius $w(z)$ and wavefront radius of curvature $R(z)$ both depend on propagation distance z. The wavefront is planar at the beam waist, shown here at $z = 0$.

tion, a sinusoidal intensity pattern with linear phase was used. Under this assumption the intensity pattern is perfectly periodic, and as a result the linewidth is invariant with position. Such linear phase pattern can only be achieved theoretically by interfering two plane waves, as shown previously in Chapter 2. Experimentally this is very difficult to realize in IL, as various issues such as particles and air turbulence will introduce phase distortion in the grating.

In the SBIL scheme, Gaussian beams are interfered, and the image grating phase is not linear. In a focused Gaussian beam, as shown as Figure 5-4, the beam radius and wavefront's radius of curvature are both functions of position. The beam's $1/e^2$ intensity radius is given by,

$$w(z) = w_o\sqrt{1 + (\frac{z_o}{z})^2}, \tag{5.6}$$

where w_o is the beam waist radius at $z = 0$ and $z_o = \frac{\pi w_o^2}{\lambda}$ is the Rayleigh range. The wavefront's radius of curvature $R(z)$ can be described as

$$R(z) = z[1 + (\frac{z_o}{z})^2]. \tag{5.7}$$

Note the radius of curvature is ∞ at $z = 0$, which is the location of the beam waist. Thus at the beam waist the wavefront of the Gaussian beam is planar, making it the ideal location to interfere.

The interference lithography optical setup for the Nanoruler is shown in Figure 5-5, depicting two Gaussian beams interfering at the beam waists. For each arm the

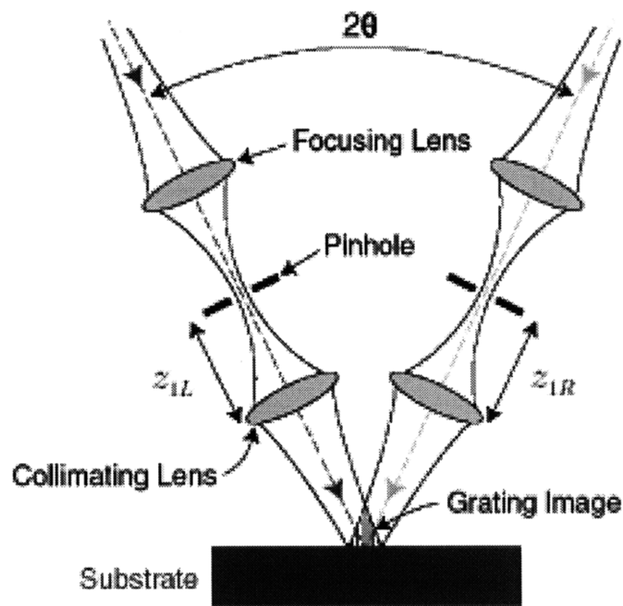


Figure 5-5: The interference lithography setup in the Nanoruler. Two Gaussian beams interfere at the beam waists.

two lenses form a $4F$ system, using the pin hole as a spatial-filter at the Fourier plane. This is critical as the high spatial-frequency distortions in the wavefront are low-pass-filtered. After the spatial-filter the only optical element is another lens, which is used to focus the Gaussian beam at the wafer plane.

An important point to note is that even though the two Gaussian beams are interfered at the beam waist, the image grating is still not perfectly linear. This is due to the oblique incidence of the beams, as the wafer plane does not match the planar wavefront at the beam waist. The resulting image grating has linear phase in the center line region, but is distorted off-center. Previously, Dr. Carl Chen has simulated that for a perfectly aligned image grating with period of 400 nm, the phase distortion across a 1 mm $1/e^2$ diameter beam to be less than 0.04 nm [15]. The phase error is negligible in this case, but becomes significant if the beams are not interfering at the waist. For example, Dr. Chen showed that if the lens of one arm is displaced by 80 μm, the phase non-linearity increases to 2 nm [15].

Due to the periodic scanning operation of the Nanoruler, the effects of spatial-

phase non-linearity is not trivial. Any phase gradients will result in possible contrast reduction, dose variation, and grating phase distortion. Another important result is that the three parameters will be position dependent in the step-over direction. It is important that these variations be modeled.

The fringe intensity pattern formed by interfering two Gaussian beams, as depicted in Figure 5-5, can be described as,

$$I(x,y) = A_R^2(x,y) + A_L^2(x,y) + 2A_R(x,y)A_L(x,y)\cos(\frac{2\pi}{\Lambda}x + \phi(x,y)), \tag{5.8}$$

where $A_R(x,y)$ and $A_L(x,y)$ are the 2D electric field amplitude profiles of the right and left beam, respectively, $\phi(x,y)$ is the non-linear phase distortion, and $\frac{2\pi}{\Lambda}$ is the spatial grating vector, k_g. Note any linear phase distortion $\Delta\phi_{linear}x$ shifts the spatial-frequency of the grating, and the result grating vector is changed to $\frac{2\pi}{\Lambda} + \Delta\phi_{linear}$. The non-linear phase distortion plays a critical role, and will be examined further later in this section. Eq. 5.8 can be expressed as,

$$I(x,y) = I_{avg}(x,y)[1 + \Gamma(x,y)\cos(\frac{2\pi}{\Lambda}x + \phi(x,y))], \tag{5.9}$$

where

$$I_{avg}(x,y) = A_R^2(x,y) + A_L^2(x,y) \tag{5.10}$$

is the 2D average intensity map, and

$$\Gamma(x,y) = \frac{2A_R(x,y)A_L(x,y)}{A_R^2(x,y) + A_L^2(x,y)} \tag{5.11}$$

is the 2D contrast map. Intuitively, the average intensity, as stated in Eq. 5.10, is obtained by summing the intensity profiles of the two beams. The contrast map depends on amplitude profiles of the beams, which are obtained by the square-root of their respective intensity profiles.

An interesting observation can be made as a result of Eq. 5.11, which addresses the question of beam overlap. The contrast at any given (x_o, y_o)-position is determined by the balance of intensity from the two beams. If the intensities are perfectly balanced, the contrast $\Gamma(x_o, y_o) = 1$. In the extreme case that the intensity of either beam is 0, then the contrast $\Gamma(x_o, y_o) = 0$, resulting in no interference fringes. This result is also quite intuitive, as at least two beams are needed for optical interference.

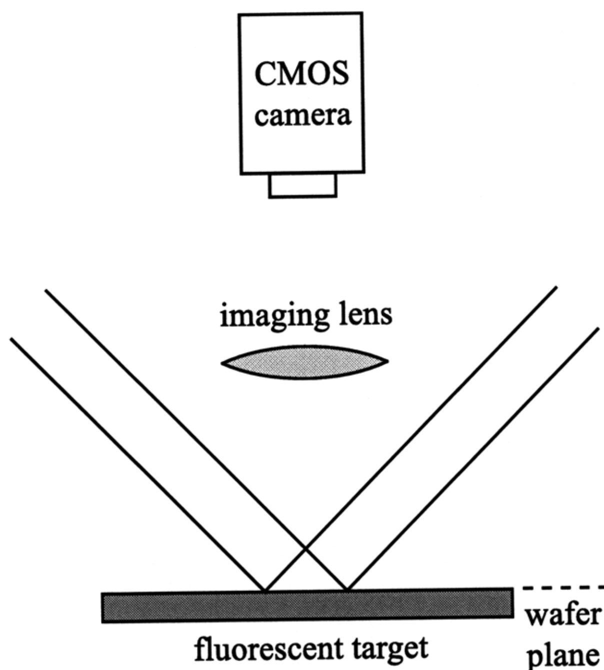


Figure 5-6: The fluorescent imaging system used to measure the intensity profiles of the beam on the wafer plane.

To illustrate these conclusions experimentally, a CMOS camera was used to image the intensity profiles at the wafer plane. The CMOS camera has difficulties imaging coherent sources, as the glass protection plate forms a Fabry-Parot cavity, and the images contain interference fringes. So instead a fluorescent target is placed at the wafer plane, and the CMOS camera is placed above. The setup is shown in Figure 5-6. Since the fluorescent image is not coherent (both temporally and spatially), the cavity will not produce interference fringes. An achromatic double with magnification of 4/3 was used as the imaging lens.

Using the CMOS imaging setup, the intensity maps of the left and right interfering arms with incident angles of 17.81 ^{o} (for $\Lambda = 574$ nm) were taken separately. The two intensity contours are plotted together in Figure 5-7(a), and the sum is plotted in Figure 5-7(b). Note that the individual beam spots are not well-matched, as the left beam appears elliptical. There is also a slight offset between the two beam spots. Note the summed intensity of the two beams appears more rounded.

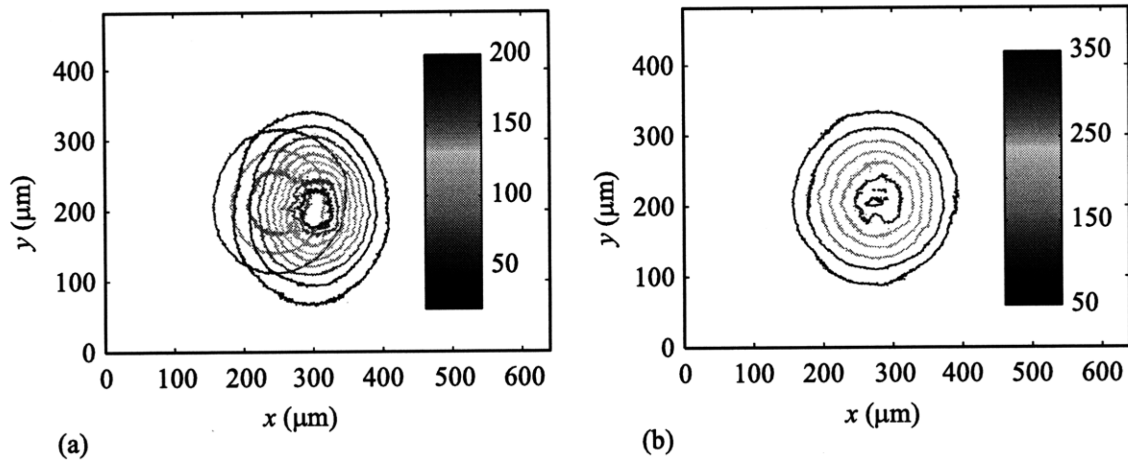


Figure 5-7: (a) The intensity maps of the left and right beams, and (b) the sum of the maps.

The contrast map, as calculated by Eq. 5.11, is shown in Figure 5-8. The contrast of the two poorly-matched beams drops off significantly from the center to the edge, with values of 0.9 and 0.7, respectively. The contrast then drops abruptly to 0 past the left edge, which is where the left beam stops. When compared to the sum of the intensity map, which is what the eye observes visually, the contrast map offers much more insight into the quality of the image grating.

The concept of contrast can also be analyzed in the frequency domain, as pointed out by Dr Juan Montoya [56]. Such a method is convenient since contrast is essentially a ratio of the AC and DC components of the intensity pattern, and thus becomes more intuitive in the frequency domain. The disadvantage of this method is that spatial information is omitted, making it impossible to obtain information as a function of position.

The frequency domain analysis, however, is especially useful in determining the optimal beam overlap of the two arms. As shown previously in Figure 5-8, two poorly overlapped beams result in contrast contours that fall off quite dramatically. In such a case the beams need to be realigned to improve contrast before lithography. It is then useful to determine the optimal overlap by examining the spectral components.

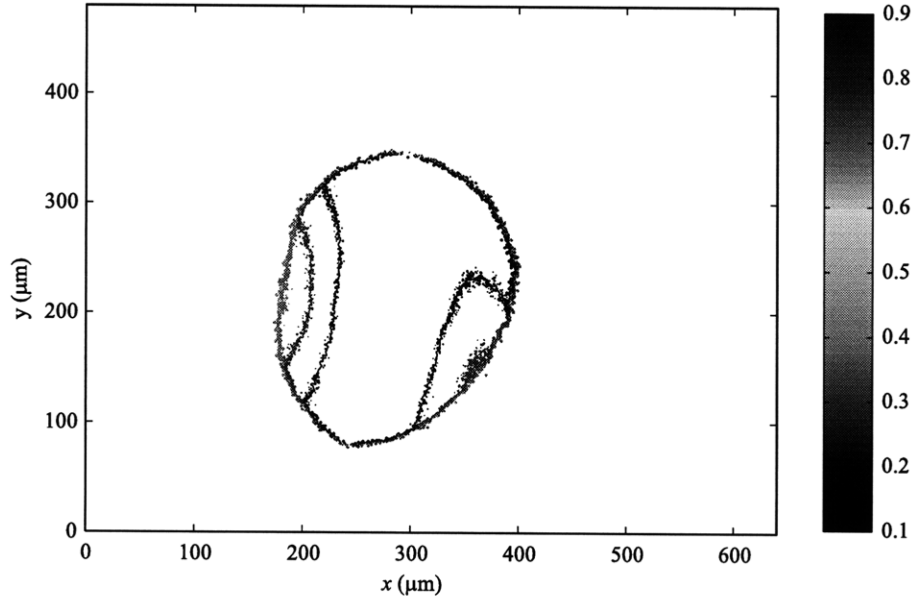


Figure 5-8: The calculated contrast map of the interference fringes.

Taking the 2D Fourier transform $\mathcal{F}\{\}$ of Eq. 5.8 and ignoring the phase term obtains

$$\mathcal{F}\{I(x,y)\} = \int\int I_{avg}(x,y)e^{-j(\omega_x x+\omega_y y)}dxdy + 2\int\int A_{RL}(x,y)\cos(k_g x)e^{-j(\omega_x x+\omega_y y)}dxdy, \tag{5.12}$$

where $A_{RL}(x,y) = A_R(x,y)A_L(x,y)$, and ω_x and ω_y are the spectral coordinates $2\pi f_x$ and $2\pi f_y$, respectively. Using Euler's formula and the shift property of Fourier transform, Eq. 5.12 can be simplified as,

$$\mathcal{F}\{I(x,y)\} = \mathcal{I}(\omega_x,\omega_y) + \mathcal{A}(\omega_x - k_g,\omega_y) + \mathcal{A}(\omega_x + k_g,\omega_y), \tag{5.13}$$

where,

$$\mathcal{I}(\omega_x,\omega_y) = \int\int I_{avg}(x,y)e^{-j(\omega_x x+\omega_y y)}dxdy, \tag{5.14}$$

and,

$$\mathcal{A}(\omega_x,\omega_y) = \int\int A_{RL}(x,y)e^{-j(\omega_x x+\omega_y y)}dxdy. \tag{5.15}$$

The Fourier transform of the intensity pattern can readily visualized as a Gaussian function (since the Fourier transform of a Gaussian function is still a Gaussian function) convolved with three scaled delta functions located at $\omega_x = 0, k_g$, and $-k_g$. Note

that since k_g is much larger than the inverse of the beam diameter, the delta functions in the frequency domain are much further apart than the width of the Gaussian function. In fact, using $\delta = 200$ nm and beam diameter of $\sim$1 mm, the length scale of k_g is 3-4 orders of magnitude larger. The broadening of the delta functions, due to the finite beam diameter, is thus extremely small.

The beam contrast can now be defined as the ratio of the magnitude at $\omega_x = \pm k_g$ and $\omega_x = 0$. Since the magnitudes at $\omega_x = \pm k_g$ are identical, the contrast is then

$$\Gamma = \frac{2\mathcal{A}(0,0)}{\mathcal{I}(0,0)}. \tag{5.16}$$

Substituting the spatial functions we obtain

$$\Gamma = \frac{2\int\int A_R(x,y)A_L(x,y)dxdy}{\int\int I_{avg}(x,y)dxdy}. \tag{5.17}$$

Note that this image contrast is a constant, and describes how well the beams are balanced. The image contrast of the beams as shown in Figure 5-7 is 0.85. This result, as well as the contrast map shown in Figure 5-8, are both indications that the beams are not well overlapped.

Contrast of the image can be significantly improved by realigning the beams. The Nanoruler has an automatic beam alignment routine, designed by Dr. Chen [15], and can change the position and angle of both arms by an amount defined by the user. Thus using Eq. 5.17 as the parameter to maximize and shifting the beams numerically, the optimal beam overlap positions can be found. A simulation script was written using Matlab to "numerically align" the beams. The script basically shifts the left beam by an array of distances in both x and y, while calculating the beam contrast as defined by Eq. 5.17. Using the poorly-overlapped beams shown previously, the image contrast as a function of left beam displacement Δx and Δy is plotted in Figure 5-9.

Using this routine the optimal x and y-displacements of the left beam can be found. To verify the improvement of the alignment, the intensity and contrast maps of the shifted beams are calculated and depicted in Figure 5-10 and Figure 5-11. The improvement is evident, as the contrast map is inceased to be greater than 0.9 over the entire beamspot.

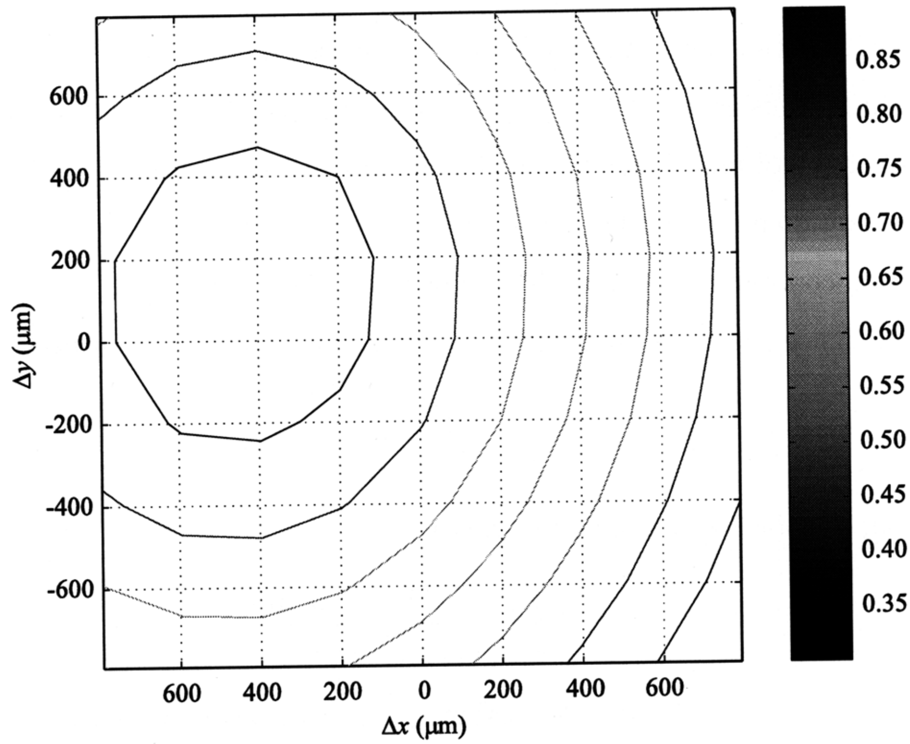


Figure 5-9: The image contrast as the left beam is numerically displaced by Δx and Δy.

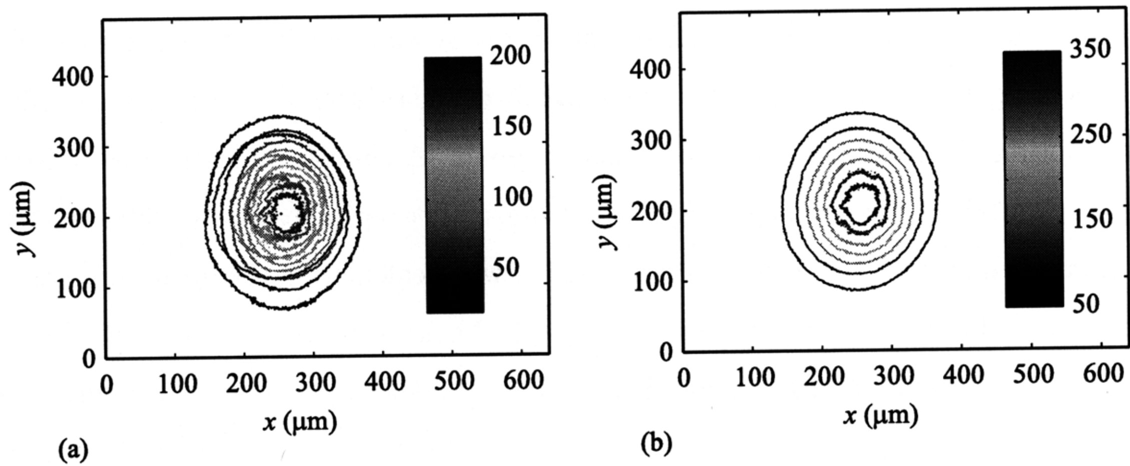


Figure 5-10: (a) The intensity maps of the numerically shifted left and right beams, and (b) the sum of the maps.

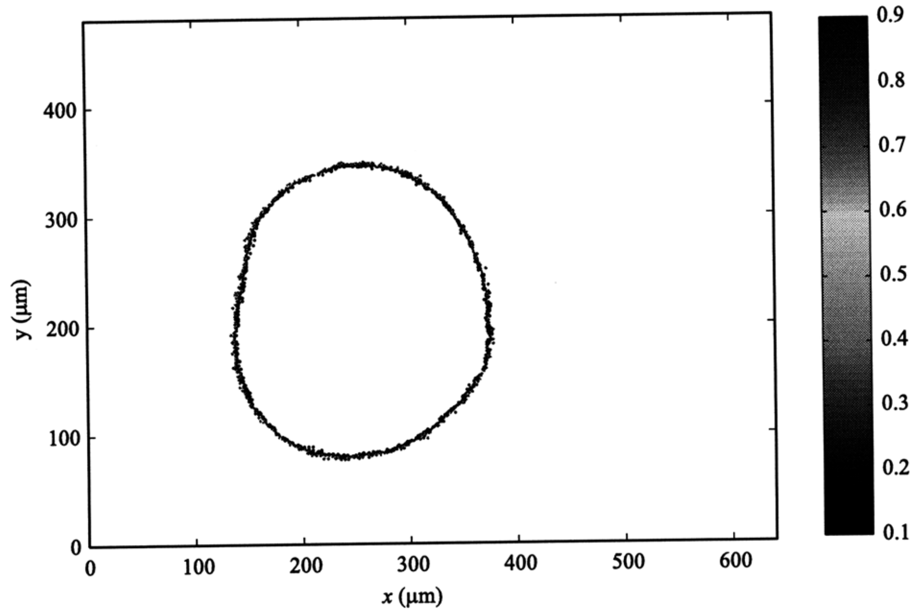


Figure 5-11: The calculated contrast map of the "numerically aligned" interference fringes.

These experimental and simulation tools are essential in determining the beam overlap of the two beams. The beam profiles were also measured for incident angles of 61.37^o, which will be used extensively for exposing 200 nm-period gratings. The beam profiles are shown in Figure 5-12. Note the due to the oblique illumination, the beam profiles are stretched in the x-direction, losing their circular symmetry.

So far in this section the beam contrast analysis has only dealt with the intensity profiles. Another factor plays a significant role, and that is the non-linear phase distortion as described in Eq. 5.8. Non-linear phase distortion, or fringe non-linearity, can be caused by various issues such as misalignment of the Gaussian optics or optics imperfection. It is impossible to obtain perfectly linear fringes experimentally, and success can only be obtained by minimizing the non-linearity.

The fact that the Nanoruler uses a scanning-beam operation ameliorates the effects of non-linear phase in the image grating. The fringes are scanned and the non-linearity is averaged by the resist. The phase distortion of the resist grating can be significantly decreased, yet due to the intensity averaging the contrast is also expected to degrade.

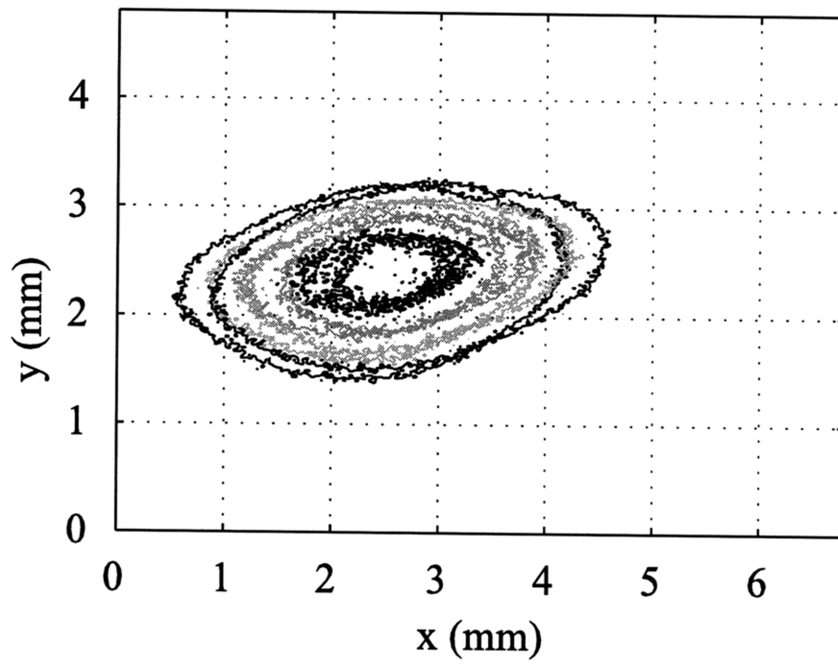


Figure 5-12: (a) The intensity maps of the left and right beams for incident angles of 61.37^o, aligned for 200 nm-period grating exposure.

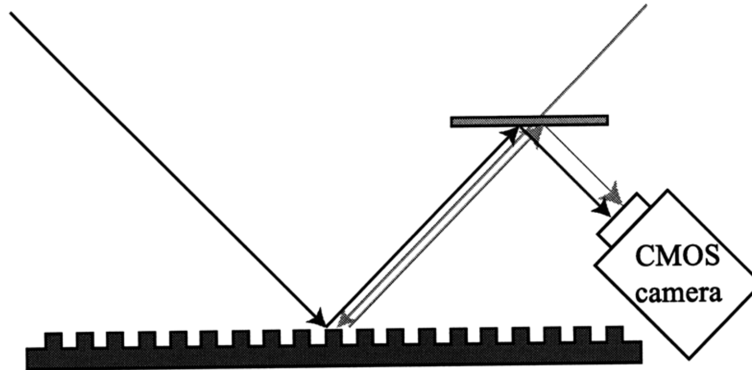


Figure 5-13: The phase-shifting interferometer used to measure the non-linear phase of the image grating.

Therefore it is important to measure the non-linear phase distortion in the image grating, as it relates directly to contrast.

The non-linear phase distortion can be measured using interferometric methods, as shown in Figure 5-13. A reference grating with the same period as the image grating is placed on the wafer plane. This configuration is known as the Littrow condition, where the back-diffracted 1st order back propagates towards the incident direction. The 0th diffracted order of the left beam and the 1st diffracted order of the right beam are interfered and captured by a CMOS camera.

A phase-shifting interferometry (PSI) scheme can be used to determine the phase of the interferogram more effectively, as described by [15]. By introducing phase shifts to one of the arms, several interferograms can be captured and using various algorithms the phase information can be retrieved. Using this setup, a phase map was obtained for 200 nm-period exposure, as shown in Figure 5-14.

5.1.3 Lithography Using Nanoruler: Mechanical Considerations.

In the previous section various optical issues were considered to determine the effects on linewidth control. Numerical tools that use beam images and non-linear phase maps to calculate beam contrast were developed. Using these tools the contrast of

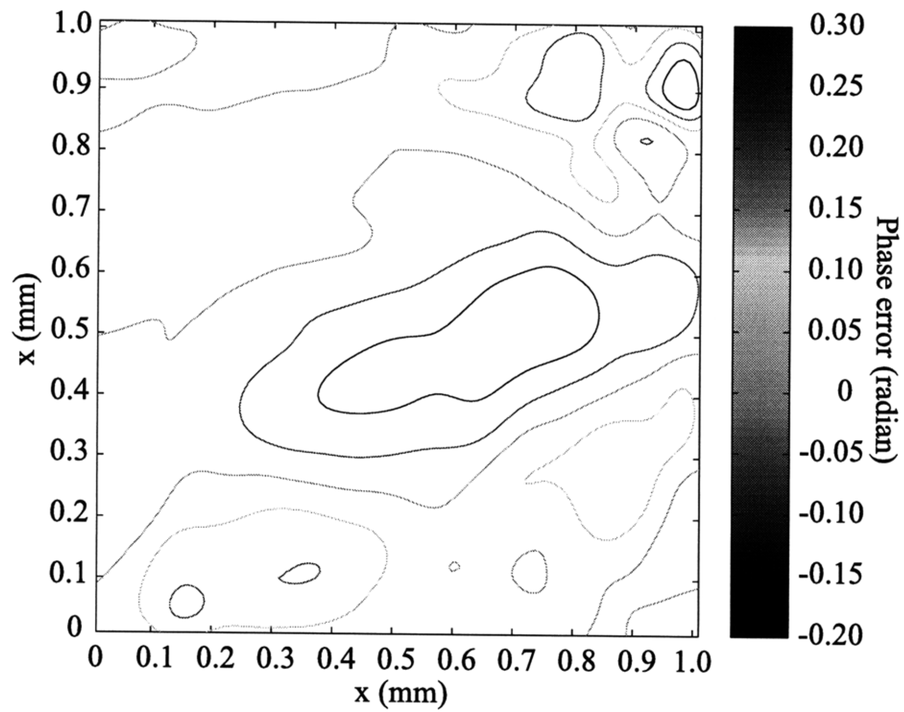


Figure 5-14: The phase map for a 200 nm-period exposure, obtained using the PSI scheme.

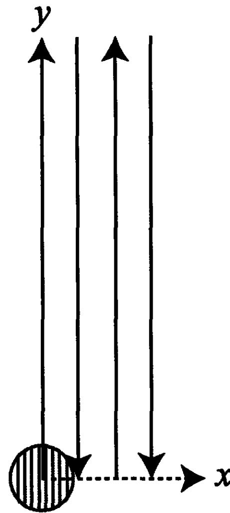


Figure 5-15: The raster scan method for parallel writing.

the image can be estimated before lithography, significantly improving process control. In this section various issues related to mechanical aspects during lithography will be examined. These can be characterized as stage scanning errors.

The lithography process of the Nanoruler involves scanning an air-bearing stage under the image grating to expose a large area, as previously described. There are a variety of scanning methods, depending on the orientation of the fringes [55]. The analysis in this section considers only a raster scan method with the image grating parallel to the scanning direction, as shown in Figure 5-15.

Using the wafer as the frame of reference, the beam spot scans in the y-direction and steps over in the x-direction. The dose as seen by the wafer can be described as a convolution of the intensity pattern and the scan profile. The dose in a raster scan can be written as

$$D(x, y) = I(x, y) \otimes \frac{1}{v_{scan}} \text{rect}(\frac{y}{L}) \text{ comb}(\frac{x}{x_s}), \tag{5.18}$$

where the rect function is scaled by the inverse of the scanning velocity $\frac{1}{v_{scan}}$, L is length of scan, and x_s is the step over distance. The convolution with the rect function represents the scanning in y-direction, while the comb represent the step over in x-direction. Since each individual scan starts before and ends after the wafer, L can be

approximated as ∞, and the rect function turns into a constant. Eq. 5.18 can then be simplified as,

$$D(x,y) = I(x,y)\frac{1}{v_{scan}} \otimes \text{ comb}(\frac{x}{x_s}). \tag{5.19}$$

In this formulation only the scan and step-over operations are considered, and there are no restrictions on the intensity function. In the remainder of this section various mechanical issues that can be described by the intensity function will be analyzed, namely the scan angle error and period measurement error.

In scanning of the image grating, it is desirable to have the fringes parallel to the scanning direction. When this condition is not satisfied, the scanning operation will result in poor exposure contrast. The scan angle error can then be defined as the offset angle between the fringes and the scanning direction y. Note in the extreme case when the product of the scan angle error and the diameter of the beam spot is greater than half the fringe period, the exposure contrast will approach zero and the resist will all be exposed. For a 200 nm-period grating and a beam spot diameter of 1 mm, the maximum scan angle error is then by simple trigonometry to be 100 μrad.

In the Nanoruler system there are a variety of routines that can be run before exposure to make sure the scan angle error is much lower than 100 μrad. A more reasonable error would be in the order of 10 μrad. However it is impossible to reduce this error to zero, and it is important to analyze its effect on contrast in order to improve linewidth control.

There are a number of way to formulate the effect of scan angle error, the most intuitive approach being to integrate the analysis into the intensity function as described in Eq. 5.19. Keeping the coordinate frame defined earlier, with y as the scan and x as the step-over directions, the scan angle error is defined in Figure 5-16 as θ_{scan}. The rotated intensity function of the image grating can then be written as

$$I(x,y) = I_{avg}[1 + \Gamma\cos(\frac{2\pi}{\Lambda}x\cos\theta_{scan} + \frac{2\pi}{\Lambda}y\sin\theta_{scan})]. \tag{5.20}$$

Factoring out the grating vector and rearranging obtains

$$I(x,y) = I_{avg}[1 + \Gamma\cos(\frac{2\pi}{\Lambda}x + \frac{2\pi}{\Lambda}x(\cos\theta_{scan} - 1) + \frac{2\pi}{\Lambda}y\sin\theta_{scan})]. \tag{5.21}$$

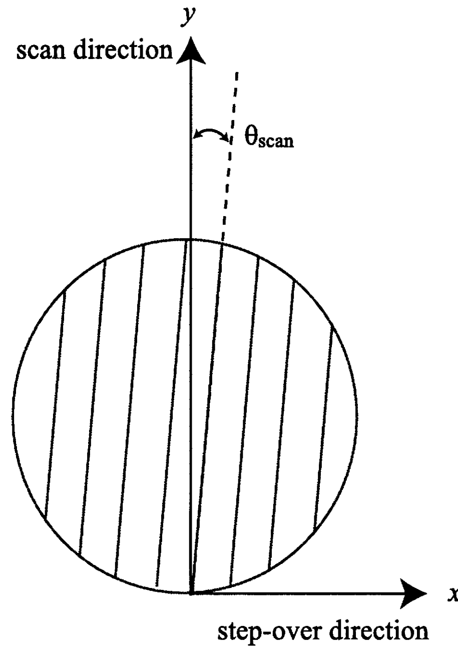


Figure 5-16: The coordinate frame with y and x as the scan and step-over directions, respectively. The scan angle error is defined as θ_{scan}.

The last two terms are phase distortions, and can be represented by a newly defined function

$$\phi_{scan}(x, y) = \frac{2\pi}{\Lambda} x(\cos\theta_{scan} - 1) + \frac{2\pi}{\Lambda} y \sin\theta_{scan}. \tag{5.22}$$

The intensity function is then a perfectly aligned grating with a linear phase distortion. The scan angle error is thus represented by a phase distortion map. This is extremely desirable, as this phase distortion can be added to the general non-linear phase distortion as described in Eq. 5.8, and no additional changes are needed in the formulation.

The other mechanical issue that is critical during scanning of the image grating is the step-over distance. The step-over distance is the x-distance between each scan, and is typically a fraction of the beam spot diameter, ranging from 40-80%, to ensure tight overlap. The step-over distance is a macroscopic parameter that can be specified. A very important criteria at the microscopic level to ensure high exposure contrast is that the step-over distance has to be an integer multiple of the fringe period [15].

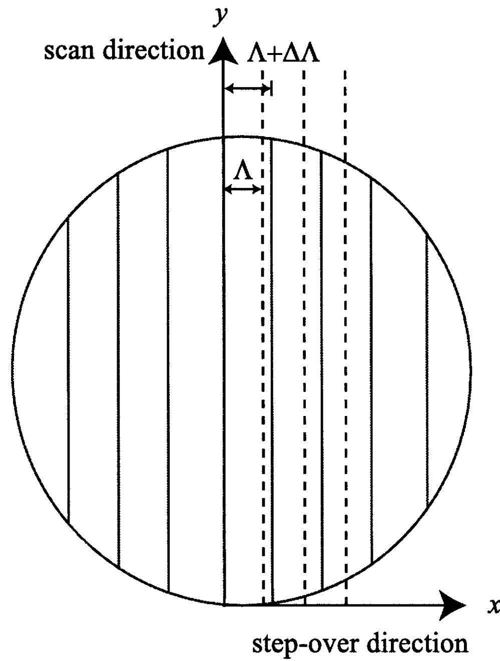


Figure 5-17: The coordinate frame with y and x as the scan and step-over directions, respectively. The period error is defined as $\delta\lambda$.

This restriction is quite intuitive, as the fringes need to be phase matched.

To calculate the step-over distance, the fringe period can be measured using a beam splitter, as described by Dr. Chen [15]. The step over is then calculated to be a number that is an integer multiple of the period and around the input step-over distance. However, if there is an error between the measured period and the actual period of the fringes, there will be a reduction in contrast. Such an error is called period measurement error, and has been studied previously by Dr. Chen [15]. The typical period measurement error is in the order of 0.001 nm, typically less than 0.005 nm. It is of interest, however, to develop a model which will fit into the formulation described in this section.

The period error is illustrated in Figure 5-17 as $\Delta\Lambda$. In this Figure the measured period is Λ, and the actual period is $\Lambda + \Delta\Lambda$. The effect can be seen readily, as the error after the first fringe is $\Delta\Lambda$, and after the N_{th} fringe $N\Delta\Lambda$. The intensity

function of the image grating can then be written as,

$$I(x,y) = I_{avg}[1 + \Gamma \cos(\frac{2\pi}{\Lambda + \Delta\Lambda}x). \tag{5.23}$$

Again, factoring out the grating vector,

$$I(x,y) = I_{avg}[1 + \Gamma \cos(\frac{2\pi}{\Lambda}x + 2\pi(\frac{1}{\Lambda + \Delta\Lambda} - \frac{1}{\Lambda})x), \tag{5.24}$$

where the last term in the cosine function represents a phase distortion. The phase distortion can be described by,

$$\phi_{period}(x,y) = 2\pi(\frac{1}{\Lambda + \Delta\Lambda} - \frac{1}{\Lambda})x. \tag{5.25}$$

The intensity function is then a grating with period equal to the measured period, and a linear phase distortion. The period error is thus represented by a phase distortion map, as in the case of scan angle error. This error can then be added to the amount of phase distortion, while keeping the formulation simple.

5.1.4 Dose, Contrast and Phase Calculation

In the last two sections various factors affecting the contrast have been examined. These include optical parameters in beam intensity profiles $I_L(x,y)$ and $I_R(x,y)$, fringe phase distortion $\phi_{fringe}(x,y)$, and mechanical parameters in scan angle error θ_{scan} and period error $\Delta\Lambda$. These five factors play significant roles in controlling the linewidth during lithography.

The goal of this section is to develop a numerical tool that simulates the contrast and dose of the exposure due to the aforementioned five factors. The desired input and output parameters are illustrated in Figure 5-18. Using this program the contrast and dose profiles can be predicted before lithography by measuring the input parameters. The user can then decide if the system is properly aligned to achieve the linewidth goal in mind.

The first step is to simulate the effect a single scan in the y-direction. The scanned dose can be obtained by performing the first convolution in Eq. 5.19,

$$D_y(x,y) = \int\int \frac{1}{v_{scan}} I(x,y)\delta(x - x')dx'. \tag{5.26}$$

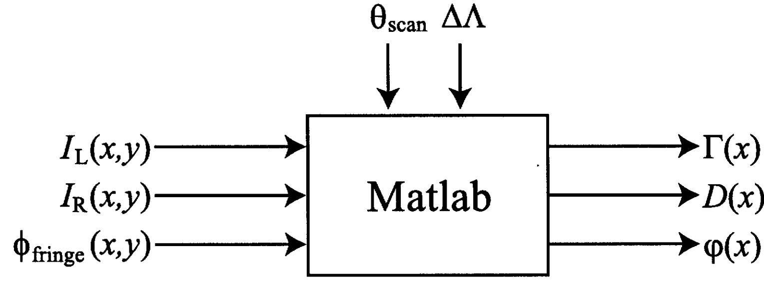


Figure 5-18: The desired input and output parameters of the simulation program.

Substituting the intensity function described by Eq. 5.8 and simplifying,

$$D_y(x) = \frac{1}{v_{scan}} \int I_{avg}(x, y) + 2A_{RL}(x, y) \cos[k_g x + \phi(x, y)] dy, \tag{5.27}$$

where $\phi(x, y)$ represents all phase distortion contributions, including non-linear phase error of the fringes and linear phase error due to scan angle or period error. Note that as a result of the summation of the intensity in y-direction, the dose profile is now independent of y. Since the input data are CMOS images expressed in pixels, Eq. 5.27 is better expressed in discrete coordinates $[x, y]$,

$$D_y[x] = \frac{\Delta y}{v_{scan}} \sum_y I_{avg}[x, y] + 2A_{RL}[x, y] \cos(k_g x + \phi[x, y]). \tag{5.28}$$

The cosine term can not be easily summed, so simplifying with angle addition formula,

$$D_y[x] = \frac{\Delta y}{v_{scan}} \sum_y I_{avg}[x, y] + 2A_{RL}[x, y] \Big(\cos(k_g x) \cos(\phi[x, y]) - \sin(k_g x) \sin(\phi[x, y]) \Big). \tag{5.29}$$

Reorganizing we obtain

$$D_y[x] = \frac{\Delta y}{v_{scan}} \Big(\sum_y I_{avg}[x, y] + P_{scan}[x] \cos(k_g x) - Q_{scan}[x] \sin(k_g x) \Big), \tag{5.30}$$

where

$$P_{scan}[x] = 2 \sum_y A_{RL}[x, y] \cos(\phi[x, y]), \tag{5.31}$$

and

$$Q_{scan}[x] = 2 \sum_y A_{RL}[x, y] \sin(\phi[x, y]). \tag{5.32}$$

An insightful, and quite intuitive, conclusion From Eq. 5.31 and Eq. 5.32 is that the effects of the phase distortion is weighted by the beam intensity. Thus it is critical that the phase distortion in the center of the beam spot be minimized, while the outer regions can be more forgiving. To calculate the contrast of the dose profile, Eq. 5.30 can be simplified by the harmonic addition theorem as,

$$D_y[x] = D_{scan,avg}[x]\Big(1 + \Gamma_{scan}[x]\cos(k_g x + \psi_{scan}[x])\Big), \tag{5.33}$$

where

$$D_{scan,avg}[x] = \frac{\Delta y}{v_{scan}} \sum_y I_{avg}[x, y], \tag{5.34}$$

is the average dose profile,

$$\Gamma_{scan}[x] = \frac{\Delta y}{v_{scan}} \frac{\sqrt{P^2_{scan}[x] + Q^2_{scan}[x]}}{D_{scan,avg}[x]}, \tag{5.35}$$

is the contrast profile, and

$$\psi_{scan}[x] = \tan^{-1}\Big(\frac{Q_{scan}[x]}{P_{scan}[x]}\Big), \tag{5.36}$$

is the phase profile of the scanned exposure dose. Eq. 5.34 and 5.35 calculate the exposure dose and contrast for a single scan in the y-direction. These calculations provide important insight into the quality of the exposure.

Using the beam profile and phase map for 200 nm-period exposure, depicted previously as Figure 5-12 and Figure 5-14, the dose and contrast of a single scanned exposure can be calculated. The result is plotted in Figure 5-19.

The upper graph plots the average dose and the fringe dose as a function of distance in x. The average scan dose, $D_{scan,avg}$, is defined earlier, and can be seen as the weight of the DC or background dose. The fringe contrast, Γ_{scan}, can be perceived as the weight of the AC term, which is the spatial modulation that the resist is recording. The lower graph depicts the calculated contrast based on the dose functions. In the case when the DC equals the AC term, the image contrast is perfect and equals to 1. More realistically, the fringe dose is always less than the average dose, resulting in contrast less than 1.

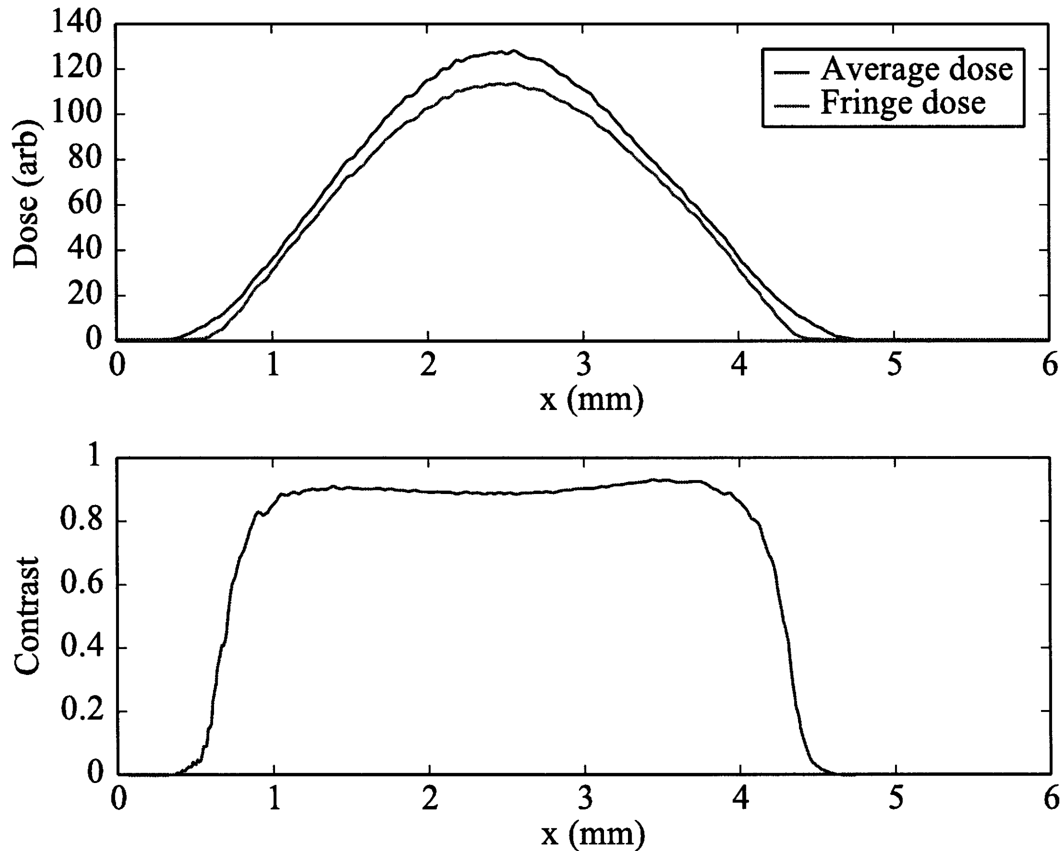


Figure 5-19: The dose and contrast of a single scanned exposure for 200 nm-period grating.

The contrast plot demonstrates high contrast near the center of the beams, while gradually tailing off towards the edge. This result is quite intuitive, since the edge of the beams are not properly overlapped, as shown in Figure 5-12. This area thus has no interference fringes, resulting in low fringe dose and contrast. Even with poor contrast near the edge, the effects can be mitigated by choosing a small step-over distance for the subsequent scans.

The next step in calculating the contrast and dose is to simulate the effect of the multiple scanning with a fixed step-over distance. The step-over scanning operation is defined previously in Eq. 5.19 as a convolution. However, since the intensity profile of the beam has a finite length, the computation can be greatly simplified. Also since the scanning operation is periodic in the x-direction, it is sufficient to calculate the

contrast and dose over a single step-over distance. Taking an initial scanned dose as previously described, $D_y[x]$, a summed total dose for multiple scans can be described as,

$$D[x] = \sum_{n=0}^{N} D_y[x \pm nx_s], \tag{5.37}$$

where x_s is the step-over distance and N is the number of adjacent scans that contribute to the initial scan. The number N is then dependent on the beam diameter and step-over distance. The greater the step-over distance the lesser effects the adjacent scans will play. It is interesting to note that mathematically for a pure Gaussian beam the tail would be infinite, and N would approach inf. But in reality the intensity becomes negligible shortly outside the $1/e^2$ radius, and thus N can be limited to a finite number depending on the step-over distance. To reduce dose variations between scans, the step-over distance is typically 0.25-0.5 of the beam diameter, resulting in a $N \sim 5$. Substituting in $D_y[x]$ as defined in Eq 5.30,

$$D[x] = \frac{\Delta y}{v_{scan}} \Big(\sum_{n=0}^{N} \sum_{y} I_{avg}[x \pm nx_s, y] + B[x]\cos(k_g x) - C[x]\sin(k_g x) \Big), \tag{5.38}$$

where,

$$B[x] = 2 \sum_{n=0}^{N} \sum_{y} A_{RL}[x \pm nx_s, y] \cos(\phi[x, y]), \tag{5.39}$$

and,

$$C[x] = 2 \sum_{n=0}^{N} \sum_{y} A_{RL}[x \pm nx_s, y] \sin(\phi[x, y]). \tag{5.40}$$

It is important to note the step-over shifts are only related to the macroscopic beam profile terms. The microscopic terms related to the image grating phase k_g are fringe-locked to the phase of the measured period, as mentioned earlier. Any phase error can be characterized by the total phase distortion term $\phi[x, y]$, as defined initially in Eq. 5.27.

Simplifying, Eq. 5.38 becomes,

$$D[x] = D_{avg}[x] \Big(1 + \Gamma[x] \cos(k_g x + \psi[x]) \Big), \tag{5.41}$$

where,

$$D_{avg}[x] = \frac{\Delta y}{v_{scan}} \sum_{n=0}^{N} \sum_{y} I_{avg}[x \pm nx_s, y], \tag{5.42}$$

is the average dose profile,

$$\Gamma[x] = \frac{\Delta y}{v_{scan}} \frac{\sqrt{B^2[x] + C^2[x]}}{D_{avg}[x]}, \tag{5.43}$$

is the contrast profile, and,

$$\psi[x] = \tan^{-1}\left(\frac{C[x]}{B[x]}\right), \tag{5.44}$$

is the phase profile of the final scanned exposure. Eq. 5.42 and Eq. 5.43 describe the dose and contrast profiles obtained from experimentally measured beam profiles and phase distortion maps. Other parameters such as scan angle error and period error can also be simulated, as well as any user defined phase distortions. These distortions can be conveniently describe by using the Zernike polynomials [49]. Using the dose and contrast profiles calculated by this model, the final linewidth of the exposed resist can be predicted.

5.2 Dose and Contrast Simulation

The derived model is extremely useful in simulating the effects of the various parameters during lithography. It examines the issue of process control, and provides insight into how well the linewidth can be controlled. Using the beam profile and phase map shown previously in Figure 5-12 and Figure 5-14, respectively, the exposure dose and contrast profiles are calculated and plotted in Figure 5-20. In this simulation the beams are aligned for a 200 nm period grating, with a step-over distance of 0.45 mm. The scan angle and period measurement errors are assumed to be zero. This is the ideal case when the only factors degrading the contrast are beam overlap and wavefront phase errors, and that the raster scanning operation of the Nanoruler is assumed to be perfect.

The simulated plots show a dose variation of $\sim$ 1% and an average contrast of $\sim$ 0.781%. The small dose variation is a welcoming result, since it will minimize linewidth variation. The contrast profile is more interesting, as the fringe image is degraded by the optical factors. It is, however, still possible to achieve duty-cycle

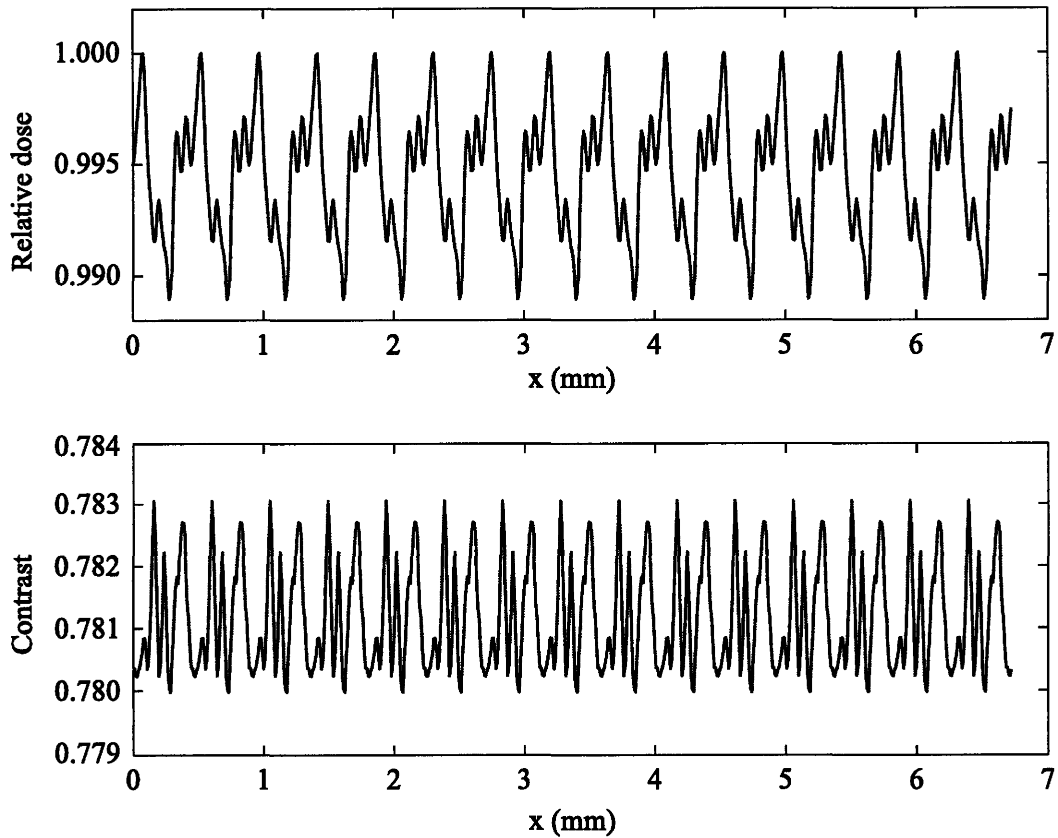


Figure 5-20: The exposure dose and contrast profile of 200 nm period grating with a step-over of 0.45 mm. The dose and contrast variations are due to the non-Gaussian profile of the beams.

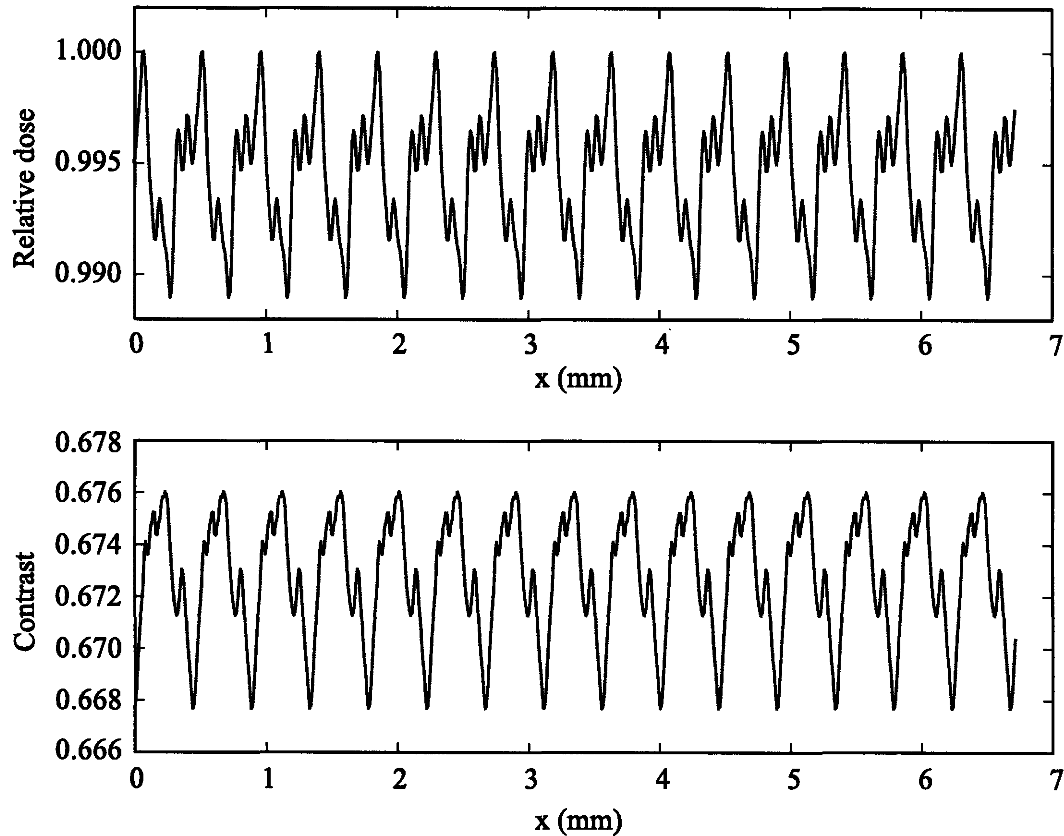


Figure 5-21: The exposure dose and contrast profile of a 200 nm period grating with a step-over of 0.45 mm. The scan angle and period measurement errors are 5 μrad and 0.005 nm, respectively. The dose and contrast variations are due to the non-Gaussian profile of the beams.

around 0.9 in a well-controlled process. The effect of the step-over distance is also evident in the plots, which are periodic over the length of 0.45 mm.

The problem arises when taking the various mechanical scanning imperfection into account. The scanning error tolerance is around 5 μrad for scan angle error and 0.005 nm for period measurement error, as defined previously. These tolerances represent the maximum estimated error that may result from the metrology process. Inserting these errors in the calculation, the simulated plots are shown in Figure 5-21. The effects of these scanning errors are evident in the plots, as the average contrast is reduced to 0.672. This level of degradation is significant, considering it is caused by metrology tolerances. The implication from these simulation results is that for a given

beam overlap and wavefront phase map, the exposure contrast will be somewhere between 0.672 and 0.781 due to scanning error tolerances. The contrast range is quite large, and the exact contrast can not be determined until after exposure. As a result, it will be difficult to control the linewidth. It is interesting to note that the dose profile is not effected by the scanning errors, which is expected since the phase distortion effects and scanning errors are extremely small.

There are also many other factors that are unrelated to the optical and mechanical issues mentioned previously, and these are more difficult to characterize and control. One such example that degrades contrast is uniform background exposure, which can originate from various sources. The interference equations derived earlier assumes monochromatic beams polarized in the TE direction. In reality, there might be some spectral leakage from the various time-frequency modulators and polarization mixing, which will result in poor optical interference. Also in a complex optical system such as the Nanoruler, multiple beams are reflecting off from various optical components and stray beams might cause unwanted background exposure of the resist.

To simulate a possible background exposure, an additional background intensity term can be introduced in Eq. 5.28, which can be easily implemented in the simulation script. This term can be represented by any profile, but for convenience a circularly symmetric Gaussian beam with diameter similar to the exposure profile is used. This is a good approximation for either spectral leakage or polarization mixing in the exposure beam. A relative intensity of 10 % for the background is used, and the single scan dose and contrast profiles are shown in Figure 5-22.

The figure shows the background dose profile in addition to average and fringe doses in the upper plot, and the single scan contrast in the bottom plot. The effects of the background dose can be readily seen. The fringe dose is significantly less than the combined average and background dose. The overall contrast has been decreased to less than 0.8, while previous contrast simulation without the background dose are above 0.9. The dose and contrast profiles can also be calculated with multiple scan exposure, as shown in Figure 5-23. This simulation assumes no scanning errors. The average contrast has been reduced to 0.633 from 0.781, a dramatic decrease.

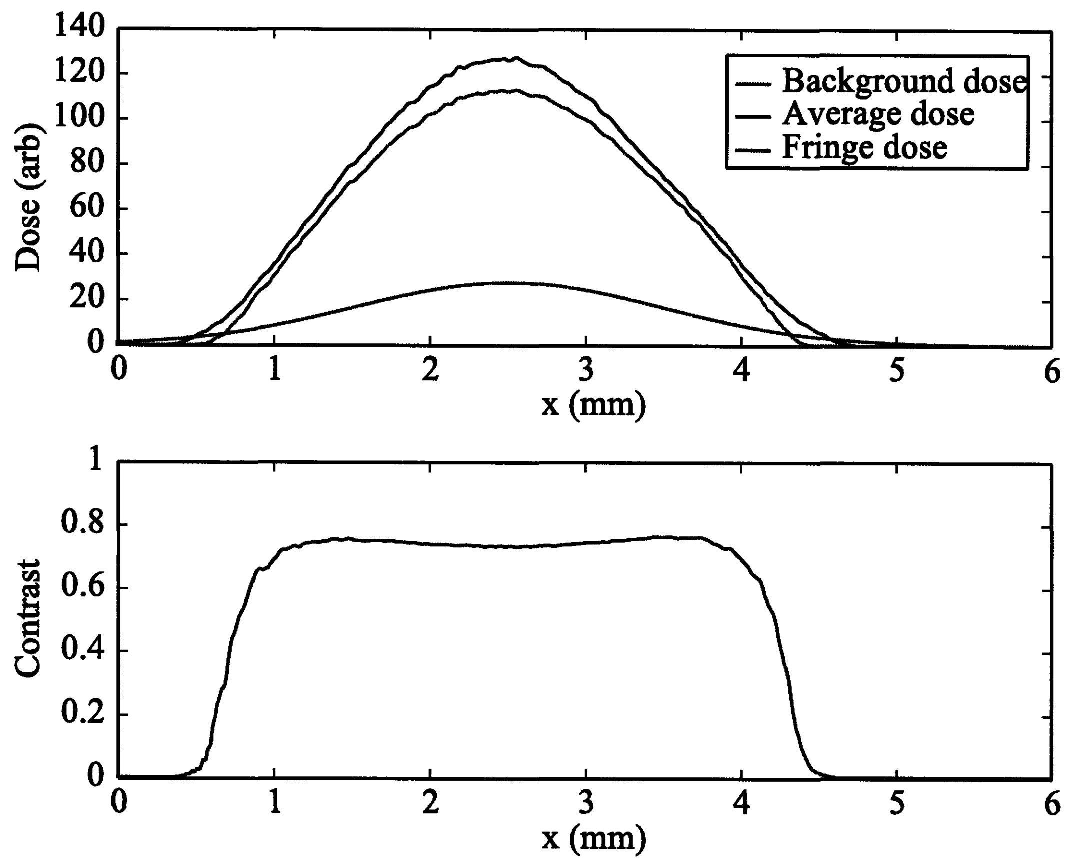


Figure 5-22: The dose and contrast of a single scanned exposure for a 200 nm-period grating with 10 % relative background intensity.

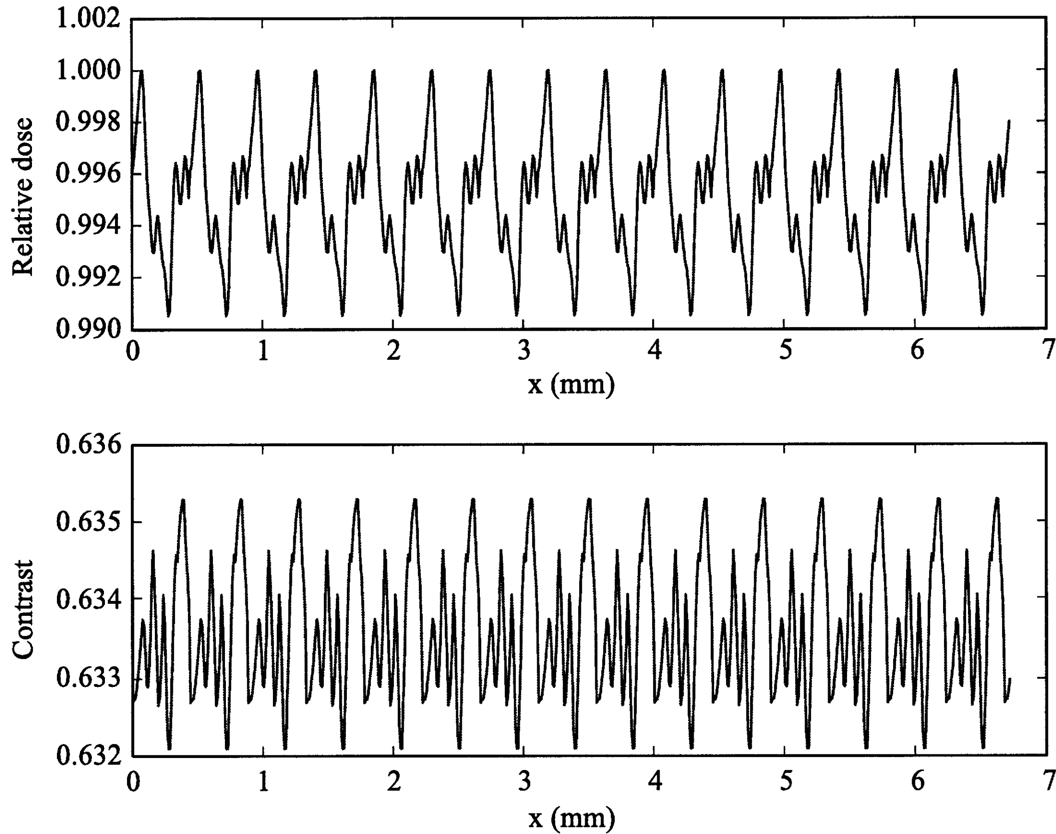


Figure 5-23: The dose and contrast profile for a 200 nm-period grating with 10 % relative background intensity.

However the 10 % relative intensity is not experimentally determined, just a basic test guideline. These results demonstrates the possible ill-effects of background exposure.

Another important factor that may degrade contrast is depth variation in the exposure plane. Using the interfering beams as the reference frame, this error is defined as shifting the resist plane in the z-axis. This can also be described as defocusing, which is a common second order aberration in optical systems. A key advantage of interference lithography is that the depth of focus is infinite, given that the beams overlap. There are some changes in wavefront profile since the interfering beams are Gaussian, but the effects are insignificant.

The sources of these depth variations are due to various factors. First of all, the interfering beams are aligned for optimized beam overlap for a 500 μm thick wafer.

If a sample with a different thickness is used, the difference in thickness will result in a constant depth variation. Also, typical commercially available wafers, depending on the polishing techniques, might have up to 10s of μm in thickness variation. It is then ideal to use double-side polished wafers which are specified to have a thickness variation under a few μm. The scanning of the stage might also contribute to depth variation by rocking the air bearing stage or the granite table. These sources are not easy to identify, and might be tedious to characterize.

The primary degenerative effect of depth variation is changing the overlap of the beams, resulting in contrast reduction. Based on simple geometry,

$$\Delta x = 2\Delta h \tan[\sin^{-1}(\frac{\lambda}{2\Lambda})], \tag{5.45}$$

where Δx is the beam displacement in the x-direction and Δh is the depth variation of the exposure plane. Using this relationship, it is fairly straightforward to shift the beams the determined amount to simulate the effect of depth variation. This can be numerically implemented by shifting one of the beams by a certain number of pixels. The calculations derived early can thus be used to examine the contrast degradation induced by depth variation.

A simulated test case can again be used to examine the depth variation effects. Assuming a depth variation of 100 μm and no scanning errors, the exposure dose and contrast profiles are shown in Figure 5-24. From these plots, the average contrast has been reduced to 0.734 from a depth-variation-free contrast of 0.781. The reduction in contrast is also significant, and great care must be taken into consideration when exposing wafers of poor quality.

In the last few sections various contrast degradation factors are examined. These include beam overlap, grating phase map, scan angle error, period measurement error, background exposure, and depth variation. Some of these issues, such as beam overlap and grating phase map, are easier to characterize, while the others are limited by error tolerances. It is of interest to investigate the combined effects of the aforementioned factors. Using the beam overlap and phase map illustrated in Figure 5-12 and Figure 5-14, respectively, the dose and contrast profiles can be simulated with

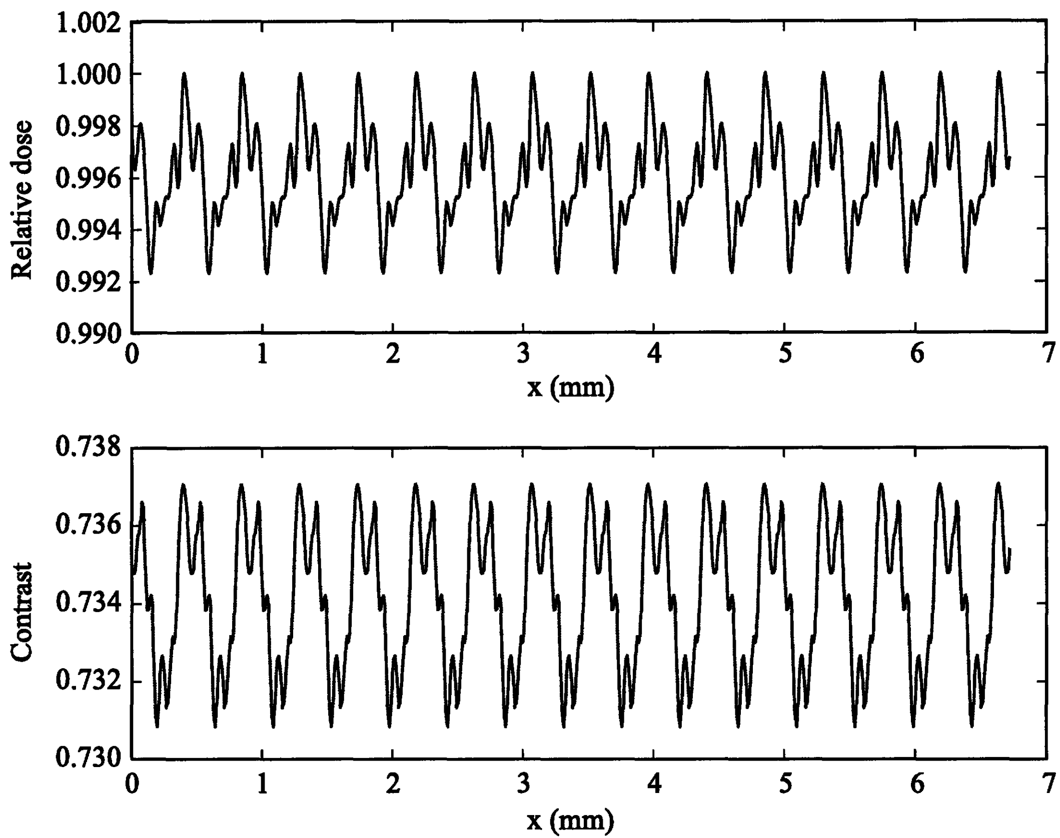


Figure 5-24: The dose and contrast profile for a 200 nm-period grating with 100 μm depth variation.

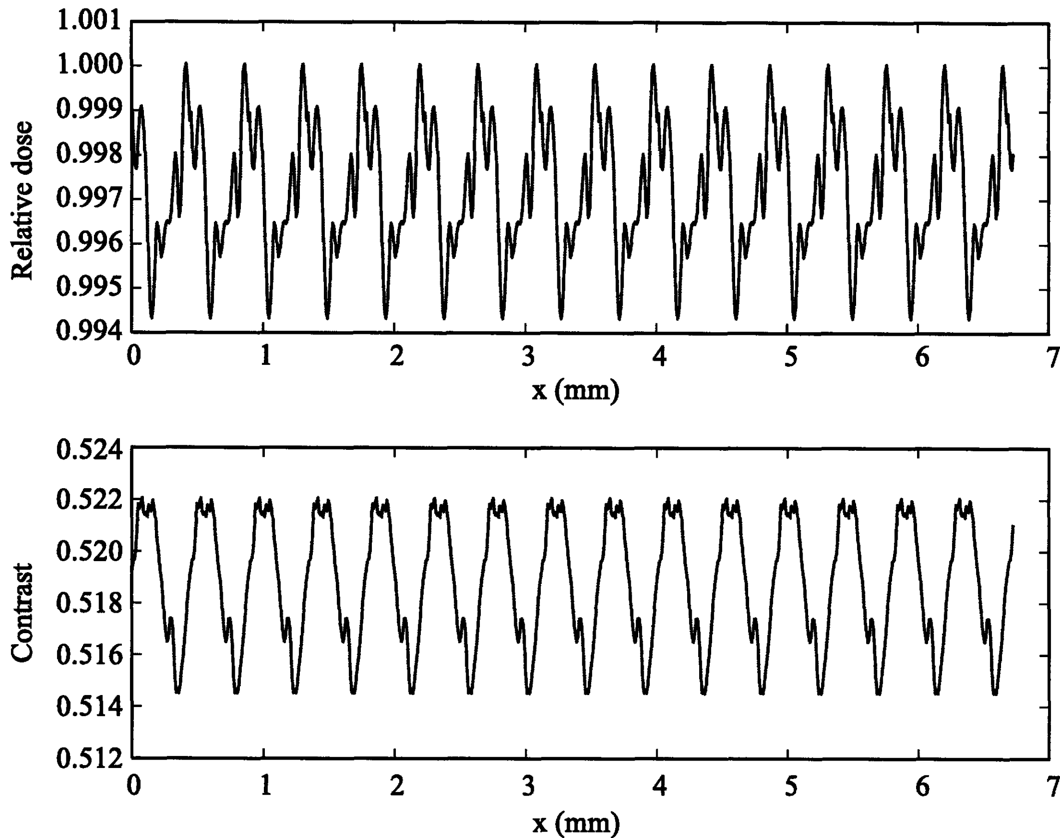


Figure 5-25: The dose and contrast profile for a 200 nm-period grating with 5 μrad of scan angle error, 0.005 nm of period measurement error, 10 % relative intensity Gaussian background exposure, and 100 μm of depth variation.

these contrast degradation factors. The simulation includes 5 μrad of scan angle error, 0.005 nm of period measurement error, 10 % relative intensity Gaussian background exposure, and 100 μm of depth variation.

The results are shown in Figure 5-25. From the plots, the average contrast reduces to 0.518 from 0.781, where there was no errors. In this model that takes into account all the possible contrast reduction factors, the contrast is greatly reduced. Given that these error tolerance levels describe the worse possible case, the real contrast is expected to fall in the range of 0.518 and 0.781. Note while the background exposure and depth variation are worst-case estimates that are not experimentally examined, the scanning errors included in the simulation are reasonable error tolerances. The

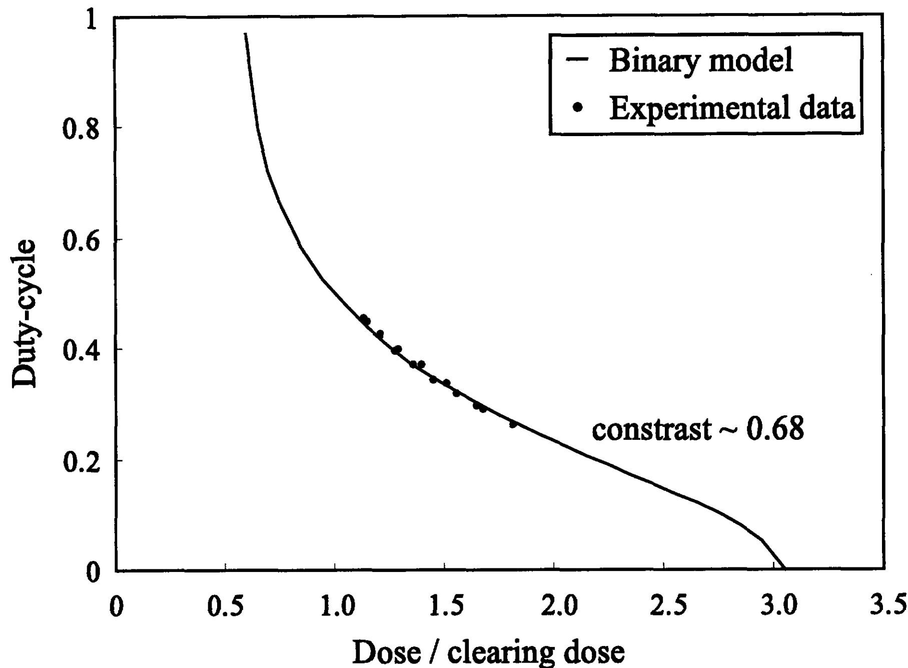


Figure 5-26: The experimental contrast as a function of clearing dose, plotted on top of the binary model with a contrast of 0.68.

scanning errors become increasingly pronounced for multiple layer exposures, as explained earlier in Chapter 4. It is also interesting to note that the dose profile is not influenced much by these contrast reduction errors.

To validate these models, numerous dose curves with positive resist were exposed to experimentally determine the contrast. In these dose curve experiments the resist is exposed with the same conditions except varying dose. After developing, the linewidth is measured by cross-section SEM, and plotted as a function of clearing dose, shown in Figure 5-26.

It is important to note that the resist thickness in these experiments was ~240 nm, which is relatively thick for a period of 200 nm. In this aspect ratio region 3-dimensional effects becomes increasingly important, and are not taken into account by the binary model. Also the model only considers two beam interference, and ignores the reflected beams from the resist-ARC interface. Those reflected beams, typically in the 1 % relative intensity range, creates standing wave in the vertical direction and

result in wavy resist wall profiles and degraded exposure contrast.

In this plot the measured duty-cycle is plotted on top of the binary model, represented by the solid line. The experimentally determined contrast is ~ 0.68, which falls within the predicted range of 0.518 and 0.781. Note that the experimental data stops below duty-cycle of around 0.25. This is because at this duty-cycle, the grating linewidth is only around 50 nm, and with 240 nm thick resist the aspect ratio is around 5. With such high aspect ratio the resist lines collapse readily during drying, and no data were taken.

From these results it is clear that even optimizing the contrast would be a difficult task, involving numerous metrology and multiple simulation steps. Even if the contrast can be optimized to achieve the duty-cycle goal, it will be a daunting task to control the linewidth to the nm-range required for the multilayer process.

5.3 Summary

In this section the dose and contrast issues during lithography were examined in detail. A careful understanding of the various contrast degradation issues is necessary for resist linewidth control. Software to simulate dose and contrast was formulated, implemented, and compared with experimental results. Input optical parameters include experimentally measured beam overlap of the interfering beams and the image grating phase map. The software then allows for scanning errors such as scan angle error and period measurement error, and other hypothetical error tolerances such as background exposure and depth variation. The software serves as an important simulation tool to determine the range of contrast that seems likely before actually exposing the resist.

The simulation software is especially useful to examine to what degree the various factors contribute to contrast degradation. It becomes possible to examine closely the influence of a specific error source to determine a reasonable error tolerance level. The simulation results for scanning errors, shown in this section, lead to a discouraging conclusion. A combination of 5 μm scan angle error and 0.005 nm period measurement

error lowers the contrast from 0.781 to 0.672. These levels of scanning errors are conservative estimates, and when exposing multilayer gratings up to 50 μm angle and 0.500 nm period errors can sometimes occur. This fact is attributed to difficulty in perfectly placing the wafer so the reference grating aligns to the fringe direction, and the change in grating period over time and processing steps.

These results lead to the conclusion that while the simulation software is useful in determining if the contrast will be high enough to allow successful exposure, it will be difficult to control linewidth to the nm-range with just overexposure. Furthermore, in order to obtain a high duty-cycle, a negative resist needs to be used, and that further reduces the linewidth controllability. To overcome this challenge, a image-reversal process using positive resist and plasma trimming will be introduced in Chapter 6.

Chapter 6

Fabrication Process for Linewidth Control

Linewidth control is one of the most critical challenge that needs to be addressed in multilayer interference lithography. In the fabrication process for 4-times spatial-frequency multiplication, two goals need to be reached. Each grating layer will need to have high duty-cycle ~0.9 and linewidth control in the nm-range. Linewidth controllability during lithography was examined in great detail in Chapter 5. The dose and contrast simulations lead to the conclusion that it will be difficult to achieve nm-level linewidth control using only exposure parameters.

To achieve high duty-cycle and nm-level linewidth control, a novel fabrication process using an image-reversal step and plasma trimming was developed. Using the image-reversal process, the tone of the exposed resist profile can be reversed, thus allowing the use of positive resist. The plasma trimming technique enables layer by layer control of grating linewidth. Using these nanofabrication methods the goals of high duty-cycle and nm-level linewidth control can be met.

The image-reversal and plasma trimming processes are outlined in great detail in this chapter. The repeatability of these processes are examined and evaluated in terms of process latitude. Lastly, the various factors contributing to the limitation of these techniques are explored. The challenges that this process address are key in the success of multilayer interference lithography.

6.1 High Duty-cycle Grating Patterning

High duty-cycle grating patterning is essential in the fabrication process for spatial-frequency multiplication. In fact, the limiting factor of the process is determined by how high the duty-cycle can be. An equation based on simple geometry can be formulated to calculate the required duty-cycle, described as

$$DC_{layer}(x) = 1 - (\frac{1 - DC_{final}}{N}), \tag{6.1}$$

where DC_{layer} is the duty-cycle of individual grating layer, DC_{final} is the duty-cycle of the final spatial-frequency multiplied grating, and N is the multiplication factor. Eq. 6.1 shows that to fabricate a two-fold spatial-frequency multiplied grating with 0.5 duty-cycle, the individual grating layer needs to have a duty-cycle of 0.75. The requirement for high duty-cycle increases with N, to ~ 0.83 and ~ 0.88 for three-fold and four-fold spatial-frequency multiplication, respectively. These calculations can shift a bit higher or lower, depending on the final desired duty-cycle. As shown in these calculations, the success of the spatial-frequency multiplication process relies heavily on the ability to achieve high duty-cycle.

High duty-cycle grating patterning can be difficult to obtain, especially if the process needs to be controllable to the nm-range. The challenges become increasing complicated as the period decreases. In the case of 200 nm-period, 20-30 nm trenches are needed. In this section several processes to fabricate high duty-cycle gratings are examined, including using negative resist with a heat treatment process, lift-off processes, and shadow evaporation techniques.

Overexposure of negative tone resist is a more direct approach to achieve high duty-cycle, as depicted in Figure 5-1. However, the commercially available negative resists at $\lambda = 351.1$ nm are all chemically-amplified. These resists have time and temperature dependent diffusion chemical reactions, initiated by a post exposure bake, and do not obey the binary resist model. In addition, since exposure is based on heat diffusion, it is difficult to control linewidth in the nm-range, especially when working in the 20-30 nm length-scale. These challenges are great obstacles in achieving high duty-cycles in the 0.9 range.

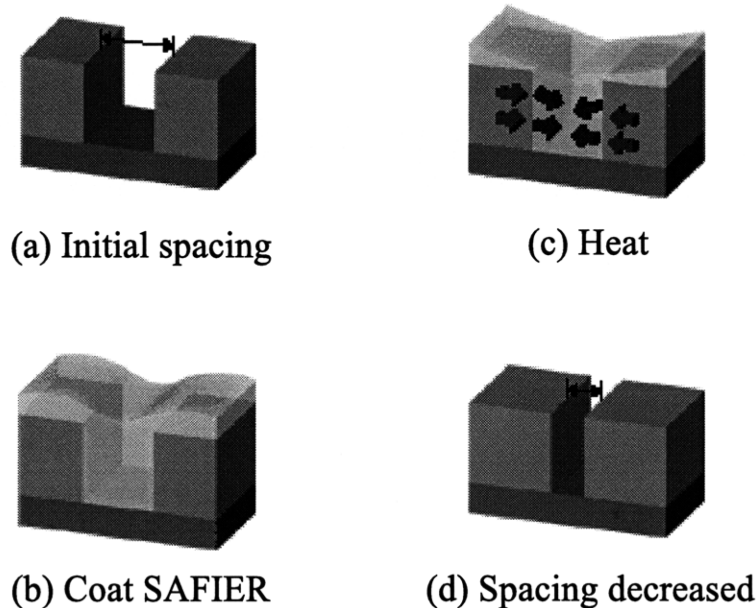


Figure 6-1: The SAFIER process to reduce the (a) initial spacing of a trench by, (b) coating the SAFIER polymer and (c) going through a thermal cycle. The reduced trench spacing is shown in (d).

6.1.1 Negative Resist with SAFIER

Fabricating narrow trenches is a similar problem that is prevalent in the semiconductor industry. Reducing the critical dimensions of trenches allows higher density integrated circuits, improving performance. In an effort to fabricate narrow trenches, Tokyo Ohka Kogyo Co. has invented a process called Shrink Assist Film for Enhanced Resolution (SAFIER) [66]. The process effectively reduces the spacing of a narrow trench by using a thermal reflow cycle.

The SAFIER process is depicted in Figure 6-1. Starting with the initial spacing, SAFIER, a proprietary polymer formulation from Ohka, is spin-coated and planarizes the trenches. The sample then goes through a heat cycle above the glass transition temperature of the resist. After removing the SAFIER material by rinsing with water, the final spacing is reduced. The mechanism of narrowing the trench spacing is reflow of the resist under the confinement of the SAFIER material. The narrowing

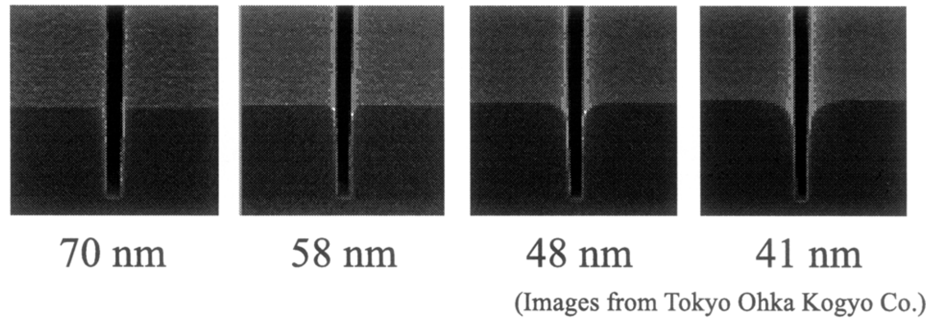


Figure 6-2: The narrowing of e-beam resist trench spacing from 70 to 41 nm after three cycles at 120° C.

is typically 10-20 % of the original spacing.

The advantages of the SAFIER process is apparent. The process is a relatively simple step that can be integrated into a fabrication flow without making any changes. Another great advantage is that the process can be repeated multiple times to further narrow the trench spacing. Experimental data from Ohka are shown in Figure 6-2, which reduced a trench spacing of an e-beam resist from 70 to 41 nm by using the SAFIER process three times at 120° C. Using this process the trench spacing of a resist pattern can be reduced, effectively increasing the duty-cycle.

Even though the SAFIER process presents an interesting approach to increasing the duty-cycle of grating patterns, there are a number of disadvantages. The most important is the number and controllability of the cycles. From a trench spacing of 60-70 nm, the estimated starting spacing with exposing with negative resist, it will take 6 or 7 cycles to reach the target of 20-30 nm. Whether the process can be repeatable is questionable. This problem becomes more pronounced since the temperature needed to reflow the negative resist is much higher at around 200° C. The process is heavily dependent on the thermal cycle, and this can be a problem at such high temperature.

Also the process is based heavily on reflow of material, so it requires high material volume. This is evident in the results shown in Figure 6-2, as the trench edges becomes increasingly round. In the case of periodic patterns, the grating profile fidelity might

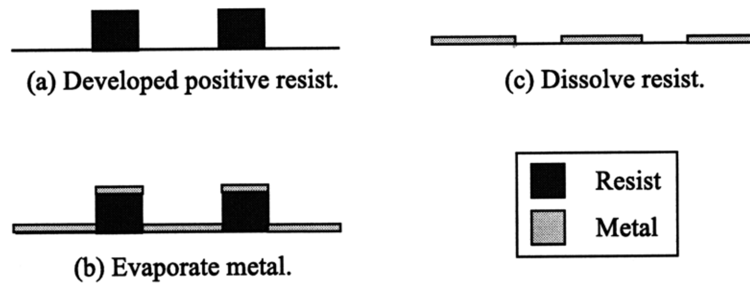


Figure 6-3: The lift-off process to reverse the tone of (a) the developed positive resist by (b) evaporating a metal, and then (c) dissolve the resist.

be greatly reduced. It is questionable if the SAFIER process will be able to reach the target trench spacing of 20-30 nm, while maintaining high profile fidelity.

6.1.2 Lift-off Process

Working with negative resist is significantly more complicated than positive resist. It does not follow the binary model and the temperature cycle that activates the chemical reaction is very sensitive, both making the linewidth difficult to control. It is then more favorable to develop a process that would work with positive resist. One such process is the lift-off process [48], which is a useful technique to reverse the tone of the resist.

An example of how a common lift-off process can achieve a high duty-cycle is shown in Figure 6-3. Starting with a developed positive resist with low duty-cycle, a metal such a Cr is deposited, shown in diagram (b). The resist, together with the metal on top, is washed away in a solution. The resist profile is designed to have a high aspect ratio or negative angle, so the metal film will be discontinuous and easy to remove. The final grating profile in metal will result in a high duty-cycle.

These type of lift-off processes provides a fabrication step which can reverse the tone of the grating profile. However, there are a number of disadvantages that provide significant challenges. One such disadvantage is the dependency of the process on the resist profile. In reality the process can be difficult to perform and is not reliable. There is also problems with the uniformity of the result, since depending on the

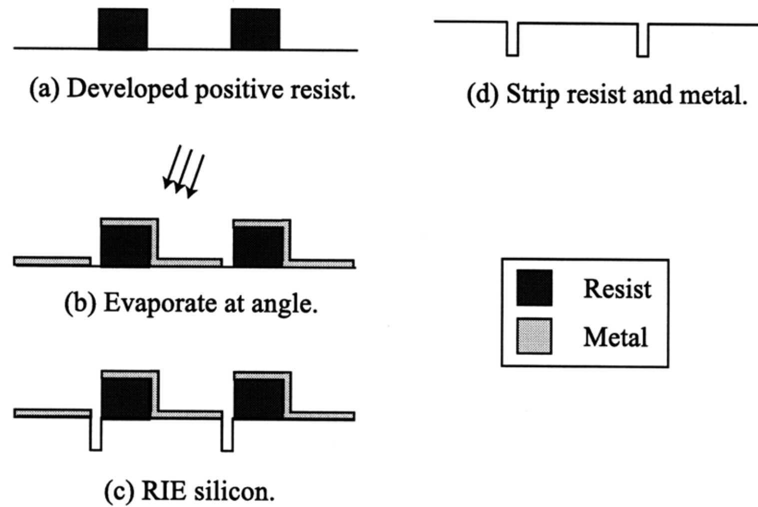


Figure 6-4: The shadow etch process starts with (a) a low duty-cycle profile in positive resist, and then (b) deposited with metal at an angle. (c) The shadowed narrow trench is then transfered into the substrate, and (d) the resist and metal are striped.

deposition technique there might be angle effects. Furthermore, the process does not address the path to 20-30 nm features, and might have difficulties working at that length scale.

6.1.3 Shadow Etch Process

Another technique that is able to fabricate high duty-cycle from positive resist is the shadow etch process [50]. The process, shown in Figure 6-4, uses directional dependence to etch narrow trenches. The process starts with a low duty-cycle pattern in positive resist, which is then evaporated by a metal at an angle. Depending on the angle and height of resist, the directional effect shadows a narrow spacing which has no deposition. The narrow spacing can then be transfered into the substrate with RIE, resulting in a high duty-cycle grating profile. The remaining resist and metal can be removed by a wet etchant.

This shadow etching technique is an interesting approach, but the heavy dependence on angle and resist geometry may make the process difficult to control. This problem becomes significantly worse over large areas, and may result in non-uniform

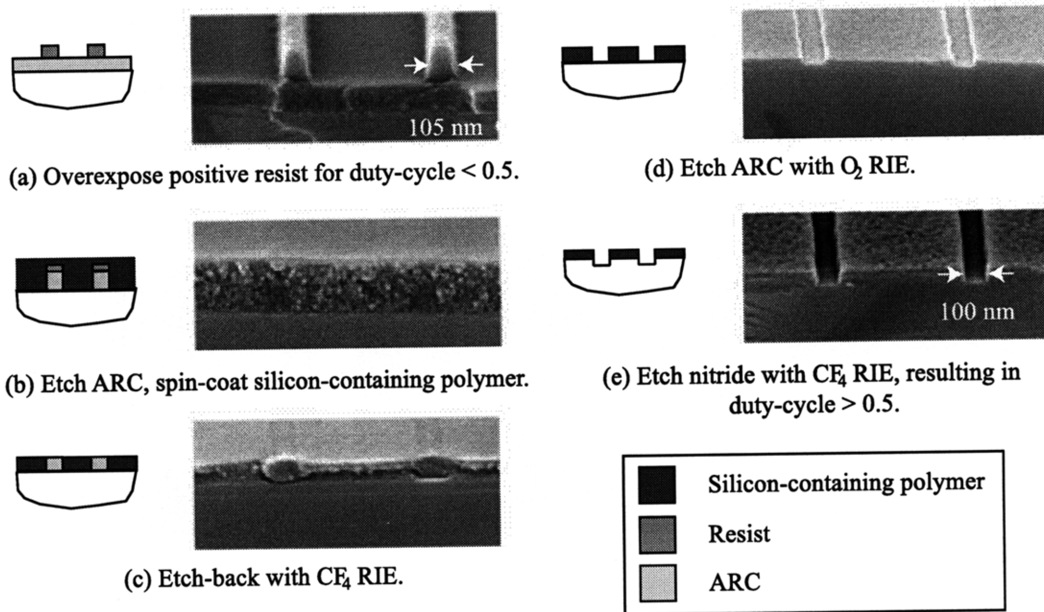


Figure 6-5: The image-reversal process starts with (a) a low duty-cycle grating profile in resist, which is then transfered into ARC. (b) A silicon-containing polymer is spin coated to planarize the ARC, which is (c) exposed by an etch back step. (d) The ARC is then etched away, and (e) the high duty-cycle pattern is transfered into the substrate.

trench spacing. Another significant problem is the since the shadowing effect is only in one direction, a different linewidth also means a constant phase offset from the original structure in resist.

6.1.4 Image-reversal Process

Image-reversal processes are also extremely useful in microfabrication. Some processes use heat treatment to reverse the image of exposed resist, others spin coat additional materials. In Nanoimprint Lithography (NIL) [20] image-reversal is important, as the process replicates the reversed image of the master mold.

Molecular Imprints, Inc. has developed a process to reverse the imprint images patterned by NIL using a spin coated silicon-containing polymer. The polymer formu-

lated by Molecular Imprints has a silicon content, thus making the material resistant to O_2 RIE. It behaves like a spin-on-glass and can be conveniently removed by a Piranha clean followed by an RCA clean. While the material was developed for reversing the NIL pattern, the material is extremely useful and has been adopted to address several goals of the spatial-frequency multiplication process.

In order to effectively and reliably fabricate high duty-cycle gratings, a novel image-reversal process for interference lithography was developed. The primary purpose of the image-reversal process is to reverse the duty-cycle of the resist grating. Thus, starting with a positive resist profile, this process transfers the pattern to a negative resist profile. This allows positive resist to be use in the spatial-frequency multiplication process.

The image-reversal process is demonstrated in Figure 6-5 with a 574 nm period grating. Starting with a low duty-cycle resist grating, the patterns are transfered into the ARC with O_2 RIE. A silicon-containing polymer is spin coated to planarize the grating pattern. An etch back step using CF_4 RIE exposes the ARC, which is cleared out with O_2 RIE. The pattern can then etched into the substrate with CF_4 RIE. This image-reversal process effectively transfers a pattern with duty-cycle DC_o to $\sim 1 - DC_o$. Note that this process will work for any period grating, even though as the period decreases pattern fidelity during the various RIE steps becomes and issue.

Positive resist in interference lithography traditionally patterns duty-cycle from $\sim 0.1 - 0.5$, depending on the grating period, process control and exposure contrast. This image-reversal process eliminates such limitation and allows duty-cycle from $\sim 0.1 - 0.8$. The biggest advantage of this process is that it allows the use of positive resist to fabricate high-duty cycle patterns, a much easier material to work with.

6.2 Line-width Control with Fabrication Processes

The image-reversal process described in the last section presents a simple solution to fabricating high duty-cycle grating patterns. However, there are two challenges that are not addressed. The first is that in order to obtain a duty-cycle in the 0.9 range

after the image-reversal process, the initial resist needs to have a 0.1 duty-cycle. This is difficult due to resist mechanical stability, especially for a small period and thick resist. The second is even if such duty-cycle can be patterned, what level of process repeatability can be achieved? These two challenges will be addressed in this section.

The main challenge of fabricating a low duty-cycle grating pattern in resist is mechanical instability. For a period of 200 nm, a 0.1 duty-cycle pattern means a linewidth of 20 nm. The stack design shows that a relatively thick resist of 200-250 nm is needed to allow effective pattern transfer. The height-to-width aspect ratio of the resist is then greater than 10, significantly high for a polymer with low stiffness. A resist profile with such high aspect ratio would collapse readily during post-develop drying due to surface tension. Experimentally for such a resist thickness, the lowest duty-cycle that will not collapse during drying is around 0.2.

To overcome the mechanical failure from surface tension, several alternatives are possible. The first is eliminating the water-air drying step by using a critical point dryer. The process involves drying in liquid CO_2 by heating to the critical temperature of 31.1°C at high pressure, in which the liquid would vaporize. The surface tension from the phase transition is essentially zero, and thus preserves the resist grating. This step allows grating profile with lower duty-cycle, which can be image-reversed, however it does not address the challenge of linewidth control. Discussions in Chapter 5 addressed in great detail the feasibility of controlling the linewidth with lithography.

Another approach is to post-develop dry the grating at higher duty-cycle such as 0.2-0.3, and then using other dry processing steps to reduce the linewidth. A very effective process for this task is high pressure oxygen plasma etching (PE). PE is non-directional and allows isotropic etching, which is ideal for a linewidth trimming process.

6.2.1 Reducing Duty-cycle with PE Trimming

The PE trimming process is outlined in Figure 6-6. The process starts with a positive resist grating profile, which is transfered into ARC with O_2 RIE. An isotropic high pressure oxygen PE is then used to reduce the linewidth of the polymer grating. Since

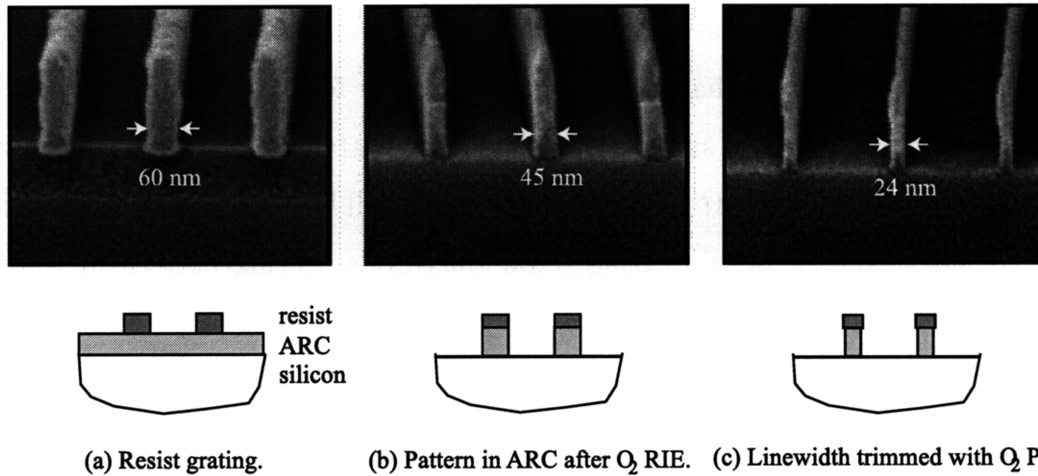


Figure 6-6: The plasma trimming process starts with (a) a grating profile in positive resist. (b) The pattern is transfered into ARC. Using a high pressure oxygen PE, (c) the grating linewidth is reduced.

ARC has a higher etch rate than resist in an oxygen plasma, the ARC results in a smaller linewidth than the resist. Therefore the ARC linewidth is designed to act as the critical dimension, while the resist serves as a masking material to preserve the profile of the ARC. Without the resist mask, the top of the ARC will erode in the PE trimming step.

A significant advantage of the PE trimming process is that it is a dry etching step. The grating with trimmed linewidth is extremely weak, and collapses readily when drying. The process thus eliminates the limitation of surface-tension-induced instability by avoiding a wet process. There is still a mechanical limit, however, which occurs when the linewidth becomes increasingly small so that the grating will collapse by its own weight.

The effect of the PE is demonstrated in the SEM images shown in Figure 6-7. The starting linewidth of the ARC grating is 45 nm, shown in Figure 6-7(a). The ARC has a thickness of around 120 nm, the while the eroded resist has around 90 nm. The grating is then etched in an oxygen PE at 200 W and 0.1 Torr for 90 s, and the linewidth has been trimmed to 24 nm, shown in Figure 6-7(b). This effectively reduces the duty-cycle of the grating pattern from 0.23 to 0.12, a significant

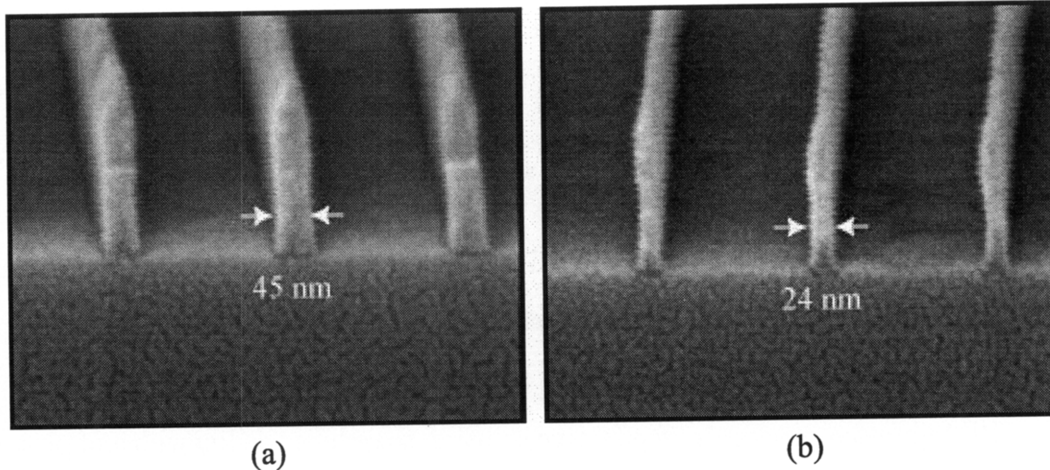


Figure 6-7: Cross-section SEM of (a) ARC grating with 45 nm linewidth. (b) The linewidth is reduced to 24 nm after 90 s PE trimming.

improvement. Combining this process with the image-reversal process, the duty-cycle can theoretically be ~ 0.9, which is in the desired range for the multilayer interference lithography process.

The boundary between the ARC and resist is difficult to see since the linewidth is relatively small, and secondary electron scattering makes the image bright. The trimmed profile also has a tree-like profile, which is because the ARC etches about 20-30% faster than resist during the PE process. Due to the resist mask, the ARC resist profile is extremely rectangular, which makes pattern transfer significantly easier and more robust.

The critical dimension after the PE trimming process is the linewidth of the ARC. The profile of the resist is typically severely eroded, and would not be the ideal profile for pattern transfer. This is critical in linewidth control during the image-reversal process. Depending on the etch-back depth, different linewidths can be obtained. For example if etch-back stops at the ARC plane, the linewidth would be smaller than if the etch-back stops at the resist plane. And since the resist profile is eroded, the linewidth would change dramatically depending on where the etch-back step ends. It is thus desirable to use only the rectangular ARC profile in the image-reversal step.

The resist can be removed in an solvent, such as N-Methylpyrrolidone (NMP),

which the cross-linked ARC would be resistant to. However, upon wetting the sample, the thin ARC grating lines would collapse while drying. In practice, the solvent used in formulating the Si-containing polymer is 1-methoxy-2-propyl-acetate (PGMEA) [54], which can dissolve the uncross-linked resist. Thus when spincoating the Si-containing polymer for the image-reversal step, it is important to dispense greater volume than necessary. The resist can then be dissolved, and only the ARC grating will remain.

The PE trimming process is an effective solution to reduce the linewidth of periodic structures. It avoids the mechanical limitations due to surface tension by being a dry process. The process integrates well with the image-reversal process, and allows high duty-cycle grating patterns necessary for the spatial-frequency multiplication process. Using these processes, a linewidth of ~20 nm has been experimentally achieved, resulting in ~0.9 duty-cycle after image-reversal.

6.2.2 Repeatability of PE Trimming

The PE trimming process, in conjunction with the image-reversal process, has addressed the challenge of fabricating high duty-cycle grating patterns. However, the question of controlling the linewidth to the nm-range still remains. It is essential that every grating layer can be fabricated consistently with the desired linewidth.

To overcome this challenge, the PE trimming process is examined in further detail. The key parameter is the etch rate, and if it is controllable and repeatable. These two issues are related, but not the same. Controllability can be defined as how stable is the etch rate in a single process, as the duty-cycle changes. While repeatability relates to the etch rate being constant from sample to sample. Both of these issues are important, and the process would be ideal if it is both controllable and repeatable.

The etch rate for a single sample over time was tested by an experiment. A sample with a 200 nm period grating profile and 45 nm initial linewidth was PE trimmed with 200 W and 0.1 Torr. The etch time varies from 20 to 90 s, in 10 s increments. The linewidth of ARC was measured with cross-section SEM images. The resulting linewidth is plotted vs the etch time, shown in Figure 6-8. The experimental data is

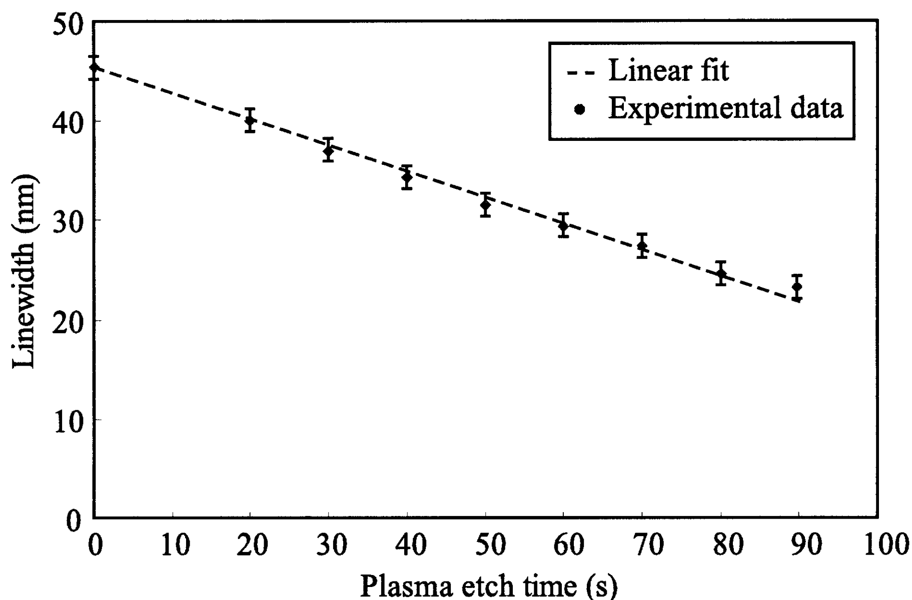


Figure 6-8: Linewidth of a 200 nm period ARC grating vs PE etch time.

fitted by a line, and the equation is shown in the figure. The etch rate is ~0.26 nm/s. From this plot, the etch rate seems extremely linear. The slight deviation lies within the uncertainty of the measurement, which is limited by the resolution of the SEM, around 1-2 nm.

The linearity of the etch rate allows the PE trimming to be controlled for a sample by adjusting the etch time. The repeatability of the etch rate, however, is a critical issue. The plasma etcher being used to perform the PE trimming is an older model, and both the power and pressure are control by manual dials. The main purpose of the system is actually to strip organic materials, and the lack of updated control may result in varying etch rates.

To test the repeatability of the etch rates, three different etch trials were conducted. The samples have exposed resist and 120 nm of ARC grating pattern on top of 40 nm of nitride. The different trials were taken on different days, which simulates closely to trimming the different layers. For each individual trial, several etch times were tested, and for each data point the sample was removed from the chamber and reloaded. The etched width, defined as the amount of linewidth that was etched, is

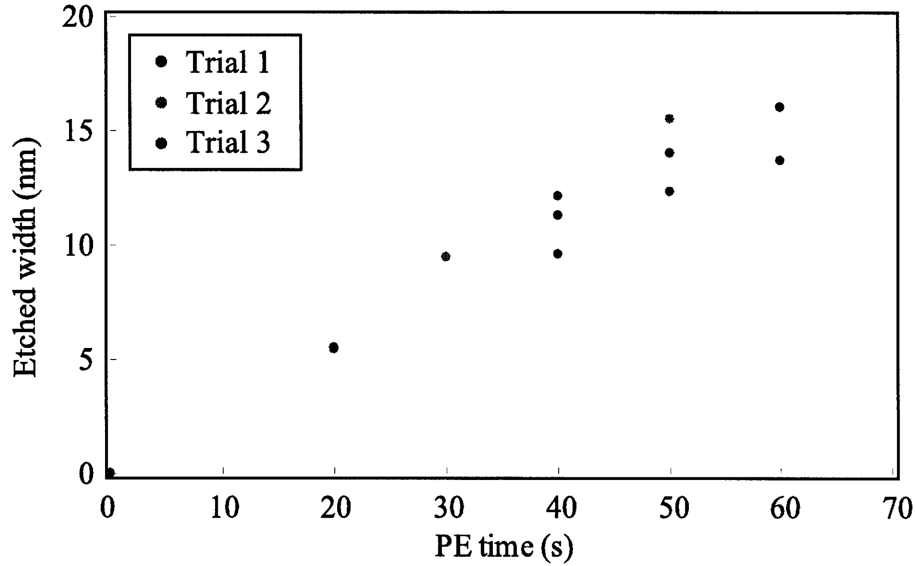


Figure 6-9: Etched width of the ARC grating for three different trials. The difference in etch rate can be observed.

plotted vs etch time in Figure 6-9.

From these data the etch rates of the different trials are shown to be not constant. The etch rates can be determined by taking the slope, and they are 0.27, 0.31, and 0.24 nm/s, respectively. The variations in etch rates were found to be as high as 30%, which would result in significant errors in linewidth control. These results conclude that it is not feasible to control the PE trimming by a single etch run.

This experiment also draws the conclusion that while the etch rate is not repeatable over days, it is consistent within a single etched sample. The etch linewidth for each trial remains very linear, as previously suggested. The daily fluctuation may result from the imprecision of power and pressure setting of the plasma etcher, since the manual dials are not readjusted within a single trial. The linearity in etch rate is an important factor to take advantage of.

From these results, the PE trim rate for a given day is unknown, but it will be linear. The solution to nm-range linewidth control is then frequent metrology for feedback. The procedure is then to measure the initial linewidth with top-view SEM after transferring the grating pattern into ARC. With that measured linewidth and

the linewidth goal, the sample can be PE trimmed for a fraction of the time needed based on the average etch rate of 27 nm/s. The sample linewidth is then measured again by top-view SEM, and using this measurement an actual etch rate can be calculated. The sample can then be further trimmed by the updated etch rate to the linewidth goal.

This process uses top-view SEM measurement as a discrete feedback signal for linewidth. In the procedure described above, two metrology steps were used. In reality more measurements should be used to improve accuracy. In the ideal case the number of metrology steps approaches infinity, and this system is essentially a continuous feedback system with a step input. For the sake of textbook argument, since the PE process can only trim, the close-loop system can not be allowed to overshoot the target linewidth. This is of course impossible to implement, since SEM metrology is extremely time consuming. A realistic procedure involves around three metrology steps, which is enough to achieve nm-range linewidth trimming accuracy.

6.2.3 Linewidth Control of High Duty-cycle Patterns

In the previous sections an image-reversal and a PE trimming process were introduced. The image-reversal process allows the use of positive resist, which has a simpler chemistry, and the PE trimming process allows controlled linewidth reduction. By combining these two processes, a high duty-cycle grating profile can be fabricated with nm-repeatability.

The process flow diagram is illustrated in Figure 6-10. The process starts with a low duty-cycle grating pattern etched into ARC. The linewidth is measured by top-view SEM, and then reduced using the PE trimming process. This step is the linewidth control cycle. The linewidth is repeatedly measured and trimmed until the target linewidth is reached. Finally the low duty-cycle grating profile is image-reversed to obtain a high duty-cycle profile.

In this process the loop of SEM metrology and PE trimming emphasizes the control of linewidth. A typical fabrication cycle with two PE trimming steps is shown in Figure 6-11(a)-(c). The image-reversal process then reverses the tone of the

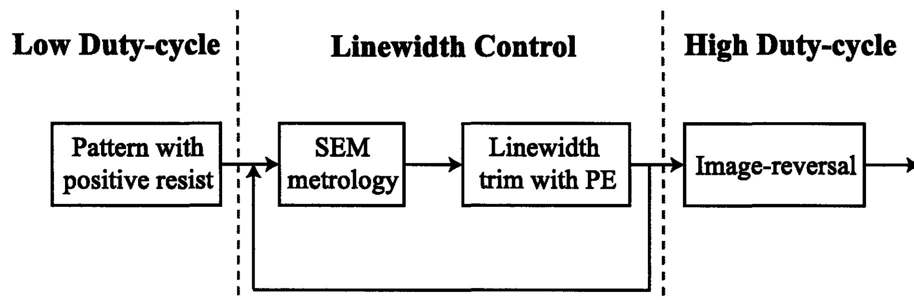


Figure 6-10: Process flow diagram for fabricating high duty-cycle grating profile with nm-repeatability.

grating pattern, shown in Figure 6-11(d). The accuracy of this process is dependent on two fundamental factors: The number of iterations and the metrology accuracy. The desired duty-cycle can be achieved with higher confidence by going through more numerous iterations. The ideal process, as previously mentioned, is for infinitesimal etch time increments.

The second factor is the measurement accuracy, which is limited by the resolution of the SEM and the measurement process. The resolution of the SEM used is around 1 nm, and that sets the fundamental limit in the accuracy of linewidth control. A common saying in metrology is "it is impossible to control what you can not measure."

The measurement process is also less than ideal. Due to the need for non-destructive measurement, the linewidth is measured with top-view SEM and no gold coating. Typically in SEM metrology, a conductive film such as gold is sputter coated onto the sample to prevent local charging of electrons, which will cause image artifacts. This is especially critical for polymer films, because a highly charged region can damage the material over time. Since the SEM metrology is only an intermediate step and further processing is required, the sample can not be gold coated.

The top-view mount is also a factor in the accuracy of linewidth control. The SEM metrology is designed to measure the linewidth of the ARC, which is masked by the resist. The two materials initially have the same linewidth after the RIE pattern transfer, but during the PE trimming the etch rate of the ARC is around 20% faster than that of resist. Even though it is impossible to measure the ARC directly, it

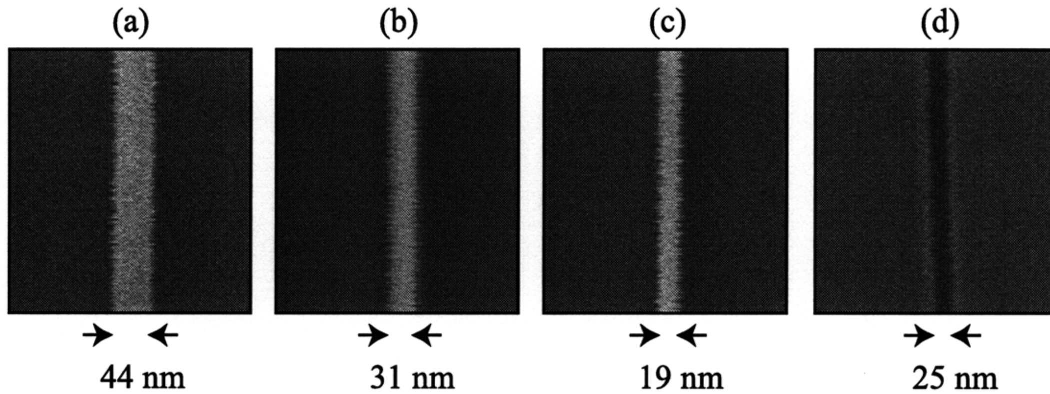


Figure 6-11: The linewidth is being (a)-(c) PE trimmed, and (d) image-reversed.

is feasible to measure the linewidth of resist and then approximate the linewidth of the ARC. The approximation becomes negligible as the etch times decreases, and for typical etch times of 60 s the linewidth difference is only 3-5 nm and can be accurately corrected.

The proposed linewidth control process has several factors limiting the fundamental accuracy. These factors include finite iteration cycle, SEM resolution, charging of samples, and ARC linewidth approximation. With these factors, the linewidth control precision is expected to be around 1-2 nm, and this describes the fundamental limit of the process. These results are reported in Chang *etal.* [13]. In reality there are more process and fabrication factors that further limit the linewidth control accuracy, which will be discussed in the next section.

6.3 Limitations of the Linewidth Control Process

The PE trimming and image-reversal processes introduced in the last section provides an effective technique, although a bit tedious, to control the linewidth of individual grating layers to 1-2 nm. In practice, there are other factors which lower the linewidth control accuracy. These factors include non-perfect reversal of image, etch control, and thickness variation.

In experiments, the image-reversal process is unable to perfectly reverse the duty-

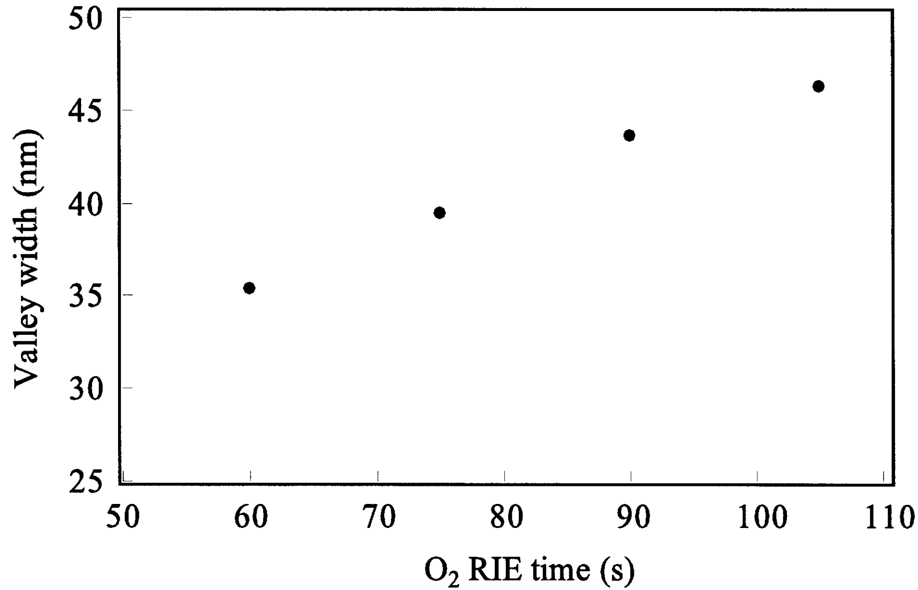


Figure 6-12: Process flow diagram for fabricating high duty-cycle grating profile with nm-repeatability.

cycle of the ARC grating profile. This is due to the erosion of the Si-containing polymer during the O_2 RIE step, depicted in Figure 6-5(d). Even though the Si-containing polymer is not chemically reactive, physical bombardment can sputter the edge and increase the linewidth. Since this effect depends on the absolute scale of the feature, it can be significant in high duty-cycle, small period patterns.

This effect is illustrated in Figure 6-12, where different O_2 RIE times were used. The plot depicts the measured width of the valley after RIE vs. the etch time. The initial linewidth of the ARC is around 27 nm before the image-reversal process. At the estimated etch time of 60 s, the linewidth has been increased to 35 nm. At shorter etch times the bottom of the ARC is not completely etched, so this increase in linewidth can not be further reduced. Additional etch times at 15 s increments were tested, and the linewidth continues to increase. At a total etch time of 105 s, the linewidth has been increased to over 45 nm.

These results show that overetching of the ARC during the image-reversal process can greatly increase the linewidth. It is then critical that the O_2 RIE etch time

be carefully monitored to improve linewidth control. Ideally an endpoint detection scheme would alleviate the problem, such as optical absorption of the volatile gas species. However such a system is not available and a simpler monitor system can be used instead.

The O_2 RIE etch time depends on the thickness of the ARC, as illustrated in Figure 6-5(e), and is determined by the CF_4 RIE etch-back. While the purpose of the etch back step is to expose the ~120 nm thick ARC, it is often overetched to reduce the aspect ratio of the trench. A typical thickness is around 80 nm, which makes the ARC easier to be removed in the next step. However, the residual thickness can fluctuate even with constant etch time, since the etch parameters are not always stable from day to day.

To determine the etch depth, a monitor sample is placed in the chamber during the etch back process. The sample is a continuous film with the same Si-polymer thickness as the product sample, and the thickness before and after the etch can be measured by an ellipsometer. Using such a method the etch depth can be monitored, which makes the process more consistent from layer to layer. Typical etch depth variations can be up to 20 nm. For the case of underetch, further CF_4 RIE is needed, and for overetch the etch time for the following O_2 RIE can be reduced. Using these etch monitoring samples, the linewidth control can be significantly improved.

Cross-section SEM of a typical linewidth control process is shown in Figure 6-13. The linewidth of the grating is still increased by the image-reversal process, but the erosion is minimized with process monitoring. In this sample the linewidth increases from 18 to 26 nm, which is a fair representation of the image-reversal process. As a result, even though the linewidth control is repeatable to a few nm, the accuracy is not as good.

Another critical issue is thickness variation of the Si-polymer film. This is typically caused during the spin-coating of the Si-polymer on the trimmed ARC grating pattern due to particles. The particles cause comet-like streaking patterns and result in local thickness variation. Particle contamination is typical during spin-coating, but in most cases the films can be removed in wet solution and the film can be recoated. However,

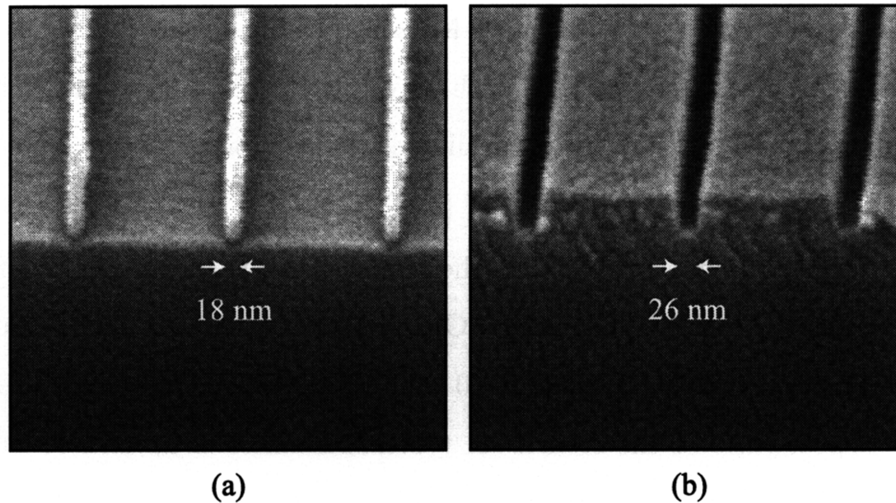


Figure 6-13: (a) The linewidth of the grating is (b) increased by the image-reversal process.

in this process the coating of the Si-polymer is irreversible. Cleaning the sample would require restarting the process for the layer. The residual resist may also increase the chance of particles if is not totally dissolved in the solvent. The particle streaking is problematic, and can result in dramatic linewidth variations or in the worse case, no grating pattern at all.

6.4 Conclusion

The proposed linewidth control method combines an image-reversal process with a PE trimming process, and is capable of fabricating the desired high duty-cycle. The image-reversal process involves spin-coating a Si-polymer which planarizes the ARC grating profile, and using various RIE etch steps to transfer the pattern into the device layer. The process allows the use of positive resist, which is significantly simpler to simulate and control.

The PE trimming process controls the linewidth of the grating pattern by using numerous SEM metrology steps. Experiments have shown that the process can effectively reduce the linewidth of ARC to 20 nm, and can be controlled since the etch rate

is linear. The PE trimming process is then coupled with the image-reversal process to fabricate a high-duty cycle grating pattern with 1-2 nm accuracy.

The limiting factors of this technique have been shown to be the number of SEM measuring iterations, the resolution of the SEM, and the measurement technique. Process control is extremely important during the image-reversal process, and monitor samples are used to measure the residual thickness.

The linewidth control process described is based on metrology and process control. Using this process the desired high duty-cycle grating pattern with ~25 nm valley width can be fabricated and controlled with nm-precision.

Chapter 7

Results and Discussion

From the previous chapters we have addressed the main challenges involved in the multilevel interference lithography process. Combining accurate overlay alignment, an optimized multilayer resist stack, and precise linewidth control, we have successfully fabricated gratings with spatial-frequency multiplication factors of two and four. In this chapter we present the fabrication and testing results for the multilevel process.

There are several key experimental results that validate the effectiveness of the proposed multilevel process. The first demonstrate the large-area phase overlay accuracy between two grating levels. This data is measured by top-view scanning electron microscopy, and demonstrates the ability of our process to align multiple grating levels. The cross-sectional profiles are also examined, and the spatial-frequency content of the fabricated grating is analyzed.

Finally, results from the third and fourth grating level exposures are presented. Using this process, we have successfully fabricated 50 nm-period gratings using light with 351.1 nm wavelength. The ultra-fine features with linewidth of 25 nm were transfered into a hard silicon nitride mask. We will also discuss future applications for multilevel interference lithography, including 2-dimensional hexagonal patterns and 3-dimensional photonic crystals.

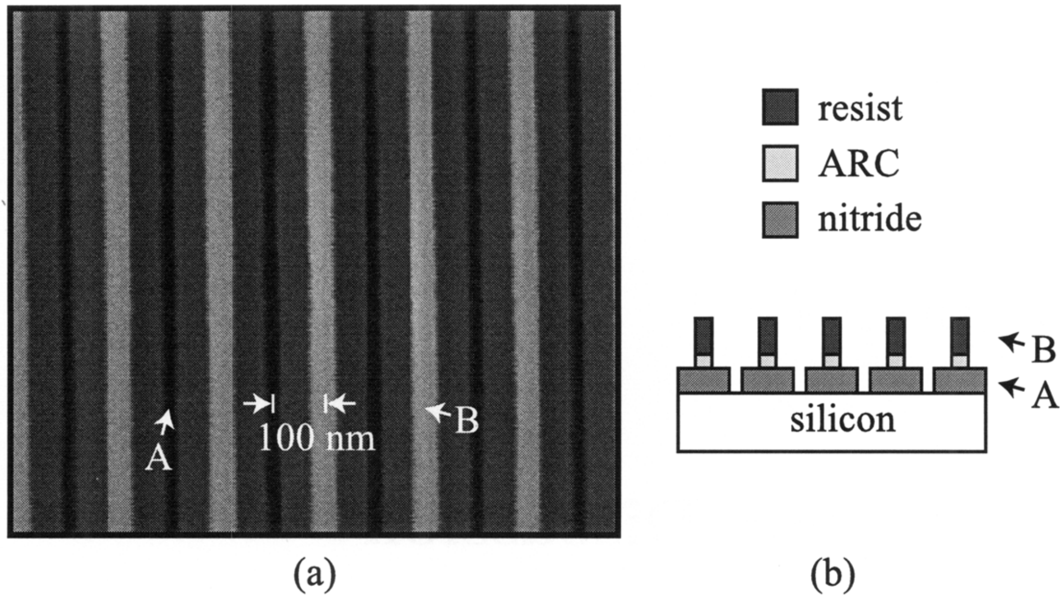


Figure 7-1: (a) Top-view SEM image measuring the overlay accuracy of the first and second grating levels, denoted as "A" and "B," respectively. (b) Illustration of the cross-sectional profile of the multilayer stack.

7.1 Phase Overlay of Two Grating Levels

One of the main challenges facing the multilevel IL process, as previously described in Chapter 3, is the overlay accuracy of multiple grating layers. In this section we examine the phase overlay accuracy of two 200 nm-period grating levels with top-view SEM. In this experiment the first grating level is etched into the underlying nitride layer, while the second grating level is just resist. It is critical to examine the phase overlay accuracy during this process step. This is because once the second grating is image-reversed and etched into the nitride layer, it will be difficult to distinguish between the grating levels. In fact the goal of high precision linewidth control, as described in Chapter 6, is that the two gratings levels are identical.

A top-view SEM image can then be taken after transferring the second grating level into ARC, following exposing the first grating level in the process, as shown in Figure 7-1(a). The geometry of the multilayer stack is better illustrated in a cross-sectional profile illustration, as shown in Figure 7-1(b). The dark lines denoted as

"A" are the thin valleys etched into nitride patterned by the first grating exposure. The bright lines denoted as "B" are resist peaks patterned by the second grating exposure.

Perfect phase overlay alignment then requires the bright lines to be positioned between the dark lines, separated by 100 nm. From the figure we can not visually detect any misalignment at this particular location. These results suggest that the overlay accuracy between the two grating lines are comparable or even better than the spatial resolution of the SEM, which is around 1-2 nm. This metrology method is extremely useful, as if provides a feedback signal for the individual exposures. If significant phase errors are detected, the second grating level in resist can be cleared and repatterned. Therefore, high phase overlay accuracy for each grating level can be ensured.

Using top-view SEM images, the overlay accuracy can be measured experimentally. However, such an image only measures the phase overlay over micrometer lengthscales. To examine phase overlay accuracy over large areas, a 2D array of SEM images can be taken, as shown in Figure 7-2. The top-view SEM images are spaced 2 mm apart in both x and y-directions. A total of 144 images over an area of 30 mm $\times$ 16 mm were taken using an automated imaging algorithm with the Raith 150 [36]. All of the images were taken in under an hour, while manual imaging of so many micrographs can take significantly longer.

These micrograph arrays are extremely useful to examine micro and nanostructures over millimeter distances. While it is difficult to measure the absolute phase of gratings among the images, relative measurements such as phase overlay errors are extremely simple. We can then use these SEM image arrays to determine the overlay accuracy of two grating levels over large areas.

A Matlab algorithm can then automatically detect the maximum and minimum pixels to determine the overlay errors between the two grating levels. The data from 144 top-view SEM images can then be used to generate a phase contour map, as shown in Figure 7-3. In this figure the grating lines are aligned horizontally, representing the x-axis. The phase map shows that there are no systematic linear phase distortions

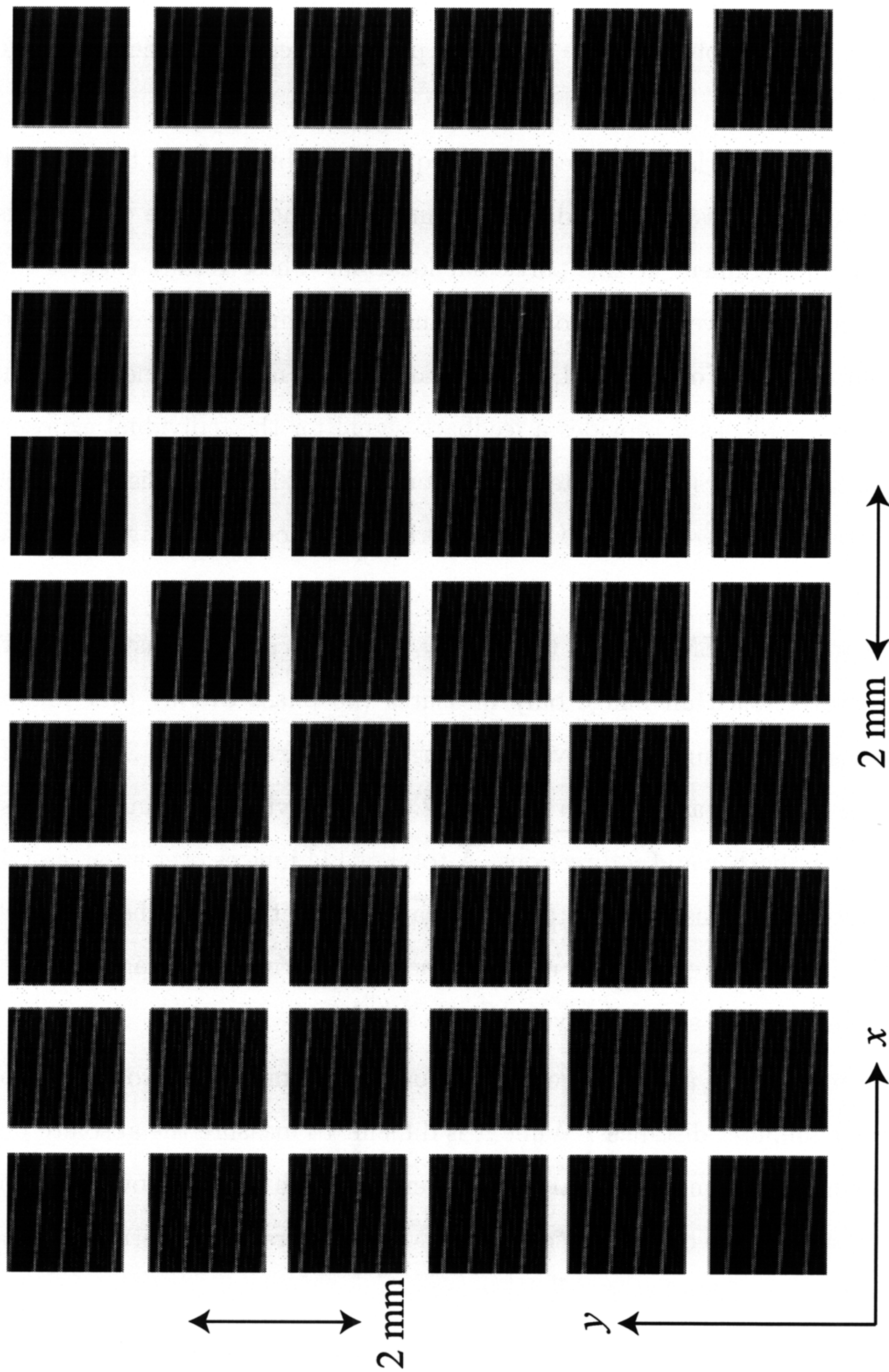


Figure 7-2: A 2D array of top-view SEM images spaced 2 mm apart to measure the phase overlay of two grating levels over large area.

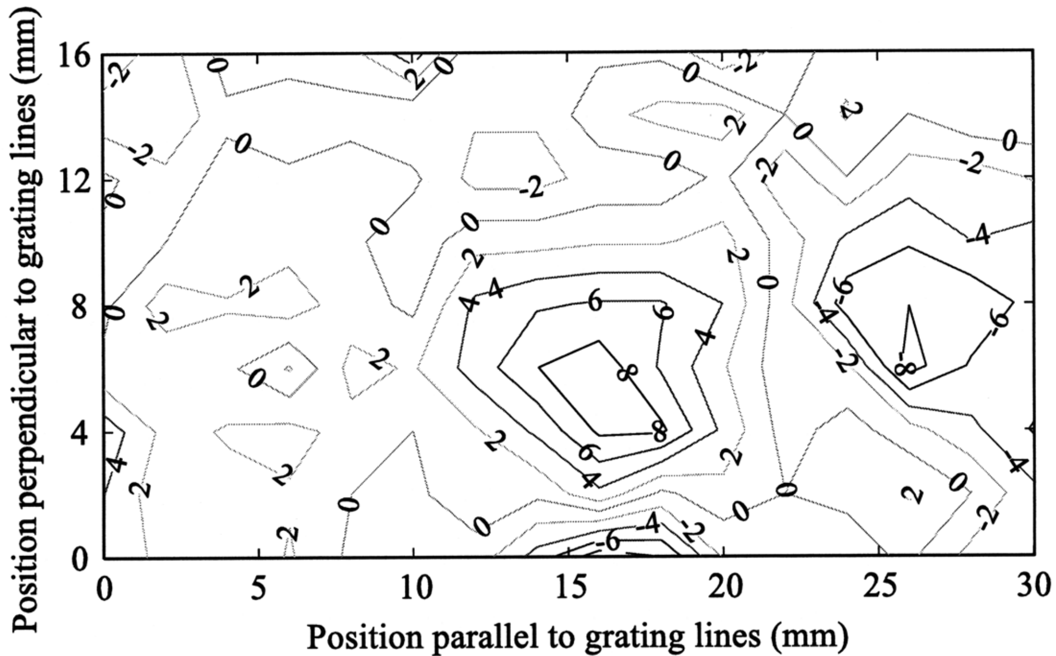


Figure 7-3: Phase map of overlay error in nanometers for two 200 nm-period grating levels. No systematic distortion is evident, however there is a local particle induced phase distortion near position (16,2).

that are associated with grating alignment. The data represents the whole device region in which the gratings were overlaid.

There is, however, clear evidence of particle-induced phase distortions near the area centered at (x = 16 mm, y = 2 mm) in Figure 7-3. The opposite sign of phase distortion above and below the area is a distinct signature of a particle, as described in Chapter 3. The magnitude of the error is relatively small, with peak distortion around 8 nm. There are also some local distortion at the right side of the phase map, which may be due to more than one particle. For the area $x < 12$ mm, the overlay accuracy is 0.6 ± 1.9 nm (1σ) over a 16×12 mm^2 area. This level of distortion is satisfactory for our multilevel IL process.

Since there is no apparent linear phase distortion, the grating alignment procedure we use is very effective. Our overlay accuracy for multiple grating levels then depends on our ability to keep the substrates free from particles. While the use of a pin chuck has significantly reduced this problem, particles remains a major concern. All of

the fabrication processes were done in class 10-1000 cleanrooms, and the substrate carefully cleaned prior to critical steps. However, in a university research setting it is difficult to completely eliminate particles from the picture. A purely mechanical process without human intervention is the only method to be completely particle-free.

An important point is that particle-induced in-plane distortion can occur in any lithography process where the substrate is constrained by a vacuum chuck. Most of the time the distortion is relatively small, depending on the size of the particle, and can not be easily detected. For interference lithography this is especially critical, especially for fine grating periods. The problem is that even if a grating structure has in-plane distortions in the order of 10s of nanometers, it is incredibly difficult to detect even with an electron microscope. The best method to detect these nanometer in-plane phase distortion is by interferometry, and in our case, a moiré-type technique.

7.2 Two-fold Spatial-frequency Multiplication

In the previous section we demonstrated our ability to overlay two 200 nm-period grating levels with accuracy in the order of nanometers. The rest of the multilevel IL process can then be used to image-reverse and transfer the second grating level to the nitride layer, resulting in two-fold spatial-frequency multiplication. There are several key factors to look for from the results, such as linewidth variation and spatial-frequency content of the profile.

A top-view micrograph of the spatial-frequency doubled grating is shown in Figure 7-4. In this figure two grating levels, each with 200 nm period, are overlaid to fabricate the 100 nm-period grating. The dark lines, around 25 nm wide, represent narrow valleys etched into the nitride layer. It is important to note that every odd and even number lines are actually etched by two different grating levels that are patterned days apart. The distinction between the two sets of lines are not visually obvious. This is due to the nanometer-level precision of the linewidth control process, as outlined in Chapter 6.

These results demonstrate that this grating has accurate alignment and repeatable

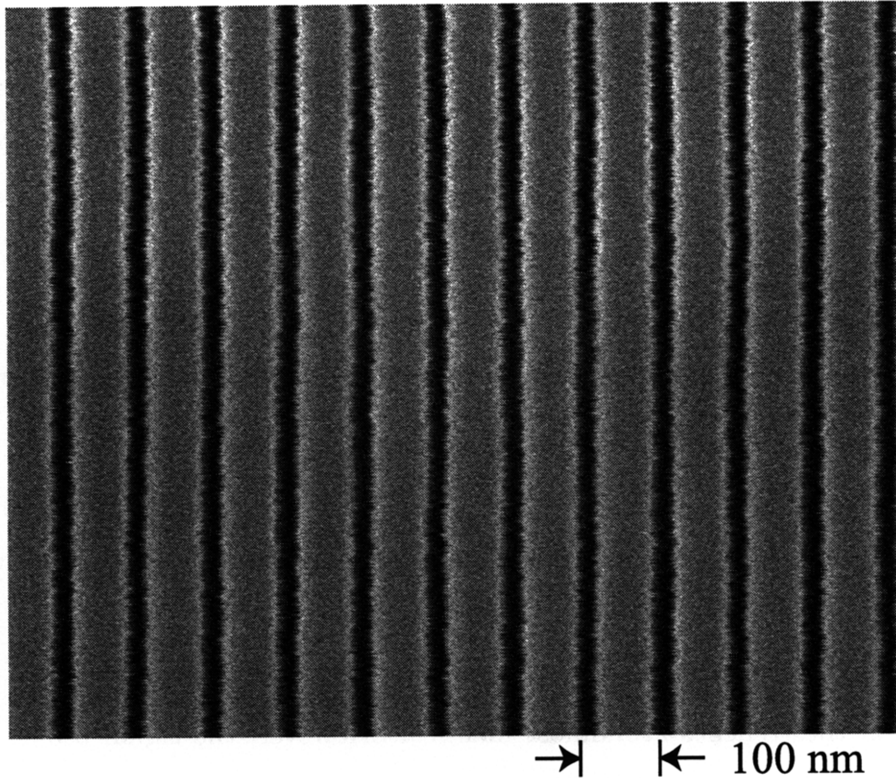


Figure 7-4: Top-view micrograph of two overlaid 200 nm-period grating levels to achieve a two-fold spatial-frequency multiplication, resulting in a grating with 100 nm period.

linewidth control. However, the fidelity of the grating profile, and more importantly the spatial-frequency content of the sample, is unknown. To examine the grating profile, cross-sectional SEM imaging is typically used. This is highly undesirable in our case, as we want to examine the profile as a function of location. Cross-sectional micrographs yield a relatively small amount of data that can not be accurately associated with location, while destroying the sample.

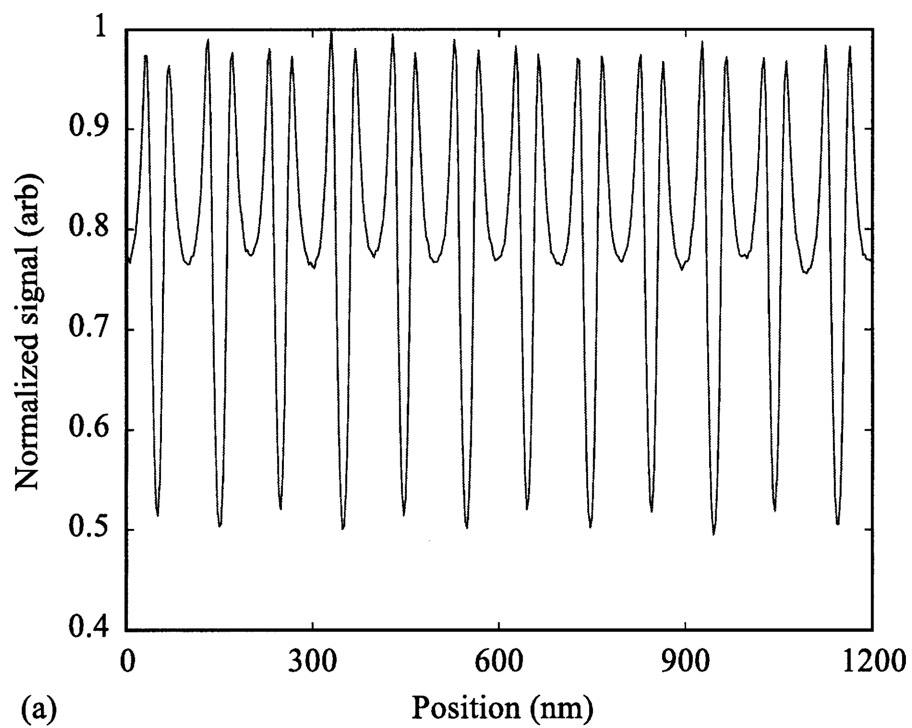
To investigate the grating profile, the relative signal intensity from top-view SEM images can be used. Our approach is then to spatially average a micrograph to produce a single low-pass-filtered linescan, and examine the profile of the signal intensity, as shown in Figure 7-5(a). In this figure the normalized signal profile is plotted from 0 to 1200 nm, which comprises six periods of the fundamental frequency. While the

signal profile does not represent the exact geometric profile of the grating, but rather a convolution of the grating and the response of the secondary electrons, it allows us to examine the difference between the trenches etched in two grating levels. Note that any type of electron imaging will have artifacts due to secondary electron scattering, especially at the edges. For our case the problem is amplified, since we are imaging deep narrow trenches. However, this method allows us to image over large areas to yield important data, and we are only interested in the differences between the two trenches.

From Figure 7-5(a), the signal profiles of the two trenches are relatively similar. The signal at the trench edges has high intensity due to edge artifacts from strong secondary electron emission, and are commonly seen. The width of the trenches measured near the peak intensity are similar to single pixels, supporting the fact that the width of the trenches are almost identical. However, the signal profile suggests there is some deviation in the depth of the trenches. While we can not determine the exact depth offset in nanometers, it is consistently present in all the images.

The power spectral density (PSD) of the signal profile can then be calculated, as shown in Figure 7-5(b). Both the fundamental and the two-fold multiplied spatial-frequency are denoted in the figure. The PSD allows us to analyze the relative power of the spectral components, and presents valuable evaluation for the spatial-frequency multiplication process. For this signal profile, the power of the fundamental spatial frequency has been suppressed to around 1/8 of the spatial-frequency doubled component.

Following this line of analysis, 144 top-view micrographs were then again taken from the previous sample, and their linescan signal profiles examined. In Figure 7-6(a) the relative power for the spatial-frequency doubled spectral component is mapped as a function of position. Note that this sample is the same as the one previously used to measure the phase overlay accuracy, as shown in Figure 7-3, except the second grating level has been etched into nitride. Due to this fact, the effect of the particle can be readily seen in this figure. As expected, the overlay error between the two grating levels have reduced the effects of the spatial-frequency multiplication. The



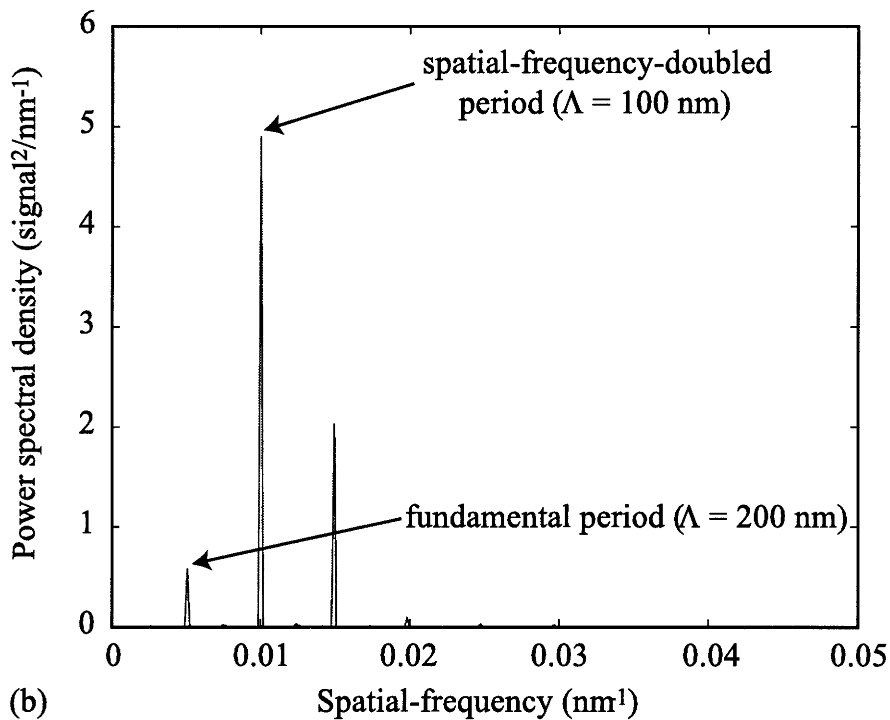
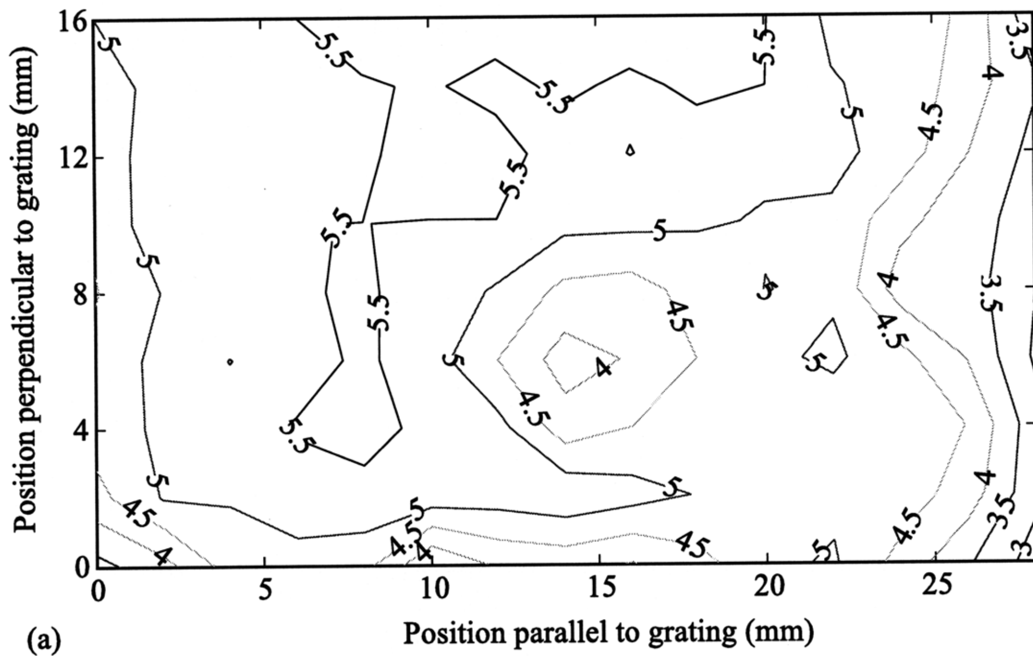


Figure 7-5: A top-view micrograph is used to extract (a) a single averaged linescan, and (b) the power spectral density.



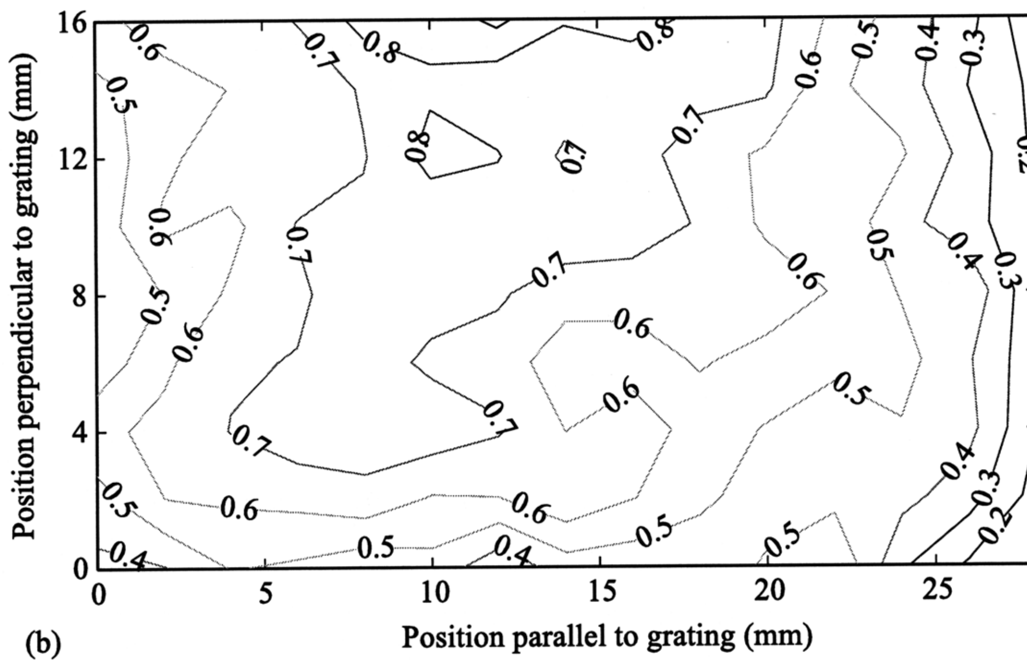


Figure 7-6: Relative power map for (a) the two-fold multiplied and (b) fundamental spatial frequency. The effects of the particle can be readily detected.

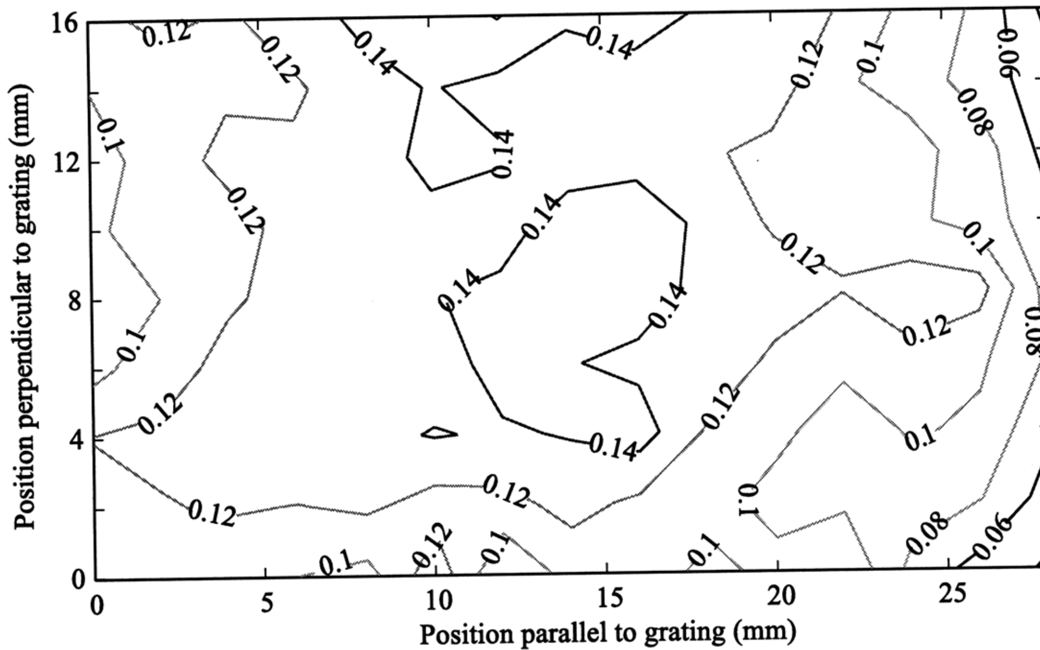


Figure 7-7: Power ratio map of the fundamental frequency and two-fold multiplied components.

power for the fundamental frequency can also be mapped, as shown in Figure 7-6(b).

We can combine both power maps, and calculate the ratios between the fundamental frequency and the frequency-doubled components, as shown in Figure 7-7. This is a good indicator on how well the fundamental spatial frequency is suppressed. The ratio is around 14% in the center region, and decreases gradually towards the outer region. The circular contours are quite interesting, and suggests a signature of the etch variations in the RIE system. The etch rates for our RIE system, and most non-manufacturing type systems, are highest in the center and drops off by ~10-20% towards the edges, forming a bull's eye shape pattern. For this reason it is difficult to control the etch depth for any etch through process, and an etch stop using another material is usually used.

We believe the source of non-perfect suppression of the fundamental frequency component is due to the variations in etch depth. This effect can be clearly observed in Figure 7-5(a). This result is not surprising, considering etch depth was disregarded as a parameter to control. This decision is due to the difficulties in controlling etch rates

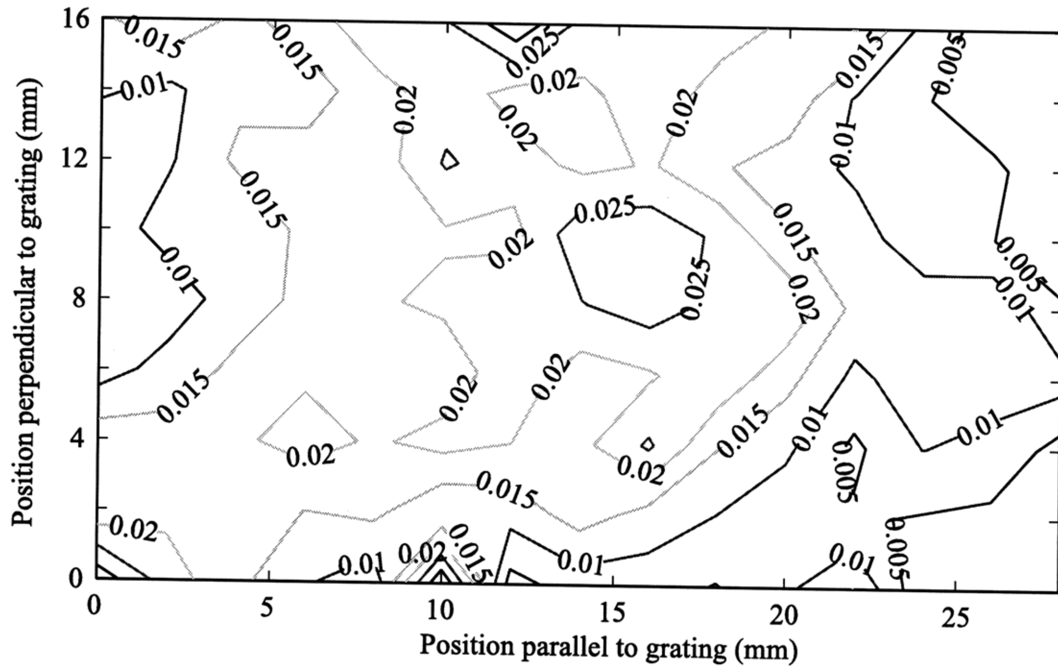


Figure 7-8: Power ratio map of the fundamental frequency and two-fold multiplied components with depth variations removed.

even within a sample, as a result of the RIE etch rate non-uniformity. Furthermore, the issue of depth variation can be easily removed by using an underlying etch stop layer, or with other processing steps.

One very important distinction is that the profile depicted in Figure 7-5(a) does not represent the true geometric profile of the grating. While the signal from the secondary electron scattering suggests that there is a depth variation, it is difficult to determine the exact value in length. There are also other contributions, such as scattering at the edges. In general it is difficult to image deep narrow trenches, in this case around 25 nm, with top-view SEM.

To test the magnitude of its contribution, depth variation can be removed from the signal profiles numerically. For example, we can take the profile shown in Figure 7-5(a) and cut off all values under 0.6 by setting them to 0.6. In practice the cutoff value is calculated individually for each profile as the signal average minus a constant. While this is not exactly scientific, it does give a rough estimate for the depth variation effects. We can then map the fundamental-to-frequency-doubled ratios, as shown in

Figure 7-8.

This figure shows that the ratio of the fundamental frequency components has been further suppressed to $\sim 2\%$. This is about a seven-fold improvement over the case that includes depth variation. These results demonstrate, in principle, the huge impact of depth variations. Note that while the magnitude of the ratio has decreased, the contour shape remains mostly unchanged.

A visual confirmation of the successful suppression of the fundamental spatial frequency can be seen in Figure 7-9. In this figure the 100 mm-diameter substrate analyzed previously is immersed in water, so the wavelength is reduced by the refractive index. The reference grating and spatial-frequency doubled areas are defined, and the two strips in between are shadowed areas that were not exposed. White light illuminates the sample at a shallow angle, and the immersion allows diffraction of blue wavelength from the 200 nm-period reference grating, as shown in the figure. The grating in the spatial-frequency-doubled region, however, has subwavelength period and no diffraction can be observed.

Recall from the overlay map shown early in Figure 7-3, some particle-induced local phase distortion occurs near the center of sample. We proceed to examine the effects of the phase distortions further in the device region. Two zoomed-in pictures of the same substrate is shown in Figure 7-10. In these pictures the color gamma of blue has been greatly increased while other colors were reduced. In this way, the diffraction of blue wavelength from the unsuppressed fundamental spatial frequency would be accentuated. In figure (a) the white light illuminates the device grating area where there is little phase distortion, so no diffraction can be observed. Note the two spots are due to the border of the shadowed region, where only one grating level was etched.

In Figure 7-3(b) the white light illuminates the area where the particle-induced phase distortion is in the device region, and a blue diffraction spot can be clearly observed. Therefore, even though the peak phase distortion caused by particle contamination is only around 8 nm, the unsuppressed fundamental spatial frequency still diffracts enough blue light to be seen. This result is rather fascinating, as it demon-

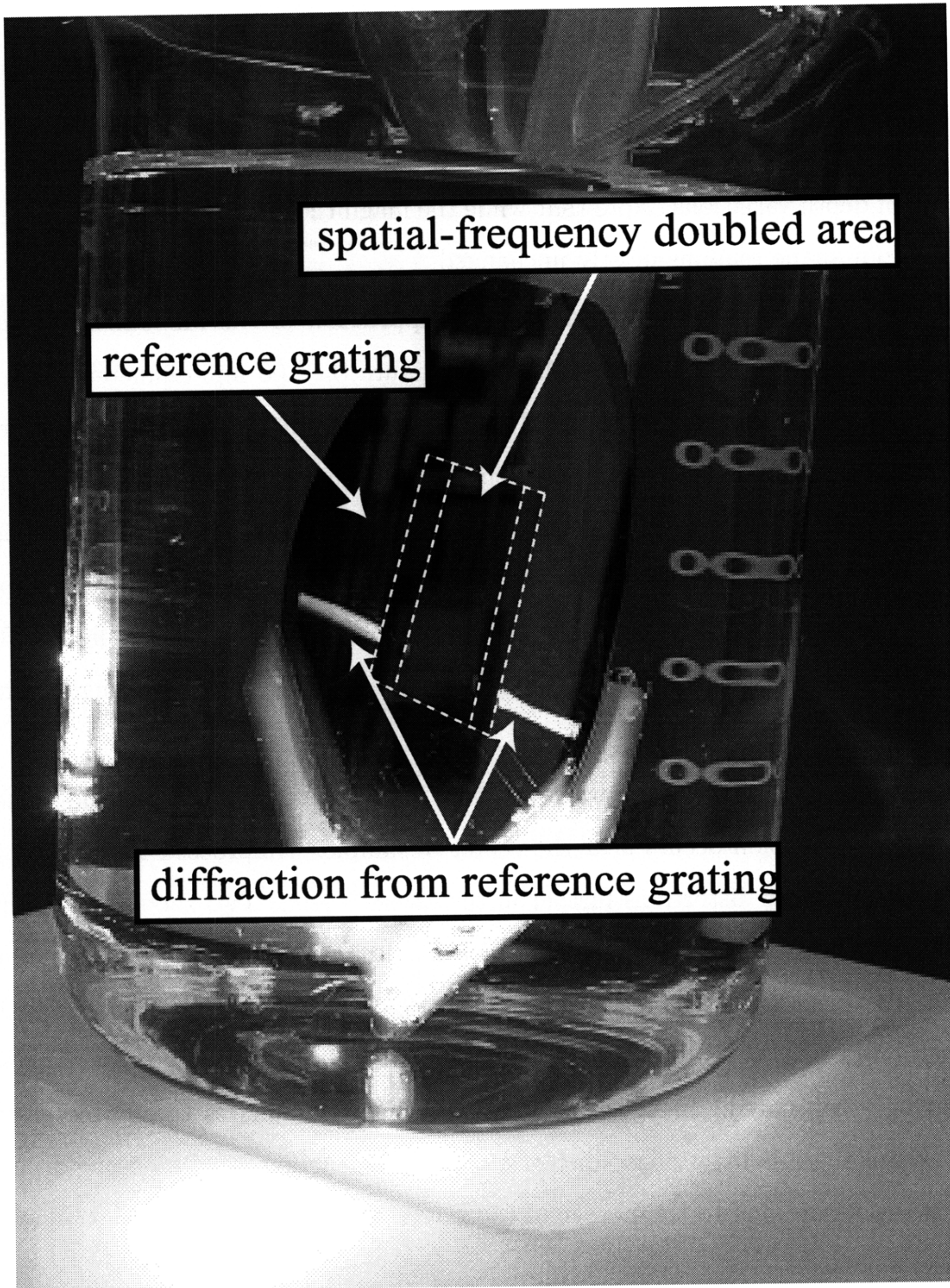


Figure 7-9: The 100 mm substrate is immersed in water and illuminated with white light to show diffraction from the reference period region. The device region has subwavelength grating and does not diffract.

(a)

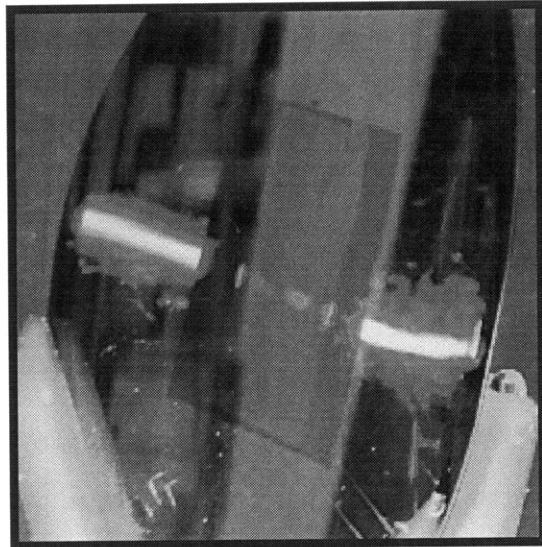

(b)

Figure 7-10: Zoomed-in pictures of the substrate with white light illuminating on device regions with (a) no phase distortion, and (b) local phase distortion induced by particle contamination. In the pictures the gamma of blue is increased to highlight the distortion.

strate the quality of the other region where diffraction can not be observed. While particle contamination proves to be a concern, the solutions are more of a practical nature, and not fundamental.

7.3 Higher-factor Spatial-frequency Multiplication

In the previous sections we have used the multilevel IL process to demonstrate two-fold spatial-frequency multiplication. Specifically we fabricated a 100 nm-period grating by overlaying two grating levels. Metrology and other testing results show that other than local distortion caused by particle contamination, most of the device grating area had high quality grating. Note that using the image-reversal process with PE trimming, as outlined in Chapter 6, we are able to fabricate grating levels with a high duty-cycle of ~ 0.9. Having such high duty-cycle allows us to achieve to high-factor spatial-frequency multiplication.

We fabricated a sample where three 200 nm-period grating levels were overlaid. For this sample the second and third grating levels are exposed at 100 nm and 50 nm offsets, respectively, to the first grating level. A top-view micrograph of this sample, etched in nitride, is shown in Figure 7-11. The linewidth of the lines are around 25 nm, as desired. While each set of 1, 4, 7, etc., 2, 5, 8, etc., and 3, 6, 9, etc. lines are patterned and etched with three different exposure steps, they are almost indistinguishable. From this figure the overlay accuracy and linewidth control are both in the order of the SEM resolution.

This particular "fence-type" geometry is quite interesting. It consists of a set of three 25 nm valleys spaced at 50 nm apart. The fourth set of lines, which would make this structure a 50 nm-period grating, is missing. From a fabrication point of view, there is no other way to fabricate this geometry unless a probe-base technique, which would suffer from having low spatial-phase coherence, is used. Therefore, there is no way this structure can be fabricated at such quality. This geometry then illustrates the versatility of the multilevel IL process, in which multiple levels of grating with any chosen linewidth can be placed at arbitrary phase-shifts.

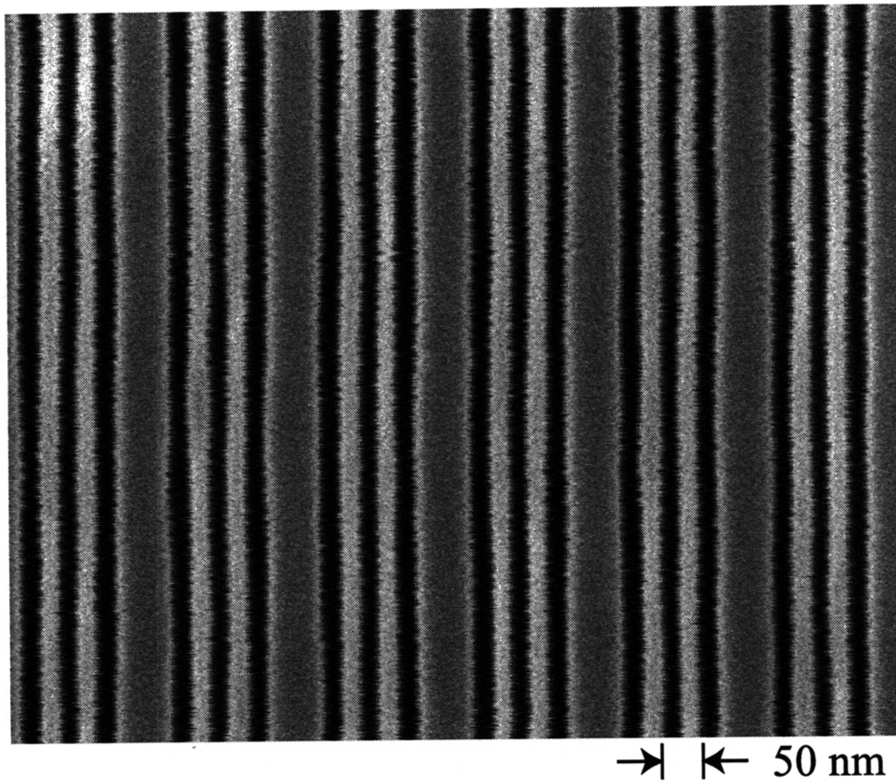


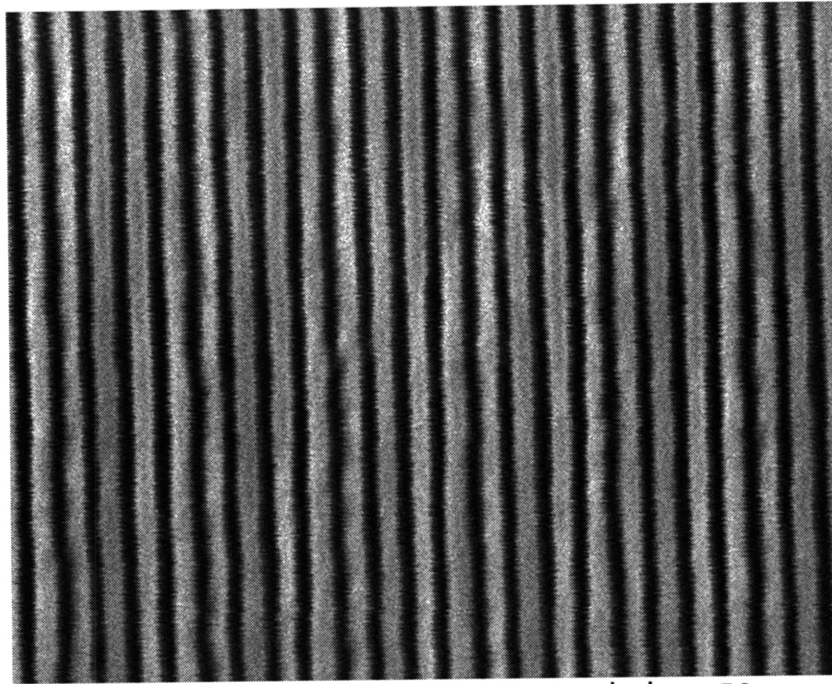
Figure 7-11: Top-view micrograph overlaying three 200 nm-period grating levels to achieve a "fence-type" grating structure.

The fourth grating level then can be exposed, resulting in a 50 nm period grating. The top-view micrograph of the sample is shown in Figure 7-12(a). The grating lines etched into nitride have a linewidth of around 25 nm. It is important to note that the 50 nm-period grating has been pattern-transfered into the nitride layer, which has a thickness of 45 nm. There appears to be some blurring in the image, which is due to both line-edge roughness (LER) and SEM noise (i.e., SEM resolution, shot noise, vibration, etc.). The images is scanned perpendicular to the grating direction, and line-to-line jitter can be observed. LER of the grating structure is an interesting parameter to analyze, but with such fine feature sizes that are not much greater than the SEM resolution, it is difficult to quantify. These results are reported in Chang *etal.* [11].

The 2D spectrum can be calculated from the image by performing a 2D FFT, as depicted in Figure 7-12(b). The spectrum is plotted as functions of spatial-frequencies relative to the fundamental frequency ($1/\Lambda$) in x and y. Other than the DC term, the most dominant spatial-frequency component term is at $4/\Lambda$. The terms at spatial-frequencies of $1/\Lambda$, $2/\Lambda$, and $3/\Lambda$ are mostly suppressed, and the other spectral components are higher-order harmonics. These results are extremely encouraging, considering the four grating levels were patterned over 3 weeks.

The area for the four-fold spatial-frequency multiplication, however, has been reduced to the order of mm^2. This is not due to any systematic overlay alignment or linewidth variation errors, but to particle contaminations. Despite numerous RCA baths to clean the substrate, there are particles in the lithography system that can not be completely eliminated. For every single grating level exposure, there is always a certain risk of particle getting caught under the substrate. When four grating levels are added together, the risk increases exponentially. The multilevel IL process is therefore not limited by the ability to align or linewidth-control the grating lines, but by our ability to keep the substrate free from particles. In principle, if particle contamination can be better controlled, this process can be scaled to fabricate much larger areas.

Another point of interest is the fabrication time for this process. As mentioned



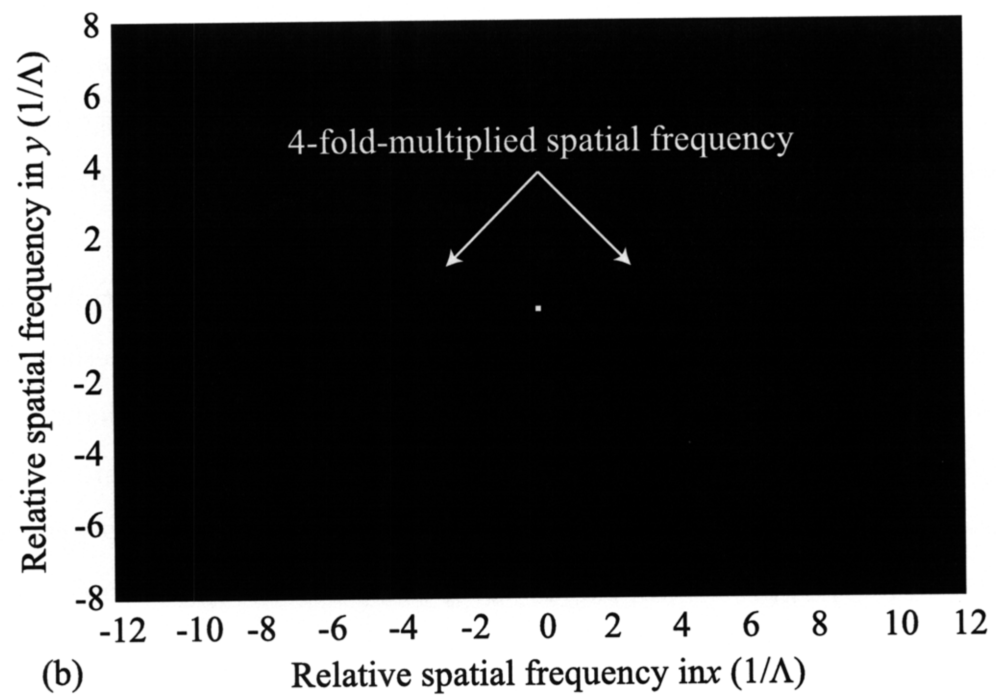


Figure 7-12: (a) Top-view micrograph overlaying four 200 nm-period grating levels to achieve a four-fold spatial-frequency multiplication, resulting in grating with 50 nm period. (b) The 2D spectrum of the 50 nm-period grating.

previously, for four grating levels the whole process takes around 3 weeks. Within a single level, the grating alignment process takes about 6 hours, while the exposure just around 10 minutes. To greatly reduce the alignment, an automated process with real-time metrology is desirable. Such a system would detect the phase of the reference grating and control the fringe of the exposure in a feedback loop. Another time consuming step is the plasma trimming step. A real-time system measuring the linewidth, such as optical diffraction setup, could also be employed. Upgrading these critical steps to automated processes would greatly reduce the processing time, and also increase the robustness of the process.

7.4 Future Work

We have demonstrated the ability of the multilevel IL process to fabricate subwavelength periodic nanostructures. Spatial-frequency multiplication, however, is only one application of the multilevel process. The technique we present is a general scheme to overlay multiple periodic nanostructures. Instead of etching every grating level to a single layer, as in our application, they can be built up to make a 2D structure.

Another interesting application of multilevel IL is to pattern multiple reference gratings oriented in different directions on the substrate. Using this method, we can phase-align multiple grating exposures to fabricate complicated 2D periodic patterns. One such example is the 2D hexagonal patterns. These geometries are interesting because they can be used as masks for directional dry etching to fabricate 3D photonic crystals [71] [16] [17] [18]. These patterns can also be used as a 2D template for self-assembly of block copolymers [5].

These patterns are typically fabricated with three or four-beam interference lithography [9]. However in such a scheme all the beams need to be phase-locked, and the equipment cost is high. Specially designed diffractive optics can also be used to fabricate 2D hexagonal structures [24] [4], but there is no freedom to change pattern parameters and the pattern area is limited. The pattern can also be fabricated by

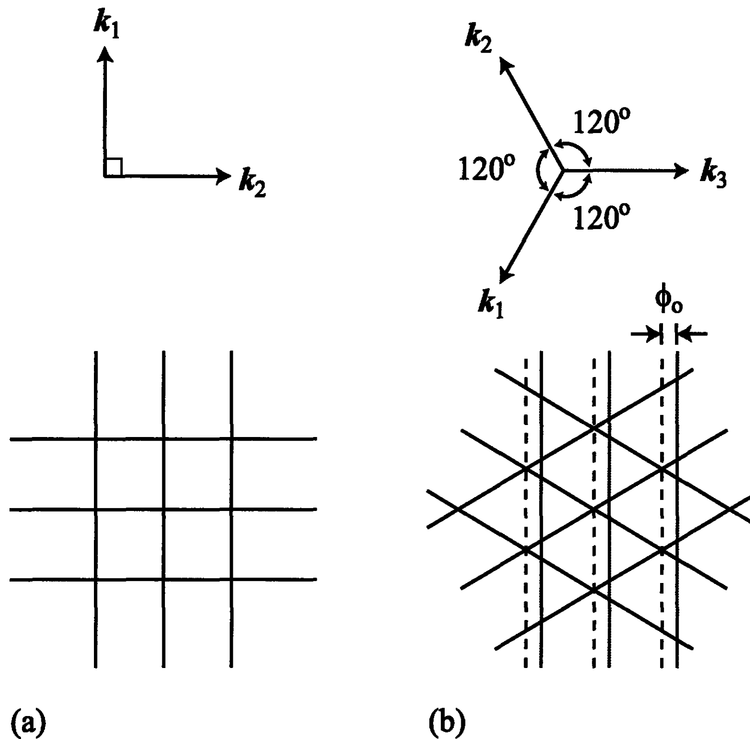


Figure 7-13: The multiple grating vectors and exposure patterns for (a) a rectangular pattern and (b) hexagonal pattern.

e-beam, which as previously described produces small areas and has low spatial-phase coherence.

Instead of using multiple-beam IL to fabricate 2D periodic pattern, our approach is to use multiple 2-beam IL exposures. In this scheme the resist is exposed by three different grating exposures before development, so that the intensity pattern is a sum of the exposures. The important issue is to phase-aligned the multiple exposures, which is an excellent application for multilevel IL process. The difficulties are highlighted in Figure 7-13. In the simplest case two grating patterns can be exposed at 90° angle offsets to pattern a rectangular grid pattern, as shown in diagram (a). The grating vectors k_1 and k_2 are illustrated, as well as the resulting pattern. Since the two grating vectors are perpendicular, no phase-alignment between the exposures is required.

The hexagonal pattern can be exposed by three grating exposures, each at 120°

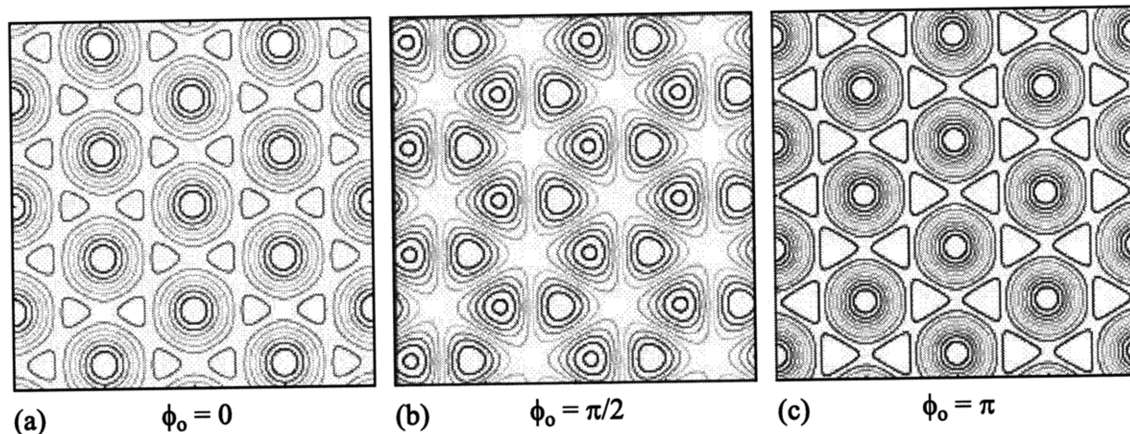


Figure 7-14: The 2D intensity profiles for three grating exposures, with the third grating having phase-offset of (a) 0, (b) $\pi/2$, and (c) π.

offset, as shown in Figure 7-13(b). Using the first two exposures, denoted by grating vectors k_1 and k_2, as the reference, the resulting pattern is illustrated. The third exposure with grating vector k_3 then needs to be aligned to the first two gratings, and the relative phase-offset can be defined as ϕ_o.

The relative phase-offset of the third grating exposure to the first two gratings is critical, as it determines the intensity profile of the 2D pattern. For third grating phase-offsets of 0, π/w, and π, the 2D intensity profiles can be simulated, as shown in Figure 7-14. These results show that the hexagonal pattern can change from intensity peaks, for the case of $\phi_o = 0$, to intensity valleys, for the case of $\phi_o = \pi$. Therefore, depending on the relative phase-offset of the third grating exposure, the exposed resist can take the form of hexagonal posts or holes.

In addition to the intensity profile, the phase-offset of the third grating can also have an effect on the contrast of the exposure. We can also simulate the image contrast of the hexagonal exposure as a function of the phase-offset, as shown in Figure 7-15. The results are quite counter-intuitive. It shows that to obtain the maximum image contrast, the third grating has to be exposed at a relative phase-offset of π instead of 0.

Multiple two-beam IL exposures is an exciting application of the multilevel IL process, as it allows us to break out of the 1D constraint and fabricate structures

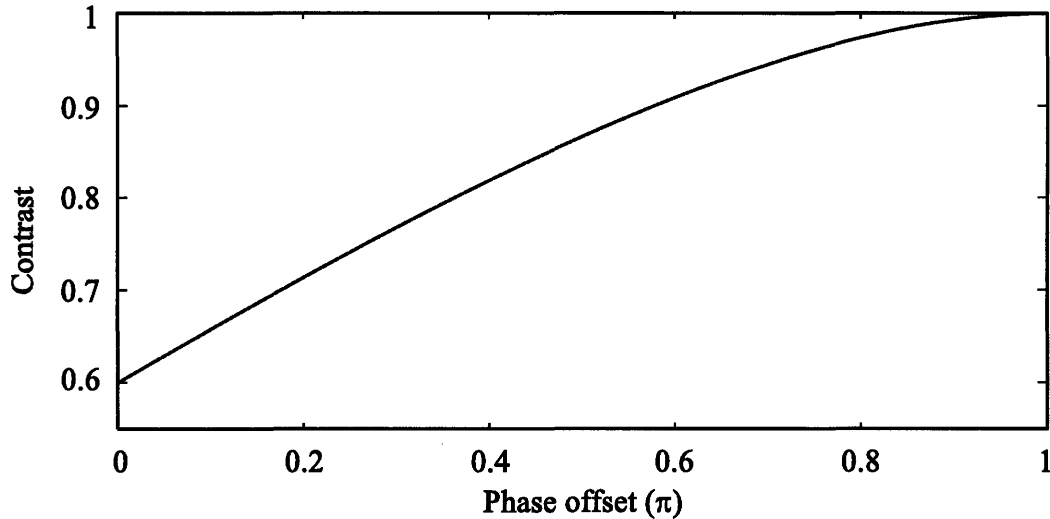


Figure 7-15: The simulated exposure contrast as a function of the phase-offset of the third grating exposure.

that are periodic in 2D. For the case of hexagonal pattern, three different reference gratings at 120° are required. This represents the most challenging aspect of this application. The overlay accuracy requirements, as it pertains to controlling the intensity profile and maximizing image contrast, are relatively more relaxed than the case of spatial-frequency multiplication.

7.5 Summary

In this chapter we presented the fabrication results for the multilevel interference lithography process. The solutions to the major challenges of phase overlay accuracy, grating linewidth control, and exposure over multilayer periodic structures, were tested experimentally.

The phase overlay accuracy was examined over large areas by using automated top-view SEM imaging and a Matlab phase detection algorithm. We were able to obtain a phase accuracy of 0.6 ± 1.9 nm (1σ) over a 16×12 mm^2 area for overlaying two grating levels. We also measured a 14% relative power of the fundamental to two-fold multiplied spatial-frequency. Much of the remaining fundamental spectral

power is due to etch variation, and removing this issue numerically the suppression of the relative fundamental frequency can be as low as 2%.

We then received confirmation of spatial-frequency multiplication by using immersion white light diffraction. Using this setup we can examine the diffraction from the reference grating but not the device grating at wafer level. Ultimately, through four-fold spatial-frequency multiplication, we succeeded in fabricating a 50 nm period grating using light with 351.1 nm wavelength.

The multilevel interference lithography is an alternative approach to increase the spatial resolution of optical lithography. Instead of reducing the wavelength and increasing the immersion index factor, we take an alternative route using high precision metrology and well-controlled fabrication processes. The fundamental concept behind spatial-frequency multiplication is that even though the resolution of a single grating level is wavelength-limited, but using multiple grating levels we can fabricate sub-wavelength periodic nanostructures.

The spatial-frequency multiplication process we described and tested experimentally is extremely useful, but it also represents one application for multilevel interference lithography. What we presented is an overall interferometric scheme in which multiple levels of periodic pattern can be controlled and placed at the desired position. Using multiple reference gratings, the process can be extended to fabricate complex 3D periodic nanostructures.

Chapter 8

Conclusion

The multilevel interference lithography (IL) is an alternative approach to increase the spatial resolution of optical lithography. There are intensive research programs in both academia and industry which are solely dedicated to reducing the wavelength and increasing the immersion index factor. These techniques are often expensive and impractical. We take an alternative route. By using high precision metrology and well-controlled fabrication processes, we are able to fabricate sub-wavelength periodic nanostructures.

The fundamental concept behind spatial-frequency multiplication is quite simple. Even though the period of a single grating level is wavelength-limited, the square profile inherently contains higher frequency harmonics. Therefore by using multiple grating levels, we can suppress the fundamental while enhancing the higher order spatial-frequencies. If instead the grating profile is sinusoidal to start with, it will have no higher order harmonics, and spatial-frequency multiplication would not work.

There are several important challenges that enables this multilevel IL process. The first is the overlay accuracy of grating levels. To resolve this issue we use heterodyne and homodyne interferometry to align every grating exposure to the reference grating. Linear phase errors are eliminated by matching the periods and angle orientation of every grating levels. Overlay accuracy of 0.6 ± 1.9 nm (1σ) was demonstrated between two grating levels over a 16 mm $\times$ 12 mm area. As a consequence, the fundamental frequency can be suppressed to 2% of the two-fold multiplied spatial-

frequency, ignoring the etch depth variation.

Note that in this scheme we align specifically to eliminate the linear and constant phase errors. Higher-order phase distortions due to issues such as particle contamination can not be compensated. At this stage our effective area is heavily influenced by particle-induced distortions, and not our alignment technique. While we have not observed these in our results, there might be also other time or environmental distortions, which might be a significant issue as we attempt to scale up the overlay area.

The ideal solution to non-linear phase distortions is real-time spatial-phase-locking. In this scheme, which can be named spatial-phase-locked interference lithography (SPLIL), the in-plane phase of the substrate is measured *insitu* and provides a feedback signal for the phase of the exposure. This scheme will eliminate alignment time, which is several hours, and compensate for any phase errors the substrate may be subjected to. Several different configurations have been discussed, such as a larger underlying grating to read with infrared, or using a patterned or "frozen" resist.

The second challenge is the linewidth control of grating levels, which has to be better than ~10 nm for the process to work. To control linewidth, we first analyze the exposure dose and image contrast of the Nanoruler. Simulation software was developed to calculate these two factors based on numerous optical and mechanical issues. The modeling results show that the contrast variation could be significant even due to alignment tolerance, and the linewidth can not be controlled to the tolerance we need. This work, however, is important in evaluating the exposure conditions of each grating level, and plays an invaluable part.

We then developed an image-reversal process using positive resist to fabricate, with nm-repeatability, the high duty-cycle grating pattern desired. In this scheme the linewidth is measured with top-view SEM and trimmed with plasma etching in an iterated cycle until the desired linewidth is reached. This algorithm functions like a feedback loop with discrete feedback signals. Using this method the linewidth of grating levels can be controlled to 1-2 nm repeatability.

The last major challenge is analyzing the optical effects of exposing resist over

multilayer periodic nanostructures. We specifically want to be able to design a resist stack that has minimized standing waves from the underlying structure. Rigorous coupled-wave analysis was implemented to analyze various stack scenarios to examine the exposure profiles in the resist layer. Combined with the binary resist model, we can predict the exact geometry of the residual resist. These tools are essential in the design of a resist stack for each grating level exposure.

After solving all the major challenges, we proceeded to overlay two 200 nm-period grating levels to fabricate a grating with 100 nm period. The grating area was 16 mm $\times$ 30 mm, which covers the whole device grating area. Ultimately, we were able to overlay four 200 nm-period grating levels to fabricate a grating with 50 nm period.

The spatial-frequency multiplication process we described and tested experimentally is extremely useful, but it also represents one application for multilevel interference lithography. What we presented is an overall interferometric scheme in which multiple levels of periodic pattern can be controlled and placed at the desired position. Using multiple reference gratings, the process can be extended to fabricate complex 3D periodic nanostructures.

Appendix A

Fabrication Process for Multilevel Interference Lithography

In this section the multilevel interference lithography fabrication process is comprehensively outlined with key parameters. Two processes are presented, the first to fabricate the reference grating, and the second to fabricate a single grating level. For two-fold spatial-frequency multiplication, the second process is used twice, while for four-fold multiplication factor it is used four times.

The fabrication processes are all done in the MIT Integrated Circuit Laboratory (ICL), NanoStructures Laboratory (NSL), and Space Nanotechnology Laboratory (SNL). Other abbreviations such as antireflection coating (ARC), reactive-ion etching (RIE), and plasma etching (PE) are used.

A.1 Fabrication Process for Reference Grating

1. In ICL, deposit silicon nitride layer.

   - Thickness = 45 nm.

   - Low-pressure Chemical Vapor Deposition (VTR tube).

2. In NSL, spincoat Brewer Science ARC i-con-16.

- Thickness = 120 nm (4.7 kRPM).
- Bake on hot plate at 185°C for 1 min.

3. In NSL, spincoat Sumitomo PFI-88A2 photoresist.

   - Thickness = 230 nm (3.00 kRPM).
   - Bake on hot plate at 95°C for 1 min.

4. In SNL, pattern reference grating with Nanoruler.

   - period = 200 nm.
   - Dose = 45 mJ/cm^2.
   - Develop TMAH (OPD 262) developer for 1 min.

5. In NSL, etch ARC layer.

   - Mask device grating region.
   - RIE with O_2, flow at 45 sccm, pressure = 7 mTorr, power = 100 W.
   - Etch time = 1 min.

6. In NSL, etch through silicon nitride layer and into silicon substrate.

   - Mask device grating region.
   - RIE with CF_4, flow at 15 sccm, pressure = 10 mTorr, power = 200 W.
   - Etch time = 3 min.

7. In NSL, RCA clean to remove residual polymer.

   - $H_2O : H_2O_2 : NH_4 = 5 : 1 : 1$.
   - Temperature = 80°C.
   - Clean time = 20 min.

A.2 Fabrication Process of Single Grating Level

1. In NSL, spincoat Brewer Science ARC i-con-16.
   - Thickness = 120 nm (4.7 kRPM).
   - Bake on hot plate at 185°C for 1 min
2. In NSL, spincoat Sumitomo PFI-88A2 photoresist.
   - Thickness = 230 nm (3.00 kRPM).
   - Bake on hot plate at 95°C for 1 min.
3. In NSL, clean polymer on top of reference grating.
   - Mask device grating region.
   - PE with O_2, pressure = 100 mTorr, power = 200 W.
   - Etch time = 5 min.
4. In SNL, align and pattern grating level with Nanoruler.
   - period = 200 nm.
   - Dose = 45 mJ/cm^2.
   - Develop TMAH (OPD 262) developer for 1 min.
5. In NSL, etch ARC layer.
   - Mask reference grating region.
   - O_2 flow at 45 sccm, pressure = 7 mTorr, power = 100 W.
   - Etch time = 1 min.
6. In NSL, measure linewidth of ARC/resist grating line.
   - Top-view SEM at 3 KeV
7. In NSL, trim linewidth of ARC/resist grating line.

- PE with O_2, pressure = 100 mTorr, power = 50 W.
- Etch time = depend on linewidth.
- Repeat steps 6-7 till target linewidth is reached.

8. In NSL, spincoat Molecular Imprint Silspin silicon-containing polymer.
   - Thickness = 200 nm (2.0 kRPM).
   - Bake on hot plate at 150°C for 1 min.
   - Starting image-reversal process.
9. In NSL, etch-back the silicon-containing polymer layer.
   - Mask reference grating region.
   - CF_4 flow at 15 sccm, pressure = 10 mTorr, power = 200 W.
   - Etch time = 3 min 15 sec.
10. In NSL, etch the thin ARC lines.
    - Mask reference grating region.
    - O_2 flow at 45 sccm, pressure = 7 mTorr, power = 100 W.
    - Etch time = 45 sec.
11. In NSL, etch the image-reversed grating into nitride.
    - Mask reference grating region.
    - CF_4 flow at 15 sccm, pressure = 10 mTorr, power = 200 W.
    - Etch time = 1 min.
12. In NSL, Piranha clean to remove silicon-containing polymer.
    - $H_2SO_4 : H_2O_2 = 3 : 1$.
    - Temperature = 80°C.
    - Clean time = 10 min.

13. In NSL, RCA clean to remove residual polymer.

- $H_2O : H_2O_2 : NH_4 = 5 : 1 : 1$.
- Temperature = 80^oC.
- Clean time = 20 min.

Appendix B

Rigorous Coupled-Wave Analysis

Rigorous coupled-wave analysis (RCWA) is an effective method for obtaining exact solutions to Maxwell's equations in periodic media [42] [51]. The theory of RCWA has been widely reported, and there are various methods of implementation to achieve numerical stability [52] [53] [47]. In the design of the multilevel interference lithography process, RCWA is an invaluable simulation technique. An overview of the RCWA multilayer transfer matrix method [53] [46], used for designing layer stacks for the multilevel interference lithography process, is presented in this section.

The geometry to consider is illustrated in Figure B-1. A beam with incident angle of θ_0 and wavelength λ_0 illuminates a multilevel structure with N layers. d_l and ε_l are the thickness and relative permittivity of the $l-th$ layer. For a layer with index variation of period Λ, depicted as layer 2, the permittivity can be expressed as $\varepsilon(x) = \varepsilon(x+\Lambda)$. Setting z_0 at $z=0$, the starting depth at each layer z_j can be defined as $\sum_{l=1}^{j} d_l$.

The electric fields in each layer is governed by the Helmholtz wave equation, given by

$$(\nabla^2 + \mu\varepsilon\omega^2)E = 0, \tag{B.1}$$

where μ is the magnetic permeability, ε is the electric permittivity, and ω is the radial frequency of light. The index of refraction n is then given by $\sqrt{\varepsilon\mu}$. For non-magnetic material, as in our application, $\mu = 1$. The relative permittivity, ε_r, can

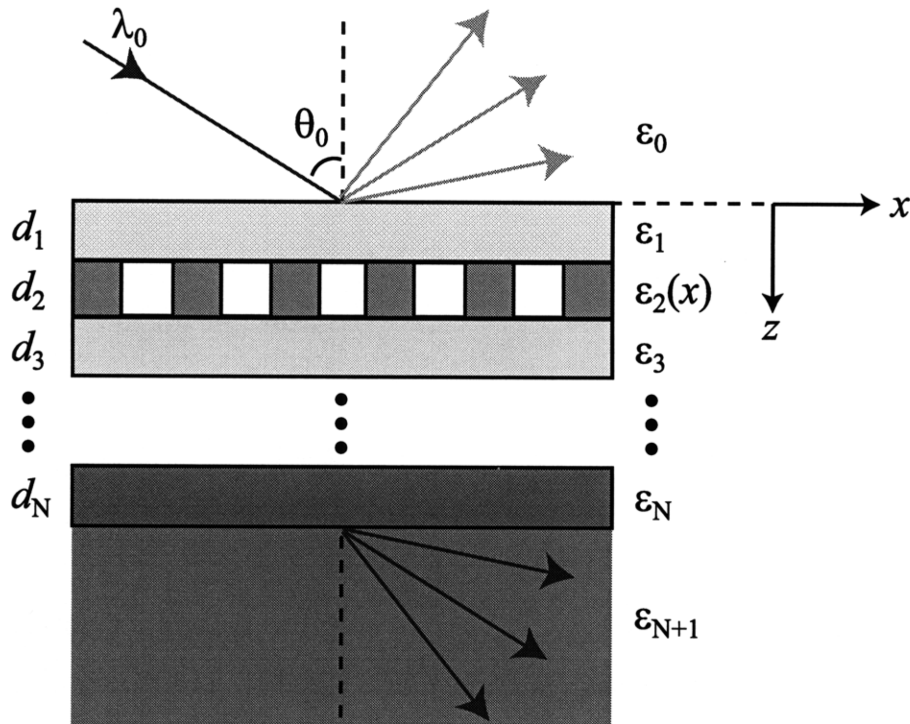


Figure B-1: Geometry of the multilevel homogenous and periodic stack.

then be defined as the permittivity relative to the permittivity of vacuum, or $\varepsilon/\varepsilon_v$. By solving the solutions to the Helmholtz equation the eigenmodes can be determined, which represent a complete and orthogonal set of solutions for each layer.

Our approach is then to determine the eigenmodes in each layer, and then solve for the phase and amplitudes of the modes by matching the boundary conditions. In this analysis the incident beam has transverse-electric (TE) polarization. As previously discussed in Chapter 2, TE polarization is desirable in interference lithography in order to achieve high exposure contrast.

B.1 Optical Fields in Homogenous Layers

The optical fields in the homogenous layers will be examined first. In a homogenous medium, the solutions to the wave equation are in the form of plane waves [72]. Since plane waves represent the eigenmodes, any optical field in a homogeneous medium

can then be described as a linear combination of plane waves.

For the $l-th$ homogeneous layer in the multilayer stack as depicted in Figure B-1, the electric field in the y-direction can be described as a sum of diffracted plane waves, given by

$$E_l(x,z) = \sum_m [A_{l,m} \exp(-jk_0 q_{l,m}(z-z_l)) + B_{l,m} \exp(jk_0 q_{l,m}(z-z_l))] \exp(-jk_{xm}x), \tag{B.2}$$

where $k_0 = 2\pi/\lambda_0$, and k_{xm} is the wave vector in the x-direction given by

$$k_{xm} = k_0(n_0 \sin\theta_0 - \frac{m\lambda_0}{\Lambda}), \tag{B.3}$$

where n_0 is the index of the zeroth layer. $q_{l,m}$ is the normalized wave vector in the z-direction given by

$$q_{l,m} = \begin{cases} \sqrt{n_l^2 - (\frac{k_{xm}}{k_0})^2}, & \text{for } k_0 n_l > k_{xm} \\ -j\sqrt{(\frac{k_{xm}}{k_0})^2 - n_l^2}, & \text{for } k_0 n_l < k_{xm} \end{cases} \tag{B.4}$$

and $A_{l,m}$ and $B_{l,m}$ represent the complex amplitudes of the $m-th$ mode in the forward and backward propagation direction, respectively. The total number of summed fields is defined as M. Note that for $l = 0$, $A_{0,m}$ is the amplitude of the single incident order, and for $l = N+1$, there are no backward propagating orders and $B_{N+1,m} = 0$.

In the case of $k_0 n_l > k_{xm}$, $q_{l,m}$ is complex and describes propagating plane waves. In the case of $k_0 n_l < k_{xm}$, $q_{l,m}$ is purely imaginary and describes evanescent waves. The number of propagating waves is determined by the grating period Λ in the grating layer. In the case where the layer stack consists of strictly homogenous levels, there is only two counter-propagating modes, i.e. $m = 0$ and $M = 1$.

B.2 Optical Fields in Periodic Layers

In the periodic layers, $l = L$, with the relative permittivity having the discrete translational property $\varepsilon_r(x) = \varepsilon_r(x+\Lambda)$. The profile is illustrated in Figure B-2. The relative permittivity can also be expanded in a sum of Fourier series,

$$\varepsilon_r(x) = \sum_h \varepsilon_h \exp(j\frac{2\pi h}{\Lambda}x), \tag{B.5}$$

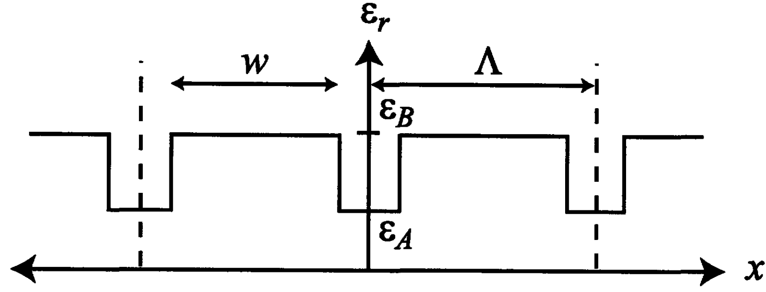


Figure B-2: Permittivity profile in the periodic layer.

where ε_h is the $h-th$ Fourier component, and the summation is over H, the number of Fourier components. For a square-wave permittivity profile with duty-cycle $f = w/\Lambda$, as depicted in Figure B-2, the coefficients are

$$\varepsilon_h = \begin{cases} n_B^2 f + n_A^2(1-f), & \text{for } h = 0 \\ (n_B^2 - n_A^2)\frac{\sin(\pi h f)}{\pi h}, & \text{for } h \neq 0 \end{cases} \tag{B.6}$$

where $n_A^2 = \varepsilon_A$ and $n_B^2 = \varepsilon_B$ for non-magnetic materials.

Such symmetry results in the solution to the wave equation being Bloch waves [40], a concept prevalent in describing electrons in periodic potentials in solid-state physics. Bloch wave solutions having discrete translational symmetry in the x-direction are expressed as

$$\psi(x) = \exp(-jkx)u(x), \tag{B.7}$$

where $\psi(x)$ is the wavefunction, $\exp(-jkx)$ is a plane wave envelope function, and $u(x)$ is a periodic function consisting of linear combinations of plane waves with the reciprocal lattice vector $G = i\frac{2\pi}{\Lambda}$, where i is an integer, as wave vectors

$$u(x) = \sum_G c_G \exp(-jGx). \tag{B.8}$$

$u(x)$ is also known as the periodic Bloch function. In this manner the electric field in the y-direction can be expressed as a sum of spatial-harmonic fields

$$E(x,z) = \sum_m S_m(z)\exp(-jk_{xm}x), \tag{B.9}$$

where $S_m(z)$ is the amplitude of the $m-th$ spatial-harmonic field. The summation is over M, defined as the number of spatial harmonics. To solve for the eigenmodes

in the periodic layer, I then insert Eq. B.9 and Eq. B.5 into Eq. B.1 to yield

$$\sum_m \Big[\frac{\partial^2 S_m}{\partial z^2} - k_{xm}^2 S_m\Big] \exp(-jk_{xm}x) + \omega^2 \varepsilon_v \sum_h \varepsilon_h \exp(j\frac{2\pi h}{\Lambda}x) \sum_m S_m \exp(-jk_{xm}x) = 0, \tag{B.10}$$

where ε_v is the absolute permittivity of vacuum. Since the number of Fourier components, H, is directly dependent on the number of spatial harmonics M, H can be set to equal to M. Simplifying the second term,

$$k_o^2 \sum_m \sum_h \varepsilon_h S_m \exp(-jk_o[n \sin\theta - \frac{\lambda_o}{\Lambda}(m+h)]x), \tag{B.11}$$

the coupling of the wave modes and the permittivity perturbation can be seen. Shifting the summation variables $m+h$ to m, and then setting $m-h=p$, this term can be written as

$$k_o^2 \sum_m \sum_p \varepsilon_{m-p} S_p \exp(-jk_{xm}x). \tag{B.12}$$

This term can be substituted back in Eq. B.10 to yield

$$\sum_m \Big[\frac{\partial^2 S_m}{\partial z^2} - k_{xm}^2 S_m + k_o^2 \sum_p \varepsilon_{m-p} S_p\Big] \exp(-jk_{xm}x) = 0. \tag{B.13}$$

For the $m-th$ spatial-harmonic field, the amplitude is then governed by,

$$\frac{\partial^2 S_m}{\partial z'^2} - \frac{k_{xm}^2}{k_o^2} S_m + \sum_p \varepsilon_{m-p} S_p = 0, \tag{B.14}$$

where $z' = k_o z$. This equation can also be expressed in matrix form,

$$\Big[\frac{\partial^2}{\partial z'^2}\Big]\big[\mathbf{S}\big] = \big[\mathbf{A}\big]\big[\mathbf{S}\big], \tag{B.15}$$

with

$$\mathbf{A} = \mathbf{E} - \mathbf{K}_x, \tag{B.16}$$

where $\mathbf{K}_x$ is a diagonal matrix with the (m,m) element equal to $|k_{xm}|^2/k_0^2$, and $\mathbf{E}$ is a matrix with the m,p element equal to the Fourier coefficient ε_{m-p}. Using the solution in the form of

$$S_m(z) = \sum_p w_{mp}[A_p \exp(-jk_0 q_m(z-z_L)) + B_p \exp(jk_0 q_m(z-z_L))], \tag{B.17}$$

Eq. B.15 can then be written as

$$\left[\mathbf{Q}\right]\left[\mathbf{W}\right]\left[\mathbf{B}\right] = \left[\mathbf{A}\right]\left[\mathbf{W}\right]\left[\mathbf{B}\right], \tag{B.18}$$

where w_{mp} is the (m, p) element of $\mathbf{W}$, and q_m is the positive square root of the $m-th$ value of diagonal matrix $\mathbf{Q}$, and $\mathbf{B}$ is an $(M \times 1)$ column matrix with the $m-th$ unit being

$$\mathbf{B}_m = \sum_p [A_p \exp(-jk_0 q_m(z - z_L)) + B_p \exp(jk_0 q_m(z - z_L))]. \tag{B.19}$$

Eq. B.18 can then be simplified

$$\left[\mathbf{A}\right]\left[\mathbf{W}\right] = \left[\mathbf{Q}\right]\left[\mathbf{W}\right]. \tag{B.20}$$

Eq. B.15 then sets up a nice eigenvalue problem, in which the eigenfunctions and eigenvalues of the matrix $\mathbf{A}$ describes the spatial-harmonics. Note $\mathbf{A}$ is an $(M \times M)$ matrix, where M is determined by the number of spatial harmonics. M is a user-specified parameter, and it is important for this value to be sufficiently large so that the simulation converges.

The corresponding eigenfunctions and eigenvalues of the characteristic matrix $\mathbf{A}$ are defined as $\mathbf{W}$ and $\mathbf{Q}$, respectively, and the spatial-harmonic field in layer $l = L$ is given by Eq. B.17. The field in the grating layer can then be represented as

$$E(x, z) = \sum_m \sum_p w_{mp}[A_p \exp(-jk_0 q_m(z - z_L)) + B_p \exp(jk_0 q_m(z - z_L))] \exp(-jk_{xm}x). \tag{B.21}$$

Using this formulism the eigenmodes in the grating layer can be determined.

I also examine the case when there are more than one fundamental period in the grating layer, as examined in the third and fourth grating exposure in Chapter 4. The permittivity profile for this case is illustrated in Figure B-3. In this figure the profile is represented by two periodic perturbations, the second with a $\Delta\phi$ phase-offset.

Since this profile still has a period of Λ, meaning the unit cell remains the same, only the efficiency of the diffracted orders changes, and not the location. As the result the profile can still be represented by

$$\varepsilon(x) = \sum_h \varepsilon_h \exp(j\frac{2\pi h}{\Lambda}x). \tag{B.22}$$

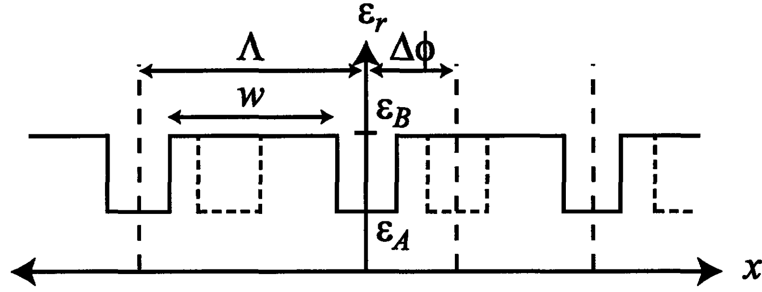


Figure B-3: Permittivity profile for multiple grating levels in the periodic layer.

Thus I can follow the same formulation, only changing the Fourier coefficients, which is now given by,

$$\varepsilon_h = \begin{cases} n_B^2(1-g+gf) + n_A^2 l(g-f), & \text{for } h=0 \\ (n_B^2 - n_A^2)\frac{\sin(\pi h f)}{\pi h}\sum_g e^{-j\frac{2\pi h}{\Lambda}\phi_g}, & \text{for } h \neq 0 \end{cases} \tag{B.23}$$

where g is the integer representing the number of grating profiles and ϕ_g is the phase-offset of the $g-th$ grating. For illustration purposes, $\phi_1 = 0$ and $\phi_2 = \Delta\phi$ in Figure B-3.

B.3 Multilevel Matrix Methods

In the previous two sections the modes in homogenous and periodic layers are examined. To solve for the optical field inside a multilevel stack, boundary conditions at the layer interfaces must be applied. For TE polarization the electric field in the y-direction must be matched at the boundaries.

The (2×2) transfer matrix method is an effective method to analyze multilevel homogenous stacks [73]. By using propagation matrices $\mathbf{P}$ to define wave propagation and dynamic matrices $\mathbf{D}$ to define refraction at layer interfaces, the boundary conditions can be applied by solving a series of matrix equations. An extended transfer matrix method for RCWA has been reported [46], and is extremely effective at analyzing multilevel periodic stacks.

In this formulation the transfer matrices have dimensions $(2M \times 2M)$, where M is the number of spatial-harmonics used to calculate the periodic layer. Note the number

of spatial-harmonics determine the accuracy of the simulation. However, increasing M also increases the size of the matrix, which increases computation cost. The trade-off is then between computation speed and simulation accuracy. It is then important to chose M as small as it allows the simulation results to converge. Typically in our simulation M is chosen to be around 40.

The propagation matrix for layer $i = 1 \rightarrow N$ is described as

$$\mathbf{P}_l = \begin{bmatrix} \mathbf{X}_l^{-1} & 0 \\ 0 & \mathbf{X}_l \end{bmatrix}, \tag{B.24}$$

where $\mathbf{X}_l$ and its inverse $\mathbf{X}_l^{-1}$ are $(M \times M)$ matrices with (m, m) element equal to $\exp(-jk_0 q_{lm} d_l)$ and $\exp(jk_0 q_{lm} d_l)$, respectively. The dynamic matrix for layer $i = 0 \rightarrow N+1$ is described as

$$\mathbf{D}_l = \begin{bmatrix} \mathbf{W}_l & \mathbf{W}_l \\ \mathbf{W}_l \mathbf{Q}_l & -\mathbf{W}_l \mathbf{Q}_l \end{bmatrix}, \tag{B.25}$$

where $\mathbf{W}_l$ is a $(M \times M)$ matrix with (m, n) element equal to w_{mn} for the grating layer and to the identity matrix $\mathbf{I}$ for the homogenous layers. $\mathbf{Q}_l$ is a $(M \times M)$ diagonal matrix with (m, m) element equal to q_m for the grating layer and to q_{lm} for the homogenous layers. By putting all the matrices together in a consecutive matrix form, the field in the $0-th$ layer can be expressed as

$$\begin{bmatrix} \mathbf{A}_0 \\ \mathbf{B}_0 \end{bmatrix} = \mathbf{D}_0^{-1} \Big(\prod_{l=1}^{N} \mathbf{D}_l \mathbf{P}_l \mathbf{D}_l^{-1} \Big) \mathbf{D}_{N+1} \begin{bmatrix} \mathbf{A}_{N+1} \\ \mathbf{0} \end{bmatrix}, \tag{B.26}$$

where $\mathbf{B}_0$ and $\mathbf{A}_{N+1}$ are $(1 \times M)$ vectors with $(m, 1)$ element equal to $B_{0,m}$ and $A_{N+1,m}$, respectively. $\mathbf{A}_0$ represents the incident beam and has only one order with unity normalized intensity at $m = 1$. Defining the recursive relations given in [46],

$$\begin{aligned} \begin{bmatrix} \mathbf{a}_N \\ \mathbf{b}_N \end{bmatrix} &= \mathbf{D}_N^{-1} \mathbf{D}_{N+1} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{a}_{l-1} \\ \mathbf{b}_{l-1} \end{bmatrix} &= \mathbf{D}_{l-1}^{-1} \mathbf{D}_l \begin{bmatrix} \mathbf{I} \\ \mathbf{X}_l \mathbf{b}_l \mathbf{a}_l^{-1} \mathbf{X}_l \end{bmatrix} \end{aligned}, \tag{B.27}$$

where $l = N$ to 1. The complex amplitudes of the forward and backward propagating orders the $S-th$ layer are then given by

$$\begin{aligned} \mathbf{A}_S &= \mathbf{X}_S \Big(\prod_{l=S-1}^{1} \mathbf{a}_l^{-1} \mathbf{X}_l \Big) \mathbf{a}_0^{-1} \mathbf{A}_0 \\ \mathbf{B}_S &= \mathbf{b}_S \mathbf{a}_S^{-1} \mathbf{A}_S \end{aligned}. \tag{B.28}$$

Using this expression I can determine important results such as diffraction efficiency, phase, and intensity distribution.

The multilayer RCWA matrix formulation described in this section was implemented using Matlab. Using this method various multilayer layer geometries that are applicable to the multilevel interference lithography process can be simulated.

Appendix C

Design of a Double-pass Acousto-optic Modulator

In this section we examine the design of a double-pass acousto-optic modulator (AOM) as an active optical component for a future generation Nanoruler. The main purpose of this device is to eliminate lateral and angular beam displacements due to changes in frequency modulation by compensating the deviation with a second pass. Using acousto-optic modulators in this configuration will allow rapid, and possibly real-time, changes in lithography and metrology modes.

There are many issues in designing such an AOM system, such as efficiency, bandwidth, and beam stability. In terms of lithography, one of the most critical issues is spectral leakage. In this section we will explore the major differences between longitudinal and shear mode acousto-optic interactions, analyze strain-induced birefringent effects, and examine polarization in the double-pass AOM. This work is described in Chang et al. [12].

C.1 Acousto-optic Effect

Acousto-optic interaction occurs when a light beam diffracts from a periodic perturbation in a material medium caused by an acoustic wave. Since the perturbation has a velocity vector, the diffracted beam is Doppler-shifted by the frequency of the

acoustic wave [72]. This phenomenon is useful for high speed frequency modulation of optical beams. There are various frequency shifting applications for acousto-optical modulators, such as displacement measuring interferometry [7]-[23], spectral tunable filters [32], and Q-switching and mode-locking in a laser cavity.

A common problem with AOMs is that the beam diffraction angle changes with the input acoustic frequency. Such an effect can be eliminated by using the AOM in a double-pass configuration, as Donley *et al.* [26] presented using a longitudinal AOM designed for the visible to near-infrared range. In such a configuration the change in diffraction angle is compensated by the second pass, and the output beam of the system has no beam displacement effects when changing the acoustic frequency. However, the double-pass system contains multiple optical interfaces, and unwanted back-reflection or scattering may result in light with frequency different from the desired at the output. Such spectral leakage is not easily detectable by eye during alignment, and can cause ill effects if ignored. For example, in interference lithography it creates non-stationary intensity patterns, resulting in a constant background exposure of photoresist and reduction of fringe contrast.

The acousto-optic effect in the Bragg's regime is illustrated in Figure C-1. In this setup, a sinusoidal RF signal with frequency f_{RF} actuates a transducer attached to the end of a crystal medium, generating a periodic strain field with velocity v and period $\Lambda = v/f_{RF}$. The strain field induces a periodic index variation, which diffracts the incident beam with frequency of f_i. In this geometry, the frequency of the diffracted beam is down-shifted to $f_i - f_{RF}$ to satisfy energy conservation. The Bragg angle inside the crystal is given by

$$\theta_B = \frac{f_{RF}\lambda}{2nv}, \tag{C.1}$$

where n is the optical index of refraction inside the medium. The separation angle, which is the angle between the zeroth and first-order diffraction exiting the crystal, is then approximately $2n\theta_B$.

There are two modes of acoustic waves that can be used in acousto-optic interaction, namely the longitudinal and the shear mode. Both modes offer their own

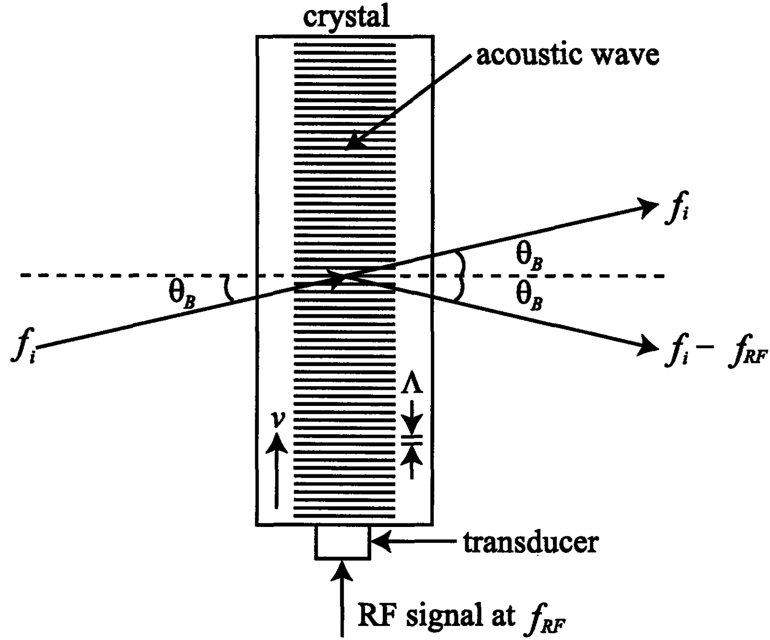


Figure C-1: Acousto-optic interaction in a crystal medium diffracts and frequency-shifts an incident beam.

advantages, and selection is based on the application. One advantage for shear mode is that in a solid medium the velocity of a shear wave is typically slower, resulting in a larger separation angle. One disadvantage is its complex shear-induced birefringence effects on the polarization of the diffracted beam. To design the optimal double-pass AOM, these issues will examined further.

The polarization effect of the acousto-optic interaction is critical to the design of a double-pass system. Depending on the type and direction of the acoustic wave, the polarization effects can be very different. To examine the polarization effect, we can calculate the linear photoelastic effect described by the tensor equation [72]

$$\Delta\eta_{ij} = p_{ijkl}S_{kl}, \tag{C.2}$$

where $\Delta\eta_{ij}$ and S_{kl} are the second-rank impermeability and strain tensors, respectively, and p_{ijkl} is the fourth-rank strain-optic tensor. For a UV laser ($\lambda = 351.1$ nm), we use fused silica, an isotropic material which has a strain-optic tensor given in a

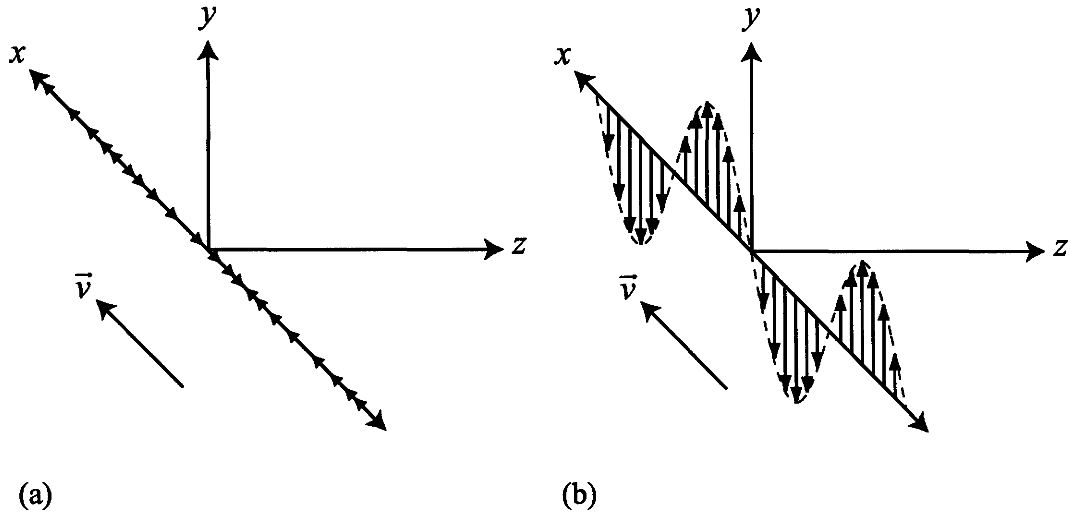


Figure C-2: Reference axis for (a) longitudinal wave, and (b) shear wave. The optical axis is in the $+z$-direction.

reduced coordinate system as

$$p_{ij} = \begin{pmatrix} p_{11} & p_{12} & p_{12} & 0 & 0 & 0 \\ p_{12} & p_{11} & p_{12} & 0 & 0 & 0 \\ p_{12} & p_{12} & p_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{44} \end{pmatrix}, \tag{C.3}$$

where p_{11} and p_{12} are the two independent material properties and $p_{44} = 0.5(p_{11} - p_{12})$.

The acoustic waves of interest are depicted in Figure C-2. Both waves travel in the $+x$-direction, and for the longitudinal wave the strain field is in x, shown in Figure C-2(a). The longitudinal acoustic wave can be described as

$$S_1 = S \sin(\Omega t - Kx), \tag{C.4}$$

where S is the strain amplitude, Ω is the frequency, and K is the wave vector. Using the strain-optic coefficients for fused silica, the optical impermeability tensor for a

longitudinal wave is given by

$$\Delta\eta_{ij} = \begin{pmatrix} p_{11}S_1 & 0 & 0 \\ 0 & p_{12}S_1 & 0 \\ 0 & 0 & p_{12}S_1 \end{pmatrix}. \tag{C.5}$$

The new index ellipsoid can then be defined as

$$x^2[\frac{1}{n^2} + p_{11}S_1] + y^2[\frac{1}{n^2} + p_{12}S_1] + z^2[\frac{1}{n^2} + p_{12}S_1] = 1, \tag{C.6}$$

and the refractive indices for the principal axes x and y can be calculated to be

$$\begin{aligned} n_x &= n - \tfrac{1}{2}n^3 p_{11} S \sin(\Omega t - Kx) \\ n_y &= n - \tfrac{1}{2}n^3 p_{12} S \sin(\Omega t - Kx) \end{aligned}. \tag{C.7}$$

From these calculations we can see that for a longitudinal acoustic wave, the index perturbations in the x and y directions are different by a constant. Therefore for a constant strain amplitude S, incident beams with polarization in the x and y-directions would have different diffraction efficiencies. However, the longitudinal acousto-optic interaction has no polarization effects on the incident beam.

The same can not be said for a shear mode acousto-optic interaction [3]-[25]. Figure C-2(b) depicts a shear acoustic wave traveling in $+x$ with strain fields in the y-direction. The shear wave can be described as

$$S_6 = S\sin(\Omega t - Kx), \tag{C.8}$$

where S_6 denotes shear in the xy directions S_{xy} in the reduced indices notation. The optical impermeability tensor for a shear wave is then given by

$$\Delta\eta_{ij} = \begin{pmatrix} 0 & \frac{1}{2}S_6(p_{11}-p_{12}) & 0 \\ \frac{1}{2}S_6(p_{11}-p_{12}) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{C.9}$$

Since the non-zero terms in Eq. C.9 represent shear terms, there will be an x-y cross term in the index ellipsoid equation. To eliminate the cross term, the x and y principal axes are rotated 45° about the z-axis. The equation for the rotated ellipsoid thus becomes

$$x'^2[\frac{1}{n^2} + \frac{1}{2}S_6(p_{11}-p_{12})] + y'^2[\frac{1}{n^2} - \frac{1}{2}S_6(p_{11}-p_{12})] + \frac{z^2}{n^2} = 1, \tag{C.10}$$

and the refractive indices for the rotated principal axes x' and y' are

$$n_{x'} = n - \frac{1}{4} n^3 S(p_{11} - p_{12}) \sin(\Omega t - Kx) n_{y'} = n + \frac{1}{4} n^3 S(p_{11} - p_{12}) \sin(\Omega t - Kx). \quad \text{(C.11)}$$

The sign difference between the x' and y'-axes in Eq. C.11 ensures that the sinusoidal perturbations in these two directions are always out-of-phase by π. Thus for light propagating along the z-direction, the shear mode AOM behaves like a half-wave plate with its optical axis oriented 45° to the x-direction for the Bragg diffracted beam. This polarization effect is not present in the longitudinal mode.

While the shear mode acousto-optic interaction is significantly more complicated than the longitudinal mode due to its polarization effects, it has a significant advantage because shear waves travels slower in a medium. Eq. C.1 shows that the Bragg angle for a shear mode AOM is larger, ensuring larger separation of the diffracted orders. This is especially critical for beams with shorter wavelength. For our lithography system where a UV source of 351.1 nm wavelength is used, fused silica is one of the few materials that is transparent at this wavelength. Acoustic shear and longitudinal waves travel at 3761 and 5960 m/s in fused silica, respectively [72]. For 100 MHz acoustic input, a shear mode AOM has separation angle of 9.34 mrad, significantly greater than 6.18 mrad for operation in the longitudinal mode.

C.2 Double-pass Configuration

For UV radiation, the shear mode acousto-optical modulator is more desirable due to a larger diffraction angle. This condition reduces the overall physical size of the system, and is critical for a complex optical system. A double-pass AOM in shear mode is then desirable for our application.

The proposed double-pass AOM is depicted in Fig. C-3. Setting the x-z plane as the plane of incidence, a TM-polarized beam is transmitted by a polarizing beamsplitter cube and focused onto the AOM crystal (Neos Technologies, model 46100). The diffracted first order, Doppler-shifted by f_{RF}, is now TE-polarized due to the shear mode acousto-optic effect. It then goes through a quarter-wave plate twice while

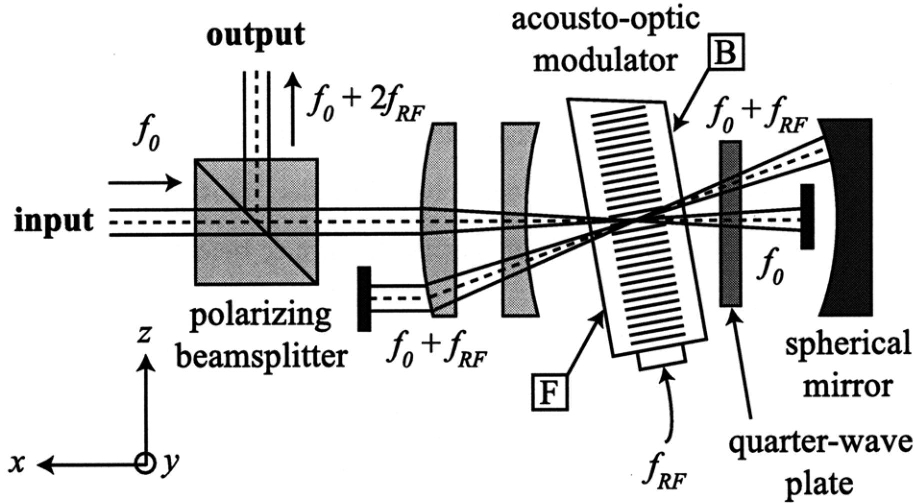


Figure C-3: Optical setup of the double-pass AOM in shear mode.

being focused back onto the AOM crystal by a spherical mirror, effectively changing the beam back to TM-polarization. The AOM diffracts the beam for the second pass, while rotating its polarization to TE and Doppler-shifting it by f_{RF}. The beam, now having a frequency of f_0+2f_{RF}, propagates back towards the polarizing beamsplitter and is reflected as the output beam. The unwanted orders are blocked by beam blocks in the setup.

The focusing lenses, separated by 61.7 mm, are plano-convex and plano-concave with foci of 81.1 mm and -21.6 mm, respectively. They are used to focus the 1 mm-diameter ($1/e^2$) beam onto the AOM 37.3 mm away. The AOM crystal material is fused silica, and the operating acoustic frequency is centered around 100 MHz, resulting in a separation angle of 9.34 mrad. A focused Gaussian beam is used in the AOM because the shear transducer creates a relatively small effective acoustic aperture of 200 μm. The strain field is also not uniform, resulting in diffraction efficiency that is position sensitive. This can be problematic if the beam angle at the input is not stable, as the lens will translate it as a lateral shift on the AOM crystal and cause the efficiency to fluctuate with time.

Focusing inside the crystal is also more desirable than other schemes, such as a cat's eye retroreflector [26], since fused silica has a high optical power threshold. Fo-

cusing UV radiation on mirrors might build up organic contamination on the reflective coating. Instead, a spherical mirror with radius 150 mm is used to focus the diverging beam back onto the AOM. The spherical mirror is placed so that the wavefront curvature matches the mirror curvature, effectively retroreflecting the beam. This position can be determined readily by propagating the complex beam parameter using the ABCD ray transfer matrix [72].

The initial beam parameter $q_o = jz_o$ is defined at the beam waist inside the AOM crystal, where $z_o = \frac{\pi\omega_o^2}{\lambda}$ and ω_o is the radius of the beam waist. The beam then propagates through thickness t in fused silica, a fused silica/air interface, and distance z in air. The transfered beam parameter as a function of z is then

$$q(z) = \frac{q_o + t}{n} + z, \tag{C.12}$$

where n is the refractive index of fused silica. The wavefront radius can then be determined from

$$R(z) = (\frac{t}{n} + z)(1 + \frac{z_0^2}{(t + nz)^2}). \tag{C.13}$$

Using system parameters $\omega_o = 65$ μm and $t = 5$ mm, the wavefront curvature matches the spherical mirror at $z = 142.1$ mm. It is important to note that this condition is satisfied exactly only at a particular acoustic frequency. When changing operating frequency the shifts in diffraction angle changes the optical path. However, within the FWHM bandwidth of our system the angular shifts are less than ~ 1 μrad, and the effects are negligible.

The AOM is driven by 6 W of acoustic power, with the double-pass efficiency as a function of input frequency shown in Fig. C-4. The figure shows a double-pass peak efficiency of 42.6% at 105 MHz and a full width at half maximum (FWHM) bandwidth of ~28 MHz. The peak double-pass efficiency is comparable to the square of the measured maximum single-pass efficiency of ~ 67%. The output beam was measured by a CMOS camera (Cohu, model 7200) to have less than 20 μm lateral and 10 μrad angular beam displacements within the FWHM bandwidth.

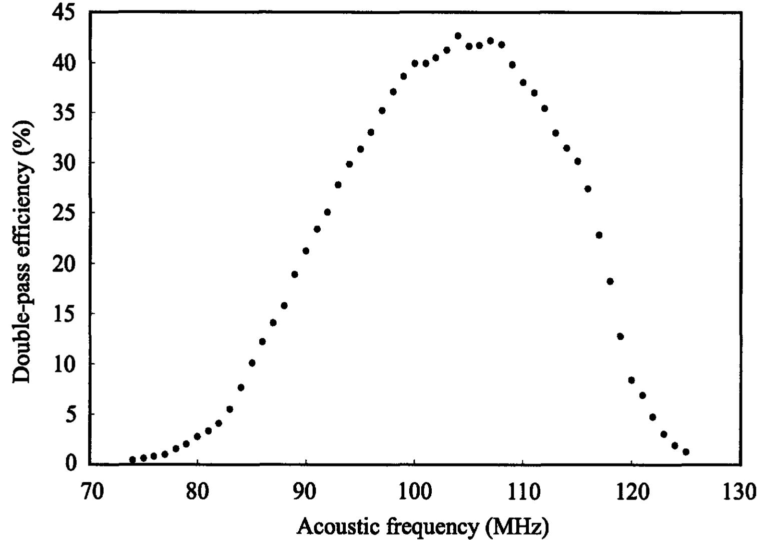


Figure C-4: Measured system efficiency of the double-pass AOM in shear mode as a function of input acoustic frequency.

C.3 Spectral Leakage

The double-pass AOM configuration eliminates the issue of beam displacement with changing acoustic frequency. However, the multiple optical components make the system prone to spectral leakage, even if perfectly aligned. The leakage originates from back-reflections of the optical interfaces which have a reflectivity of $\sim 1\%$ with anti-reflection coatings. The polarizing beamsplitter, as shown in Fig. C-3, acts like a polarizer and only reflects $\sim 1\%$ of TM-polarized output light. It is then desirable that the unwanted reflections are TM-polarized to minimize spectral leakage in the output.

To evaluate the effects of leakage, the spectral components of the beam at the output of the double-pass AOM setup are examined. The reference frame in Fig. C-3 is being used in this analysis. Assuming only frequency shifts caused by the acousto-optic effect, and that the beam is a TE-polarized plane wave propagating in z, the electric field can be written as the sum of spectral components

$$E_y(z,t) = \sum_m A_m\, exp[-j2\pi(f_0 + mf_{RF})t + jkz] = \sum_m g_m(f_0 + mf_{RF}), \qquad \text{(C.14)}$$

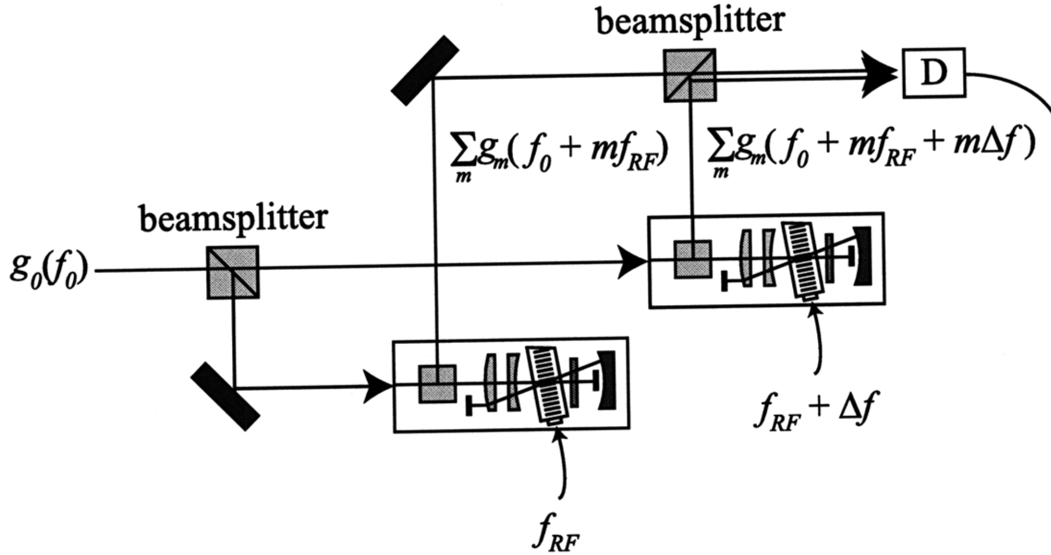


Figure C-5: Experimental setup to measure the magnitude of spectral leakage components.

where k is the wave vector, A_m are the amplitudes for each spectral component, and m is an integer number representing the number of times the beam undergoes acousto-optic interaction. The intensity of the m-th spectral component is thus $|A_m|^2$, and the double-pass signal component is described by m=2. For convenience, Eq. C.14 can be expressed as a sum of frequency components $g_m(f_o + mf_{RF})$. The output beams, having different frequencies, will beat and form a heterodyne intensity signal that contains the frequency differences of the components. The heterodyne signal can then be measured by a photodetector to examine the level of leakage. However, using this method the heterodyne frequencies will always be multiples of f_{RF}, and it will be difficult to determine the absolute frequency of each leakage component. Another consideration is that the operating acoustic frequency f_{RF} is around 100 MHz, and expensive electronics with high speed A/D will be required.

In order to analyze the spectral leakage more accurately, a setup with two double-pass AOMs operating at offset acoustic frequencies was used, as illustrated in Fig. C-5. A beam with optical frequency f_0 is split into two by a beamsplitter cube, each beam entering a double-pass AOM setup. The input acoustic frequencies for the two

AOMs are f_{RF} and $f_{RF} + \Delta f$, where Δf is a constant frequency offset chosen in the kHz range. The outputs of the double-pass AOMs are combined by a second beamsplitter and then measured by a photodetector. Due to the frequency offset Δf, the heterodyne signal contains high frequencies in the hundreds of MHz range and low frequencies in the kHz range. The higher frequencies are blocked by a low-pass-filter, while the lower frequencies can be analyzed by a spectrum analyzer (Hewlett-Packard model 35670A). The filtered heterodyne intensity signal in this setup is then given by

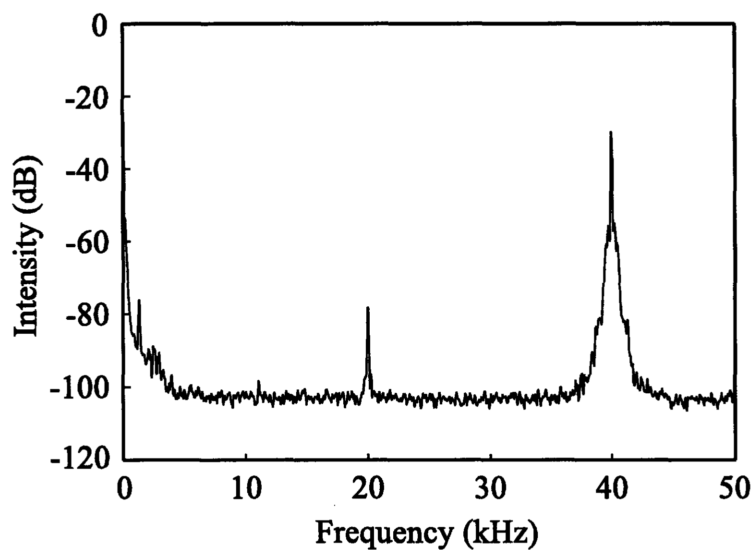
$$I(t) = I_0 + \sum_m |A_m|^2 \cos(2\pi m \Delta f t), \tag{C.15}$$

where I_0 is the background DC intensity. Eq. C.15 shows that by using this technique, the m-th spectral component is represented as having a frequency of $m\Delta f$, and can be measured independently. An important assumption in this formulation is that the spectral leakage characteristics in the two double-pass AOM setups are the same. While this condition can not be exact, the assumption is reasonable since the two setups differ only by alignment tolerances and have comparable performance to a few percent. Using this experimental setup, the relative intensity of each spectral component in the double-pass AOM setup can be measured.

For an acoustic frequency input f_{RF} of 105 MHz and Δf of 20 kHz, the measured spectrum of the heterodyne intensity signal is plotted in Fig. C-6(a). The figure illustrates the desired double-pass signal at 40 kHz and a single-pass leakage component at 20 kHz with $\sim 0.02\%$ relative intensity. Higher orders have insignificant intensity so are ignored in further analysis. The broad bandwidth around DC and signal frequency are due to the lack of vibration isolation in the experimental setup. Using Eq. C.14, the intensity $I = |E_y(z,t)|^2$ is

$$I(t) = |A_1|^2 + |A_2|^2 + 2A_1A_2\cos(2\pi f_{RF} t), \tag{C.16}$$

where A_1 and A_2 are the amplitudes of the single-pass leakage and double-pass signal, respectively. Using $|A_1|^2 \sim 0.002|A_2|^2$ as measured, the leakage will interfere with the signal beam and result in $2A_1A_2/(A_1^2 + A_1^2) \sim 9\%$ relative intensity fluctuations at a frequency of f_{RF}. In order to enable better process control in lithography, this leakage-induced fluctuation needs to be reduced to a few percent.



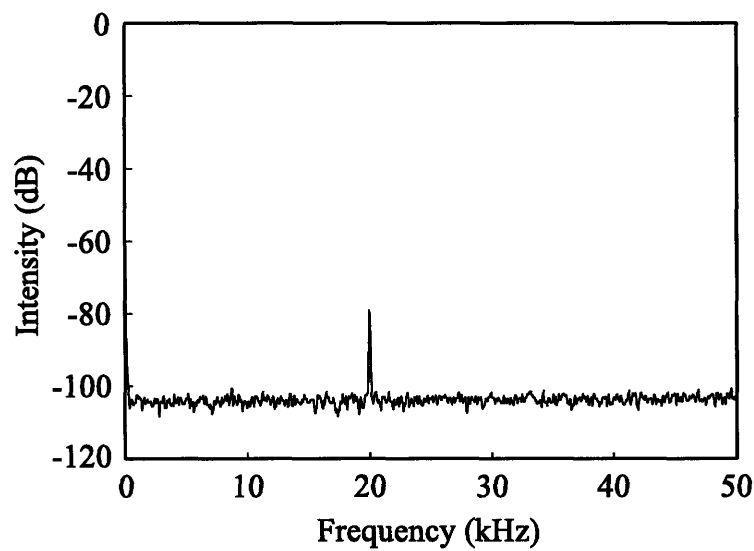


Figure C-6: Semi-logarithmic plot of measured intensity spectra with nominal acoustic frequency input of 105 MHz and Δf of 20 kHz for (a) a double-pass AOM, and (b) with the beam path after the AOMs blocked.

To trace the sources of the single-pass leakage, another set of data, shown in Fig. C-6(b), were taken with the beam paths after the AOM blocked in both setups. The leakage remained, suggesting that the majority of leakage is not coming from the optics after the AOM. The leakage is then determined to be predominately originated from the back AOM/air interface, shown in Fig. C-3 as surface "B". Due to the shear acousto-optic effect, the diffracted beam after the first-pass through the AOM is TE-polarized. It then back-reflects from the back AOM/air interface and results as leakage at the output. While this beam path is a significant source of leakage, it is not the only source; other weaker sources may exist in the system.

Since this leakage beam does not go through a second pass to compensate for the diffraction angle, it moves at the output when the acoustic frequency is changed. This can be observed by the CMOS camera with a long integration time. Due to this characteristic, for a certain frequency range the single-pass leakage beam is displaced from the signal beam at the output. In this case the leakage can be easily eliminated by using spatial filters. However, since the leakage sweeps with frequency, it may not be spatially-filtered at all operating frequencies. It is important that the overall leakage intensity be reduced.

To reduce the single-pass leakage, a half-wave plate with its optical axis oriented at 22.5^o to the z-axis in Fig. C-3 was placed after the polarizing beamsplitter. The quarter-wave plate after the AOM is also rotated by 45°. The purpose of the half-wave plate is to rotate the incoming TM-polarized beam by 45°, aligning it to the principal optical axis of the shear mode AOM. The beam then does not undergo any polarization effects in the AOM, and the polarization of the back AOM/air reflection will be rotated back to TM. As a result, the single-pass leakage is transmitted through the polarizing beamsplitter.

To verify the effects of the half-wave plate over the FWHM bandwidth of the double-pass AOM, the heterodyne experiment was repeated with nominal acoustic frequency inputs of 85-115 MHz and a constant Δf of 20 kHz. The results are plotted in Fig. C-7. In this figure the magnitudes of the 40 kHz double-pass signal and the 20 kHz single-pass leakage components are plotted as functions of the nominal frequency.

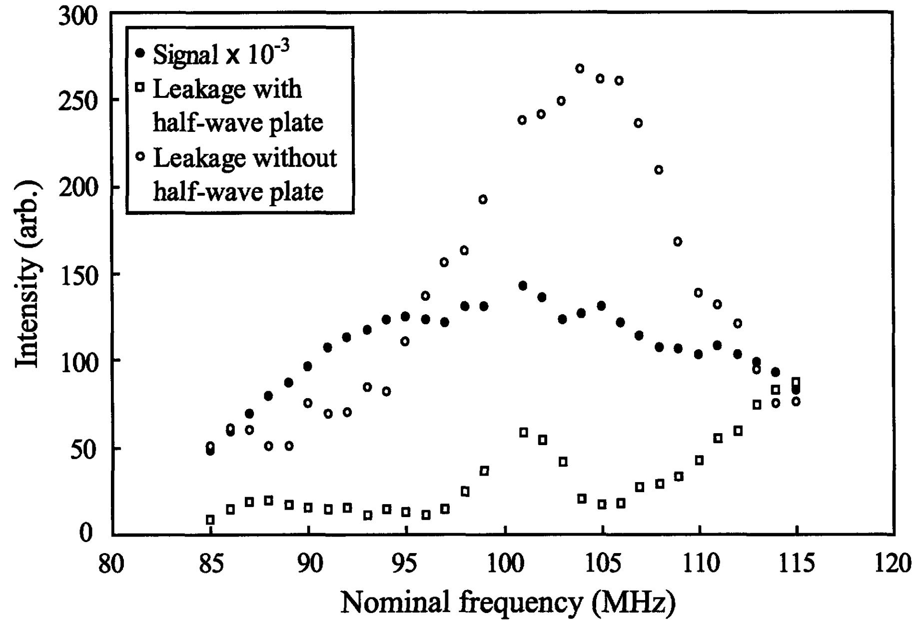


Figure C-7: Relative intensity of the double-pass signal and single-pass leakage components over the FWHM band of the double-pass AOM. The signal intensity is scaled by 10^{-3}.

The signal magnitude is scaled by a factor of 10^{-3}. The spectral leakage is reduced to $\sim 0.01 - 0.04\%$ relative intensity over the FWHM bandwidth. The reduction of spectral leakage is apparent, improving by a factor of up to $\sim$20 at a nominal frequency of 105 MHz. Note while leakage from a single dominant source is reduced by the addition of the half-wave plate, leakage from other weaker undetermined sources remain. This is evident at the edge of the FWHM band around 115 MHz, where the half-wave plate had negligible effect on presumably another leakage source. The residual leakage level is calculated to induce $2-4\%$ intensity fluctuations on the signal beam.

In this section we present a double-pass AOM in shear mode designed for high power operations in the UV. Such an optical system eliminates variations in beam position and angle during frequency tuning. Spectral leakage plays a critical role in shear mode acousto-optical interaction, and is analyzed experimentally using an experimental setup with two double-pass AOMs perturbed at offset acoustic frequen-

cies. The leakage was found to be predominantly from a single source, and can be reduced by rotating the polarization of the incident beam. Tested with $\lambda = 351.1$ nm, the shear mode double-pass AOM demonstrated a peak efficiency of 42.6% with relative leakage intensity of $\sim 0.01 - 0.04\%$ over the FWHM bandwidth of 28 MHz.

Bibliography

[1] M. Ahn, R. K. Heilmann, and M. L. Schattenburg. Fabrication of ultrahigh aspect ratio freestanding gratings on silicon-on-insulator wafers. *J. Vac. Sci. Technol. B*, 25:2593–2597, 2007.

[2] M. Ahn, R. K. Heilmann, and M. L. Schattenburg. Fabrication of 200 nm-period blazed transmission gratings on silicon-on-insulator wafers. *Summited to J. Vac. Sci. Technol. B*, 2008.

[3] V. I. Balakshy and J. A. Hassan. "polarization effects in acousto-optic interaction". *Opt. Eng.*, 32:746–751, 1993.

[4] V. Berger, O. Gauthier-Lafaye, and E. Costard. Photonic band gaps and holography. *J. Appl. Phys.*, 79:60–64, 1997.

[5] I. Bita, J. K. W. Yang, Y. S. Jung, C. A. Ross, E. L. Thomas, and K. K. Berggren. Graphoepitaxy of self-assembled block copolymers on two dimensional periodic patterned templates. *accepted to Science*, 2008.

[6] T. M. Bloomstein, M. F. Marchant, S. Deneault, D. E. Hardy, and M. Rothschild. 22-nm immersion interference lithography. *Opt. Exp.*, 14:6434–6443, 2006.

[7] N. Bobroff. "recent advances in displacement measuring interferometry". *Meas. Sci. Technol.*, 4:907–926, 1993.

[8] S. R. J. Brueck. Optical and interferometric lithography–nanotechnology enablers. *Proc. IEEE*, 93:1704–1721, 2005.

[9] L. Z. Cai, X. L. Yang, and Y. R. Wang. All fourteen bravais lattices can be formed by interference of four noncoplanar beams. *Opt. Lett.*, 27:900–902, 2002.

[10] C.-H. Chang, J. C. Montoya, M. Akilian, A. Lapsa, R. K. Heilmann, M. L. Schattenburg, M. Li, K. A Flanagan, A. P. Rasmussen, J. F. Seely, J. M. Laming, B. Kjornrattanawanich, and L. I. Goray. High fidelity blazed grating replication using nanoimprint lithography. *J. Vac. Sci. Technol. B*, 22:3260–3264, 2004.

[11] C.-H. Chang, Y. Zhao, R. K. Heilmann, and M. L. Schattenburg. Fabrication of 50 nm period gratings with multilevel interference lithography. *Opt. Lett.*, 33:1572–1574, 2008.

[12] Chih-Hao Chang, R. K. Heilmann, M. L. Schattenburg, and P. Glenn. Design of a double-pass shear mode acousto-optic modulator. *Rev. Sci. Instrum.*, 79:033104, 2008.

[13] Chih-Hao Chang, Y. Zhao, R. K. Heilmann, and M. L. Schattenburg. Spatial-frequency multiplication using multilevel interference lithography. *Summited to J. Vac. Sci. Technol. B*, 2008.

[14] C. G. Chen, P. T. Konkola, R. K. Heilmann, G. S. Pati, and M. L. Schattenburg. Image metrology and system controls for scanning beam interference lithography. *J. Vac. Sci. Technol. B*, 19:2335–2341, 2001.

[15] Carl Gang Chen. *Beam Alignment and Image Metrology for Scanning Beam Interference Lithography–Fabricateing Gratings with Nanometer Phase Accuracy.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2003.

[16] C. C. Cheng and A. Scherer. Fabrication of photonic band-gap crystals. *J. Vac. Sci. Technol. B*, 13:2696–2700, 1995.

[17] C. C. Cheng, A. Scherer, V. Arbet-Engels, and E. Yablonovitch. Lithographic band gap tuning in photonic band gap crystals. *J. Vac. Sci. Technol. B*, 14:4110–4114, 1996.

[18] C. C. Cheng, A. Scherer, R.-C. Tyan, Y. Fainman, G. Witzgall, and E. Yablonovitch. New fabrication techniques for high quality photonic crystals. *J. Vac. Sci. Technol. B*, 15:2764–2767, 1997.

[19] J. Y. Cheng, C. A. Ross, E. L. Thomas, H. I. Smith, and G. J. Vancso. Fabrication of nanostructures with long-range order using block copolymer lithography. *Appl. Phys. Lett.*, 81:3657–3659, 2002.

[20] Stephen Y. Chou, Peter R. Krauss, and Preston J. Renstrom. Imprint of sub-25 nm vias and trenches in polymers. *Appl. Phys. Lett.*, 67:3114–3116, 1995.

[21] M. F. Crommie, C. P. Lutz, and D. M. Eigler. Confinement of electrons to quantum corrals on a metal surface. *Science*, 262:218–220, 1993.

[22] B. Cui, Z. Yu, H. Ge, and S. Y. Chou. Large area 50 nm period grating by multiple nanoimprint lithography and spatial frequency doubling. *Appl. Phys. Lett.*, 90:043118, 2007.

[23] F. C. Demarest. "high-resolution, high-speed, low data age uncertainty, heterodyne displacement measuring interferometer electronics". *Meas. Sci. Technol.*, 9:1024–1030, 1998.

[24] I. B. Divliansky, A. Shishido, I.-C. Khoo, T. S. Mayer, D. Pena, S. Nishimura, C. D. Keating, and T. E. Mallouk. Fabrication of two-dimensional photonic crystals using interference lithography and electrodeposition of cdse. *Appl. Phys. Lett.*, 79:3392–3294, 2001.

[25] E. A. Donley, T. P. Heavner, F. Levi, M. O. Tataw, and S. R. Jefferts. "half-wave plate behaviour of ultrasonic waves light modulators". *Opt. Comm.*, 8:397–400, 1973.

[26] E. A. Donley, T. P. Heavner, F. Levi, M. O. Tataw, and S. R. Jefferts. "double-pass acousto-optic modulator system". *Rev. Sci. Intrum.*, 76:063112, 2005.

[27] J. Ferrera, M. L. Schattenburg, and H. I. Smith. Analysis of distortion in interferometric lithography. *J. Vac. Sci. Technol. B*, 14:4009–4013, 1996.

[28] R. P. Feynman. There's plenty of room at the bottom. *Engineering and Science*, 23, 1960.

[29] R. P. Feynman. Cramming more components onto integrated circuits. *Electronics Magazine*, 38, 1965.

[30] Joseph W. Goodman. *Introduction to Fourier Optics.* McGraw-Hill Science/Engineering/Math, 1996.

[31] D. Hambach, G. Schneider, and E. M. Gullikson. Efficiency high-order diffraction of extreme-ultraviolet light and soft x-rays by nanostructured volume gratings. *Opt. Lett.*, 26:1200–1202, 2001.

[32] S. E. Harris, S. T. K. Nieh, and D. K. Winslow. "electronically tunable acousto-optic filter". *Appl. Phys. Lett.*, 15:325–326, 1969.

[33] R. K. Heilmann, M. Ahn, E. M. Gullikson, and M. L. Schattenburg. Blazed high-efficiency x-ray diffraction via transmission through arrays of nanometer-scale mirros. *Opt. Exp.*, 16:8658–8669, 2008.

[34] R. K. Heilmann, C. G. Chen, P. T. Konkola, and M. L. Schattenburg. Dimensional metrology for nanometre-scale science and engineering: towards sub-nanometre accurate encorders. *Nanotechnology*, 15:S504–S511, 2004.

[35] R. K. Heilmann, P. T. Konkola, C. G Chen, G. S. Pati, and M. L. Schattenburg. Digital heterodyne interference fringe control system. *J. Vac. Sci. Technol. B*, 19:2342–2346, 2001.

[36] http://www.raith.com/. *Raith 150.*

[37] John D. Joannopoulos, Robert D. Meade, and Joshua N. Winn. *Photonic Crystals.* Princeton University Press, 1995.

[38] Y. Kanamori, M. Sasaki, and K. Hane. Broadband antireflection gratings fabricated upon silicon substrates. *Opt. Lett.*, 24:1422–1424, 1999.

[39] D. W. Keith, M. L. Schattenburg, H. I. Smith, and D. E. Pritchard. Diffraction of atoms by a transmission grating. *Phys. Rev. Lett.*, 61:1580–1583, 1988.

[40] Charles Kittel. *Introduction to Solid State Physics.* John Wiley & Sons, Inc, 2004.

[41] KLA-Tencor Corporation. *PROLITH^{TM}*.

[42] K. Knop. Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves. *J. Opt. Soc. Am.*, 68:1206–1210, 1978.

[43] P. T. Konkola, C. G. Chen, R. K. Heilmann, C. Joo, J. C. Montoya, C.-H. Chang, and M. L. Schattenburg. Nanometer-level repeatable metrology using the nanoruler. *J. Vac. Sci. Technol. B*, 21:3097–3110, 2003.

[44] Paul Thomas Konkola. *Design and Analysis of a Scanning Beam Interference Lithography System for Patterning Gratings with Nanometer-level Distortions.* PhD thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 2003.

[45] M. P. Kowalski, R. K. Heilmann, M. L. Schattenburg, C.-H. Chang, F. B. Berendse, and W. R. Hunter. Near-normal-incidence extreme-ultraviolet efficiency of a flat crystalline anisotropically etched blazed grating. *Appl. Opt.*, 45:1676–1679, 2004.

[46] W. Lee and F. L. Degertekin. Rigorous coupled-wave analysis of multilayered grating structures. *J. Lightwave Technol.*, 22:2359–2363, 2004.

[47] L. Li. Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings. *J. Opt. Soc. Am. A*, 13:1024–1035, 1996.

[48] Marc Madou. *Fundamentals of Microfabrication.* CRC Press, 1997.

[49] Daniel Malacara. *Optical Shop Testing.* John Wiley & Sons, Inc, 1992.

[50] MIT internal discussion. *Shadow tech process.*

[51] M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am.*, 71:811–818, 1981.

[52] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord. Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *J. Opt. Soc. Am. A*, 12:1068–1076, 1995.

[53] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord. Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach. *J. Opt. Soc. Am. A*, 12:1077–1086, 1995.

[54] Molecular Imprints, Inc. *Silspin HDMS data sheet.*

[55] J. C. Montoya, C.-H. Chang, R. K. Heilmann, and M. L. Schattenburg. Doppler writing and linewidth control for scanning beam interference lithography. *J. Vac. Sci. Technol. B*, 23:2640–2645, 2005.

[56] Juan Montoya. *Towards Nano-accuracy in Scanning Beam Interference Lithography.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2006.

[57] R. D. Piner, J. Zhu, F. Xu, S. Hong, and C. A. Mirkin. "dip-pen" nanolithography. *Science*, 283:661–663, 1999.

[58] M. Qi, E. Lidorikis, P. T. Rakich, S. G. Johnson, J. D. Joannopoulos, E. P. Ippen, and H. I. Smith. A three-dimensional optical photonic crystal with designed point defects. *Nature*, 429:538–542, 2004.

[59] C. A. Ross, S. Haratani, F. J. Casta no, Y. Hao, M. Hwang, M. Shima, J. Y. Cheng, B. Vögeli, M. Farhound, M. Walsh, and H. I. Smith. Magnetic behavior of lithographically patterned particle arrays. *J. Appl. Phys.*, 91:6848–6853, 2002.

[60] M. L. Schattenburg, R. J. Aucoin, and R. C. Fleming. Optically matched trilevel resist process for nanostructure fabrication. *J. Vac. Sci. Technol. B*, 13:3007–3011, 1995.

[61] M. L. Schattenburg, C. Chen, P. N. Everett, J. Ferrera, P. Konkola, and H. I. Smith. Sub-100 nm metrology using interferometrically produced fiducials. *J. Vac. Sci. Technol. B*, 17:2692–2697, 1999.

[62] J. F. Seely, L. I. Goray, B. Kjornrattanawanich, J. M. Laming, G. E. Holland, K. A. Flanagan, R. K. Heilmann, C.-H. Chang, M. L. Schattenburg, and A. P. Rasmussen. Efficiency of a grazing-incidence off-plane grating in the soft-x-ray region. *Appl. Opt.*, 45:1680–1687, 2006.

[63] H. H. Solak. Nanolithography with coherent extreme ultraviolet light. *J. Phys. D: Appl. Phys.*, 39:R171–R188, 2006.

[64] H. H. Solak, C. David, J. Gobrecht, V. Golovkina, F. Cerrina, S. O. Kim, and P. F. Nealey. Sub-50 nm period patterns with euv interference lithography. *Microelectron. Eng.*, 67:56–62, 2003.

[65] H. H. Solak, D. He, W. Li, S. Singh-Gasson, F. Cerrina, B. H. Sohn, X. M. Yang, and P. Nealey. Exposure of 38 nm period grating patterns with extreme ultraviolet interferometric lithography. *Appl. Phys. Lett.*, 75:2328–2330, 1999.

[66] Y. Sugeta. "the 23rd annual tokyo ohka seminar. *Tokyo, Japan*, 2002.

[67] M. Switkes, T. M. Bloomstein, and M. Rothschild. Patterning of sub-50 nm dense features with space-invariant 157 nm interference lithography. *Appl. Phys. Lett.*, 77:3149–3151, 2000.

[68] G. M. Whitesides, J. P. Mathias, and C. T. Seto. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science*, 254:1312–1319, 1991.

[69] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional dna crystals. *Nature*, 394:539–544, 1998.

[70] E. Yablonovitch. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.*, 58:2059–2062, 1987.

[71] E. Yablonovitch, T. J. Gmitter, and K. M. Leung. Photonic band structure: The face-centered-cubic case employing nonspherical atoms. *Phys. Rev. Lett.*, 67:2295–2298, 1991.

[72] Amnon Yariv and Pochi Yeh. *Optical Waves in Crystals.* John Wiley & Sons, Inc, 1984.

[73] Pochi Yeh. *Optical Waves in Layered Media.* John Wiley & Sons, Inc, 2005.

[74] Z. Yu, W. Wu, L. Chen, and S. Y. Chou. Fabrication of large area 100 nm pitch grating by spatial frequency doubling and nanoimprint lithography for subwavelength optical applications. *J. Vac. Sci. Technol. B*, 19:2816–2819, 2001.

[75] Y. Zhao, C.-H. Chang, R. K. Heilmann, and M. L. Schattenburg. Phase control in multiexposure spatial frequency multiplication. *J. Vac. Sci. Technol. B*, 25:2439–2443, 2007.

[76] Yong Zhao. *Ultra-high Precision Scanning Beam Interference Lithography and its Application–Spatial Frequency Multiplication.* PhD thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 2008.

[77] Zygo Corporation. *ZMI 2000 Manual.*